

จิตติมา ผสมญาติ : การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบย์เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ ( A COMPARISON OF BAYESIAN SELECTION METHODS FOR BEST REGRESSION EQUATION WITH CONJUGATE NORMAL PRIOR DISTRIBUTION.)

อ. ที่ปรึกษา : รศ.ดร. วีระพร วีระถาวร, 165 หน้า. ISBN 974-17-4282-7

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบย์เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ โดยจะเปรียบเทียบวิธีการคัดเลือกสมการถดถอย 3 วิธี ได้แก่ วิธีการเฉลี่ยตัวแบบของเบย์ (Bayesian Model Averaging method) โดยการหาค่าประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo model composition (MC<sup>3</sup>)) เมื่อพิจารณาการแปลงที่เหมาะสมของตัวแปรอิสระ (Model Uncertainty via Simultaneous Variable and Transformation Selection)(BMA<sub>SVT</sub>) วิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุด (Optimal Predictive Model Selection: median probability model)(OPM) และวิธีการถดถอยแบบขั้นบันได (Stepwise Regression method)(SR) เกณฑ์ที่ใช้ในการตัดสินใจ คือ เกณฑ์ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Average of Mean Square Error )(AMSE) และเกณฑ์ที่ใช้ประกอบการตัดสินใจ คือ เกณฑ์ค่าอัตราส่วนผลต่างของค่าเฉลี่ยค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Ratio of Different Average Mean Square Error ) (RDAMSE) การแจกแจงของค่าคลาดเคลื่อนที่ศึกษา คือ การแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐาน ( $\sigma$ ) เท่ากับ 0.25 0.50 และ 2.50 ตามลำดับ ขนาดตัวอย่างที่ศึกษา ( $n$ ) คือ 15 30 50 และ 100 จำนวนตัวแปรอิสระที่ศึกษาคือ 3 5 8 10 12 และ 15 ค่าคงที่สำหรับวิธี BMA<sub>SVT</sub> และวิธี OPM ( $\sigma_B / \tau, c$ ) ที่ศึกษามี 4 ระดับ คือ (1,5) (1,10) (10,100) และ (10,500) ข้อมูลที่ใช้ในการวิจัยได้จากการจำลองแบบด้วยเทคนิคมอนติคาร์โล กระทำซ้ำ 500 รอบ ในแต่ละสถานการณ์ ซึ่งผลการวิจัยได้ข้อสรุปดังนี้

การเปรียบเทียบค่า AMSE ของทั้ง 3 วิธีเรียงลำดับจากน้อยไปมาก ได้แก่ วิธี BMA<sub>SVT</sub> OPM และ SR ตามลำดับ สำหรับทุกสถานการณ์ วิธี OPM จะให้ค่า AMSE สูงกว่าวิธี BMA<sub>SVT</sub> เพียงเล็กน้อยโดยเฉพาะอย่างยิ่งเมื่อค่าคงที่ ( $\sigma_B / \tau, c$ ) มีค่าต่ำ ๆ ส่วนวิธี SR มีค่า AMSE แตกต่างจากวิธี BMA<sub>SVT</sub> และวิธี OPM อย่างชัดเจนในทุกสถานการณ์ ปัจจัยที่มีผลต่อค่า AMSE ของทุกวิธี คือ ขนาดตัวอย่าง ค่าส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนและจำนวนตัวแปรอิสระ โดยที่ค่า AMSE จะแปรผันตามค่าส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนและจำนวนตัวแปรอิสระ แต่จะแปรผกผันกับขนาดตัวอย่าง

นอกจากนั้นค่า AMSE ของ 2 วิธีซึ่งเป็นวิธีการภายใต้แนวทางของเบย์ คือ วิธี BMA<sub>SVT</sub> และวิธี OPM จะแปรผันตามค่าคงที่ ( $\sigma_B / \tau, c$ ) โดยที่ค่า AMSE ของวิธี SR ไม่เปลี่ยนแปลงเนื่องจากไม่ได้นำค่าคงที่ ( $\sigma_B / \tau, c$ ) มาพิจารณา

## 4482197026 : MAJOR STATISTICS

KEY WORD : BAYESIAN MODEL AVERAGING / MARKOV CHAIN MONTE CARLO MODEL COMPOSITION / CHANGE-POINT TRANSFORMATION / OPTIMAL PREDICTIVE MODEL SELECTION / STEPWISE REGRESSION

JITTIMA PHASOMYARD : A COMPARISON OF BAYESIAN SELECTION METHODS FOR BEST REGRESSION EQUATION WITH CONJUGATE NORMAL PRIOR DISTRIBUTION.)

THESIS ADVISOR : ASSOC. PROF. THEERAPORN VERATHAWORN, 165 pp. ISBN 974-17-4282-7

The objective of this research is to compare the Bayesian selection methods for best regression equation with conjugate normal prior distribution. The three Bayesian selection methods for best regression equation in this comparison composed of Bayesian Model Averaging method using Markov Chain Monte Carlo model composition via Simultaneous Variable and Transformation Selection ( $BMA_{SVT}$ ), Optimal Predictive Model Selection (OPM) and Stepwise Regression method (SR). The average of mean square error (AMSE) and the ratio of different average mean square error (RDAMSE) were used as the criteria in this project. The details of the data were represented as follows. In this study, the distribution of random errors are normal distribution with mean equal to 0 and standard deviation equal to 0.25, 0.50 and 2.50, respectively. The sizes of the samples are varied, which composed of 15, 30, 50 and 100 samples. The numbers of independent variables in regression model are 3, 5, 8, 10, 12, and 15. The Bayesian hyperparameters ( $\sigma_\beta / \tau, c$ ) for  $BMA_{SVT}$  and OPM are (1,5), (1,10), (10,100) and (10,500), respectively. Using the Monte Carlo simulation technique with 500 repetitions for each case generated all data.

The analyzed results of data were demonstrated as follow. The comparisons of the AMSE from three methods ranging from minimum to maximum were  $BMA_{SVT}$ , OPM and SR for all cases. The OPM gave AMSE slightly higher than  $BMA_{SVT}$  especially when the Bayesian hyperparameters ( $\sigma_\beta / \tau, c$ ) had low values. The SR gave AMSE clearly different from  $BMA_{SVT}$  and OPM in all cases. The factors that affected AMSE of all methods are sample size, the standard deviation of random errors and the number of independent variables. The AMSE of all method were proportionate to the standard deviation of random errors and the number of independent variables although they were inversely proportionate to sample size.

Furthermore, the AMSE of two methods under Bayesian Approach,  $BMA_{SVT}$  and OPM were proportionate to the Bayesian hyperparameters ( $\sigma_\beta / \tau, c$ ) and whereas the AMSE of SR was constant because we did not consider the Bayesian hyperparameters ( $\sigma_\beta / \tau, c$ ).