

การเปรียบเทียบตัวประมาณปริมาตรสำหรับการถดถอยแบบปริมาตรโดยเน้นแนวคิดแบบเบส



นางสาวเบญจวรรณ ชัยกิจ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2545

ISBN 974-17-1647-8

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

29 ธ.ค. 2549

120975673

A COMPARISON OF RIDGE ESTIMATORS FOR RIDGE REGRESSION
EMPHASIZING ON BAYESIAN CONCEPT

BENCHAWAN CHAIKIT

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Statistics

Department of Statistics

Faculty of Commerce and Accountancy


Chulalongkorn University

Academic Year 2002


ISBN 974-17-1647-8

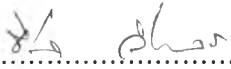
หัวข้อวิทยานิพนธ์	การเปรียบเทียบตัวประมาณไรต์สำหรับการถดถอยแบบบริดจ์ โดยเน้นแนวคิดแบบเบส
โดย	นางสาวเบญจวรรณ ชัยกิจ
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.ธีระพร วีระถาวร

คณะแพทยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้แก่นักศึกษานิพนธ์
ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต



..... คณบดีคณะแพทยศาสตร์และการบัญชี
(ผู้ช่วยศาสตราจารย์ ดร.วิรัช อภิเมธีธำรง)

คณะกรรมการสอบวิทยานิพนธ์


..... ประธานกรรมการ
(รองศาสตราจารย์ ศิริพร สาเกตทอง)


..... อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร.ธีระพร วีระถาวร)


..... กรรมการ
(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)


..... กรรมการ
(รองศาสตราจารย์ นพรัตน์ รุ่งอุทัยศิริ)

เบญจวรรณ ชัยกิจ : การเปรียบเทียบตัวประมาณริดจ์ สำหรับการถดถอยแบบริดจ์โดยเน้นแนวคิดแบบเบส์
(A COMPARISON OF RIDGE ESTIMATORS FOR RIDGE REGRESSION EMPHASIZING ON BAYESIAN
CONCEPT) อ.ที่ปรึกษา : รศ.ดร.ธีระพร วีระถาวร, 278 หน้า, ISBN 974-17-1647-8

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบตัวประมาณริดจ์สำหรับการถดถอยแบบริดจ์ ซึ่งใช้วิธีการหาค่า k จาก 3 วิธี คือ วิธีโฮเอินร์นและเคนนาร์ต (HK) วิธีค้นหาข้อมูลแบบลำดับ (SEQ) และวิธีเบส์ (BAY) โดยกระทำภายใต้เงื่อนไขของการแจกแจงของความคลาดเคลื่อนเป็นแบบปกติ ณ ระดับส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 0.05, 0.5, 1 และ 3 ขนาดตัวอย่าง 10, 30, 50 และ 100 และระดับความสัมพันธ์ของตัวแปรอิสระเท่ากับ 0.1, 0.3, 0.5, 0.7, 0.9 และ 0.99 โดยพิจารณา 2 กรณี กล่าวคือ กรณีที่ 1 ตัวแปรอิสระมีความสัมพันธ์กัน 1 กลุ่ม (จำนวนตัวแปรอิสระเท่ากับ 3 และ 5 ตัวแปร) และกรณีที่ 2 ตัวแปรอิสระมีความสัมพันธ์กัน 2 กลุ่ม (ตัวแปรอิสระในกลุ่มที่ 1 (จำนวนตัวแปรอิสระเท่ากับ 2 หรือ 3 ตัวแปร) และกลุ่มที่ 2 (จำนวนตัวแปรอิสระเท่ากับ 2 ตัวแปร)) ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากการจำลองข้อมูลโดยเทคนิคมอนติคาร์โล กระทำซ้ำ ๆ กัน 1,000 ครั้งในแต่ละสถานการณ์ที่กำหนด เพื่อเปรียบเทียบค่าเฉลี่ยของความคลาดเคลื่อนยกกำลังสอง (AMSE) ของตัวประมาณริดจ์แบบต่าง ๆ ข้างต้น ซึ่งผลการวิจัยสรุปได้ดังนี้

1) กรณีที่ 1 ตัวแปรอิสระมีความสัมพันธ์กัน 1 กลุ่ม

- จำนวนตัวแปรอิสระเท่ากับ 3 โดยทั่วไปวิธี BAY ให้ค่า AMSE น้อยที่สุด ทุกระดับความสัมพันธ์และทุกขนาดตัวอย่าง โดยที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีค่าเท่ากับ 0.05, 0.5 และ 1 ส่วนวิธี SEQ ให้ค่า AMSE น้อยที่สุดโดยทั่วไปเมื่อระดับความสัมพันธ์มีค่าสูงมาก และส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีค่าเท่ากับ 3

- จำนวนตัวแปรอิสระเท่ากับ 5 โดยทั่วไปวิธี BAY ให้ค่า AMSE น้อยที่สุด ทุกระดับความสัมพันธ์และทุกขนาดตัวอย่าง โดยที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีค่าเท่ากับ 0.05 และ 0.5 และถ้าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีค่าเท่ากับ 1 วิธี BAY ให้ค่า AMSE น้อยที่สุดเมื่อขนาดตัวอย่างมีค่าเท่ากับ 30, 50 และ 100 ส่วนวิธี SEQ ให้ค่า AMSE น้อยที่สุดโดยทั่วไปเมื่อระดับความสัมพันธ์มีค่าสูงมาก และส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีค่าเท่ากับ 3

2) กรณีที่ 2 ตัวแปรอิสระมีความสัมพันธ์กัน 2 กลุ่ม

โดยทั่วไปวิธี BAY ให้ค่า AMSE น้อยที่สุด เมื่อระดับความสัมพันธ์ของตัวแปรอิสระในแต่ละกลุ่มมีค่าต่ำ และปานกลาง และทุกขนาดตัวอย่าง โดยที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีค่าเท่ากับ 0.05 แต่ถ้าระดับความสัมพันธ์ของตัวแปรอิสระในกลุ่มที่ 1 หรือกลุ่มที่ 2 มีค่าสูง และส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีค่าเท่ากับ 0.5 และ 1 วิธี BAY ให้ค่า AMSE น้อยที่สุดเมื่อขนาดตัวอย่างมีค่าเท่ากับ 30, 50 และ 100 ส่วนวิธี SEQ ให้ค่า AMSE น้อยที่สุดโดยทั่วไปเมื่อระดับความสัมพันธ์มีค่าสูงมาก และส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีค่าเท่ากับ 3

ค่า AMSE จะแปรผันตามจำนวนตัวแปรอิสระที่มีความสัมพันธ์กัน ระดับความสัมพันธ์ของตัวแปรอิสระ และส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน โดยเรียงลำดับจากมากไปน้อย แต่ค่า AMSE จะแปรผกผันกับจำนวนตัวแปรอิสระและขนาดตัวอย่าง โดยเรียงลำดับจากมากไปน้อย

ภาควิชา สถิติ..... ลายมือชื่อนิสิต..... *เบญจวรรณ ชัยกิจ*
สาขาวิชา สถิติ..... ลายมือชื่ออาจารย์ที่ปรึกษา..... *ธีระพร วีระถาวร*
ปีการศึกษา 2545..... ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

4282287626: MAJOR STATISTICS

KEY WORD: RIDGE REGRESSION / MULTICOLLINEARITY / BAYESIAN

BENCHAWAN CHAIKIT: A COMPARISON OF RIDGE ESTIMATORS FOR RIDGE REGRESSION

EMPHASIZING ON BAYESIAN CONCEPT: THESIS ADVISOR: ASSOC.PROF. TEERAPORN

WEERATHAWORN, Ph.D., 278 pp. ISBN 974-17-1647-8

The objective of this study is to compare the ridge regression estimators with three methods for finding value of k in ridge regression are Hoerl and Kennard method, Sequential Search method and Bayesian method. The comparison was done under conditions of the distribution of random errors is Normal distribution with standard deviations are 0.05, 0.5, 1 and 3, sample sizes are 10, 30, 50 and 100, and the levels of multicollinearity are 0.1, 0.3, 0.5, 0.7, 0.9 and 0.99. The study represents in two parts, namely: the independent variables are related in one group, the numbers of independent variable are three and five, and the independent variables are related in two groups, the numbers of independent variable are two or three in first group and the numbers of independent variable are two in second group. The data of this experiment were generated through the Monte Carlo Simulation Technique. The experiment was repeated 1,000 times under each condition to calculate the average mean square error (AMSE) of ridge regression estimators. Results of this study can be summarized as follow

1. In the case of the independent variables are related in one group

- When the numbers of independent variable are three, Bayesian method has the lowest AMSE for the every levels of multicollinearity and sample sizes with standard deviations are 0.05, 0.5 and 1. Sequential Search method has the lowest AMSE for the very high level of multicollinearity with standard deviation is 3.

- When the numbers of independent variable are five, Bayesian method has the lowest AMSE for the every levels of multicollinearity and sample sizes with standard deviations are 0.05 and 0.5, and when sample sizes are 30, 50 and 100 with standard deviation is 1. Sequential Search method has the lowest AMSE for the very high level of multicollinearity with standard deviation is 3.

2. In the case of the independent variables are related in two groups

Bayesian method has the lowest AMSE for the low and medium level of multicollinearity and every sample sizes with standard deviation 0.05, and when the level of multicollinearity in the first or the second group is high and sample sizes are 30, 50 and 100 with standard deviations are 0.5 and 1. Sequential Search method has the lowest AMSE for the very high level of multicollinearity with standard deviation is 3.

The AMSE varies directly with the numbers of independent variable are related, the level of multicollinearity and the standard deviation of error, but the AMSE varies inversely with the numbers of independent variable and sample size.

DepartmentStatistics.....Student's signature.....*Benchanwan Chaikit*
Field of studyStatistics.....Advisor's signature.....*Teeraporn Weerathaworn*
Academic year.....2002.....Co-advisor's signature.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความกรุณาและความช่วยเหลืออย่างดียิ่งจาก รองศาสตราจารย์ ดร.ธีระพร วีระถาวร อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งกรุณาให้คำปรึกษา คำแนะนำ และข้อคิดเห็นต่างๆ ในการวิจัย ตลอดจนแก้ไขข้อบกพร่องต่างๆ เป็นอย่างดี ผู้วิจัยขอกราบขอบพระคุณอย่างสูง อีกทั้งผู้วิจัยขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ร้อยเอกมานพ วราภักดิ์ ที่ได้ให้คำปรึกษาด้านการเขียนโปรแกรมภาษาฟอร์แทรนและการจำลองแบบ และเนื่องจากทุนการวิจัยครั้งนี้บางส่วนได้รับมาจากทุนอุดหนุนการวิจัยของบัณฑิตวิทยาลัย จึงขอขอบคุณบัณฑิตวิทยาลัยมา ณ ที่นี้ด้วย

ท้ายนี้ผู้วิจัยขอกราบขอบพระคุณ บิดา มารดา และพี่น้อง ที่ส่งเสริมและสนับสนุนทางด้านการศึกษา พร้อมทั้งให้กำลังใจที่มั่นคงแก่ผู้วิจัยเสมอมาจนสำเร็จการศึกษา

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญตาราง.....	ช
สารบัญรูป.....	๓
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	4
1.3 สมมติฐานของการวิจัย.....	4
1.4 ขอบเขตของการวิจัย.....	4
1.5 เกณฑ์ที่ใช้ในการพิจารณา.....	6
1.6 คำจำกัดความที่ใช้ในการวิจัย.....	7
1.7 ประโยชน์ที่คาดว่าจะได้รับ.....	7
บทที่ 2 ทฤษฎีและสถิติที่เกี่ยวข้องในการวิจัย.....	8
2.1 ตัวแบบทั่วไป.....	8
2.2 การประมาณค่าสัมประสิทธิ์โดยวิธีการถดถอยแบบบริดจ์.....	10
2.3 ค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองของตัวประมาณบริดจ์.....	12
2.4 การประมาณค่าคงที่ k ในการถดถอยแบบบริดจ์.....	15
2.4.1 วิธีไฮเออร์นและเคนนาร์ด (Hoerl & Kennard method).....	15
2.4.2 วิธีค้นหาข้อมูลแบบลำดับ (Sequential search method).....	16
2.4.3 วิธีเบย์ส์ (Baysian method).....	19
บทที่ 3 วิธีดำเนินการวิจัย.....	27
3.1 แผนการวิจัย.....	27
3.2 ขั้นตอนการวิจัย.....	28
แผนผังการเขียนโปรแกรม.....	35

สารบัญ (ต่อ)

	หน้า
บทที่ 4 ผลการวิจัย.....	38
4.1 การเปรียบเทียบตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณจากวิธีการถดถอย แบบริดจ์ เมื่อตัวแปรอิสระมีความสัมพันธ์กัน 1 กลุ่ม.....	39
4.2 การเปรียบเทียบตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณจากวิธีการถดถอย แบบริดจ์ เมื่อตัวแปรอิสระมีความสัมพันธ์กัน 2 กลุ่ม.....	174
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....	247
5.1 สรุปผลการวิจัย.....	247
5.2 ข้อเสนอแนะ.....	256
รายการอ้างอิง.....	257
ภาคผนวก.....	259
ภาคผนวก ก.....	260
ประวัติผู้เขียนวิทยานิพนธ์.....	278

สารบัญตาราง

ตาราง	หน้า
3.1 แสดงลักษณะการทำงานของโปรแกรมทั้งหมดที่ใช้ในการวิจัย.....	33
4.1.1 การเปรียบเทียบค่า AMSE และ PDMSE ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 3 (X1, X2) มีความสัมพันธ์กัน และ $\sigma = 0.05$	40
4.1.2 การเปรียบเทียบค่า AMSE และ PDMSE ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 3 (X1, X2) มีความสัมพันธ์กัน และ $\sigma = 0.5$	43
4.1.3 การเปรียบเทียบค่า AMSE และ PDMSE ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 3 (X1, X2) มีความสัมพันธ์กัน และ $\sigma = 1$	46
4.1.4 การเปรียบเทียบค่า AMSE และ PDMSE ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 3 (X1, X2) มีความสัมพันธ์กัน และ $\sigma = 3$	49
4.1.5 การเปรียบเทียบค่า AMSE และ PDMSE ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 3 (X1, X2, X3) มีความสัมพันธ์กัน และ $\sigma = 0.05$	52
4.1.6 การเปรียบเทียบค่า AMSE และ PDMSE ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 3 (X1, X2, X3) มีความสัมพันธ์กัน และ $\sigma = 0.5$	55
4.1.7 การเปรียบเทียบค่า AMSE และ PDMSE ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 3 (X1, X2, X3) มีความสัมพันธ์กัน และ $\sigma = 1$	58
4.1.8 การเปรียบเทียบค่า AMSE และ PDMSE ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 3 (X1, X2, X3) มีความสัมพันธ์กัน และ $\sigma = 3$	61
4.2.1 การเปรียบเทียบค่า AMSE และ PDMSE ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 5 (X1, X2, X3) มีความสัมพันธ์กัน และ $\sigma = 0.05$	65
4.2.2 การเปรียบเทียบค่า AMSE และ PDMSE ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 5 (X1, X2, X3) มีความสัมพันธ์กัน และ $\sigma = 0.5$	68
4.2.3 การเปรียบเทียบค่า AMSE และ PDMSE ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 5 (X1, X2, X3) มีความสัมพันธ์กัน และ $\sigma = 1$	71
4.2.4 การเปรียบเทียบค่า AMSE และ PDMSE ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 5 (X1, X2, X3) มีความสัมพันธ์กัน และ $\sigma = 3$	74
4.2.5 การเปรียบเทียบค่า AMSE และ PDMSE ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 5 (X1, X2, X3, X4) มีความสัมพันธ์กัน และ $\sigma = 0.05$	77

สารบัญรูป

รูป		หน้า
1	กราฟแสดงความสัมพันธ์ระหว่างความแปรปรวน (Variance) ความเอนเอียง ยกกำลังสอง (bias^2) และค่าคงที่ k	14