

การเปรียบเทียบประสิทธิภาพของวิธีที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุ

นายปรุ ม กลั่นน้ำทิพย์



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาคตามหลักสูตรปริญญาศึกษาศาสตรมหาบัณฑิต

ภาควิชาสถิติ

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2535

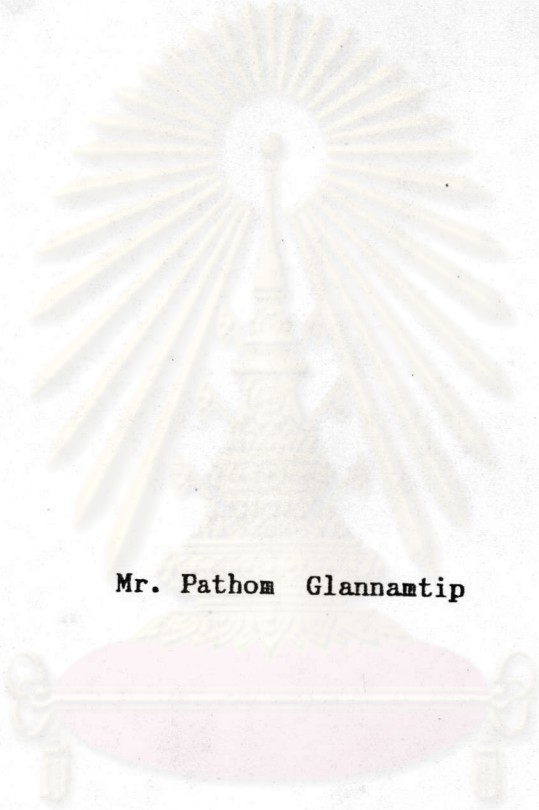
ISBN 974-581-562-4

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

018558

117191069

A COMPARISON ON THE EFFICIENCY OF METHODS
FOR ESTIMATING MULTIPLE REGRESSION COEFFICIENT



Mr. Pathom Glannantip

ศูนย์วิทยทรัพยากร
กองส่งเสริมมหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science

Department of Statistics

Graduate School

Chulalongkon University

1992

ISBN 974-581-562-4

หัวข้อวิทยานิพนธ์ : การเปรียบเทียบประสิทธิภาพของวิธีที่ใช้ในการประมาณค่าสัมประสิทธิ์
การถดถอยพหุ
ชื่อ นิสิต : นายปฐม กลั่นน้ำทิพย์
ภาควิชา : สถิติ
อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.สรชัย พิศาลบุตร

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัยอนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาโทบัณฑิต

.....*ดร.ถาวร วัชรภักย์*.....คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร.ถาวร วัชรภักย์)

คณะกรรมการสอบวิทยานิพนธ์

.....*ดร.ผกาดี ศิริรังษี*.....ประธานกรรมการ
(รองศาสตราจารย์ ผกาดี ศิริรังษี)

.....*ดร.สรชัย พิศาลบุตร*.....อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร.สรชัย พิศาลบุตร)

.....*ดร.นพรัตน์ รุ่งอุทัยศิริ*.....กรรมการ
(รองศาสตราจารย์ นพรัตน์ รุ่งอุทัยศิริ)

.....*ร.อ.มานพ วรภักดิ์*.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ร.อ.มานพ วรภักดิ์)



พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว

ปฐม กลิ่นน้ำทิพย์ : การเปรียบเทียบประสิทธิภาพของวิธีที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุ (A COMPARISON ON THE EFFICIENCY OF METHODS FOR ESTIMATING MULTIPLE REGRESSION COEFFICIENT) อาจารย์ที่ปรึกษา : รศ. ดร. สรชัย พิศาลบุตร 109 หน้า. ISBN 974-581-562-4

การวิจัยนี้มีวัตถุประสงค์ที่จะศึกษาวิธีที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุ เมื่อความผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ การแจกแจงแบบเบ้และการแจกแจงแบบดับเบิลเอกซ์โปเนนเชียล โดยเปรียบเทียบวิธีการข้างล่างสองน้อยที่สุด วิธีบูตสเตรปและวิธีตัวประมาณชนิด M ซึ่งใช้เกณฑ์ความแกร่งของ Huber เกณฑ์การเปรียบเทียบที่ใช้ คือ ค่าเฉลี่ยความผิดพลาดกำลังสอง (MSE) การแจกแจงของค่าความผิดพลาดที่ศึกษาคือ การแจกแจงหางยาวกว่าการแจกแจงปกติ ใช้การแจกแจงแบบปกติปลอมปนที่มีสเกลแฟคเตอร์ 3 และ 10 ใช้เปอร์เซ็นต์การปลอมปนที่ 1% 5% 10% 25% เป็นกรณีศึกษา การแจกแจงแบบเบ้ใช้การแจกแจงแบบลอกลอนอร์มอลที่ $\mu = 0$ และ $\sigma = 1$ และการแจกแจงแบบแกมมาที่มีค่าสัมประสิทธิ์ความแปรปรวน (C.V.) 58% 70% และ 100% เป็นกรณีศึกษา และการแจกแจงแบบดับเบิลเอกซ์โปเนนเชียล ที่ $\alpha = 0$, $\beta = 5$, 10 โดยใช้จำนวนตัวแปรอิสระและขนาดตัวอย่างดังนี้คือ จำนวนตัวแปรอิสระ = 2 เมื่อขนาดตัวอย่าง = 4 จำนวนตัวแปรอิสระ = 3 เมื่อขนาดตัวอย่าง = 5, 10 และ 20 จำนวนตัวแปรอิสระ = 5 เมื่อขนาดตัวอย่าง = 50 และ 100 ในการทดลองนี้ทำการจำลองเหตุการณ์ต่าง ๆ ด้วยเครื่องคอมพิวเตอร์ โดยเทคนิคมอนติคาร์โลกระทำซ้ำ 200 ครั้ง แต่ในกรณีที่ข้อมูลมีการแจกแจงแบบเบ้ จะกระทำซ้ำ 100 ครั้ง และการสุ่มตัวอย่างแบบโบส์คินที่ใช้สำหรับวิธีบูตสเตรป กระทำซ้ำ 100 ครั้ง ผลการศึกษาพบว่า

1. กรณีที่ค่าความผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ การแจกแจงของค่าความผิดพลาดที่มีค่าผิดปกติจะเกิดขึ้นในรูปแบบต่าง ๆ ซึ่งกำหนดโดยสเกลแฟคเตอร์ และเปอร์เซ็นต์การปลอมปน พบว่า วิธีตัวประมาณชนิด M ที่ใช้เกณฑ์ความแกร่งของ Huber เป็นวิธีที่ดีที่สุด โดยมีปัจจัยที่มีอิทธิพลเรียงลำดับจากมากไปน้อยคือ สเกลแฟคเตอร์ที่ 10 และ 3 เปอร์เซ็นต์การปลอมปน 25% 10% 5% และ 1%

2. กรณีที่ค่าความผิดพลาดมีการแจกแจงแบบเบ้ ได้แก่การแจกแจงแบบลอกลอนอร์มอลและการแจกแจงแบบแกมมา พบว่าวิธีการข้างล่างสองน้อยที่สุด วิธีตัวประมาณชนิด M และ วิธีบูตสเตรปให้ค่าการประมาณสัมประสิทธิ์การถดถอยพหุได้ใกล้เคียงกัน เมื่อใช้เทคนิคการแปลงข้อมูลที่อยู่ในรูปยกกำลังของ Box และ Cox ในการแปลงข้อมูลให้เข้าสู่ภาวะปกติก่อนการประมาณค่า

3. กรณีที่ค่าความผิดพลาดมีการแจกแจงแบบดับเบิลเอกซ์โปเนนเชียล พบว่า วิธีตัวประมาณชนิด M เป็นวิธีที่ดีที่สุดเมื่อตัวอย่างมีขนาดเล็ก แต่เมื่อขนาดตัวอย่างมีขนาดใหญ่ วิธีการข้างล่างสองน้อยที่สุด เป็นวิธีที่ดีที่สุด

ภาควิชา สถิติ.....
สาขาวิชา สถิติ.....
ปีการศึกษา 2534.....

ลายมือชื่อนิสิต
ลายมือชื่ออาจารย์ที่ปรึกษา
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

C022995 : MAJOR STATISTICS

KEY WORD : MULTIPLES REGRESSION COEFFICIENT/LONGTAIL DISTRIBUTION/SKEWED DISTRIBUTION/DOUBLE EXPONENTIAL DISTRIBUTION

PATHOM GLANNAMTIP : A COMPARISON ON THE EFFICIENCY OF METHODS FOR ESTIMATING MULTIPLE REGRESSION COEFFICIENT. THESIS ADVISOR : ASSO. PROF. SORACHAI PHISARNBUTRA, Ph.D. 109 pp. ISBN 974-581-562-4

The objective of this thesis is to study the estimation of multiple regression coefficients when residuals have long tailed distribution than normal distribution, skewed distribution and double exponential distribution by comparing the Ordinary least square method, Bootstrap method and M-estimator using Huber' robust criteria. They are compared by using mean square errors. The distribution used in case of longer tailed distribution are Scale Contaminated normal distribution with scale factors of 3 and 10 and percent contaminations of 1, 5, 10 and 25. Other case of skewed distribution are Lognormal distribution with $\mu = 0$ and $\sigma^2 = 1$, and Gamma distribution with coefficient of variance of 58%, 70% and 100%. Double exponential distribution with $\alpha = 0$, $\beta = 5, 10$. This study used the number of independent variables of 2 for the sample size of 4, the number of independent variables of 3 for the sample size of 5, 10 and 20 and the number of independent variable of 5 for the sample size of 50 and 100. Data is obtained through simulation using Monte Carlo technique, and repeating 200 times for the longer tailed distribution than normal distribution, repeating 100 times for skewed distribution, and resampling 100 times for Bootstrap Method

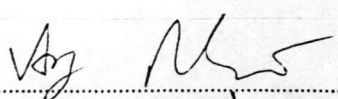
The results of this study are as follow :

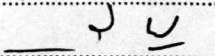
1. In case, residuals have longer tailed distribution than normal distribution. Shape of the residuals distributions are determined by percent contamination and scale factor for contaminate normal distribution. The M-estimator using Huber' robust criteria gives better results than the Ordinary least square and Bootstrap method when scale factor and percent contamination increases.
2. In case, residuals have skewed distribution, such as, Lognormal and Gamma distribution. The result is that the Ordinary least square method, Bootstrap and M-estimator method give closed results when using Box and Cox's Power Transformation.
3. In case, residuals have double exponential distribution. The result is that the M-estimator gives the best efficiency when small sample size and Ordinary least square gives the best when the samples size are large.


ภาควิชา สถิติ

สาขาวิชา สถิติ

ปีการศึกษา 2534

ลายมือชื่อนิติ 

ลายมือชื่ออาจารย์ที่ปรึกษา 

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม 

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาของ รศ. ดร. สรชัย พิศาลบุตร รองคณบดีฝ่ายวิจัย คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ให้คำแนะนำและให้คำปรึกษา ตลอดจนแก้ไขข้อบกพร่องต่าง ๆ เป็นอย่างดียิ่ง ตลอดมา ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้ และขอกราบขอบพระคุณ รศ. ผกาวดี ศิริรังษี รศ. นพรัตน์ รุ่งอุทัยศิริ และ ผศ. ร.อ.มานพ วรภักดิ์ ที่ให้ข้อเสนอแนะ และการแก้ไขปรับปรุงข้อบกพร่องต่างๆ

ขอขอบคุณ คุณอุคม สาระรัตน์ คุณลักขณา สาระรัตน์ ที่ช่วยจัดพิมพ์วิทยานิพนธ์ และเจ้าหน้าที่ศูนย์คอมพิวเตอร์ มหาวิทยาลัยขอนแก่นทุกท่านที่อำนวยความสะดวก และช่วยเหลือ เป็นอย่างดีในการใช้เครื่องคอมพิวเตอร์ สำหรับการวิจัยนี้

สุดท้ายนี้ ขอขอบพระคุณ คุณพ่อ คุณแม่ คุณสารภี กลั่นน้ำทิพย์ และ เด็กหญิงนพร กลั่นน้ำทิพย์ ที่ส่งเสริม สนับสนุนการเรียนของผู้วิจัย ทั้งด้านกำลังใจและกำลังใจทรัพย์ตลอดมา และขอกราบขอบพระคุณอาจารย์ทุกท่านที่ได้ประสิทธิ์ประสาทวิชาความรู้ แก่ผู้วิจัยมาโดยตลอด

ปรม กลั่นน้ำทิพย์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญตาราง.....	ง
บทที่ 1 บทนำ	
1.1 ความสำคัญและความเป็นมาของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	3
1.3 ข้อยกเว้นเบื้องต้น.....	3
1.4 ขอบเขตการวิจัย.....	4
1.4.1 ลักษณะการแจกแจงของความผิดพลาดที่ศึกษา.....	4
1.4.2 จำนวนตัวแปรอิสระและขนาดตัวอย่าง.....	7
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	8
1.6 คำจำกัดความ.....	8
บทที่ 2 ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง	
2.1 ข้อสมมติทั่วไปของตัวแบบเชิงเส้น.....	12
2.2 วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุ.....	14
2.2.1 วิธีกำลังสองน้อยที่สุดแบบธรรมดา (Ordinary least square method).....	14
2.2.2 วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (General least square method).....	17
2.2.3 วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (Weighted least square method).....	18
2.2.4 วิธีของตัวประมาณชนิด M (M-estimator method).....	19
2.2.5 การประมาณค่าพารามิเตอร์โดยวิธีบูตสเตรป.....	21

สารบัญ (ต่อ)

	หน้า
2.3 การแปลงข้อมูลที่อยู่ในรูปยกกำลัง (Power Transformation)...	26
2.4 ผลงานวิจัยที่เกี่ยวข้อง.....	29
บทที่ 3 วิธีดำเนินการวิจัย	
3.1 วิธีมอนติคาร์โล (Monte Carlo method).....	32
3.2 แผนดำเนินการวิจัย.....	33
3.3 ขั้นตอนของการดำเนินงานวิจัย.....	34
3.4 ผังงานการดำเนินงานวิจัย.....	35
บทที่ 4 ผลการวิจัย	
4.1 การประมาณค่าสัมประสิทธิ์การถดถอยพหุเมื่อความผิดพลาด มีการแจกแจงแบบหางยาวกว่าการแจกแจงปกติ.....	45
4.2 การประมาณค่าสัมประสิทธิ์การถดถอยพหุเมื่อความผิดพลาด มีการแจกแจงแบบเบ้.....	49
4.3 การประมาณค่าสัมประสิทธิ์การถดถอยพหุเมื่อความผิดพลาด มีการแจกแจงแบบดับเบิลเอ็กซ์โปเนนเชียล.....	54
บทที่ 5 สรุปผลการวิจัยและการอภิปรายผล	
5.1 ผลสรุปการเปรียบเทียบวิธีการประมาณทั้ง 3 วิธี.....	55
5.2 การอภิปรายผล.....	57
5.3 ข้อเสนอแนะ.....	59
บรรณานุกรม.....	63
ภาคผนวก ก.....	67
ภาคผนวก ข.....	82
ประวัติผู้วิจัย.....	

สารบัญตาราง

ตารางที่	หน้า
3.1	แสดงค่าสเกลแฟคเตอร์และเปอร์เซ็นต์การปลอมปนทั้งหมด ที่ใช้ในการวิจัย..... 33
3.2	แสดงขนาดตัวอย่างและจำนวนตัวแปรอิสระทั้งหมด..... 34
4.1.1	แสดงค่าเฉลี่ยความผิดพลาดกำลังสองในการประมาณค่า สัมประสิทธิ์การถดถอยหาคู่ด้วยวิธี OLS MAD และ Boot โดยที่ความผิดพลาดมีการแจกแจงแบบปกติปลอมปน..... 46
4.2.1	แสดงค่าเฉลี่ยความผิดพลาดกำลังสองในการประมาณค่า สัมประสิทธิ์การถดถอยหาคู่ด้วยวิธี OLS MAD และ Boot โดยที่ความผิดพลาดมีการแจกแจงแบบลอกนอร์มอล..... 50
4.2.2	แสดงค่าเฉลี่ยความผิดพลาดกำลังสองในการประมาณค่า สัมประสิทธิ์การถดถอยหาคู่ด้วยวิธี OLS MAD และ Boot โดยที่ความผิดพลาดมีการแจกแจงแบบแกมมา..... 52
4.3.1	แสดงค่าเฉลี่ยความผิดพลาดกำลังสองในการประมาณค่า สัมประสิทธิ์การถดถอยหาคู่ด้วยวิธี OLS MAD และ Boot โดยที่ความผิดพลาดมีการแจกแจงแบบคัมเบิ้ลเอ็กซ์โปเนนเชียล..... 54