

การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยแบบไม่ติดกลุ่ม

นางสาวบุญจิรา มากอ้น

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศิลปศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2545

ISBN 974-17-1734-2

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A COMPARISON ON THE MODEL SELECTION CRITERIA FOR NONNESTED  
REGRESSION MODELS



Miss Bunjira Makond

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

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 974-17-1734-2


หัวข้อวิทยานิพนธ์  
โดย  
สาขาวิชา  
อาจารย์ที่ปรึกษา

การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยแบบไม่ติดกลุ่ม  
นางสาวบุญจิรา มากอ้น  
สถิติ  
รองศาสตราจารย์ ดร. สุปล ดุรงค์วัฒนา

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้รับวิทยานิพนธ์  
ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต


  
..... คณบดีคณะพาณิชยศาสตร์และการบัญชี  
(ผู้ช่วยศาสตราจารย์ ดร. วิรัช อภิเมธีธำรง)

คณะกรรมการสอบวิทยานิพนธ์

  
..... ประธานกรรมการ  
(รองศาสตราจารย์ ศิริพร สาเกตทอง)

  
..... อาจารย์ที่ปรึกษา  
(รองศาสตราจารย์ ดร. สุปล ดุรงค์วัฒนา)

  
..... กรรมการ  
(รองศาสตราจารย์ นพรัตน์ รุ่งอุทัยศิริ)

  
..... กรรมการ  
(อาจารย์ ดร. อรุณี กำลัง)



บุญจิรา มากอัน : การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยแบบไม่ติดกลุ่ม.

(A COMPARISON ON THE MODEL SELECTION CRITERIA FOR NONNESTED REGRESSION MODELS) อ. ที่ปรึกษา : รศ. ดร. สุพล ดุรงค์วัฒนา , 91 หน้า. ISBN 974-17-1734-2.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความถูกต้องของเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นแบบไม่ติดกลุ่ม ซึ่งมีตัวแบบทั่วไปดังนี้

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$$

เมื่อ  $\underline{Y}$  แทน เวกเตอร์ของตัวแปรตามขนาด  $(n \times 1)$   $\underline{X}$  แทน เมทริกซ์ของตัวแปรอิสระที่ควบคุมให้คงที่ขนาด  $(n \times p)$   $\underline{\beta}$  แทน เวกเตอร์ของพารามิเตอร์ของสัมประสิทธิ์การถดถอยขนาด  $(p \times 1)$   $\underline{\varepsilon}$  แทน เวกเตอร์ของความคลาดเคลื่อนขนาด  $(n \times 1)$  โดย  $\underline{\varepsilon} \sim N(0, \sigma^2 I)$   $\sigma^2 I$  แทน เมทริกซ์ความแปรปรวนร่วมขนาด  $(n \times n)$   $n$  แทน ขนาดตัวอย่าง  $p$  แทน จำนวนพารามิเตอร์ของสัมประสิทธิ์การถดถอย เกณฑ์การคัดเลือกตัวแบบสำหรับการวิจัยครั้งนี้ คือ เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ(Akaike's Information Criterion : AIC) กำหนดอยู่ในรูปสมการ  $AIC = n \log(\hat{\sigma}^2) + 2p$  และเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์ (Bayesian Information Criterion : BIC) กำหนดอยู่ในรูปสมการ  $BIC = n \log(\hat{\sigma}^2) + \log(n)p$  เมื่อ  $\hat{\sigma}^2 = \frac{SSE}{n}$  โดย SSE แทน ผลรวมค่าคลาดเคลื่อนกำลังสอง การคัดเลือกตัวแบบของเกณฑ์ AIC(BIC) คือ ตัวแบบที่ให้ค่า AIC(BIC) ต่ำสุด จะเป็นตัวแบบที่เหมาะสมที่สุด สำหรับการคัดเลือกตัวแบบแต่ละครั้งเกณฑ์ AIC(BIC) จะคัดเลือกตัวแบบได้ถูกต้องถ้าตัวแบบที่ได้มีค่า MSE ต่ำสุด การวิจัยครั้งนี้ทำการจำลองแบบโดยเทคนิคมอนติคาร์โลด้วยโปรแกรม S-PLUS 2000 โดยใช้ตัวแปรอิสระเริ่มต้นเป็น 2 3 และ 4 ตัว ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติมีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐาน เป็น 1 5 10 และ 15 ขนาดตัวอย่างเป็น 25 50 75 และ 100 ระดับความสัมพันธ์ของตัวแปรอิสระคือ 0 0.5 และ 0.99 ค่าสัดส่วนของการเลือกผิดที่กำหนดขึ้นเป็น 1% 5% และ 10% ระดับนัยสำคัญที่ใช้คือ 0.01 และ 0.05 เกณฑ์ที่นำมาใช้ในการเปรียบเทียบคือ ค่าสัดส่วนของการคัดเลือกตัวแบบผิด และผลการทดสอบสมมติฐาน ผลการวิจัยสรุปได้ดังนี้

1. เมื่อจำนวนตัวแปรอิสระเริ่มต้นเป็น 2 ตัวแปร เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ(AIC) และเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์(BIC) มีค่าสัดส่วนของการคัดเลือกตัวแบบผิดเป็น 0 เท่ากัน ทุกระดับความสัมพันธ์ของตัวแปรอิสระ ขนาดตัวอย่าง และส่วนเบี่ยงเบนมาตรฐาน

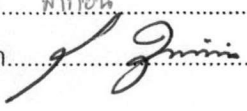
2. เมื่อจำนวนตัวแปรอิสระเริ่มต้นเป็น 3 และ 4 ตัวแปร เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ(AIC) มีค่าสัดส่วนของการเลือกผิดต่ำกว่าเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์(BIC) สำหรับทุกสถานการณ์ โดยเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ(AIC) และเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์(BIC) จะมีค่าสัดส่วนของการเลือกผิดสูงขึ้น เมื่อจำนวนตัวแปรอิสระมีจำนวนมากขึ้น และจะมีค่าลดลงเมื่อขนาดตัวอย่างมีจำนวนมากขึ้น

ภาควิชา สถิติ

สาขาวิชา สถิติ

ปีการศึกษา 2545

ลายมือชื่อนิสิต.....บุญจิรา มากอัน.....

ลายมือชื่ออาจารย์ที่ปรึกษา.....

##4382257926 : MAJOR STATISTICS

KEY WORD : AIC / BIC / MODEL SELECTION / NONNESTED REGRESSION MODEL

BUNJIRA MAKOND : A COMPARISON ON THE MODEL SELECTION CRITERIA FOR NONNESTED REGRESSION MODELS. THESIS ADVISOR : ASSOC. PROF. SUPOL DURONGWATANA , Ph.D. 91 pp. ISBN 974-17-1734-2.

The objective of this study is to compare the accuracy of model selection criteria for non-nested linear regression model. The general linear regression model is shown as follow :

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$$

where  $\underline{Y}$  is an  $(n \times 1)$  vector of dependent variables,  $\underline{X}$  is an  $(n \times p)$  matrix of independent variables,  $\underline{\beta}$  is a  $(p \times 1)$  vector of the regression coefficient,  $\underline{\varepsilon}$  is an  $(n \times 1)$  vector of errors, where  $\underline{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$ ;  $\sigma^2 \mathbf{I}$  is an  $(n \times n)$  matrix of variance-covariance,  $n$  is a sample sizes and  $p$  is the number of unknown parametric regression coefficients. Model selection criteria for this research are Akaike's Information Criterion(AIC) which is defined by equation  $AIC = n \log(\hat{\sigma}^2) + 2p$  and Bayesian Information Criterion(BIC) which is defined by equation  $BIC = n \log(\hat{\sigma}^2) + \log(n)p$ , where  $\hat{\sigma}^2 = \frac{SSE}{n}$ , SSE is sum square error. The model which has minimum (AIC)BIC is the preferred model. For each selection, AIC(BIC) can select accurate model if the model has minimum MSE value. In this study, the datas are simulated by S-PLUS 2000 package using Monte Carlo technique. The numbers of independent variables are 2, 3 and 4. The errors are normal distribution with mean 0 and standard deviations are 1, 5, 10 and 15. The sample sizes are 25, 50, 75 and 100. The levels of correlation among independent variables are 0, 0.5 and 0.99. Given proportion of miss selection values are 1%, 5% and 10%. Significance levels are 0.01 and 0.05. Proportion of miss selection values and result of testing hypothesis are used to compare the accuracy of two model selection criteria. The results of this study can be summarized as follow :

1. When the number of independent variables is 2, AIC and BIC give proportion of miss selection values are 0 for every levels of correlation among independent variables, sample size and standard deviations.
2. when the numbers of independent variables are 3 and 4, AIC give proportion of miss selection values less than BIC for all cases of this research. AIC and BIC give proportion of miss selection values increase when the number of independent variables increase and decrease when sample sizes increase.

Department Statistics  
Field of study Statistics  
Academic year 2002

Student's signature.....*Bunjira Makond*.....  
Advisor's signature.....*S.P. Durongwatana*.....



## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความสามารถของรองศาสตราจารย์ ดร. สุพล  
 ดุรงค์วัฒนา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้การปรึกษา คำแนะนำ ตลอดจนช่วยเหลือ  
 บกพร่องต่าง ๆ ด้วยดีตลอดมา จนกระทั่งวิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ ผู้วิจัยขอกราบ  
 ขอบพระคุณเป็นอย่างสูง ไว้ ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ศิริพร สาททอง รองศาสตราจารย์  
 นพรัตน์ รุ่งอุทัยศิริ และ อาจารย์ ดร. อรุณี กำลิ่ง ในฐานะประธานกรรมการ และกรรมการสอบ  
 วิทยานิพนธ์ ที่กรุณาตรวจแก้วิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น รวมถึงคณาจารย์ประจำภาค  
 วิชาสถิติทุกท่านที่ได้ประสิทธิประสาทความรู้ให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษา

ทำนนี้ ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดา มารดา ซึ่งได้ให้การสนับสนุนในทุก ๆ  
 ด้าน และเพื่อน ๆ ที่ให้กำลังใจแก่ผู้วิจัยเสมอมาจนสำเร็จการศึกษา นอกจากนี้ผู้วิจัยได้รับการ  
 สนับสนุนทุนการศึกษาจากโครงการพัฒนาอาจารย์วิทยาเขตสารสนเทศ มหาวิทยาลัยสงขล  
 นครินทร์ และทุนอุดหนุนการวิจัยของบัณฑิตวิทยาลัย จึงขอขอบคุณมา ณ ที่นี้

ศูนย์วิทยทรัพยากร  
 จุฬาลงกรณ์มหาวิทยาลัย

## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฌ
สารบัญภาพ.....	ฎ
บทที่	
1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	3
1.3 สมมติฐานของการวิจัย.....	4
1.4 ข้อตกลงเบื้องต้น.....	4
1.5 คำจำกัดความที่ใช้ในการวิจัย.....	5
1.6 ขอบเขตของการวิจัย.....	5
1.7 ประโยชน์ของการวิจัย.....	12
1.8 เกณฑ์ที่ใช้ในการตัดสินใจ.....	12
1.9 ขั้นตอนในการดำเนินการวิจัย.....	13
2 แนวคิดและทฤษฎี.....	14
2.1 การวิเคราะห์การถดถอยเชิงเส้น.....	14
2.2 การประมาณด้วยภาวะน่าจะเป็นสูงสุด.....	16
2.3 การพิจารณาความเหมาะสมของตัวแบบ.....	17
2.4 เกณฑ์ที่นำมาใช้ในการคัดเลือกตัวแบบ.....	17
2.4.1 เกณฑ์การคัดเลือกตัวแบบโดยข้อเสนอแนะของอาโคเค.....	17
2.4.1.1 วิธีการคัดเลือกตัวแบบของเกณฑ์การคัดเลือกตัวแบบโดย ข้อเสนอแนะของอาโคเค.....	19
2.4.2 เกณฑ์การคัดเลือกตัวแบบโดยข้อเสนอแนะของเบล์.....	19
2.4.2.1 วิธีการคัดเลือกตัวแบบของเกณฑ์การคัดเลือกตัวแบบโดย ข้อเสนอแนะของเบล์.....	21

## สารบัญ(ต่อ)

บทที่	หน้า
2.5 การทดสอบสมมติฐานเกี่ยวกับค่าสัดส่วน.....	22
2.5.1 การทดสอบสมมติฐานเปรียบเทียบระหว่างค่าสัดส่วนของการคัดเลือกตัวแบบผิด กับค่าสัดส่วนของการคัดเลือกตัวแบบผิดที่กำหนดขึ้น.....	22
2.5.2 การทดสอบสมมติฐานเปรียบเทียบระหว่างค่าสัดส่วนของการคัดเลือกตัวแบบผิด จากเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาโคเคะกับเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์.....	23
3. วิธีดำเนินการวิจัย.....	25
3.1 หลักการของวิธีการมอนติคาร์โล.....	25
3.2 ขั้นตอนในการดำเนินการวิจัย.....	26
4. ผลการวิจัย.....	34
4.1 ผลการวิจัยของการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นแบบไม่ติดกลุ่ม กรณีจำนวนตัวแปรอิสระเริ่มต้นเป็น 2 ตัวแปร.....	36
4.2 ผลการวิจัยของการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นแบบไม่ติดกลุ่ม กรณีจำนวนตัวแปรอิสระเริ่มต้นเป็น 3 ตัวแปร.....	44
4.3 ผลการวิจัยของการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นแบบไม่ติดกลุ่ม กรณีจำนวนตัวแปรอิสระเริ่มต้นเป็น 4 ตัวแปร.....	56
5. สรุปผลการวิจัยและข้อเสนอแนะ.....	68
5.1 สรุปผลการวิจัย.....	69
5.2 ข้อเสนอแนะ.....	71
รายการอ้างอิง.....	73
ภาคผนวก.....	75
ประวัติผู้เขียนวิทยานิพนธ์.....	91





## สารบัญญัตราสาร(ต่อ)

ตาราง	หน้า
4.12 แสดงผลการเปรียบเทียบค่าสัดส่วนของการคัดเลือกตัวแบบผิด กรณีจำนวนตัวแปรอิสระที่ใช้เริ่มต้นเป็น 3 ตัวแปร เมื่อ $p = 0.99$ .....	55
4.13 แสดงค่าสัดส่วนของการคัดเลือกตัวแบบผิด ของการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นแบบไม่ติดกลุ่ม กรณีจำนวนตัวแปรอิสระที่ใช้เริ่มต้นเป็น 4 ตัวแปร เมื่อ $p = 0$ .....	57
4.14 แสดงค่าสัดส่วนของการคัดเลือกตัวแบบผิด ของการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นแบบไม่ติดกลุ่ม กรณีจำนวนตัวแปรอิสระที่ใช้เริ่มต้นเป็น 4 ตัวแปร เมื่อ $p = 0.5$ .....	59
4.15 แสดงค่าสัดส่วนของการคัดเลือกตัวแบบผิด ของการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นแบบไม่ติดกลุ่ม กรณีจำนวนตัวแปรอิสระที่ใช้เริ่มต้นเป็น 4 ตัวแปร เมื่อ $p = 0.99$ .....	61
4.16 แสดงผลการเปรียบเทียบค่าสัดส่วนของการคัดเลือกตัวแบบผิด กรณีจำนวนตัวแปรอิสระที่ใช้เริ่มต้นเป็น 4 ตัวแปร เมื่อ $p = 0$ .....	65
4.17 แสดงผลการเปรียบเทียบค่าสัดส่วนของการคัดเลือกตัวแบบผิด กรณีจำนวนตัวแปรอิสระที่ใช้เริ่มต้นเป็น 4 ตัวแปร เมื่อ $p = 0.5$ .....	66
4.18 แสดงผลการเปรียบเทียบค่าสัดส่วนของการคัดเลือกตัวแบบผิด กรณีจำนวนตัวแปรอิสระที่ใช้เริ่มต้นเป็น 4 ตัวแปร เมื่อ $p = 0.99$ .....	67

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## สารบัญภาพ

ภาพประกอบ	หน้า
3.1 แผนผังแสดงขั้นตอนการดำเนินการวิจัย.....	30
3.2 แผนผังเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ.....	31
3.3 แผนผังเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์.....	32
3.4 แผนผังการทดสอบสมมติฐานของค่าสัดส่วนของการคัดเลือกตัวแบบผิด.....	33
4.1 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นแบบไม่ติดกลุ่ม ด้วยการพิจารณาค่าสัดส่วนของการเลือกตัวแบบผิด สำหรับตัวแปรอิสระ 3 ตัวแปร เมื่อมี $p = 0$ .....	46
4.2 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นแบบไม่ติดกลุ่ม ด้วยการพิจารณาค่าสัดส่วนของการเลือกตัวแบบผิด สำหรับตัวแปรอิสระ 3 ตัวแปร เมื่อมี $p = 0.5$ .....	48
4.3 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นแบบไม่ติดกลุ่ม ด้วยการพิจารณาค่าสัดส่วนของการเลือกตัวแบบผิด สำหรับตัวแปรอิสระ 3 ตัวแปร เมื่อมี $p = 0.99$ .....	50
4.4 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นแบบไม่ติดกลุ่ม ด้วยการพิจารณาค่าสัดส่วนของการเลือกตัวแบบผิด สำหรับตัวแปรอิสระ 4 ตัวแปร เมื่อมี $p = 0$ .....	58
4.5 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นแบบไม่ติดกลุ่ม ด้วยการพิจารณาค่าสัดส่วนของการเลือกตัวแบบผิด สำหรับตัวแปรอิสระ 4 ตัวแปร เมื่อมี $p = 0.5$ .....	60
4.6 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นแบบไม่ติดกลุ่ม ด้วยการพิจารณาค่าสัดส่วนของการเลือกตัวแบบผิด สำหรับตัวแปรอิสระ 4 ตัวแปร เมื่อมี $p = 0.99$ .....	62