

สมลักษณ์ ศิริชื่นวิจิตร : การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของตัวแบบถดถอยพหุนาม
(A COMPARISON ON PARAMETERS - ESTIMATION OF POLYNOMIAL REGRESSION
MODEL) อ. ที่ปรึกษา : รองศาสตราจารย์ ดร.วิระพร วีระถาวร , 99 หน้า. ISBN 974-53-2729-8

การวิจัยครั้งนี้มีวัตถุประสงค์ที่จะเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการวิเคราะห์ความถดถอยพหุนาม โดยจะเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ 3 วิธี ได้แก่ วิธีกำลังสองน้อยสุดสามัญ (Ordinary Least Squares method (OLS)) วิธีริจด์สามัญ (Ridge Ordinary Least Squares method (ROLS)) วิธีริจด์ที่มีค่าสัมบูรณ์น้อยสุด (Ridge Least Absolute Value method (RLAV)) เกณฑ์การเปรียบเทียบที่ใช้คือค่าเฉลี่ยรากของค่าคลาดเคลื่อนกำลังสอง (Average Root Mean Squares Error (AMSE)) และใช้อัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าคลาดเคลื่อนกำลังสอง (Ratio of Different Average Root Mean Squares Error (DIFF)) เพื่อเปรียบเทียบประสิทธิภาพ การแจกแจงของความคลาดเคลื่อนสุ่มในตัวแปรตามที่ศึกษาคือการแจกแจงแบบปกติที่มีค่าเฉลี่ย 0 และความแปรปรวนเป็น 4, 6, 8 และ 10 ส่วนการแจกแจงของตัวแปรอิสระที่ศึกษาคือการแจกแจงแบบปกติที่มีค่าเฉลี่ย 5 และความแปรปรวนเป็น 4 ขนาดตัวอย่างที่ใช้คือ 15, 30, 60, 120 และ 240 และกำลังสูงสุดของตัวแปรอิสระที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบถดถอยพหุนาม (MB) คือ 2, 3, 4, 5 และ 6 ข้อมูลที่ใช้ในการวิจัยใช้โปรแกรมคอมพิวเตอร์จำลองด้วยเทคนิคมอนติคาร์โลกระทำซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์ ซึ่งผลการวิจัยสรุปได้ดังนี้

ค่าเฉลี่ยรากของค่าคลาดเคลื่อนกำลังสอง (AMSE) แปรผันตามปัจจัยต่อไปนี้เรียงลำดับจากมากไปหาน้อย ได้แก่ กำลังสูงสุดของตัวแปรอิสระที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบถดถอยพหุนาม (MB) และความแปรปรวนของความคลาดเคลื่อนสุ่มในตัวแปรตาม (σ^2) แต่ AMSE แปรผกผันกับขนาดตัวอย่างซึ่งอิทธิพลของปัจจัยดังกล่าวเป็นดังนี้

1. กรณีขนาดตัวอย่างมีค่าน้อย ($n \leq 30$)

ทุกเลขที่กำลังสูงสุดของตัวแปรอิสระที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบถดถอยพหุนาม (MB) วิธีประมาณค่าพารามิเตอร์ที่ดีที่สุดคือ RLAV รองมาคือ ROLS และ OLS ตามลำดับ ทุกความแปรปรวนของความคลาดเคลื่อน (σ_e^2)

2. กรณีขนาดตัวอย่างมีค่าปานกลาง ($n = 60$)

กรณีเลขที่กำลังสูงสุดของตัวแปรอิสระที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบถดถอยพหุนามมีค่าน้อย ($MB \leq 3$) และความแปรปรวนของความคลาดเคลื่อนมีขนาดเล็ก ($\sigma_e^2 \leq 6$) พบว่าประสิทธิภาพของตัวประมาณ OLS ต่างจาก RLAV ไม่ถึงหนึ่งเท่า ดังนั้นวิธีประมาณค่าพารามิเตอร์ที่เหมาะสมในกรณีนี้คือ OLS ส่วนกรณีอื่น ๆ วิธีประมาณค่าพารามิเตอร์ที่ดีที่สุดได้แก่ RLAV รองมาคือ ROLS และ OLS ตามลำดับ สำหรับทุกความแปรปรวนของความคลาดเคลื่อน (σ_e^2)

3. กรณีขนาดตัวอย่างมีค่าค่อนข้างมาก ($n = 120$)

กรณีเลขที่กำลังสูงสุดของตัวแปรอิสระที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบถดถอยพหุนามมีค่าน้อย ($MB \leq 3$) ทุกความแปรปรวนของความคลาดเคลื่อน (σ_e^2) และเลขที่กำลังสูงสุดของตัวแปรอิสระ ($MB = 4$) ความแปรปรวนของความคลาดเคลื่อนมีขนาดเล็ก ($\sigma_e^2 \leq 6$) พบว่าประสิทธิภาพของ OLS ต่างจาก RLAV ไม่ถึงหนึ่งเท่า ดังนั้นวิธีประมาณค่าพารามิเตอร์ที่เหมาะสมในกรณีนี้คือ OLS ส่วนกรณีอื่น ๆ วิธีประมาณค่าพารามิเตอร์ที่ดีที่สุดคือ RLAV รองมาคือ ROLS และ OLS ตามลำดับ สำหรับทุกความแปรปรวนของความคลาดเคลื่อน (σ_e^2)

4. กรณีขนาดตัวอย่างมีค่ามาก ($n = 240$)

กรณีเลขที่กำลังสูงสุดของตัวแปรอิสระที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบถดถอยพหุนามมีค่าน้อยถึงปานกลาง ($MB \leq 4$) และทุกความแปรปรวนของความคลาดเคลื่อน (σ_e^2) พบว่าประสิทธิภาพของวิธี OLS ต่างจาก RLAV ไม่ถึงหนึ่งเท่า ดังนั้นวิธีประมาณค่าพารามิเตอร์ที่เหมาะสมในกรณีนี้คือ OLS ส่วนกรณีที่เลขที่กำลังสูงสุดของตัวแปรอิสระที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบถดถอยพหุนามมีค่ามาก ($MB \geq 5$) วิธีประมาณค่าพารามิเตอร์ที่ดีที่สุดคือ RLAV รองมาคือ ROLS และ OLS ตามลำดับ สำหรับทุกความแปรปรวนของความคลาดเคลื่อน (σ_e^2)

4682452126 : MAJOR STATISTICS

KEY WORD : POLYNOMIAL REGRESSION/ RIDGE LEAST ABSOLUTE VALUE/ LEAST ABSOLUTE VALUE/
RIDGE REGRESSION

SOMLAK SIRICHUANWIJIT : A COMPARISON ON PARAMETERS – ESTIMATION OF
POLYNOMIAL REGRESSION. THESIS ADVISOR : ASSOC. PROF. THEERAPORN VERATHAWORN,
Ph.D. 99 pp. ISBN 974-53-2729-8

The purpose of this research is to compare the parameter estimator in polynomial regression models by Ordinary least squares method (OLS), ridge ordinary least squares method (ROLS) and ridge least absolute value method (RLAV). The criterion of comparison is average mean squares error (AMSE) and use ratio of different average mean squares error (DIFF) to compare the efficiency of these methods. The distribution of random errors are normal distribution with mean equal to 0 and variance equal to 4, 6, 8 and 10, respectively. The sample sizes used in this study are 15, 30, 60, 120 and 240, respectively, highest degree of independent variables for dependent variable building in model (MB) are 2, 3, 4, 5 and 6, respectively. The data for this experiment are generated through the Monte Carlo simulation technique and repeating 1,000 times for each case. The results of this research are as follow :

The average mean squares error (AMSE) vary with , most to least , respectively, highest degree of independent variables for dependent variable building in model (MB), variance of random errors (σ_ϵ^2) but AMSE is converse to sample size (n).

1. In case of sample size is low ($n \leq 30$).

For all MB and all variance of random errors (σ_ϵ^2) the RLAV method is the best, the ROLS method and the OLS method are secondary method respectively.

2. In case of sample size is medium ($n = 60$).

If MB is low ($MB \leq 3$), variance of random errors is low ($\sigma_\epsilon^2 \leq 6$), the lowest AMSE is the RLAV method and the highest AMSE is the OLS method but the DIFF is less than one times. Therefore OLS method is suitable. For the other case, the RLAV method is the best, the ROLS method and the OLS method, respectively.

3. In case of sample size is quite more ($n = 120$).

If MB is low ($MB \leq 3$), for all variance of random errors (σ_ϵ^2), and If MB = 4, variance of random errors is low ($\sigma_\epsilon^2 \leq 6$), the lowest AMSE is the RLAV method and the highest AMSE is the OLS method but the DIFF is less than one times. Therefore OLS method is suitable. For the other case, the RLAV method is the best, the ROLS method and the OLS method, respectively.

4. In case of sample size is high ($n = 240$).

If MB is low to medium, for all variance of random errors, the lowest AMSE is the RLAV method and the highest AMSE is the OLS method but the DIFF is less than one times. Therefore OLS method is suitable. If MB is so high ($MB \geq 5$), for all variance of random errors (σ_ϵ^2) the RLAV method is the best, the ROLS method and the OLS method, respectively.