การประมาณสัมประสิทธิ์การถดถอยพหุ เมื่อความผิดพลาดมีการแจกแจงแบบเบ้และ มีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ

นางสาวปราณี รัตนัง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
ภาควิชาสถิติ
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
พ.ศ. 2531

ISBN 974-569-223-9

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

013917

I16486 16X

THE ESTIMATION OF MULTIPLE REGRESSION COEFFICIENTS WHEN RESIDUALS HAVE SKEWED DISTRIBUTION AND LONGER TAILED DISTRIBUTION THAN NORMAL DISTRIBUTION

Miss Pranee Rattanung

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science
Department of Statistics
Graduate School
Chulalongkorn University
1988

ISBN 974-569-223-9

หัวข้อวิทยานิพนธ์ การประมาณสัมประสิทธิ์การถดถอยพหุ เมื่อความผิดพลาดมีการแจกแจงแบบแย้ และมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ

โดย นางสาวปราณี รัตนัง

ภาควิชา สถิติ

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.ธีระพร วีระถาวร

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร.ถาวร วัชราภัย)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(รองศาสตราจารย์ ดร.สุชาดา กีระนันทน์)

..... อาจารย์ที่ปรึกษา
(ผู้ช่วยศาสตราจารย์ ดร.ธีระพร วีระถาวร)

..... กรรมการ
(รองศาสตราจารย์ ดร.สรชัย พิศาลบุตร)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ผกาวดี ดิริรังษี)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ปราณี รัตนัง : การประมาณสัมประสิทธิ์การถดถอยพหุ เมื่อความผิดพลาดมีการแจกแจงแบบเบ้และ มีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ (THE ESTIMATION OF MULTIPLE REGRESSION COEFFICIENTS WHEN RESIDUALS HAVE SKEWED DISTRIBUTION AND LONGER TAILED DISTRIBUTION THAN NORMAL DISTRIBUTION) อ. ที่ปรึกษา : ผศ. ดร. ธีระพร วีระถาวร, 129 หน้า.

การวิจัยครั้งนี้มีวัตถุประสงค์ที่จะศึกษาการประมาณสัมประสิทธิ์การถดถอยพหุ เมื่อความผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติและการแจกแจงแบบเบ้ โดยเปรียบเทียบวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay เกณฑ์การเปรียบเทียบที่ใช้ คือ ค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสอง และค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสอง การแจกแจงของค่าผิดพลาดที่ศึกษาคือ การแจกแจงแบบหางยาวกว่าแจกแจงแบบปกติ ซึ่งจะใช้การแจกแจงปกติปลอมปนที่มีสเกลแฟคเตอร์เท่ากับ 3 และ 10 และใช้เปอร์เซ็นต์การปลอมปนเท่ากับ 1 5 10 และ 25 และการแจกแจงที่ซึ่งมีระดับความเป็นอิสระเท่ากับ 4 และ 8 ส่วนกรณีการแจกแจงแบบเบ้จะใช้การแจกแจงลอกนอร์มอล $\mu = 0$ และ $\sigma^2 = 1$ การแจกแจงแกมมาซึ่งมีค่าสัมประสิทธิ์ของความแปรปรวน 32% 58% 70% และ 100% การแจกแจงไวบูลล์ซึ่งมีค่าสัมประสิทธิ์ของความแปรปรวน 17% 35% 52% และ 100% โดยกำหนดจำนวนตัวแปรอิสระเท่ากับ 3 สำหรับขนาดตัวอย่างเท่ากับ 20 และจำนวนตัวแปรอิสระเท่ากับ 5 และ 10 สำหรับขนาดตัวอย่าง 50 100 และ 150 สำหรับข้อมูลลักษณะต่าง ๆ ที่ใช้ในการทดลองนี้ จำลองด้วยโปรแกรมคอมพิวเตอร์โดยใช้เทคนิคมอนติคาร์โล และกระทำซ้ำกัน 200 ครั้ง ในแต่ละกรณี ยกเว้นบางกรณีของการแจกแจงแกมมาและไวบูลล์ ณ ค่าสัมประสิทธิ์ความแปรปรวน 32% และ 17% ซึ่งกระทำซ้ำ 100 ครั้ง

ผลจากการศึกษาเปรียบเทียบค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสองและค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสอง สรุปผลได้ดังนี้

1. กรณีที่ค่าผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ การแจกแจงของค่าผิดพลาดที่มีค่าผิดปกติจะเกิดขึ้นในรูปแบบต่าง ๆ ซึ่งกำหนดโดยสเกลแฟคเตอร์และเปอร์เซ็นต์ของการปลอมปนสำหรับการแจกแจงแบบปกติปลอมปน และระดับความเป็นอิสระสำหรับการแจกแจงแบบที จะได้ว่า สเกลแฟคเตอร์ เปอร์เซ็นต์ของการปลอมปน จะมีอิทธิพลจากมากไปน้อยที่ทำให้วิธี M-estimator ซึ่งใช้เกณฑ์ของ Ramsay ดีกว่าวิธีกำลังสองน้อยที่สุด ระดับความเป็นอิสระก็มีอิทธิพลที่ทำให้ได้ข้อสรุปเช่นเดียวกัน

2. กรณีที่ค่าผิดพลาดมีการแจกแจงแบบเบ้ ได้แก่ การแจกแจงที่มีรูปแบบต่าง ๆ เช่น ลอกนอร์มอล แกมมาและไวบูลล์ จะได้ว่าวิธีกำลังสองน้อยที่สุดและวิธี M-estimator สามารถใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุได้ใกล้เคียงกัน เมื่อใช้เทคนิคการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox ในการแปลงข้อมูลให้เข้าสู่ภาวะปกติ

ภาควิชา สถิติ
สาขาวิชา สถิติ
ปีการศึกษา 2530

ลายมือชื่อนิสิต .....
ลายมือชื่ออาจารย์ที่ปรึกษา .....

PRANEE RATTANUNG : THE ESTIMATION OF MULTIPLE REGRESSION COEFFICIENTS WHEN RESIDUALS HAVE SKEWED DISTRIBUTION AND LONGER TAILED DISTRIBUTION THAN NORMAL DISTRIBUTION. THESIS ADVISOR : ASSIS. PROF. THEERAPORN VERATHAWORN, Ph.D., 129 PP.

The objective of this thesis is to study the estimation of multiple regression coefficients when residuals have longer tailed distribution than normal distribution and skewed distribution by comparing the ordinary least square method with M-estimator using Ramsay's robust criteria. They are compared by using the average of relative value of mean square error and the average of absolute value of differrent ratio of mean square errors. The distributions 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, and t distribution with degree of freedom of 4 and 8. Other cases of skewed distribution are Lognormal distribution with $\mu = 0$ and $\sigma^2 = 1$, and Gamma distribution with coefficient of variations of 32%, 58%, 70%, and 100%. This study used the number of independent variables of 3 for a sample size of 20 and the number of independent variables of 5 and 10 for a sample size of 50, 100 and 150. Data is obtained through simulation using Monte Carlo Technique, repeated 200 times, except for some cases of Gamma distribution and Weibull distribution with coefficient of variation of 32% and 17%, respectively, repeated 100 times.

The results of this study are as follow :

1. In case, residuals have longer tailed distribution than normal distribution. Shapes of the residuals distributions are determined by scale factor and percent contamination for scale-contaminated normal distribution and determined by the degree of freedom for t distribution. The M-estimator using Ramsay's robust criteria gives better results than the ordinary least square when scale factor and percent contamination increases, and scale factor give more better results than percent contamination. The same result is obtained in t distribution which the degree of freedom is the main factor.

2. In case, residuals have skewed distribution, such as, Lognormal, Gamma and Weibull distribution. The result is that the ordinary least square method and M-estimator method give closed results when using Box and Cox's Power Transformation.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา สถิติ ลายมือชื่อนิสิต .....

สาขาวิชา สถิติ

ปีการศึกษา 2530 ลายมือชื่ออาจารย์ที่ปรึกษา .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาของผู้ช่วยศาสตราจารย์ ดร. ธีระพร วีระถาวร อาจารย์ประจำภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ให้คำแนะนำ ปรึกษา ตลอดจนแก้ไขข้อบกพร่องต่าง ๆ เป็นอย่างดีมาโดยตลอด ซึ่งผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

นอกจากนี้ ผู้วิจัยขอกราบขอบพระคุณรองศาสตราจารย์ ดร.สรชัย พิศาลบุตร และอาจารย์ ดร.สุพล ดุรงค์วัฒนา ที่ให้คำแนะนำเพิ่มเติมเสมอมา

ขอบพระคุณผู้ที่ให้ความช่วยเหลือตลอดจนคำแนะนำต่าง ๆ ไว้ ณ โอกาสนี้ด้วย ทุก ๆ ท่าน

สุดท้ายนี้ขอกราบขอบพระคุณ คุณพ่อ คุณแม่ และพี่ ๆ ที่ส่งเสริมและสนับสนุนการเรียนของผู้วิจัยตลอดมา และขอกราบขอบพระคุณท่านอาจารย์ทุกท่าน ที่ได้ประสิทธิประสาทความรู้แก่ผู้วิจัยมาโดยตลอด

ปราณี รัตนัง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญตาราง ..... ฌ

สารบัญรูป ..... ฎ

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของการวิจัย 4

1.3 ข้อตกลงเบื้องต้น 5

1.4 ขอบเขตของการวิจัย 5

1.5 ประโยชน์ที่คาดว่าจะได้รับ 9

1.6 วิธีดำเนินการวิจัย 10

1.7 คำศัพท์ต่าง ๆ ที่ใช้ในการวิจัย 10

2 ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง 13

2.1 ข้อสมมติทั่วไปของตัวแบบการถดถอยเชิงเส้น 14

2.2 วิธีกำลังสองน้อยที่สุด (Least square method) 16

2.2.1 วิธีกำลังสองน้อยที่สุดแบบธรรมดา

(Ordinary least square method) 16

2.2.2 วิธีกำลังสองน้อยที่สุดแบบทั่วไป

(Generalized least square method) 18

2.2.3 วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก

(Weighted least square method) 18

สารบัญ (ต่อ)

บทที่		หน้า
	2.3 วิธีที่ใช้ตัวประมาณชนิด M (M-estimator method)	19
	2.4 การแปลงข้อมูลที่ใช้การยกกำลัง (Power Transformation)	26
	2.5 ผลงานวิจัยที่เกี่ยวข้อง	29
3	วิธีดำเนินการวิจัย	32
	3.1 วิธีมอนติคาร์โล (Monte Carlo method)	32
	3.2 แผนการทดลอง	33
	3.3 ขั้นตอนในการวิจัย	34
4	ผลการวิจัย	47
	4.1 การทดสอบความเป็นอิสระซึ่งกันและกันของตัวแปรอิสระ หรือการทดสอบ Multicollinearity	48
	4.2 การประมาณสัมประสิทธิ์การถดถอยพหุเมื่อความผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ	50
	4.3 การประมาณสัมประสิทธิ์การถดถอยพหุเมื่อความผิดพลาดมีการแจกแจงแบบเบ้	67
	4.4 สรุปผลการวิเคราะห์ของการประมาณสัมประสิทธิ์การถดถอยพหุระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator	79
5	สรุปผลการวิเคราะห์และการอภิปรายผล	85
	5.1 ผลสรุปการเปรียบเทียบวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator	86
	5.2 การอภิปรายผล	88
	5.3 ข้อเสนอแนะ	89
บรรณานุกรม		91
ภาคผนวก		93
ประวัติผู้เขียน		129

สารบัญตาราง

ตารางที่		หน้า
2.1	ฟังก์ชันเกณฑ์ของความแกร่งชนิดต่าง ๆ	22
3.1	แสดงค่าสเกลแฟคเตอร์และเปอร์เซนต์การปลอมปนทั้งหมดที่ใช้ในการวิจัย	33
3.2	แสดงขนาดตัวอย่างและจำนวนตัวแปรอิสระทั้งหมดที่ใช้ในการวิจัย	34
3.3	แสดงลักษณะการทำงานของโปรแกรมทั้งหมดที่ใช้ในการวิจัย	44
4.1	แสดงค่า F ที่คำนวณได้ในการทดสอบ Multicollinearity	49
4.2	ตารางค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสองระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator เมื่อค่าผิดพลาดมีการแจกแจงเป็นแบบปกติปลอมปน ซึ่งใช้สเกลแฟคเตอร์เป็น 3 โดยจำแนกตามจำนวนตัวแปรอิสระ ขนาดตัวอย่าง เปอร์เซนต์ของการปลอมปน และวิธีที่ให้ค่า MSE น้อยกว่า	52
4.3	ตารางค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสองระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator เมื่อค่าผิดพลาดมีการแจกแจงเป็นแบบปกติปลอมปน ซึ่งใช้สเกลแฟคเตอร์เป็น 3 โดยจำแนกตามจำนวนตัวแปรอิสระ ขนาดตัวอย่าง เปอร์เซนต์ ของการปลอมปน และวิธีที่ให้ค่า MSE น้อยกว่า	53
4.4	ตารางค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสองระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator เมื่อค่าผิดพลาดมีการแจกแจงเป็นแบบปกติปลอมปน ซึ่งใช้สเกลแฟคเตอร์เป็น 10 โดยจำแนกตามจำนวนตัวแปรอิสระ ขนาดตัวอย่าง เปอร์เซนต์ของการปลอมปน และวิธีที่ให้ค่า MSE น้อยกว่า	57

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4.5 ตารางค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสองระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator เมื่อค่าผิดพลาดมีการแจกแจงเป็นแบบปกติปลอมปน ซึ่งใช้สเกลแฟคเตอร์เป็น 10 โดยจำแนกตามจำนวนตัวแปรอิสระ ขนาดตัวอย่าง เปอร์เซนต์ของการปลอมปน และวิธีที่ให้ค่า MSE น้อยกว่า 58

4.6 ตารางค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสองระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator เมื่อค่าผิดพลาดมีการแจกแจงเป็นแบบที โดยจำแนกตามจำนวนตัวแปรอิสระ ขนาดตัวอย่าง ระดับของความเป็นอิสระ (degree of freedom) และวิธีที่ให้ค่า MSE น้อยกว่า 63

4.7 ตารางค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสองระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator เมื่อค่าผิดพลาดมีการแจกแจงแบบที โดยจำแนกตามจำนวนตัวแปรอิสระ ขนาดตัวอย่าง ระดับของความเป็นอิสระ และวิธีที่ให้ค่า MSE น้อยกว่า 64

4.8 ตารางค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสองระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator เมื่อค่าผิดพลาดมีการแจกแจงเป็นแบบลอกนอร์มอล ($\mu=0$, $\sigma^2=1$ 1) จำแนกตามจำนวนตัวแปรอิสระ ขนาดตัวอย่าง และวิธีที่ให้ค่า MSE น้อยกว่า 68

4.9 ตารางค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสองระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator เมื่อค่าผิดพลาดมีการแจกแจงเป็นแบบ ลอกนอร์มอล $\mu=0$, $\sigma^2=1$ จำแนกตามจำนวนตัวแปรอิสระ ขนาดตัวอย่าง และวิธีที่ให้ค่า MSE น้อยกว่า 69

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
4.10	ตารางค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสองระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator เมื่อค่าผิดพลาดมีการแจกแจงเป็นแบบแกมมา โดยจำแนกตามจำนวนตัวแปรอิสระ ขนาดตัวอย่าง ระดับของค่าสัมประสิทธิ์ของความแปรปรวน (coefficient of Variation) และวิธีที่ให้ค่า MSE น้อยกว่า	72
4.11	ตารางค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสองระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator เมื่อค่าผิดพลาดมีการแจกแจงเป็นแบบแกมมา โดยจำแนกตามจำนวนตัวแปรอิสระ ขนาดตัวอย่าง ระดับของค่าสัมประสิทธิ์ของความแปรปรวน (coefficient of variation) และวิธีที่ให้ค่า MSE น้อยกว่า	73
4.12	ตารางค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสองระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator เมื่อค่าผิดพลาดมีการแจกแจงเป็นแบบไวบูลล์ โดยจำแนกตามจำนวนตัวแปรอิสระ ขนาดตัวอย่าง ระดับของค่าสัมประสิทธิ์ของความแปรปรวน (coefficient of Variation) และวิธีที่ให้ค่า MSE น้อยกว่า	77
4.13	ตารางค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสองระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator เมื่อค่าผิดพลาดมีการแจกแจงเป็นแบบไวบูลล์ โดยจำแนกตามจำนวนตัวแปรอิสระ ขนาดตัวอย่าง ระดับของค่าสัมประสิทธิ์ของความแปรปรวน (coefficient of variation) และวิธีที่ให้ค่า MSE น้อยกว่า	78

สารบัญรูป

รูปที่		หน้า
2.1	แสดงฟังก์ชันของเกณฑ์วิธีที่แกร่ง	23
2.2	แสดงฟังก์ชันอิทธิพล (ψ) ของเกณฑ์วิธีที่แกร่ง	23
2.3	แสดงฟังก์ชันน้ำหนักของวิธีที่แกร่ง LS, H_2, $E_{0.3}$, $W_{1.399}$, H17A และแสดงภาพรวมทุกวิธี	24
4.1	เส้นโค้งของการแจกแจงแบบปกติปลอมปนที่มีสเกลแฟคเตอร์เท่ากับ 3 และใช้เปอร์เซนต์ของการปลอมปน 5	50
4.2	เส้นโค้งของการแจกแจงแบบปกติปลอมปนที่มีสเกลแฟคเตอร์เท่ากับ 3 และใช้เปอร์เซนต์ของการปลอมปน 25	50
4.3	เส้นโค้งของการแจกแจงแบบปกติปลอมปนที่มีสเกลแฟคเตอร์เท่ากับ 10 และใช้เปอร์เซนต์ของการปลอมปน 5	50
4.4	เส้นโค้งของการแจกแจงแบบปกติปลอมปนที่มีสเกลแฟคเตอร์เท่ากับ 10 และใช้เปอร์เซนต์ของการปลอมปน 25	51

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย