

การเปรียบเทียบค่าพยากรณ์ที่ได้จากตัวแบบที่คัดเลือกตัวแปรด้วยวิธีเบสส์เซียน วิธีการกำจัดตัวแปรแบบถอยหลัง และวิธีการถดถอยแบบขั้นบันได ในการวิเคราะห์ความถดถอยพหุนามแบบลำดับขั้น



นางสาวนุชรินทร์ ทิพยวรรณกร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2540

ISBN 974 - 639 - 175 - 5

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

A COMPARISON OF PREDICTED VALUES OF DEPENDENT VARIABLE USING MODELS
SELECTED BY BAYESIAN VARIABLE SELECTION , BACKWARD ELIMINATION
AND STEPWISE REGRESSION WITH HIERACHICAL POLYNOMIAL REGRESSION

Miss Noocharin Tippayawannakorn

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Statistics

Department of Statistics

Graduate School

Chulalongkorn University

Academic Year 1997

ISBN 974 – 639 – 175 - 5

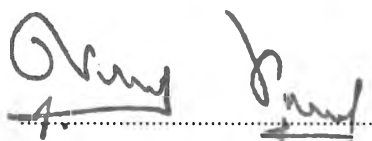
หัวข้อวิทยานิพนธ์ การเปรียบเทียบค่าพยากรณ์ที่ได้จากตัวแบบที่คัดเลือกตัวแปรด้วยวิธีเบสส์เซียน วิธีการกำจัดตัวแปรแบบถอยหลัง และวิธีการถดถอยแบบขั้นบันได ในการวิเคราะห์ความถดถอยพหุนามแบบลำดับชั้น

โดย นุชรินทร์ ทิพย์วรรณกร


ภาควิชา สถิติ


อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร. ชีระพร วีระถาวร


บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

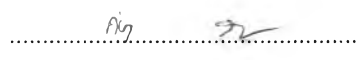

..... คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ นายแพทย์ สุภวัฒน์ ชุตินวงศ์)

คณะกรรมการสอบวิทยานิพนธ์


..... ประธานกรรมการ
(รองศาสตราจารย์ มัลลิกา บุญนาค)


..... อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร. ชีระพร วีระถาวร)


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


..... กรรมการ
(รองศาสตราจารย์ ดร. กัลยา วานิชย์บัญชา)

พิมพ์ต้นฉบับจากบทคัดย่อวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว

นุชรินทร์ ทิพย์วรรณการ : การเปรียบเทียบค่าพยากรณ์ที่ได้จากตัวแบบที่คัดเลือกตัวแปรด้วยวิธีเบย์เซียน วิธีการกำจัดตัวแปรแบบดอยหลัง และวิธีการลดดอยแบบขั้นบันได ในการวิเคราะห์ความถดถอยพหุนามแบบลำดับชั้น (A COMPARISON OF PREDICTED VALUES OF DEPENDENT VARIABLE USING MODELS SELECTED BY BAYESIAN VARIABLE SELECTION , BACKWARD ELIMINATION AND STEPWISE REGRESSION WITH HIERACHICAL POLYNOMIAL REGRESSION)

อาจารย์ที่ปรึกษา : รศ. ดร. วีระพร วีระถาวร ; 255 หน้า. ISBN 974-639-175-5.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบค่าพยากรณ์จากตัวแบบที่คัดเลือกจาก 3 วิธี คือ วิธีเบย์เซียน(BS) วิธีการกำจัดตัวแปรแบบดอยหลัง (BE) และวิธีการลดดอยแบบขั้นบันได (SW) ในการวิเคราะห์ความถดถอยเชิงเส้นเมื่อตัวแปรอิสระมีความสัมพันธ์พหุนามแบบลำดับชั้น เกณฑ์ที่ใช้ในการตัดสินใจคือ เกณฑ์ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) กำหนดจำนวนตัวแปรอิสระสูงสุดที่ใช้ในการสร้างตัวแปรตาม (MB) และจำนวนตัวแปรอิสระสูงสุดที่ใช้ในการสร้างตัวแบบเริ่มต้น (MI) ไม่เกิน 6 ตัวแปร อันดับสูงสุดของพจน์ที่ศึกษาไม่เกิน 2 ค่าคลาดเคลื่อนที่ใช้มีการแจกแจงแบบปกติด้วยค่าเฉลี่ย 0 และส่วนเบี่ยงเบนมาตรฐาน 5 10 20 และ 25 ขนาดตัวอย่าง (n) ที่ศึกษาคือ 25 50 75 และ 100 ระดับนัยสำคัญ (α) ที่ใช้ คือ 0.01 และ 0.05 และค่าคงที่ของวิธี BS คือ $(\sigma_\beta/\tau, c)$ มีค่าเป็น (1,5) (1,10) (10,100) และ (10,500) ข้อมูลที่ใช้ในการวิจัยได้จากการจำลองด้วยเทคนิคมอนติคาร์โล กระทำซ้ำ 200 รอบในแต่ละสถานการณ์

ผลการวิจัยได้ข้อสรุปดังนี้ เมื่อใช้ตัวแบบที่เหมาะสม คือ $MI - MB = 0$ และตัวแบบที่มีจำนวนตัวแปรมากเกินไป $MI - MB > 0$ วิธี BS จะให้ค่าพยากรณ์ที่มีความคงเส้นคงวา และมี AMSE ต่ำกว่าวิธีอื่น แต่เมื่อใช้ตัวแบบที่มีจำนวนตัวแปรน้อยเกินไป คือ $MI - MB < 0$ วิธี BS จะให้ค่าพยากรณ์ที่ดีกว่าวิธีอื่น เมื่อค่าส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนมีค่าต่ำและขนาดตัวอย่างใหญ่ แต่เมื่อระดับนัยสำคัญเพิ่มขึ้น วิธี SW และ BE จะให้ค่าพยากรณ์ที่ดีขึ้น ในกรณีที่ค่าคงที่ $(\sigma_\beta/\tau, c)$ ของวิธี BS มีค่าต่ำ ๆ จะทำให้ค่า AMSE ของวิธี BS ต่ำกว่าวิธีอื่น

ปัจจัยที่มีผลต่อค่า AMSE ของแต่ละวิธีเรียงตามลำดับจากมากไปน้อย คือ ค่า AMSE ของทุกวิธีจะแปรผันตามส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน ค่า MB และค่า $MI - MB$ แต่จะแปรผกผันกับขนาดตัวอย่าง สำหรับปัจจัยอื่นที่มีผลกระทบต่อ AMSE ของวิธี BE และ SW คือ ระดับนัยสำคัญ โดยที่ค่า AMSE ของทั้งสองวิธีนั้นจะแปรผกผันกับระดับนัยสำคัญ

ภาควิชา สถิติ
สาขาวิชา สถิติ
ปีการศึกษา 2540

ลายมือชื่อนิสิต นุชรินทร์ ทิพย์วรรณการ
ลายมือชื่ออาจารย์ที่ปรึกษา
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

C723889 : MAJOR STATISTICS

KEY WORD: Bayesian Variable Selection / Backward Elimination / Stepwise Regression / Gibbs Sampling / Hierarchical Polynomial Regression

NOOCHARIN TIPPAYAWANNAKORN : A COMPARISON OF THE PREDICTED VALUES OF DEPENDENT VARIABLE USING MODELS SELECTED BY BAYESIAN VARIABLE SELECTION, BACKWARD ELIMINATION AND STEPWISE REGRESSION WITH HIERACHICAL POLYNOMIAL REGRESSION. THESIS ADVISOR : ASSOC. PROF. THEERAPORN VERATHAWORN, Ph.D. 255 pp. ISBN 974-639-175-5.

The objective of this research is to compare the predicted values of dependent variable using models selected by Bayesian variable selection (BS), Backward elimination (BE) and Stepwise regression (SW) with Hierarchical polynomial regression. Under consideration in this research are the following: The highest numbers of independent variables for dependent variable generating (MB) and initial model building (MI) are 1, 2, 3, 4, 5 and 6. The highest exponent of hierarchical polynomial term is 2nd. Random error distributions are normal with mean zero and standard deviations 5, 10, 20 and 25. Sample sizes are 25, 50, 75 and 100. Significant levels are 0.01 and 0.05. Bayesian hyperparameters ($\sigma_\beta/\tau, c$) are (1,5), (1,10), (10,100) and (10,500). The criterion of comparison is average mean square error (AMSE). The data are obtained through simulation using a Monte Carlo technique and repeating 200 times for each case.

The results of this research are as follows: When using optimal model, $MI - MB = 0$, or overfitted model, $MI - MB > 0$, BS gives consistent predicted values of dependent variable and minimum AMSE. When using underfitted model, $MI - MB < 0$, BS gives minimum AMSE in cases where the standard deviation of the random error is small and the sample size is large. Higher significant level, supports SW and BE to give minimum AMSE. Lower hyperparameters of BS are ($\sigma_\beta/\tau, c$) which support BS gives minimum AMSE

Standard deviation of random error, number of variables in MB and the difference in the number of variables in MI and MB ($MI - MB$) affected AMSE, respectively. AMSE is decreased when sample size increased. The higher significant level decreased affected the AMSE of BE and SW decreased but did not affect the AMSE of BS.

ภาควิชา.....สถิติ.....
สาขาวิชา.....สถิติ.....
ปีการศึกษา.....2540.....

ลายมือชื่อนิสิต.....สุรินทร์ ทิพวรรณการ.....
ลายมือชื่ออาจารย์ที่ปรึกษา.....[Signature].....
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....-.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความสามารถของรองศาสตราจารย์ ดร. วีระพร วีระถาวร อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้ให้คำปรึกษาและแนะนำตลอดจนแก้ไขข้อบกพร่องต่าง ๆ เป็นอย่างดีมาโดยตลอด ผู้วิจัยใคร่ขอขอบพระคุณเป็นอย่างสูง

ขอกราบขอบพระคุณ รองศาสตราจารย์มัลลิกา บุนนาค รองศาสตราจารย์ ดร. สรชัย พิศาลบุตร และรองศาสตราจารย์ ดร.กัลยา วาณิชย์บัญชา ในฐานะประธานและกรรมการสอบวิทยานิพนธ์ที่ได้กรุณาตรวจแก้ไขให้วิทยานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น ขอกราบขอบพระคุณท่านอาจารย์ภาควิชาสถิติ แห่งจุฬาลงกรณ์มหาวิทยาลัย และสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือทุกท่านที่ได้ประสิทธิประสาทวิชาความรู้แก่ผู้วิจัย ขอขอบพระคุณ Edward I. George, Professor of Statistics, Department of MSIS, University of Texas, Austin และ Panlop J. Zeepongsekul, Professor of Mathematics, Department of Mathematics, RMIT University, Australia ที่ให้คำแนะนำที่เป็นประโยชน์อย่างยิ่งเกี่ยวกับรายละเอียดต่าง ๆ ในการทำงานวิจัยนี้

ขอขอบพระคุณ คุณพ่อ คุณแม่ ผู้ส่งเสริมและสนับสนุนด้านการเรียนของผู้วิจัยเสมอมาจนสำเร็จการศึกษา ขอขอบพระคุณสมาคมราชกรีฑาสโมสรที่อนุเคราะห์ทุนอุดหนุนการศึกษาแก่ผู้วิจัย ขอขอบคุณเพื่อน ๆ พี่ ๆ และน้อง ๆ ที่ให้กำลังใจเสมอมา

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญ	ช
สารบัญตาราง	ฅ
สารบัญรูป	ฉ
บทที่	
1 บทนำ	1
2 ทฤษฎีที่เกี่ยวข้อง.....	8
3 วิธีดำเนินการวิจัย	26
4 ผลการวิจัย	38
5 สรุปผลการวิจัยและข้อเสนอแนะ	212
รายการอ้างอิง	222
ภาคผนวก	223
ประวัติผู้วิจัย	255

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
4.43 แสดงค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและอัตราส่วนผลต่างของค่าเฉลี่ยของ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยเมื่อ $MB=6$ $\alpha=0.01$ $c=100$ และ $\sigma_\beta/\tau=10$	191
4.44 แสดงค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและอัตราส่วนผลต่างของค่าเฉลี่ยของ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยเมื่อ $MB=6$ $\alpha=0.01$ $c=500$ และ $\sigma_\beta/\tau=10$	194
4.45 แสดงค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและอัตราส่วนผลต่างของค่าเฉลี่ยของ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยเมื่อ $MB=6$ $\alpha=0.05$ $c=5$ และ $\sigma_\beta/\tau=1$	198
4.46 แสดงค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและอัตราส่วนผลต่างของค่าเฉลี่ยของ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยเมื่อ $MB=6$ $\alpha=0.05$ $c=10$ และ $\sigma_\beta/\tau=1$	202
4.47 แสดงค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและอัตราส่วนผลต่างของค่าเฉลี่ยของ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยเมื่อ $MB=6$ $\alpha=0.05$ $c=100$ และ $\sigma_\beta/\tau=10$	205
4.48 แสดงค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและอัตราส่วนผลต่างของค่าเฉลี่ยของ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยเมื่อ $MB=6$ $\alpha=0.05$ $c=500$ และ $\sigma_\beta/\tau=10$	208
5.1 แสดงวิธีที่ให้ค่า AMSE ต่ำที่สุด ณ สถานการณ์ต่าง ๆ จำแนกตาม MB, MI, n, c และ σ_β/τ	215

สารบัญรูป

รูปที่	หน้า
1 แสดงความสัมพันธ์ระหว่าง Parent และ Inherit	8
2 แสดงความหนาแน่นส่วนริมของการแจกแจง $N(0, \sigma^2 + t^2)$ และ $N(0, \sigma^2 + c^2 t^2)$ เมื่อ $(\sigma, \tau, c) = (1, 5), (1, 10), (10, 100)$ และ $(10, 500)$	17
3 แสดงผังงานแสดงขั้นตอนในการทำวิจัย	28
4 แสดงผังการสร้างตัวแบบด้วยวิธีกำจัดตัวแปรแบบถอยหลัง	31
5 แสดงผังงานแสดงการสร้างตัวแบบด้วยวิธีถดถอยแบบขั้นบันได	33
6 แสดงผังงานแสดงวิธีการสร้างตัวแบบด้วยวิธีการคัดเลือกตัวแปรของเบส์เซียน	36
7 แสดงแผนผังผลสรุปการเลือกวิธีการสร้างตัวแบบในการวิเคราะห์ ความถดถอยพหุนามแบบลำดับขั้น	215