บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันการวิจัยในด้านต่างๆ ไม่ว่าจะเป็นด้านเศรษฐศาสตร์ บริหารธุรกิจ สังคมศาสตร์ วิทยาศาสตร์ ผู้วิจัยอาจสนใจผลลัพธ์ที่เป็นไปได้ 2 อย่างคือสำเร็จหรือไม่สำเร็จ ดังตัวอย่างของผู้ประกอบการอุตสาหกรรมที่ต้องการทราบผลการผลิตสินค้าที่ได้ว่าเป็นสินค้าที่ใช้ได้หรือใช้ไม่ได้เป็นจำนวนเท่าไร เพื่อใช้ในการปรับปรุงและเป็นประโยชน์ต่อการผลิตสินค้าอุตสาหกรรม ในด้านการกีฬา เช่น การยิงปืนนั้นก็สามารถนำไปใช้ในการตรวจสอบการยิงแต่ละนัดว่าถูกเป้าหรือไม่ถูกเป้าเพื่อตรวจนับความแม่นยำในการแข่งขัน เป็นต้น ซึ่งแต่ละด้านจำเป็นต้องใช้วิธีการสถิติเข้ามาช่วยในการค้นคว้าวิจัยโดยเฉพาะการหาคำตอบเพื่อคาดคะเนเหตุการณ์ที่จะเกิดขึ้นในอนาคตหรือที่เรียกว่าการพยากรณ์ ซึ่งผู้วิจัยจำนวนมากจะใช้วิธีการวิเคราะห์ความถดถอย (regression analysis)

การวิเคราะห์การถดถอยใช้ในการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม โดยความสัมพันธ์อาจมีได้ทั้งในรูปแบบที่เป็นเชิงเส้นและไม่เป็นเชิงเส้น วิธีการถดถอยที่นักวิจัยส่วนใหญ่นิยมใช้กันอย่างแพร่หลายคือการวิเคราะห์การถดถอยเชิงเส้น (linear regression analysis)

การนำตัวแบบการถดถอยเชิงเส้น (linear regression model) เพื่อไปใช้ในการพยากรณ์นั้นพบว่าถ้าใช้ตัวแปรอิสระเพียงตัวเดียวในตัวแบบการถดถอยเชิงเส้นจะเรียกว่า การวิเคราะห์ความถดถอยเชิงเส้นอย่างง่าย (simple linear regression analysis) ซึ่งการวิเคราะห์นี้อาจจะทำให้การพยากรณ์คลาดเคลื่อนจากความเป็นจริงหรืออาจเชื่อถือได้ไม่มากนักเนื่องจากตัวแปรอิสระที่มีอิทธิพลต่อตัวแปรตามมีเพียงหนึ่งตัว เพราะฉะนั้นผู้วิจัยอาจต้องใช้ตัวแปรอิสระมากกว่าหนึ่งตัวแปรเพื่อเพิ่มความแม่นยำในการพยากรณ์ซึ่งจะเรียกว่าการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ (multiple linear regression) แต่บางครั้งการใช้ตัวแปรอิสระมากเกินไปในตัวแบบถดถอยอาจไม่ให้ผลดีเสมอไป ในทางกลับกันอาจจะเพิ่มความคลาดเคลื่อนในการพยากรณ์ให้สูงกว่าการใช้ตัวแปรอิสระเพียงตัวเดียวในตัวแบบถดถอยก็เป็นได้ ดังนั้นเพื่อให้การพยากรณ์มีความถูกต้องการใช้วิธีการคัดเลือกตัวแบบที่เหมาะสมจึงมีความจำเป็นอย่างมาก การวิจัยในการคัดเลือกตัวแบบการถดถอยเชิงเส้นภายใต้เงื่อนไขของตัวแบบติดกลุ่ม (nested model) ผู้วิจัยสามารถใช้เกณฑ์ในการ

คัดเลือกตัวแบบโดยการใช้ข้อมูลในอดีตของพารามิเตอร์มาช่วยในการประเมินค่าเพื่อลดความคลาดเคลื่อนให้ต่ำลงและได้แบ่งการแจกแจงก่อนเป็น 2 กรณีดังนี้[1]

กรณีที่ 1 (กรณีพิเศษ) เนื่องจากเรื่องที่วิจัยเป็นเรื่องเกี่ยวกับผลที่เป็นไปได้ 2 อย่างคือสำเร็จหรือไม่สำเร็จ ดังนั้นควรพิจารณาการแจกแจงเบตา (Beta distribution) เป็นการแจกแจงก่อน (prior distribution) ซึ่งกรณีพิเศษของการแจกแจงเบตา ($\alpha = \beta = 1$) มีรูปแบบเป็นการแจกแจงสม่ำเสมอ (uniform distribution)

กรณีที่ 2 (กรณีทั่วไป) ตัวแปรสุ่มที่มีการแจกแจงเบอร์นูลลี (Bernoulli distribution) เป็นเรื่องเกี่ยวกับผลที่เป็นไปได้ 2 อย่างคือสำเร็จหรือไม่สำเร็จ จึงเป็นเหตุให้พิจารณาการแจกแจงก่อนสังยุคเป็นการแจกแจงเบตา (conjugate beta prior distribution) เพื่อให้ง่ายต่อการคำนวณความน่าจะเป็นภายหลัง (posterior probability) ของกลุ่มตัวแปรอิสระ โดยเลือกกลุ่มของตัวแปรอิสระที่มีความน่าจะเป็นภายหลังสูงสุด ซึ่งในปีค.ศ.1997 จอร์จ (George) และแมคคัลลอช (McCulloch)ได้เสนอวิธี Markov Chain Monte Carlo มาช่วยในการเลือกกลุ่มของตัวแปรอิสระ[2]

แนวความคิดที่ใช้การแจกแจงก่อนของพารามิเตอร์มาพิจารณานี้เรียกว่า การวิเคราะห์เชิงเบส์ (Bayesian analysis) การนำแนวทางของเบส์มาใช้ในการหาเกณฑ์หรือวิธีการคัดเลือกตัวแบบการถดถอยเชิงเส้นที่ดีที่สุด จึงเป็นแนวทางที่นักสถิติให้ความสนใจกันมากในปัจจุบัน เพราะมีแนวโน้มทำให้ได้วิธีการคัดเลือกตัวแบบที่มีประสิทธิภาพและเหมาะสมกว่าวิธีการภายใต้แนวทางอื่นๆ

ในปี ค.ศ. 1995 โฮเอ็ททิง ราฟเทอร์รีและเมดิแกน[3]ได้เสนอวิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging method (BMA)) สำหรับตัวแบบการถดถอยเชิงเส้นซึ่งวิธีนี้มีการคำนึงถึงหลักการเกี่ยวกับความไม่แน่นอนของตัวแบบ (model uncertainty) วิธีการเฉลี่ยตัวแบบของเบส์จะพิจารณาค่าความน่าจะเป็นภายหลัง (posterior probability) สำหรับทุกๆตัวแบบที่เราสนใจ และนำตัวแบบทุกตัวแบบที่เราสนใจมาเฉลี่ยกันโดยใช้ความน่าจะเป็นภายหลังของแต่ละตัวแบบเป็นตัวถ่วงน้ำหนัก เพื่อหาค่าพยากรณ์ที่เหมาะสมโดยการหาองค์ประกอบของตัวแบบด้วย

---

[1] Mu Zhu and Arthur Y. Lu, "The Counter-intuitive Non-informative Prior for the Bernoulli Family," Journal of Statistics Education (Number 2004).

[2] สุรางค์ ถาวร, "การศึกษาวิธีการเลือกสมการถดถอยเชิงเส้นที่ดีที่สุดในกรณีตัวแปรอิสระมีพหุสัมพันธ์กันด้วยวิธีเบย์เซียน," (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์,2542).

[3] Adrian E.Raftery, Jennifer A.Hoeting and David Madigan, Bayesian Simultaneous Variable and Transformation Selection in Linear Regression (June 1995).

เทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo model composition (MC^3)) เมื่อพิจารณาการคัดเลือกตัวแปรและการแปลงไปพร้อมๆกัน ซึ่งทำให้ค่าพยากรณ์เหมาะสมขึ้นทั้งในเรื่องของคะแนนการพยากรณ์ทั้งหมด (overall predictive score) และความครอบคลุมของช่วงการพยากรณ์ (the coverage of prediction intervals)

ในปีค.ศ. 2002 บาร์บีรี (Barbieri) และเบอร์เกอร์ (Berger) [4] ได้เสนอวิธีการคัดเลือกตัวแบบของเบส์สำหรับการคัดเลือกระหว่างกลุ่มของตัวแบบเชิงเส้นปกติ(selection among normal linear models) ตัวแบบที่เหมาะสมที่สุดอาจไม่ใช่ตัวแบบที่มีค่าความน่าจะเป็นภายหลังสูงสุด (the model with highest posterior probability) แต่กลับเป็นตัวแบบที่ค่าความน่าจะเป็นภายหลังมัธยฐาน (the median probability model) และเกณฑ์ที่ใช้ในการเลือกตัวแบบที่เหมาะสมที่สุดจากกลุ่มของตัวแบบที่มีความน่าจะเป็นภายหลังมัธยฐานคือความสูญเสียอันเกิดจากความผิดพลาดยกกำลังสอง (square error loss) ซึ่งจะเลือกตัวแบบที่มีค่าดังกล่าวต่ำสุด

ในปี ค.ศ. 2003 จิตติมา ผสมญาติ [5] เปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ โดยเปรียบเทียบวิธีการคัดเลือกสมการถดถอย 3 วิธี ได้แก่ วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging method) โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo model composition (MC^3)) เมื่อพิจารณาการแปลงที่เหมาะสมของตัวแปรอิสระ (model Uncertainty via Simultaneous Variable and Transformation Selection (BMA_{SVT})) วิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุด (Optimal Predictive Model Selection:median probability model (OPM)) และวิธีการถดถอยแบบขั้นบันได (Stepwise Regression method (SR)) เกณฑ์ที่ใช้ในการตัดสินใจ คือเกณฑ์ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Average of Mean Square Error (AMSE)) และใช้ค่าอัตราส่วนผลต่างของค่าเฉลี่ยค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Ratio of Different Average Mean Square Error (RDAMSE)) เพื่อประกอบการเปรียบเทียบประสิทธิภาพของวิธีการข้างต้นและสามารถแสดงข้อสรุปผลการวิจัยได้ดังนี้

---

[4] Maria M.Barbieri and James O.Berger "Optimal predictive model selection," Technical Report 02-02 (April 2002).

[5] จิตติมา ผสมญาติ, "การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ," (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย,2546).

การเปรียบเทียบค่า AMSE ของทั้ง 3 วิธีเรียงลำดับจากน้อยไปหามาก ได้แก่ วิธี BMA_{SVT} OPM และSR ตามลำดับ สำหรับทุกสถานการณ์ วิธี OPM จะได้ค่า AMSE สูงกว่าวิธี BMA_{SVT} เพียงเล็กน้อยโดยเฉพาะอย่างยิ่งเมื่อค่าคงที่ $\left(\frac{\sigma_\beta}{\tau}, c\right)$ มีค่าต่ำๆ ส่วนวิธี SR มีค่า AMSE แตกต่างจากวิธี BMA_{SVT} และวิธี OPM อย่างชัดเจนในทุกสถานการณ์

ในการวิจัยครั้งนี้ผู้วิจัยได้สนใจทำการศึกษาเปรียบเทียบวิธีการที่ใช้ในการคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณ 3 วิธีคือ

(1) วิธีการเฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่เมื่อพิจารณาการแปลงที่เหมาะสมของตัวแปรอิสระ(BMA_{SVT})

(2) วิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุด(OPM)

(3) วิธีการถดถอยแบบขั้นบันได (SR)

ซึ่งสองวิธีแรกเป็นวิธีการภายใต้แนวทางของเบส์ ส่วนวิธีที่ 3 เป็นวิธีการพื้นฐานซึ่งนิยมใช้กันอย่างแพร่หลายและมีประสิทธิภาพดีในการคัดเลือกสมการถดถอยที่ดีที่สุดเมื่อการวิเคราะห์ความถดถอยเป็นไปตามข้อตกลงเบื้องต้น[6]ซึ่งผู้วิจัยได้นำมาเปรียบเทียบกับวิธีการภายใต้แนวทางของเบส์เพื่อศึกษาดูแนวโน้มว่าวิธีการคัดเลือกตัวแบบภายใต้แนวทางของเบส์จะมีความเหมาะสมกว่าวิธีการพื้นฐานหรือไม่เมื่อใช้วงค์คู่สังยุคแบบเบตา ดังนั้นถ้าผลการวิจัยพบว่าวิธีการใดมีประสิทธิภาพในการพยากรณ์มากที่สุดก็ควรจะนำไปใช้ในการคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณที่ดีที่สุดต่อไป

1.2 วัตถุประสงค์การวิจัย

1) เพื่อศึกษาการคัดเลือกสมการถดถอยที่ดีที่สุดในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ 3 วิธีคือ

1.1 วิธีการ เฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟเมื่อพิจารณาการแปลงที่เหมาะสมของตัวแปรอิสระ (BMA_{SVT})

1.2 วิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุด (OPM)

1.3 วิธีการถดถอยแบบขั้นบันได (SR)

2) เพื่อเปรียบเทียบความถูกต้องของการพยากรณ์จากตัวแบบที่ได้จากทั้ง 3 วิธีการ

---

[6] จิตติมา ผสมญาติ, "การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ," (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย,2546) หน้า 4.

1.3 ข้อตกลงเบื้องต้น

1) รูปแบบทั่วไปของสมการถดถอยเชิงเส้น(linear regression analysis)มีรูปแบบทั่วไปดังนี้

(1.1) $$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{y}$ เป็นเวกเตอร์ของตัวแปรตามขนาด $n \times 1$

X เป็นเมทริกซ์ของตัวแปรอิสระขนาด $n \times (p+1)$

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของพารามิเตอร์ที่ไม่ทราบค่าขนาด $(p+1) \times 1$

$\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของความคลาดเคลื่อนสุ่มขนาด $n \times 1$

n เป็นขนาดตัวอย่าง

และ p เป็นจำนวนตัวแปรอิสระ

2) ตัวแปรอิสระแต่ละตัวถือว่าเป็นค่าคงที่

3) ความคลาดเคลื่อนสุ่มเป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น0และส่วนเบี่ยงเบนมาตรฐานเป็น σ (หรือ $N(0,\sigma^2)$) เหมือนกันและเป็นอิสระซึ่งกันและกัน

1.4 สมมติฐานของการวิจัย

1) วิธีการคัดเลือกสมการถดถอยที่ดีที่สุดภายใต้แนวทางของเบส์ในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณนั้น จะให้ค่าพยากรณ์ที่มีความถูกต้องและแม่นยำมากกว่าวิธีการพื้นฐาน เพราะวิธีพื้นฐานหรือในที่นี้คือวิธีการขั้นบันได(SR)จะใช้หลักการในการคัดเลือกตัวแปรเข้าออกจากตัวแบบผ่านการทดสอบสมมติฐานเท่านั้น ในขณะที่วิธีการภายใต้แนวทางของเบส์ทั้ง 2 วิธีคือ วิธี BMA_{SVT}และวิธี OPM ได้มีการพิจารณาถึงความเหมาะสมของตัวแบบด้วยความน่าจะเป็นภายหลังของตัวแบบ โดยเฉพาะวิธี BMA_{SVT} ได้มีการคำนึงถึงความน่าจะเป็นภายหลังของตัวแบบและมีการพิจารณาการแปลงที่เหมาะสมของตัวแปรอิสระ

2) วิธีการเฉลี่ยตัวแบบของเบส์เมื่อพิจารณาการแปลงค่าตัวแปรอิสระที่เหมาะสม (BMA_{SVT})จะให้ค่าพยากรณ์ที่มีความถูกต้องและแม่นยำมากกว่าวิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุด(OPM) เพราะวิธี OPM พิจารณาถึงความเหมาะสมของตัวแบบด้วยความน่าจะเป็นภายหลังของตัวแบบ ส่วนวิธี BMA_{SVT} คำนึงถึงความน่าจะเป็นภายหลังของตัวแบบและพิจารณาการแปลงที่เหมาะสมของตัวแปรอิสระ

1.5 ขอบเขตของการวิจัย

1) ตัวแบบของการถดถอยเชิงเส้นพหุคูณที่น่าสนใจศึกษามีรูปแบบดังสมการ (1.1) ซึ่งรูปแบบการถดถอยนอกจากจะเป็นเชิงเส้นในพารามิเตอร์แล้ว ยังเป็นเชิงเส้นในตัวแปรอิสระด้วย

2) ตัวแบบการถดถอยในการวิจัยครั้งนี้เป็นตัวแบบติดกลุ่ม (nested model)

3) ขนาดตัวอย่าง[7]ที่ศึกษามี 4 ขนาดคือ 10 30 50 และ 100

4) จำนวนตัวแปรอิสระ[8]ที่ศึกษามี 4 ระดับคือ 3 5 10 และ 15 ตัวแปร โดยตัวแปรอิสระแต่ละตัวแปรสร้างจากการแจกแจงปกติ $N(0,1)$

5) ในการวิจัยครั้งนี้จะศึกษาเมื่อความคลาดเคลื่อนสุ่มมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.25 0.50 และ 2.50 ตามลำดับ[9]

6) ในการวิจัยครั้งนี้ใช้การแจกแจงก่อนแบบคู่สังยุคเป็นการแจกแจงแบบเบตา(conjugate beta prior distribution) เพราะการใช้ข้อมูลในอดีตของพารามิเตอร์มาช่วยในการประเมินค่าจะลดความคลาดเคลื่อนให้ต่ำลงได้และจะแบ่งการแจกแจงก่อนออกเป็น 2 กรณีได้ดังนี้

<u>กรณีที่ 1</u> (กรณีพิเศษ) เนื่องจากเรื่องที่วิจัยเป็นเรื่องเกี่ยวกับผลที่เป็นไปได้ 2 อย่างคือสำเร็จหรือไม่สำเร็จ ดังนั้นควรพิจารณาการแจกแจงเบตา (Beta distribution) เป็นการแจกแจงก่อน (prior distribution) ซึ่งกรณีพิเศษของการแจกแจงเบตา ($\alpha = \beta = 1$) เป็นการแจกแจงสม่ำเสมอ (uniform distribution)

<u>กรณีที่ 2</u> (กรณีทั่วไป) ตัวแปรสุ่มที่มีการแจกแจงเบอร์นูลลี (Bernoulli distribution) เป็นเรื่องเกี่ยวกับผลที่เป็นไปได้ 2 อย่างคือสำเร็จหรือไม่สำเร็จ จึงเป็นเหตุให้พิจารณาการแจกแจงก่อนสังยุคเป็นการแจกแจงเบตา (conjugate beta prior distribution)เพื่อให้ง่ายต่อการคำนวณความน่าจะเป็นภายหลัง (posterior probability) ของกลุ่มตัวแปรอิสระ โดยเลือกกลุ่มของตัวแปรอิสระที่มีความน่าจะเป็นภายหลังสูงสุด

7) ค่าคงที่สำหรับวิธี BMA_{SVT}และวิธี OPM มี 4 ระดับคือ$\left(\frac{\sigma_\beta}{\tau}, c\right)$มีค่าเป็น (1,5) (1,10) (10,100) และ (10,500) ทั้งนี้เพราะ[10]*ค่าคงที่* (1,5) *จะให้การแจกแจงแบบปกติที่มีการกระจายของพารามิเตอร์แคบเกินไปเมื่อเปรียบเทียบกับค่าคงที่* (1,10) *และค่าคงที่* (10,500) *จะให้การแจกแจงแบบปกติที่มีการกระจายของพารามิเตอร์กว้างเกินไปเมื่อเปรียบเทียบกับค่าคงที่* (10,100)*จึงเป็นเหตุให้ค่า AMSE จะเพิ่มขึ้นเมื่อค่าคงที่สูงขึ้น เนื่องจากค่าคงที่กำหนดลักษณะการกระจายของพารามิเตอร์สัมประสิทธิ์การถดถอยสุ่มได้ ดังนั้นค่าคงที่ที่สูงขึ้นจะทำให้การกระจายของ*

---

[7] เมื่อขนาดตัวอย่างเพิ่มขึ้นจะช่วยลดความเบี่ยงเบนที่ไม่ทราบสาเหตุลงได้

[8] เมื่อมีจำนวนตัวแปรอิสระเพิ่มขึ้นโอกาสที่จะได้ตัวแบบที่ไม่เหมาะสมก็จะมีมากขึ้นตามไปด้วย

[9] ค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มมีค่าเพิ่มขึ้นมาก ($\sigma \geq 0.50$) ค่าAMSEจะมีแนวโน้มเพิ่มขึ้นมากตามไปด้วย เพราะค่าAMSEเป็นตัวประมาณค่าความแปรปรวนของค่าคลาดเคลื่อน (σ^2)

[10] นิทัสน์ สุขสุวรรณ, "การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดภายใต้แนวทางของเบส์ในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ," (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย,2545), หน้า 20.

พารามิเตอร์สัมประสิทธิ์การถดถอยมีการกระจายมากขึ้น ทำให้ค่าที่สุ่มได้มีความแม่นยำลดลงซึ่งส่งผลให้ค่า AMSE มีค่าสูงขึ้น

8) ข้อมูลที่ใช้ในการศึกษาวิจัยครั้งนี้ใช้โปรแกรมคอมพิวเตอร์จำลองด้วยเทคนิคมอนติคาร์โล กระทำซ้ำ 500 รอบในแต่ละสถานการณ์

1.6 ขั้นตอนในการศึกษาวิจัย

1) กำหนดลักษณะการแจกแจงของค่าคลาดเคลื่อน ขนาดตัวอย่าง จำนวนตัวแปรอิสระ ระดับนัยสำคัญและค่าคงที่สำหรับวิธี BMA_{SVT} และวิธี OPM

2) สร้างข้อมูลตัวแปรตามจากตัวแปรอิสระและความคลาดเคลื่อนที่มีลักษณะการแจกแจงตามที่ต้องการศึกษา โดยให้ตัวแปรตามมีความสัมพันธ์เชิงเส้นในพารามิเตอร์กับตัวแปรอิสระ

3) สร้างตัวแบบโดยวิธีทั้ง 2 วิธีคือ

3.1วิธีการเฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟเมื่อพิจารณาการแปลงที่เหมาะสมของตัวแปรอิสระ(BMA_{SVT})

3.2 วิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุด(OPM)

4) หาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของแต่ละตัวแบบ

5) สรุปผลในรูปของตารางและรูปภาพ

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1) ผลที่ได้จากการศึกษาจะเป็นแนวทางในการคัดเลือกตัวแบบที่เหมาะสมเพื่อใช้ในการพยากรณ์สำหรับการวิเคราะห์ความถดถอยเชิงพหุคูณ

2) เพื่อให้ได้ข้อเสนอแนะในการศึกษาวิธีการคัดเลือกตัวแบบการถดถอยที่ดีที่สุดวิธีอื่นๆ ภายใต้แนวทางของเบส์สำหรับการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณต่อไป

1.8 เกณฑ์ที่ใช้ในการตัดสินใจ[11]

เกณฑ์ที่ใช้ในการตัดสินใจว่าวิธีการคัดเลือกตัวแบบการถดถอยวิธีใดจะมีประสิทธิภาพมากที่สุดพิจารณาจากค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย(Average of Mean Square Error (AMSE)) และใช้ค่าอัตราส่วนผลต่างของค่าเฉลี่ยค่าคลาดเคลื่อนกำลังสองเฉลี่ย(Ratio of Different Average Mean Square Error (RAMSE)) เพื่อเปรียบเทียบประสิทธิภาพของวิธีการซึ่งมีสูตรดังนี้

---

[11] จิตติมา ผสมญาติ, "การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ," (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย,2546) หน้า 7.

$$MSE_k = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-(p+1)}$$

ดังนั้น $$AMSE = \frac{\sum_{k=1}^{n} MSE_k}{500}$$

เมื่อ y_i เป็นค่าสังเกตที่ i

$\hat{y}_i$ เป็นค่าพยากรณ์ที่ i

p เป็นจำนวนตัวแปรอิสระ

n เป็นขนาดตัวอย่าง

และ MSE_k เป็นค่าคลาดเคลื่อนกำลังสองเฉลี่ยของการทำซ้ำรอบที่ k

ส่วน $$RDAMSE_j = \frac{(AMSE_j - AMSE_{\min})}{AMSE_{\min}} \times 100\%$$

เมื่อ $AMSE_j$ เป็นค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากวิธีที่ j ; $j = 1,2,...,n$

และ $AMSE_{\min}$ เป็นค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยที่มีค่าต่ำสุดจากทุกวิธี