การประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยที่ค่าประมาณสเกลเปลี่ยนไป

นายทรงพันธุ์ ชุณหสวัสดิกุล

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
ภาควิชาสถิติ
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
พ.ศ. 2532

ISBN 974-576-982-7

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

015866

MULTIPLE REGRESSION COEFFICIENT ESTIMATION WITH CHANGED SCALE ESTIMATES

Mr. Songphan Choonhaswadikul

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

1989

ISBN 974-576-982-7

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสีเขียวนี้เพียงแผ่นเดียว

ทรงพันธุ์ ชุณหสวัสดิกุล : การประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยที่ค่าประมาณสเกลเปลี่ยนไป
(MULTIPLE REGRESSION COEFFICIENT ESTIMATION WITH CHANGED SCALE ESTIMATES)
อาจารย์ที่ปรึกษา : ผศ.ดร.ธีระพร วีระถาวร , หน้า ISBN 974-576-982-7

การวิจัยครั้งนี้มีวัตถุประสงค์ที่จะศึกษาการประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยที่ค่าประมาณสเกลเปลี่ยนไป โดยจะเปรียบเทียบวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator ซึ่งใช้ the standard deviation of location, the median absolute deviaton และ the modified biweight A-estimator (ที่ c = 9 และ 10) โดยใช้เกณฑ์ความแกร่งของ Huber เมื่อ b = 2 และใช้เกณฑ์การเปรียบเทียบของอัตราส่วนผลต่างของค่าเฉลี่ยความผิดพลาดกำลังสอง และอัตราส่วนผลต่างความผิดพลาดกำลังสอง การแจกแจงของค่าผิดพลาดที่ศึกษาคือการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติและการแจกแจงแบบเบ้ การแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติจะใช้การแจกแจงแบบปกติปลอมปนที่มีสเกลแฟกเตอร์ 5 10 และ 15 และใช้เปอร์เซนต์การปลอมปน 5 10 20 และ 30 และการแจกแจงแบบทีซึ่งใช้ระดับความเป็นเสรี 4 และ 8 ส่วนการแจกแจงแบบเบ้จะใช้การแจกแจงแบบลอกนอร์มอล (ที่ $\mu = 0$ และ $\sigma^2 = 1$) การแจกแจงแบบแกมมา ซึ่งมีค่าสัมประสิทธิ์ความแปรปรวน (C.V.) 70% และ 100% และการแจกแจงแบบไวบูลล์ ซึ่งมีค่าสัมประสิทธิ์ความแปรปรวน (C.V.) 52% และ 100% เมื่อการแจกแจงของค่าผิดพลาดไม่ใช่การแจกแจงแบบทีจะใช้ตัวแปรอิสระ 3 สำหรับตัวอย่างขนาด 20 และจำนวนตัวแปรอิสระ 5 และ 10 สำหรับตัวอย่างขนาด 50 100 และ 150 แต่เมื่อการแจกแจงของค่าผิดพลาดมีการแจกแจงแบบทีจะใช้ตัวแปรอิสระ 3 5 และ 10 สำหรับตัวอย่างขนาด 20 เท่านั้น ในการทดลองนี้ทำการจำลองเหตุการณ์ต่างๆ ด้วยเครื่องคอมพิวเตอร์โดยเทคนิคมอนติคาร์โลและกระทำซ้ำ 200 ครั้งในแต่ละกรณี ยกเว้นกรณีที่ข้อมูลมีการแจกแจงแบบเบ้ซึ่งกระทำซ้ำ 100 ครั้ง

ผลจากการศึกษาเปรียบเทียบอัตราส่วนผลต่างของค่าเฉลี่ยความผิดพลาดกำลังสองและอัตราส่วนผลต่างความผิดพลาดกำลังสองสามารถสรุปได้ดังนี้

1. กรณีที่ค่าผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ

การแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติในรูปแบบต่างๆ ซึ่งกำหนดโดย เปอร์เซนต์การปลอมปนและสเกลแฟกเตอร์สำหรับการแจกแจงแบบปกติปลอมปน และระดับความเป็นเสรีสำหรับการแจกแจงแบบที จะได้ว่าสเกลแฟกเตอร์มีอิทธิพลเพิ่มขึ้น เรียงลำดับจากมากไปน้อย 15 10 และ 5 เปอร์เซนต์การปลอมปนมีอิทธิพลเพิ่มขึ้น เรียงลำดับจากมากไปน้อย 30 20 10 และ 5 และระดับความเป็นเสรีมีอิทธิพลเพิ่มขึ้น เรียงลำดับจากมากไปน้อย 4 และ 8 ทำให้ the mediam absolute deviation เป็นตัวประมาณสเกลซึ่งใช้เกณฑ์ความแกร่งของ Huber เป็นวิธีการประมาณค่าที่ดีที่สุด

2. กรณีที่ค่าผิดพลาดมีการแจกแจงแบบเบ้

เช่น การแจกแจงแบบแกมมา ไวบูลล์ และลอกนอร์มอล วิธี M-estimator ซึ่งใช้ the median absolute deviation เป็นตัวประมาณสเกลให้ผลในการประมาณค่าได้ดีที่สุดสำหรับการแจกแจงแบบแกมมา เมื่อค่าสัมประสิทธิ์ความแปรปรวน = 100% แต่เมื่อค่าสัมประสิทธิ์ความแปรปรวน = 70% วิธีกำลังสองน้อยที่สุดจะให้ผลในการประมาณค่าดีที่สุดซึ่งอาจมีสาเหตุมาจากค่าสังเกตมีการแจกแจงใกล้เคียงการแจกแจงแบบปกติ และเมื่อใช้เทคนิคการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox ในการแปลงข้อมูลเข้าสู่ภาวะปกติ วิธี M-estimator ซึ่งใช้ the median absolute deviation เป็นตัวประมาณสเกลให้ผลในการประมาณค่าได้ดีที่สุดสำหรับการแจกแจงแบบไวบูลล์ เมื่อค่าสัมประสิทธิ์ความแปรปรวน = 100% และ 52% แต่เมื่อค่าสัมประสิทธิ์ความแปรปรวน = 52% วิธีกำลังสองน้อยที่สุดจะให้ผลดีใกล้เคียงกับวิธี M-estimator ซึ่งใช้ the median absolute deviation เป็นตัวประมาณสเกล และทุกวิธีการจะให้ผลในการประมาณค่าได้ใกล้เคียงกันสำหรับการแจกแจงแบบลอกนอร์มอล เมื่อใช้เทคนิคการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox ในการแปลงข้อมูลเข้าสู่ภาวะปกติ

ภาควิชา สถิติ ลายมือชื่อนิสิต .....

สาขาวิชา สถิติ

ปีการศึกษา 2532 ลายมือชื่ออาจารย์ที่ปรึกษา .....

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสีเขียวนี้เพียงแผ่นเดียว

SONGPHAN CHOONHASWADIKUL : MULTIPLE REGRESSION ESTIMATION WITH CHANGED SCALE ESTIMATE. THESIS ADVISOR : ASSIS. PROF. THEERAPORN VERATHAWORN, Ph.D., PP.

This thesis aims at studying multiple regression estimation with changed scale estimate evaluated by comparing the ordinary least square method with M-estimator method using the standard deviation of location, the median absolute deviation and the modified biweight A-estimator (c=9 and 10) in which the criterion of robust of Huber when b = 2 is used. The ratio of different average mean square error and the ratio of different mean square error are considered as the criterion of comparision. The distribution of residuals in this study are longer tailed distribution than normal distribution than normal distribution and skewed distribution. The distributions used in case of long tailed distribution are scale-contaminated distribution with scale factor of 5,10 and 15 and percent contamination of 5,10,20 and 30, and t-distribution with degree of freedom of 4 and 8. Other case of skewed distribution are lognormal distribution ($\mu = 0$ and $\sigma^2 = 1$), gamma distribution with coefficient of variation of 70% and 100% and weibull distribution with coefficient of variation of 52% and 100%. When the distribution is not t-distribution well will use the number of independent variable of 3 for the sample size of 20 and the number of independent variables of 5 and 10 for the sample sizes of 50, 100 and 150, but when the distribution is t-distribution we will use the number of independent variables of 3,5 and 10 for the sample size of 20. Data is obtained through simulation using Monte Carlo technique, and repeating 200 times for the longer tailed distribution than normal distribution and repeating 100 times for skewed distribution.

The results of this study are as follow:

1. In case residuals have longer tailed distribution than normal distribution;

Shapes of residuals distribution are determined by percent contamination and scale factor for contaminated normal distribution and the degree of freedom for t-distribution. The M-estimator method using the median absolute deviation as the scale factor with Huber's robust criterion gives the best result when scale factor increases from 5,10 and 15 and percent of contamination increases from 5,10,20 and 30. The same result is obtained in t-distribution when the degree of freedom decreases from 8 to 4.

2. Incase residuals have skewed distribution:

Such as gamma, weibull and lognormal distribution. The M-estimator method using the median absolute deviation as the scale estimator gives the best result for gamma distribution with the coefficient of variation 100%. But when the coefficient of variation is 70%, the ordinary least square method gives the best result, probably since the observation is closed to normal distribution and using Box and Cox's Power Transformation. The M-estimator method using the median absolute deviation as as the scale estimator gives the best result for weibull distrubution with coefficient of variation of 100% and 52% and specially when the coefficient of variation is 52%, the ordinary least square gives the similar result as the former one. Every method gives similar results for lognormal distrubition when using Box and Cox's Power Transformation.

ภาควิชา สถิติ ลายมือชื่อนิสิต .....

สาขาวิชา สถิติ

ปีการศึกษา 2532 ลายมือชื่ออาจารย์ที่ปรึกษา .....

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

หัวข้อวิทยานิพนธ์ : การประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยที่ค่าประมาณสเกลเปลี่ยนไป

ชื่อนิสิต : นายทรงพันธุ์ ชุณหสวัสดิกุล

ภาควิชา : สถิติ

อาจารย์ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร. ธีระพร วีระถาวร

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีบัณฑิตวิทยาลัย

(ศาสตราจารย์ ดร. ถาวร วัชราภัย)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(รองศาสตราจารย์ ดร. สรชัย พิศาลบุตร)

..... อาจารย์ที่ปรึกษา

(ผู้ช่วยศาสตราจารย์ ดร. ธีระพร วีระถาวร)

..... กรรมการ

(รองศาสตราจารย์ สำองศรี พิทยารัตน์)

..... กรรมการ

(รองศาสตราจารย์ มัลลิกา บุนนาค)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาของ ผศ. ดร. ธีระพร วีระถาวร อาจารย์ประจำภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ให้คำแนะนำ และให้คำปรึกษาตลอดจนแก้ไขข้อบกพร่องต่างๆ เป็นอย่างดีมาโดยตลอด ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ. โอกาสนี้ และขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. สรชัย พิศาลบุตร รองศาสตราจารย์ ดร. สุชาดา กีระนันท์ และอาจารย์ ดร. สุพล ดุรงค์วัฒนา ในเรื่องการค้นคว้าและยืมตำรา และขอขอบคุณทุกท่านที่ให้ความช่วยเหลือตลอดจนให้คำแนะนำต่างๆ

สุดท้ายนี้ขอขอบพระคุณ คุณพ่อ คุณแม่ พี่และน้องทุกๆ ท่านที่ส่งเสริม และสนับสนุนการเรียนของผู้วิจัยตลอดมา และขอกราบขอบพระคุณอาจารย์ทุกท่านที่ได้ประสิทธิ์ประสาทความรู้แก่ผู้วิจัยมาโดยตลอด

ทรงพันธุ์ ชุณหสวัสดิกุล

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญตาราง ..... ญ

สารบัญรูป ..... ฐ

บทที่ 1 บทนำ

1.1 ความสำคัญและความเป็นมาของปัญหา 1

1.2 วัตถุประสงค์ของการวิจัย 4

1.3 ข้อตกลงเบื้องต้น 4

1.4 ขอบเขตการวิจัย 5

1.4.1 ลักษณะการแจกแจงของความคลาดเคลื่อนที่ศึกษา 5

1.4.2 จำนวนตัวแปรอิสระและขนาดตัวอย่าง 9

1.4.3 การจำลองประชากรที่ศึกษาจากตัวแบบเชิงเส้น 10

1.5 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย 10

1.6 วิธีดำเนินการวิจัย 10

1.7 ศัพท์ต่างๆ ที่ใช้ในการวิจัย 11

บทที่ 2 ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง

2.1 ข้อสมมติทั่วไปของตัวแบบเชิงเส้น 15

2.2 วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุแบบต่างๆ 16

2.2.1 วิธีกำลังสองน้อยที่สุดแบบธรรมดา

(Ordinary least square method) 16

2.2.2 วิธีกำลังสองน้อยที่สุดแบบทั่วไป

(General least square method) 19

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ (ต่อ)

หน้า

2.2.3 วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (Weighted least square method) 21

2.2.4 วิธีประมาณค่าแบบ M (M-estimate method) 22

2.3 ตัวประมาณค่าสเกล (Scale-Estimator) 24

2.4 การแปลงข้อมูลที่อยู่ในรูปยกกำลัง (Power Transformation) .. 27

2.5 ผลงานวิจัยที่เกี่ยวข้อง 31

บทที่ 3 วิธีดำเนินการวิจัย

3.1 วิธีมอนติคาร์โล (Monte Carlo method) 38

3.2 แผนดำเนินการวิจัย 38

3.3 ขั้นตอนของการดำเนินงานวิจัย 39

3.4 ผังงานการดำเนินงานวิจัย 41

บทที่ 4 ผลการวิจัย

4.1 การประมาณค่าสัมประสิทธิ์การถดถอยพหุ เมื่อความผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงปกติ 55

4.2 การประมาณค่าสัมประสิทธิ์การถดถอยพหุ เมื่อความผิดพลาดมีการแจกแจงแบบเบ้ 79

4.3 สรุปผลการวิเคราะห์ของการประมาณค่าสัมประสิทธิ์การถดถอยพหุ ด้วยวิธี OLS STD MAD และ MOD (ที่ c = 9 และ 10) 86

บทที่ 5 สรุปผลการวิเคราะห์ และการอภิปรายผล

5.1 ผลสรุปการเปรียบเทียบวิธีการประมาณทั้ง 5 วิธี 91

5.5 การอภิปรายผล 94

5.6 ข้อเสนอแนะ 95

บรรณานุกรม 97

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ (ต่อ)

หน้า

ภาคผนวก ก 100

ภาคผนวก ข 112

ภาคผนวก ค 120

ประวัติผู้เขียน 156

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตารางที่		หน้า
1.7.1	แสดงฟังก์ชันความแกร่งและตัวประมาณสเกลที่สามารถใช้ได้	12
3.1	แสดงค่าสเกลแฟกเตอร์และเปอร์เซนต์การปลอมปนทั้งหมดที่ใช้ในการวิจัย	39
4.1.1.1	แสดงผลการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี OLS STD MAD และ MOD (ที่ c=9 และ 10) โดยที่การแจกแจงของความผิดพลาดมีการแจกแจงแบบปกติปลอมปน ณ Scale P m และ n ที่กำหนดตามลำดับ	57
4.1.1.2	แสดงผลการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี OLS STD MAD และ MOD (ที่ c=9 และ 10) โดยที่การแจกแจงของความผิดพลาดมีการแจกแจงแบบปกติปลอมปน ณ Scale P m และ n ที่กำหนดตามลำดับ	60
4.1.1.3	แสดงผลการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี OLS STD MAD และ MOD (ที่ c=9 และ 10) โดยที่การแจกแจงของความผิดพลาดมีการแจกแจงแบบปกติปลอมปน ณ Scale P m และ n ที่กำหนดตามลำดับ	63
4.1.1.4	แสดงผลการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี OLS STD MAD และ MOD (ที่ c=9 และ 10) โดยที่การแจกแจงของความผิดพลาดมีการแจกแจงแบบปกติปลอมปน ณ Scale P m และ n ที่กำหนดตามลำดับ	66
4.1.1.5	แสดงผลการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี OLS STD MAD และ MOD (ที่ c=9 และ 10) โดยที่การแจกแจงของความผิดพลาดมีการแจกแจงแบบปกติปลอมปน ณ Scale P m และ n ที่กำหนดตามลำดับ	69

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
4.1.1.6	แสดงผลการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี OLS STD MAD และ MOD (ที่ c=9 และ 10) โดยที่การแจกแจงของความผิดพลาดมีการแจกแจงแบบปกติปลอมปน ณ Scale P m และ n ที่กำหนดตามลำดับ	72
4.1.1.7	แสดงผลการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี OLS STD MAD และ MOD (ที่ c=9 และ 10) โดยที่การแจกแจงของความผิดพลาดมีการแจกแจงแบบปกติปลอมปน ณ Scale P m และ n ที่กำหนดตามลำดับ	75
4.1.2.1	แสดงผลการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี OLS STD MAD และ MOD (ที่ c=9 และ 10) โดยที่การแจกแจงของความผิดพลาดมีการแจกแจงแบบที ณ DF m และ n ที่กำหนดตามลำดับ	78
4.2.1.1	แสดงผลการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี OLS MAD และ MOD (ที่ c=10) โดยที่การแจกแจงของความผิดพลาดมีการแจกแจงแบบแกมมา ณ Alpha Beta m และ n ที่กำหนดตามลำดับ	81
4.2.2.1	แสดงผลการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี OLS MAD และ MOD (ที่ c=10) โดยที่การแจกแจงของความผิดพลาดมีการแจกแจงแบบไวบูลล์ ณ Alpha Beta m และ n ที่กำหนดตามลำดับ	83

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
4.2.3.1	แสดงผลการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี OLS MAD และ MOD (ที่ c=10) โดยที่การแจกแจงของความผิดพลาดมีการแจกแจงแบบลอกนอร์มอล ณ $\mu = 0, \sigma^2 = 1$,m และ n ที่กำหนดตามลำดับ	85
5.2.1	แสดงผลสรุปการเลือกวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุของค่าผิดพลาดของการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติและการแจกแจงแบบเบ้	95

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญรูป

รูปที่		หน้า
1.4.1	แสดงเส้นโค้งของการแจกแจงแบบปกติปลอมปน ณ p และ c	5
1.4.2	แสดงเส้นโค้งของการแจกแจงแบบที ณ DF = 4 และ 8	6
1.4.3	แสดงเส้นโค้งของการแจกแจงแบบลอกนอร์มอล	7
1.4.4	แสดงเส้นโค้งของการแจกแจงแบบแกมมา ณ $\beta = 5$ และ $\alpha = 1$ และ 2	8
1.4.5	แสดงเส้นโค้งของการแจกแจงแบบไวบูลล์ ณ $\beta = 5$ และ $\alpha = 1$ และ 2	9
1.7.1	แสดงกราฟของเกณฑ์ความแกร่ง Huber ณ k เมื่อ k คือจุดเปลี่ยนเว้า	11
3.1	แสดงผังงานสำหรับการประมาณค่าสัมประสิทธิ์การถดถอยพหุเมื่อค่าประมาณเสกลเปลี่ยนไป	42

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย