

การศึกษาเปรียบเทียบการประมาณค่าสัมประสิทธิ์ความถดถอย
ระหว่างวิธีกำลังสองน้อยที่สุดและวิธีริคจ์



นางสาวศุภณิพรรณ วายุภักตร์

004123

ศูนย์วิทยทรัพยากร

จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

ภาควิชาสถิติ

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. ๒๕๒๔

A Comparative study of the estimation of regression coefficients
between the method of least squares and ridge regression

Miss Doodsaneeapan Vayuphak

A Thesis Submitted in Partial Fulfilment of the Requirement

for the Degree of Master of Commerce

Department of Statistics

Graduate School

Chulalongkorn University

1981

หัวข้อวิทยานิพนธ์ การศึกษาเปรียบเทียบการประมาณค่าสัมประสิทธิ์ความถดถอย
ระหว่างวิธีกำลังสองน้อยที่สุดและวิธีรีดจ์

โดย นางสาวดุชนิพรรณ วายุภักตร์

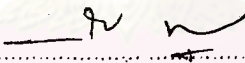
ภาควิชา สถิติ

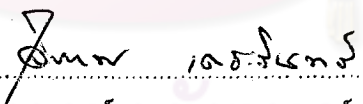
อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. สุชาดา กิระนันท์

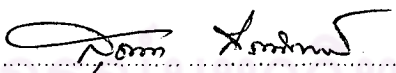
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยเป็นส่วนหนึ่ง
ของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต


..... คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร. สุประดิษฐ์ บุนนาค)

คณะกรรมการสอบวิทยานิพนธ์


..... ประธานกรรมการ
(รองศาสตราจารย์ ดร. สรชัย พิศาลบุตร)


..... กรรมการ
(รองศาสตราจารย์ สุภาพ เดชะรินทร์)


..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. สุชาดา กิระนันท์)

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

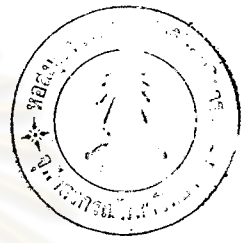
หัวข้อวิทยานิพนธ์ การศึกษาเปรียบเทียบการประมาณค่าสัมประสิทธิ์ความถดถอย
 ระหว่างวิธีกำลังสองน้อยที่สุดและวิธีรีดจ์

ชื่อนิสิต นางสาวศุภณีพรรณ วายุภักตร์

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. สุชาดา กิระนันท์

ภาควิชา สถิติ

ปีการศึกษา ๒๕๒๔



บทคัดย่อ

ในกรณีเกิดสภาพไม่เหมาะสมในตัวแปรอิสระการประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยที่สุดจะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูง ซึ่งวิธีการนี้อาจทำให้ค่าประมาณพารามิเตอร์ β ที่ได้มีคุณภาพไม่ดีเท่าที่ควร การวิจัยครั้งนี้ทำการศึกษาวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีรีดจ์ รีเกรสชัน ซึ่งเป็นวิธีที่อาจลดค่าเฉลี่ยความคลาดเคลื่อนกำลังสองให้ต่ำกว่าวิธีกำลังสองน้อยที่สุด และจะเปรียบเทียบการประมาณค่าสัมประสิทธิ์ความถดถอยระหว่างวิธีกำลังสองน้อยที่สุด และวิธีรีดจ์ รีเกรสชัน

ผลการศึกษาปรากฏว่าวิธีรีดจ์ รีเกรสชัน ให้ค่าประมาณพารามิเตอร์ β ที่เอนเอียงและให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด เมื่อเปรียบเทียบกับวิธีกำลังสองน้อยที่สุด และเมื่อเปรียบเทียบค่าประมาณของตัวแปรตามระหว่างวิธีทั้ง ๒ นี้ ปรากฏว่าไม่มีวิธีใดจะดีกว่าอีกวิธีการหนึ่ง และเมื่อนำค่าประมาณของแต่ละวิธีมาทดสอบกับค่าสังเกตของตัวแปรตามปรากฏว่าให้ค่าประมาณโดยวิธีดังกล่าวนี้ไม่แตกต่างจากค่าสังเกตของตัวแปรตาม และวิธีการของรีดจ์ รีเกรสชัน ไม่สะดวกที่จะนำมาใช้กับข้อมูลทั่วไปได้เหมือนกับวิธีกำลังสองน้อยที่สุด ทั้งนี้เพราะวิธีรีดจ์ รีเกรสชัน เป็นวิธีการที่ยุ่งยากและไม่ประหยัดเวลาในการคำนวณค่าสถิติต่าง ๆ แต่ถ้าข้อมูลที่นำมาศึกษาตัวแปรอิสระมีลักษณะเข้าใกล้ Non-orthogonal ควรใช้วิธีรีดจ์ รีเกรสชัน ประมาณค่าสัมประสิทธิ์ความถดถอยดีกว่าใช้วิธีกำลังสองน้อยที่สุด เพราะสามารถให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด เมื่อเปรียบเทียบกับวิธีกำลังสองน้อยที่สุด

Thesis Title A comparative study of the estimation of regression coefficients between the method of least squares and ridge regression

Name Miss Doodsaneeapan Vayuphak

Thesis Advisor Assistant Professor Suchada Kiranandana, Ph.D.

Department Statistics

Academic Year 1981

Abstract

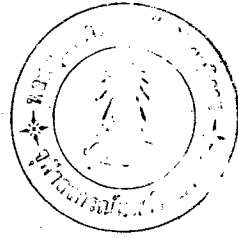
The maximum mean squares error will result when the least squares method is used to determine the regression coefficients of ill condition of the independent variables. In fact, the approximate value of parameter β , is not good enough in quality. This research shows how to evaluate the regression coefficients by the method of Ridge Regression. This method has presumably more effectiveness in minimizing the mean squares error than the method of the least squares. The study, in addition, shows the evaluation of regression coefficients between the method of least squares and Ridge Regression.

The result of this research by Ridge Regression method gives the biased estimator of parameter β and minimizes the mean squares error when compared to the least squares method. The results of these two methods have equal pros and cons when using them to examine the estimated value of both methods with the

observed variable. Given the estimate values of each method, the value of the above method is not different from the observed variable of the dependent variable. The Ridge Regression method is not suitable for general data like the least squares method. The Ridge Regression method is very difficult, complicated and time consuming for the purpose of statistic value determination. On the other hand, if the data concerning the determination of the independent variables approach to non-orthogonal, the Ridge Regression is the method for the estimation of the regression coefficient. This method is better than the least squares method because it gives the minimum mean squares error, compared to the least squares method.



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



กิติกรรมประกาศ

ผู้วิจัยขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. สุชาติ กิระนันท์
หัวหน้าภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่กรุณา
ให้คำแนะนำในการวิจัยและให้ความช่วยเหลือในการปรับปรุงแก้ไขข้อบกพร่องต่าง ๆ
อย่างดียิ่งตลอดมา ซึ่งทำให้การวิจัยครั้งนี้สำเร็จลุล่วงไปได้ดี

และขอขอบคุณทุกท่านที่กรุณาให้คำแนะนำและสนับสนุนด้วยดีมาโดยตลอด

ดุษณีพรรณ วาญักตร

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ข
รายการตารางประกอบ	ญ
รายการรูปประกอบ	ฉ
บทที่	
๑ บทนำ	๑
ที่มาของปัญหา	๑
วัตถุประสงค์ของการวิจัย	๔
ขอบเขตของการวิจัย	๔
ข้อตกลงเบื้องต้น	๔
สมมติฐานในการวิจัย	๕
ประโยชน์ที่คาดว่าจะได้รับ	๕
๒ ทฤษฎีเกี่ยวกับรีดจ์ รีเกรสชัน	๖
ตัวแบบทั่วไป	๗
คุณสมบัติของตัวประมาณค่า β ด้วยวิธีกำลังสองน้อยที่สุด	๘
การประมาณค่า β โดยวิธีรีดจ์ รีเกรสชัน	๑๒
รีดจ์ เทอเรซ	๑๔
คุณสมบัติของผลบวกของความคลาดเคลื่อนกำลังสองของ	
รีดจ์ รีเกรสชัน	๒๒
Generalized Ridge Regression	๓๐

สารบัญ (ต่อ)

บทที่	หน้า
๓ วิธีดำเนินการวิจัยและผลการวิจัย	๓๖
หลักการดำเนินการวิจัย.....	๓๖
ผลการวิจัย	๓๗
ผลการวิจัยการประมาณค่าพารามิเตอร์ β โดยวิธีรีดจ์	
รีเกรสชั่น	๔๒
การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง	
ทั้ง ๒ วิธี	๔๖
ความสัมพันธ์ระหว่างตัวประมาณ $\hat{\beta}^*$ และ K	๔๘
ความสัมพันธ์ระหว่างค่า $E[L_1^2(0)]$ และ $E[L_1^2(K)]$	
เมื่อ $K \in (0, \infty]$	๖๗
๔ สรุปผลและข้อเสนอแนะ	๗๑
บรรณานุกรม	๗๔
ภาคผนวก	๗๖
ประวัติผู้เขียน	๙๒

รายการตารางประกอบ

ตารางที่		หน้า
๑	แสดงการเปรียบเทียบการประมาณค่า β ระหว่างวิธีกำลังสองน้อยที่สุดและวิธีรีดจ์ รีเกรสชัน	๓๕
๒	แสดงการประมาณค่าพารามิเตอร์ β ด้วยวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ ของข้อมูลชุดที่ ๑	๓๘
๓	แสดงการประมาณค่าพารามิเตอร์ β โดยวิธี Step-Wise Multiple Regression ของข้อมูลชุดที่ ๑	๓๙
๔	แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีรีดจ์ รีเกรสชันของข้อมูลชุดที่ ๑	๔๐
๕	แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ ของข้อมูลชุดที่ ๒	๔๑
๖	แสดงการประมาณค่าพารามิเตอร์ β โดยวิธี Step-Wise Multiple Regression ของข้อมูลชุดที่ ๒	๔๒
๗	แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีรีดจ์ รีเกรสชันของข้อมูลชุดที่ ๒	๔๓
๘	แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีที่มีตัวแปรอิสระทุกตัวในตัวแบบ ของข้อมูลชุดที่ ๓	๔๔
๙	แสดงการประมาณค่าพารามิเตอร์ β โดยวิธี Step-Wise Multiple Regression ของข้อมูลชุดที่ ๓	๔๕
๑๐	แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีรีดจ์ รีเกรสชันของข้อมูลชุดที่ ๓	๔๖
๑๑	แสดงการประมาณค่าพารามิเตอร์ β โดยวิธี ที่มีตัวแปรอิสระทุกตัวในตัวแบบ ของข้อมูลชุดที่ ๔	๔๗
๑๒	แสดงการประมาณค่าพารามิเตอร์ β โดยวิธี Step-Wise Multiple Regression ของข้อมูลชุดที่ ๔	๔๘
๑๓	แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีรีดจ์ รีเกรสชันของข้อมูลชุดที่ ๔	๔๙

รายการตารางประกอบ (ต่อ)

ตารางที่		หน้า
๑๔	แสดงการประมาณค่าพารามิเตอร์ β โดยวิธีที่มีตัวแปรอิสระทุกตัว ในตัวแบบ ของข้อมูลชุดที่ ๕	๕๐
๑๕	แสดงการประมาณค่าพารามิเตอร์ β โดยวิธี Step-Wise Multiple Regression ของข้อมูลชุดที่ ๕	๕๑
๑๖	แสดงการประมาณค่าพารามิเตอร์ โดยวิธีริคจ์ รีเกรสชันของข้อมูลชุดที่ ๕	๕๒
๑๗	แสดงการเปรียบเทียบ $E[L_1^2(0)]$ และ $E[L_1^2(k)]$ ระหว่างวิธีกำลังสองน้อยที่สุดและวิธีริคจ์ รีเกรสชัน ตามลำดับ และแสดงเปอร์เซ็นต์ของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ ลดลง เมื่อเปรียบเทียบระหว่าง ๒ วิธี	๕๗
๑๘	แสดงการเปรียบเทียบค่าตัวประมาณพารามิเตอร์ เมื่อคำนวณ ด้วยวิธีทั้ง ๒	๖๐
๑๙	แสดงค่าประมาณของตัวแปรตามที่เกิดจากการประมาณค่าระหว่าง วิธีกำลังสองน้อยที่สุดและวิธีริคจ์ รีเกรสชัน	๖๓
๒๐	เปรียบเทียบการทดสอบสมมติฐานของค่าประมาณของตัวแปรตาม ที่ประมาณค่าจากวิธีกำลังสองน้อยที่สุดและวิธีริคจ์ รีเกรสชัน กับ ค่าสังเกตของตัวแปรตามสำหรับข้อมูลชุดที่ ๑ - ๕	๖๖
๒๑	แสดงคุณสมบัติของการประมาณค่า ด้วยวิธีกำลังสองน้อยที่สุด และวิธีริคจ์ รีเกรสชัน เมื่อนำไปใช้กับข้อมูลทั่วไปในทางปฏิบัติ	๖๙

รายการรูปประกอบ

รูปที่		หน้า
๑	แสดงการเปรียบเทียบการกระจายตัวประมาณค่าพารามิเตอร์ β เมื่อประมาณค่าโดยวิธีกำลังสองน้อยที่สุดและวิธีริตจ์ ริเกรสชั่น	๑๔
๒	แสดงความสัมพันธ์ระหว่าง β^* และ K เมื่อ $K \in (0, 1]$	๑๗
๓	แสดงการ Minimize ค่า β^* เมื่อ $K > 0$	๑๘
๔	แสดงการกระจายของตัวประมาณของพารามิเตอร์ทั้ง ๒ วิธี	๒๔
๕	แสดงกราฟของ $E[L_1^2(K)]$	๒๕
๖	แสดงความสัมพันธ์ระหว่าง $E[L_1^2(K)]_i$ กับค่า K ของข้อมูลทั้งหมด n ชุด เมื่อ $i =$ จำนวนชุดของข้อมูลที่มีค่าเท่ากับ ๑, ๒, ๓, ๔, ๕	๔๓
๗	แสดงความสัมพันธ์ระหว่างค่าความแปรปรวนของ β^* กับค่า K เมื่อ $K \in (0, 1]$	๕๔
๘	แสดงความสัมพันธ์ระหว่างความเอนเอียงกำลังสองกับค่า K	๕๕
๙	แสดงความสัมพันธ์ระหว่าง β^* กับ K เมื่อ $K \in (0, 1]$ ของข้อมูลชุดที่ ๑	๕๘
๑๐	แสดงการเปลี่ยนแปลงค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของข้อมูลชุดที่ ๔	๖๘