



การเปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์ด้วยวิธีแลชโซในการถดถอยเชิงเส้นที่มี
มิติสูง



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติประยุกต์ แผนก ก แบบ ก 2 ระดับปริญญามหาบัณฑิต

ภาควิชาสถิติ

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

ปีการศึกษา 2560

ลิขสิทธิ์ของบัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

การเปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์ด้วยวิธีแลชโซในการถดถอย
เชิงเส้นที่มีมิติสูง



โดย
นางสาวพัชราภรณ์ พรดำเนินสวัสดิ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติประยุกต์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
ภาควิชาสถิติ
บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร
ปีการศึกษา 2560
ลิขสิทธิ์ของบัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

PERFORMANCE COMPARISONS OF LASSO METHODS FOR PARAMETER
ESTIMATION IN HIGH DIMENSIONAL LINEAR REGRESSION



A Thesis Submitted in partial Fulfillment of Requirements
for Master of Science (APPLIED STATISTICS)
Department of STATISTICS
Graduate School, Silpakorn University
Academic Year 2017
Copyright of Graduate School, Silpakorn University

58304203 : สถิติประยุกต์ แผน ก แบบ ก 2 ระดับปริญญาโท

นางสาว พัชราภรณ์ พรดำเนินสวัสดิ์: การเปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์ด้วยวิธีแลซโซในการถดถอยเชิงเส้นที่มีมิติสูง อาจารย์ที่ปรึกษาวิทยานิพนธ์ : ดร.กรรณิกาน์ หิรัญกลี

การวิเคราะห์การถดถอยเชิงเส้นพหุคูณแบบพินอลไลซ์เป็นวิธีหนึ่งที่ยอมรับกันอย่างแพร่หลายในการประมาณค่าพารามิเตอร์เมื่อข้อมูลมีมิติสูง ซึ่งการกำหนดฟังก์ชันพินอลที่แตกต่างกันเป็นผลให้ตัวประมาณแบบพินอลไลซ์แตกต่างกัน งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณพินอลไลซ์ชนิด L_1 และ L_2 ในกรณีข้อมูลมีมิติสูง 4 วิธี ได้แก่ วิธีริดจ์, LASSO, Adaptive LASSO และ Relaxed LASSO ภายใต้ตัวแบบถดถอยพหุคูณ 4 กรณี ได้แก่ ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระเป็นเชิงตั้งฉากบางส่วน, ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก, ตัวแบบถดถอยบางเบาที่ความสัมพันธ์ของตัวแปรอิสระมีโครงสร้างแบบแบ่งกลุ่ม และตัวแบบถดถอยไม่บางเบาและเมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก งานวิจัยเป็นการศึกษาการจำลองด้วยโปรแกรม R ในแต่ละสถานการณ์ ทำซ้ำ 500 รอบ เกณฑ์ที่ใช้ในการเปรียบเทียบ คือ ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์, ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์การถดถอย และความผิดพลาดของการคัดเลือกตัวประมาณ

ผลการวิจัย พบว่า 1) ในกรณีตัวแบบถดถอยบางเบาและเมทริกซ์สหสัมพันธ์ของตัวแปรอิสระเป็นเชิงตั้งฉากบางส่วน ไม่เป็นเชิงตั้งฉาก และโครงสร้างแบบแบ่งกลุ่ม วิธี Relaxed LASSO มีประสิทธิภาพของการพยากรณ์และประสิทธิภาพของการประมาณสูงที่สุด และวิธี Ridge regression มีประสิทธิภาพต่ำที่สุด และวิธี Adaptive LASSO มีประสิทธิภาพสูงกว่าวิธี LASSO เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูงและขนาดตัวอย่างขนาดใหญ่ 2) ในกรณีตัวแบบถดถอยไม่บางเบาและเมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก วิธี Ridge regression มีประสิทธิภาพของการพยากรณ์และประสิทธิภาพของการประมาณสูงที่สุด รองลงมาคือวิธี LASSO เมื่อเปรียบเทียบวิธี Adaptive LASSO และ Relaxed LASSO พบว่า วิธี Adaptive LASSO มีประสิทธิภาพสูงกว่าวิธี Relaxed LASSO เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูง

58304203 : Major (APPLIED STATISTICS)

MISS PATCHARAPORN PORNDUMNERNSAWAT : PERFORMANCE COMPARISONS OF LASSO METHODS FOR PARAMETER ESTIMATION IN HIGH DIMENSIONAL LINEAR REGRESSION THESIS ADVISOR : KANNIGARH HIRUNKASI, PH.D.

Penalized linear regression analysis is one of the widely used methods to estimate the parameters in high-dimensional data. The differences of parameters estimation in penalized linear regression are based on the penalty functions. The purpose of this research aim to compare the performances of L_1 and L_2 penalized estimation methods in high-dimensional linear regression models including ridge regression, LASSO, adaptive LASSO and relaxed LASSO under four different regression models such as partial orthogonal sparse model, non-orthogonal sparse model, grouping structure sparse model and non-orthogonal and non-sparse model. The simulation study is conducted by R programming using 500 replications for each situation setting. The performances of penalized regression estimators are compared by three criteria, median of the prediction mean square error, median of the estimation mean square error and the incorrect variable selection in models.

The results of this study are following: 1) In cases of partial orthogonal sparse model, non-orthogonal sparse model, grouping structure sparse model, relaxed LASSO has the highest prediction and estimation accuracy but ridge regression performs worst. In addition, the performance of adaptive LASSO estimation is better than LASSO when predictors are high correlate and sample size is large. 2) For non-orthogonal and non-sparse model, the highest performance penalized estimation method is ridge regression and the secondary is LASSO. Comparing next two methods, adaptive LASSO has better performance than relaxed LASSO when predictors are high correlate.

กิตติกรรมประกาศ

การดำเนินงานวิจัยและเรียบเรียงวิทยานิพนธ์นี้สำเร็จลุล่วงได้เป็นอย่างดีเนื่องจากได้รับความอนุเคราะห์จากอาจารย์กรรณิกาณ์ หิรัญกลี ซึ่งเป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์ และอาจารย์ปิยพล ไพจิตร ซึ่งเป็นอาจารย์ที่ปรึกษาร่วม ที่กรุณาให้คำแนะนำ คำปรึกษา แนวคิด ตลอดจนการตรวจทานและแก้ไขข้อบกพร่องต่างๆ เป็นอย่างดี จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี ผู้วิจัยจึงขอกราบขอบพระคุณอาจารย์เป็นอย่างสูงด้วยความซาบซึ้งในสำนึกในพระคุณ

ขอขอบคุณผู้ช่วยศาสตราจารย์ ดร.ไพโรจน์ ขาวสิทธิวงษ์ ที่กรุณาเป็นประธานกรรมการในการสอบวิทยานิพนธ์ และผู้ช่วยศาสตราจารย์ ดร.วินัย โพธิ์สุวรรณ ที่กรุณาเป็นผู้ทรงคุณวุฒิ สำหรับคำแนะนำ ตรวจสอบ และชี้แนะ ทำให้วิทยานิพนธ์ฉบับนี้สมบูรณ์มากยิ่งขึ้น

ผู้วิจัยขอขอบพระคุณคณาจารย์ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากรทุกท่านที่ได้ประสิทธิ์ประสาทองค์ความรู้ คำแนะนำ พร้อมทั้งให้กำลังใจตลอดระยะเวลาในการศึกษา ขอขอบพระคุณคุณนงลักษณ์ เอี้ยวเจริญ และคุณเดือนเพ็ญ เกาต์วง ที่ให้ความช่วยเหลือและอำนวยความสะดวกด้านงานเอกสารแก่ผู้วิจัย

ขอขอบพระคุณโครงการพัฒนากำลังคนด้านวิทยาศาสตร์ (ทุนเรียนดีวิทยาศาสตร์แห่งประเทศไทย) ที่ให้การสนับสนุนด้านทุนทรัพย์ในการศึกษาและการทำวิจัย

ขอบคุณรุ่นพี่ รุ่นน้อง และเพื่อนๆ ภาควิชาสถิติ และภาควิชาอื่นๆ ที่ให้คำแนะนำ ให้ความช่วยเหลือ และเป็นกำลังใจทำให้วิทยานิพนธ์นี้สำเร็จลุล่วงไปด้วยดี

สุดท้ายนี้ขอขอบพระคุณคุณลุง คุณป้า ที่เป็นผู้ปกครองของผู้ทำวิจัย ขอขอบคุณครอบครัว ที่สนับสนุนการศึกษา การทำวิจัยครั้งนี้ ตลอดจนให้ความรัก กำลังใจ แรงผลักดันและแรงสนับสนุนในทุกด้านแก่ผู้ทำวิจัย

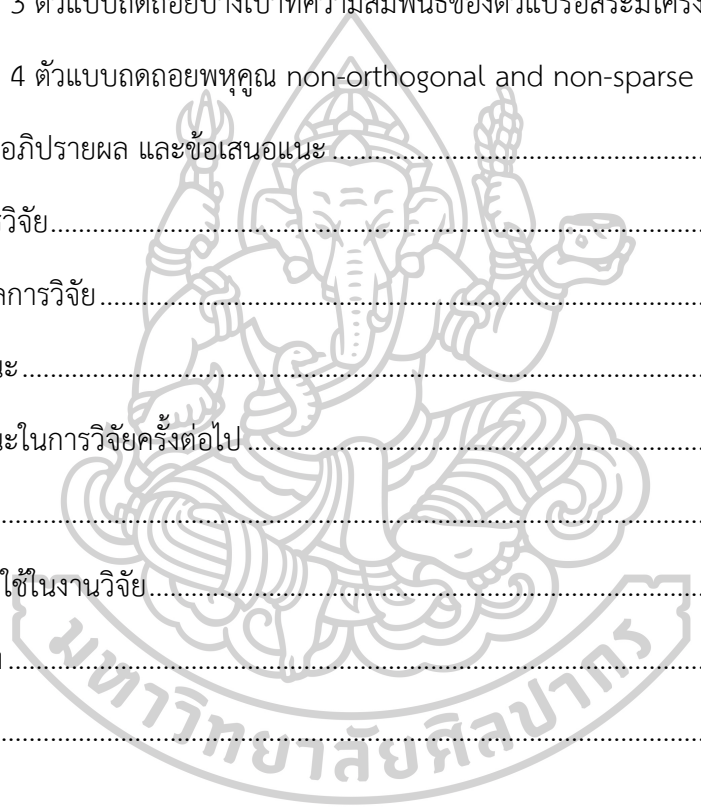
พัชราภรณ์ พรดำเนินสวัสดิ์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ญ
สารบัญภาพ.....	ฐ
บทที่ 1 บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา.....	1
วัตถุประสงค์การทำวิจัย.....	3
ขอบเขตของงานวิจัย.....	3
นิยามศัพท์เฉพาะ.....	5
ประโยชน์ที่คาดว่าจะได้รับ.....	5
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	6
ทฤษฎีที่เกี่ยวข้อง.....	6
1. ตัวแบบการถดถอยเชิงเส้นพหุคูณ (Multiple linear regression model).....	6
2. การถดถอยเชิงเส้นพหุคูณด้วยวิธีกำลังสองน้อยที่สุด (OLS regression).....	7
3. การถดถอยเชิงเส้นพหุคูณด้วยวิธีริดจ์ (Ridge Regression).....	15
4. การถดถอยเชิงเส้นพหุคูณด้วยวิธีแลซโซ (Least absolute selection and shrinkage operator regression : LASSO regression).....	26
5. การถดถอยเชิงเส้นพหุคูณด้วยวิธีแลซโซปรับได้ (Adaptive LASSO regression).....	40
6. การถดถอยเชิงเส้นพหุคูณด้วยวิธีรีแลกซ์ แลซโซ (Relaxed LASSO regression).....	43

7. การหาพารามิเตอร์การปรับแต่ง (Tuning parameter) โดยวิธี Cross-validation	49
งานวิจัยที่เกี่ยวข้อง	50
บทที่ 3 วิธีดำเนินการวิจัย	53
ขอบเขตของการวิจัยและการทดลอง	53
เกณฑ์ที่ใช้ในการพิจารณา.....	57
ขั้นตอนการวิจัย	58
ขั้นตอนการวิจัยของวิธีการหาตัวประมาณพารามิเตอร์ด้วยวิธีริคต์.....	58
ขั้นตอนการวิจัยของวิธีการหาตัวประมาณพารามิเตอร์ด้วยวิธีแลชโซ.....	60
ขั้นตอนการวิจัยของวิธีการหาตัวประมาณพารามิเตอร์ด้วยวิธีแลชโซปรับได้.....	61
ขั้นตอนการวิจัยของวิธีการหาตัวประมาณพารามิเตอร์ด้วยวิธีรีแลกซ์แลชโซ	62
บทที่ 4 ผลการวิเคราะห์ข้อมูล	64
ความถูกต้องของการพยากรณ์ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริคต์, LASSO, Adaptive LASSO และ Relaxed LASSO.....	65
กรณีที่ 1 ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระเป็นเชิงตั้งฉากบางส่วน	66
กรณีที่ 2 ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก	68
กรณีที่ 3 ตัวแบบถดถอยบางเบาที่ความสัมพันธ์ของตัวแปรอิสระมีโครงสร้างแบบแบ่งกลุ่ม	70
กรณีที่ 4 ตัวแบบถดถอยไม่บางเบาและเมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก	71
ความถูกต้องของการประมาณค่าพารามิเตอร์ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริคต์, LASSO, Adaptive LASSO และ Relaxed LASSO.....	73
กรณีที่ 1 ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระเป็นเชิงตั้งฉากบางส่วน	74
กรณีที่ 2 ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก	76
กรณีที่ 3 ตัวแบบถดถอยบางเบาที่ความสัมพันธ์ของตัวแปรอิสระมีโครงสร้างแบบแบ่งกลุ่ม	77

กรณีที่ 4 ตัวแบบถดถอยไม่บางเบาและเมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก	78
.....	
ความถูกต้องในการคัดเลือกตัวแปรอิสระ	79
กรณีที่ 1 ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระเป็นเชิงตั้งฉากบางส่วน	80
.....	
กรณีที่ 2 ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก	86
กรณีที่ 3 ตัวแบบถดถอยบางเบาที่ความสัมพันธ์ของตัวแปรอิสระมีโครงสร้างแบบแบ่งกลุ่ม	91
กรณีที่ 4 ตัวแบบถดถอยพหุคูณ non-orthogonal and non-sparse model.....	94
บทที่ 5 สรุป อภิปรายผล และข้อเสนอแนะ	98
สรุปผลการวิจัย	98
อภิปรายผลการวิจัย	102
ข้อเสนอแนะ	103
ข้อเสนอแนะในการวิจัยครั้งต่อไป	103
ภาคผนวก	104
โปรแกรมที่ใช้ในงานวิจัย	105
รายการอ้างอิง	118
ประวัติผู้เขียน	120



สารบัญตาราง

หน้า

ตารางที่ 1 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Med_PMSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริตจ์, LASSO, Adaptive LASSO และ Relaxed LASSO กรณี partial orthogonal sparse model.....	67
ตารางที่ 2 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Med_PMSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริตจ์, LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal sparse model	69
ตารางที่ 3 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Med_PMSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริตจ์, LASSO, Adaptive LASSO และ Relaxed LASSO กรณี grouping structure sparse model	71
ตารางที่ 4 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Med_PMSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริตจ์, LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal and non-sparse model.....	72
ตารางที่ 5 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย (Med_MSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริตจ์, LASSO, Adaptive LASSO และ Relaxed LASSO กรณี partial orthogonal sparse model.....	74
ตารางที่ 6 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย (Med_MSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริตจ์, LASSO, Adaptive LASSO และ Relaxed LASSO กรณี partial non-orthogonal sparse model.....	76
ตารางที่ 7 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย (Med_MSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริตจ์, LASSO, Adaptive LASSO และ Relaxed LASSO กรณี grouping structure sparse model.....	77
ตารางที่ 8 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย (Med_MSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริตจ์, LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal and non-sparse model.....	78

ตารางที่ 9 ค่าเฉลี่ยของ IC1 (Mean_IC1) และค่าร้อยละของ IC1 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (%IC1) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี partial orthogonal sparse model 80

ตารางที่ 10 ค่าเฉลี่ยของ IC2 (Mean_IC2) และค่าร้อยละของ IC2 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (%IC2) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี partial orthogonal sparse model 82

ตารางที่ 11 ค่าเฉลี่ยของจำนวนตัวแปรอิสระในสมการถดถอย (Mean_covariate) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี partial orthogonal sparse model 84

ตารางที่ 12 ค่าเฉลี่ยของ IC1 (Mean_IC1) และค่าร้อยละของ IC1 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (%IC1) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal sparse model 86

ตารางที่ 13 ค่าเฉลี่ยของ IC2 (Mean_IC2) และค่าร้อยละของ IC2 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (%IC2) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal sparse model 88

ตารางที่ 14 ค่าเฉลี่ยของจำนวนตัวแปรอิสระในสมการถดถอย (Mean_Covariate) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal sparse model 90

ตารางที่ 15 ค่าเฉลี่ยของ IC1 (Mean_IC1) และค่าร้อยละของ IC1 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (%IC1) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี grouping structure sparse model 91

ตารางที่ 16 ค่าเฉลี่ยของ IC2 (Mean_IC2) และค่าร้อยละของ IC2 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (%IC2) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี grouping structure sparse model 92

ตารางที่ 17 ค่าเฉลี่ยของจำนวนตัวแปรอิสระในสมการถดถอย (Mean_Covariate) ของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยแบบ LASSO, Adaptive LASSO และ Relaxed LASSO กรณี grouping structure sparse model 93

ตารางที่ 18 ค่าเฉลี่ยของ IC1 (Mean_IC1) และค่าร้อยละของ IC1 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (%IC1) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal and non-sparse model..... 94

ตารางที่ 19 ค่าเฉลี่ยของจำนวนตัวแปรอิสระในสมการถดถอย (Mean_Covariate) ของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยแบบ LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal and non-sparse model..... 96

ตารางที่ 20 เปรียบเทียบประสิทธิภาพของวิธีการประมาณ จำแนกตามตัวแบบการถดถอยพหุคูณ 101



สารบัญภาพ

หน้า

ภาพที่ 1	กราฟแสดงพื้นที่ผิวตอบสนองและคอนทิวรัลของผลบวกกำลังสองของส่วนเหลือจากสมการถดถอยพหุคูณของค่าประมาณของสัมประสิทธิ์การถดถอย	17
ภาพที่ 2	กราฟแสดงริจต่ำสุดและเส้นทางของพิกัดของริจต่ำสุดของผลบวกกำลังสองของส่วนเหลือ ภายใต้ข้อจำกัด.....	18
ภาพที่ 3	กราฟแสดงความสัมพันธ์ระหว่างความแปรปรวนร่วม ค่ากำลังสองของความเอนเอียง และค่า k	25
ภาพที่ 4	กราฟแสดงวิธีการหาตัวประมาณ LASSO และเส้นทางของพิกัดของสัมประสิทธิ์การถดถอยของตัวประมาณ LASSO ภายใต้ข้อจำกัด.....	28
ภาพที่ 5	แสดงความสัมพันธ์ของการสัมประสิทธิ์การถดถอยโดยวิธีต่าง ๆ กับสัมประสิทธิ์การถดถอยวิธีกำลังสองน้อยที่สุด ในกรณีของเมทริกซ์ตั้งฉากปกติ เมื่อ $\gamma = 2$	33
ภาพที่ 6	การประมาณค่าพารามิเตอร์สัมประสิทธิ์การถดถอยแบบ LASSO และแบบบริดจ์ ในกรณี $p=2$	35
ภาพที่ 7	คู่อันดับของสัมประสิทธิ์การถดถอยวิธี LASSO และ วิธีริจ ภายใต้ข้อจำกัดของพารามิเตอร์ของทั้งสองวิธี สำหรับตัวแปรอิสระ 2 ตัว โดยเส้นประสัมพันธ์กับค่า $\rho = 0, 0.23, 0.45, 0.68$ และ 0.90 จากล่างขึ้นบน	36
ภาพที่ 8	รูปแบบของการสัมประสิทธิ์การถดถอยโดยวิธีต่าง ๆ ในกรณี orthonormal design เมื่อ $\lambda=2$ และ	42
ภาพที่ 9	รูปแบบของการสัมประสิทธิ์การถดถอยโดยวิธี Relaxed LASSO ในกรณี orthonormal design เมื่อ $\lambda=1$ และ $\phi = 0, 1/3, 2/3$ และ 1	46
ภาพที่ 10	ความสัมพันธ์ระหว่างพารามิเตอร์พีนอลตี้ λ และตัวประมาณ LASSO และ Relaxed LASSO สำหรับข้อมูลที่มีตัวแปรอิสระ 3 ตัว	48
ภาพที่ 11	แสดงการปรับค่าสัมประสิทธิ์การถดถอยวิธี LASSO ในตัวอย่างข้อมูลมะเร็งต่อมลูกหมาก โดยแต่ละเส้นแสดงความสัมพันธ์ระหว่างสัมประสิทธิ์การถดถอยและพารามิเตอร์ ของวิธี LASSO คือเส้นประแสดงตัวแบบสำหรับ $s = 0.44$ ซึ่งถูกเลือกโดยวิธี generalized cross-validation.....	50

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ความก้าวหน้าทางเทคโนโลยีในปัจจุบันช่วยให้สามารถจัดเก็บข้อมูลขนาดใหญ่ในรูปแบบอิเล็กทรอนิกส์ได้อย่างมีประสิทธิภาพ โดยมีการจัดเก็บข้อมูลในด้านต่างๆ ซึ่งข้อมูลบางประเภทมีตัวแปรจำนวนมากและมีความซับซ้อน ซึ่งเรียกข้อมูลที่มีตัวแปรอิสระจำนวนมากและมีจำนวนตัวแปรอิสระ (p) มากกว่าขนาดตัวอย่าง (n) หรือ $p > n$ ว่า ข้อมูลมิติสูง (High dimensional data) กรณีที่ข้อมูลมีมิติสูง ซึ่งมีจำนวนตัวแปรอิสระจำนวนมากมหาศาลนั้นมีโอกาสที่ทำให้ตัวแปรอิสระเพียงบางส่วนมีความสัมพันธ์กับตัวแปรตามและตัวแปรอิสระส่วนใหญ่ไม่มีความสัมพันธ์กับตัวแปรตาม หรือเมื่อตัวแปรอิสระมีจำนวนมาก ทำให้โอกาสที่ตัวแปรอิสระจะมีความสัมพันธ์เชิงเส้นสูงนั้นมีมาก หรือที่เรียกว่า เกิดความสัมพันธ์เชิงเส้นแบบพหุ (Multicollinearity) ส่งผลให้ตัวแบบการถดถอยเชิงเส้นที่ได้เป็นตัวแบบบางเบา (sparse model) คือ เป็นตัวแบบการถดถอยเชิงเส้นที่มีตัวแปรอิสระจำนวนมาก และมีค่าสัมประสิทธิ์การถดถอยส่วนใหญ่เป็นศูนย์และบางส่วนไม่เป็นศูนย์ หรือมีโอกาสที่ตัวแปรอิสระส่วนใหญ่มีความสัมพันธ์กับตัวแปรตาม จึงส่งผลให้ตัวแปรอิสระส่วนใหญ่อยู่ในตัวแบบการถดถอยเชิงเส้น ซึ่งเรียกว่า ตัวแบบไม่บางเบา (non-sparse model) ตัวอย่างข้อมูลมิติสูง เช่น ข้อมูลเกี่ยวกับไมโครแอเรย์ของดีเอ็นเอ (DNA microarray) ซึ่งเป็นการศึกษาการแสดงออกของกลุ่มยีน เป็นต้น มีผู้วิจัยหลายท่านที่นำวิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณแบบพินอลไลซ์มาประยุกต์ใช้กับการวิเคราะห์การถดถอยกรณีข้อมูลมิติสูง เช่น งานวิจัยของฮัง, มา และซาง (Huang, Ma and Zhang, 2008: 1603-1618) นำข้อมูลเกี่ยวกับการแสดงออกของยีนในดวงตาและโรคที่เกี่ยวข้องกับดวงตาของสัตว์เลี้ยงลูกด้วยน้ำนมรุ่นลูกเพศผู้อายุ 12 สัปดาห์ จำนวน 120 ตัว ที่เกิดจากรุ่นพ่อแม่ที่ผสมข้ามสายพันธุ์ แล้วนำเนื้อเยื่อจากดวงตาของรุ่นลูกมาวิเคราะห์ ซึ่งไมโครแอเรย์ใช้สำหรับวิเคราะห์ RNA ประกอบด้วยยีนมากกว่า 31,042 ชุดที่แตกต่างกัน แต่ฮัง, มา และซางสนใจเฉพาะยีนที่สัมพันธ์กับยีน TRIM32 ที่แสดงโรคเกี่ยวกับดวงตา ซึ่งมีจำนวน 18,975 ยีน ดังนั้นข้อมูลที่ได้จะเป็นข้อมูลมิติสูง ฮัง, มา และซาง ได้นำข้อมูลนี้มาประยุกต์ใช้กับวิธีการวิเคราะห์ข้อมูลวิธี Adaptive LASSO เป็นต้น เมื่อข้อมูลที่น่ามาวิเคราะห์นั้นเป็นข้อมูลมิติสูง การสร้างตัวแบบการพยากรณ์และกระบวนการคัดเลือกตัวแปรอิสระจึงได้รับความสนใจเป็นอย่างมาก

การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) เป็นวิธีการทางสถิติที่ใช้ศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระ X_1, \dots, X_p และตัวแปรตาม Y ในรูปแบบของตัวแบบการถดถอยเชิงเส้น $Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi} + \varepsilon_i, i = 1, \dots, n$ โดยมีสัมประสิทธิ์การถดถอย β_1, \dots, β_p แสดงความสัมพันธ์ระหว่างตัวแปรอิสระ X_1, \dots, X_p และตัวแปรตาม Y ตามลำดับ และ ε_i คือ ความคลาดเคลื่อนสุ่ม โดยที่ตัวแบบถดถอยเชิงเส้นสามารถเขียนอยู่ในรูปแบบของเมทริกซ์ คือ $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ เมื่อ $\mathbf{Y} = (Y_1, \dots, Y_n)'$ เป็นเวกเตอร์ของค่าสังเกตของตัวแปรตาม และ \mathbf{X} เป็นเมทริกซ์ของตัวแปรอิสระขนาด $n \times (p + 1)$ และ $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$ เป็นเวกเตอร์ของพารามิเตอร์ และ $\boldsymbol{\varepsilon}$ เป็นเวกเตอร์ของความคลาดเคลื่อนสุ่ม ตัวประมาณสัมประสิทธิ์การถดถอยที่นิยมใช้กันอย่างแพร่หลายคือ ตัวประมาณด้วยวิธีกำลังสองน้อยที่สุด (Ordinary Least Square : OLS) ซึ่งเป็นตัวประมาณที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนสุ่มมีค่าน้อยที่สุด ซึ่งตัวประมาณนี้สามารถเขียนอยู่ในรูป $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ ในกรณีที่ข้อมูลที่ใช้ในการวิเคราะห์เป็นข้อมูลที่มีมิติสูง จะส่งผลให้เกิดปัญหาในการประมาณค่าสัมประสิทธิ์การถดถอย เนื่องจากเมทริกซ์ $\mathbf{X}'\mathbf{X}$ จะเป็นเมทริกซ์เอกฐาน (singular matrix) จึงไม่สามารถหาตัวประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยที่สุดได้ หรือมีโอกาสเกิดความสัมพันธ์เชิงเส้นแบบพหุ ส่งผลให้ตัวประมาณโดยวิธีกำลังสองน้อยที่สุดนั้นมีความแปรปรวนสูง และทำให้การประมาณค่าพารามิเตอร์ไม่ถูกต้อง นอกจากนี้ การแปลผลลัพธ์ของตัวแบบมีความยากและซับซ้อน เนื่องจากมีจำนวนตัวแปรอิสระเป็นจำนวนมากอยู่ในตัวแบบ ดังนั้นตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุดจึงเป็นตัวประมาณที่ไม่เหมาะสมในกรณีข้อมูลมิติสูง หรือเกิดความสัมพันธ์เชิงเส้นแบบพหุ

วิธีการวิเคราะห์การถดถอยเชิงเส้นสำหรับข้อมูลมิติสูงมีหลายวิธี เช่น วิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณแบบพินอลไลซ์ (Penalized Multiple Linear Regression) และวิธีเบย์ส (Bayesian approach) เป็นต้น ในงานวิจัยนี้จะขอกล่าวถึงวิธีการวิเคราะห์การถดถอยเชิงเส้นพหุคูณแบบพินอลไลซ์เท่านั้น ซึ่งเป็นวิธีหนึ่งที่นิยมใช้กันอย่างแพร่หลายในการประมาณค่าพารามิเตอร์เมื่อข้อมูลมิติสูง โดยมีหลักการคือหาตัวประมาณของเวกเตอร์พารามิเตอร์ $\boldsymbol{\beta}$ ที่ทำให้ $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ มีค่าต่ำสุด ภายใต้ข้อจำกัด (constraint) ของฟังก์ชันของ $\boldsymbol{\beta}$ ที่เรียกว่า ฟังก์ชันพินอลตี้ (penalty function) เขียนแทนด้วย $P(\|\boldsymbol{\beta}\|_q^2)$ เมื่อ $\|\boldsymbol{\beta}\|_q$ เป็น L_q นอร์ม (L_q norm) ของ $\boldsymbol{\beta}$ นั่นคือ

$$\hat{\boldsymbol{\beta}} = \operatorname{argmin}_{\boldsymbol{\beta}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \quad \text{ภายใต้ข้อจำกัด } P(\|\boldsymbol{\beta}\|_q^2) < t \quad \text{เมื่อ } t$$
 เป็นค่าคงที่ ซึ่งสามารถเขียนอยู่ในรูปสมการที่สอดคล้องกัน คือ

$$\hat{\boldsymbol{\beta}} = \operatorname{argmin}_{\boldsymbol{\beta}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \lambda P(\|\boldsymbol{\beta}\|_q^2)$$

เมื่อ λ คือ ตัวคูณลากรองจ์ (Lagrange multiplier) การกำหนดฟังก์ชันพินอลตี้ต่างกัน เป็นผลให้ตัวประมาณแบบพินอลไลซ์แตกต่างกัน

จากที่กล่าวมาข้างต้น จะเห็นว่า ข้อมูลที่ใช้ในการวิเคราะห์ส่วนใหญ่เป็นข้อมูลที่มีมิติสูง โดยตัวแบบการถดถอยเชิงเส้นพหุคูณกรณีข้อมูลมิติสูงอาจเป็นตัวแบบบางเบาหรือตัวแบบไม่บางเบา ซึ่งการวิเคราะห์การถดถอยเชิงเส้นพหุคูณแบบพินอลไลซ์เป็นวิธีหนึ่งที่ใช้ในการประมาณค่าพารามิเตอร์เมื่อข้อมูลมีมิติสูง ตัวประมาณแบบพินอลไลซ์แตกต่างกัน ส่งผลให้ตัวประมาณมีประสิทธิภาพภายใต้เงื่อนไขและสถานการณ์ที่ต่างกัน ดังนั้นผู้วิจัยจึงมีความสนใจที่จะศึกษาเปรียบเทียบคุณสมบัติของวิธีการประมาณการถดถอยเชิงเส้นพหุคูณแบบพินอลไลซ์ ในตัวแบบการถดถอยเชิงเส้นที่มีมิติสูงในกรณีตัวแบบบางเบาและกรณีตัวแบบไม่บางเบา ภายใต้สถานการณ์ที่แตกต่างกันหลายเงื่อนไขที่ครอบคลุมข้อจำกัดคุณสมบัติของตัวประมาณแต่ละวิธี

วัตถุประสงค์การทำวิจัย

เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณแบบพินอลไลซ์ 4 วิธี ในกรณีที่ข้อมูลมีมิติสูง ได้แก่ วิธีริตจ์, LASSO, Adaptive LASSO และ Relaxed LASSO โดยพิจารณาจากความถูกต้องในการพยากรณ์ ความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอย และความถูกต้องในการคัดเลือกตัวแปรอิสระเข้าในตัวแบบ

ขอบเขตของงานวิจัย

ในงานวิจัยนี้เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสัมประสิทธิ์การถดถอย 4 วิธี ในกรณีที่ข้อมูลมีมิติสูง ได้แก่ วิธีริตจ์, LASSO, Adaptive LASSO และ Relaxed LASSO โดยกำหนดขอบเขตของการวิจัย ดังนี้

1. ขนาดตัวอย่าง (n) ที่ใช้ในการวิจัยครั้งนี้คือ 50, 100 และ 200
2. จำนวนตัวแปรอิสระ (p) ที่ใช้ในการวิจัยครั้งนี้คือ 2 เท่าและ 4 เท่าของขนาดตัวอย่าง นั่นคือ 100, 200, 400 และ 800 ตัวแปร
3. เมทริกซ์ของตัวแปรอิสระ (X) ขนาด $n \times p$ มีการแจกแจงปรกติหลายตัวแปร (Multivariate normal distribution) ที่มีเวกเตอร์ค่าเฉลี่ยเท่ากับเวกเตอร์ศูนย์ และเมทริกซ์ความแปรปรวนร่วม $\Sigma_{p \times p}$
4. ความคลาดเคลื่อนสุ่ม ε มีการแจกแจงปรกติ ที่มีค่าเฉลี่ยเป็นศูนย์ และความแปรปรวนคงที่ โดยกำหนดความแปรปรวน $\sigma^2 = 1.5^2$
5. เวกเตอร์ตัวแปรตาม (Y) สร้างจากตัวแบบการถดถอยเชิงเส้นพหุคูณ $Y = X\beta + \varepsilon$
6. กำหนดสถานการณ์ของการทดลองสำหรับแต่ละค่า n และ p โดยกำหนดรูปแบบความสัมพันธ์ของตัวแปรอิสระ X ในแต่ละกรณี ดังนี้

6.1) ความสัมพันธ์ของตัวแปรอิสระ X เป็นแบบเมทริกซ์เชิงตั้งฉากบางส่วน (partial orthogonal matrix) กล่าวคือ ตัวแปรอิสระแบ่งออกเป็น 2 กลุ่ม โดยตัวแปรอิสระกลุ่มแรกคือตัวแปรอิสระจำนวน 15 ตัวที่สัมพันธ์กัน และตัวแปรอิสระกลุ่มที่สองคือตัวแปรอิสระตัวที่เหลือจำนวน $p - 15$ ตัวที่สัมพันธ์กัน โดยตัวแปรอิสระกลุ่มที่ 1 และกลุ่มที่ 2 เป็นอิสระกัน

6.2) ความสัมพันธ์ของตัวแปรอิสระ X ไม่เป็นเมทริกซ์เชิงตั้งฉาก (non-orthogonal matrix) กล่าวคือ ตัวแปรอิสระแบ่งออกเป็น 2 กลุ่ม โดยตัวแปรอิสระกลุ่มแรกคือตัวแปรอิสระจำนวน 15 ตัวที่สัมพันธ์กัน และตัวแปรอิสระกลุ่มที่สองคือตัวแปรอิสระตัวที่เหลือจำนวน $p - 15$ ตัวที่สัมพันธ์กัน โดยตัวแปรอิสระกลุ่มที่ 1 และกลุ่มที่ 2 มีความสัมพันธ์กัน

6.3) ความสัมพันธ์ของตัวแปรอิสระ X มีโครงสร้างแบบแบ่งกลุ่ม (grouping structure) กล่าวคือ ตัวแปรอิสระจำนวน p ตัว แบ่งออกเป็น 4 กลุ่ม โดยตัวแปรอิสระแต่ละกลุ่มอิสระต่อกัน และตัวแปรอิสระกลุ่มที่ 1, 2 และ 3 มีความสัมพันธ์ภายในกลุ่มกันสูง

7. ตัวแบบการถดถอยแบ่งเป็น 2 แบบ ได้แก่ ตัวแบบการถดถอยแบบบางเบา คือ การกำหนดค่าสัมประสิทธิ์การถดถอยจำนวน 15 ตัวไม่เท่ากับศูนย์ และจำนวน $p-15$ ตัวที่เหลือให้เท่ากับศูนย์ เพื่อกำหนดให้ตัวแปรอิสระส่วนน้อยอยู่ในตัวแบบ และตัวแบบการถดถอยแบบไม่บางเบา คือ การกำหนดค่าสัมประสิทธิ์การถดถอยทุกตัวไม่เท่ากับศูนย์ เพื่อกำหนดให้ตัวแปรอิสระทุกตัวอยู่ในตัวแบบ

8. เกณฑ์ที่ใช้ในการพิจารณาเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริคต์, LASSO, Adaptive LASSO และ Relaxed LASSO ได้แก่

8.1) ประสิทธิภาพของการพยากรณ์โดยวัดจากค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์

8.2) ประสิทธิภาพของการประมาณโดยวัดจากค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์การถดถอย

8.3) ความความผิดพลาดของการเลือกตัวประมาณ โดยความผิดพลาดในการคัดเลือกตัวแปรมี 2 แบบ คือ กรณีที่ค่าพารามิเตอร์ไม่เท่ากับศูนย์ แต่ตัวแปรอิสระไม่ถูกคัดเลือกเข้าสู่ตัวแบบ (IC1) และกรณีที่ค่าพารามิเตอร์เท่ากับศูนย์ แต่ตัวแปรอิสระถูกคัดเลือกเข้าสู่ตัวแบบ (IC2) แล้วจึงพิจารณาค่าเฉลี่ย, ค่าร้อยละของความผิดพลาด และจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ย

9. การศึกษาวิจัยครั้งนี้ใช้โปรแกรม R ในการจำลองข้อมูลและการประมาณค่าพารามิเตอร์ โดยการซ้ำ 500 รอบในแต่ละสถานการณ์

นิยามศัพท์เฉพาะ

ประสิทธิภาพของตัวประมาณ (Performance of estimator) คือ การวัดสมบัติของตัวประมาณสัมประสิทธิ์การถดถอยใน 3 ด้าน คือ ความถูกต้องในการคัดเลือกตัวแปรอิสระเข้าในตัวแบบ ความถูกต้องในการพยากรณ์ และความถูกต้องในการคัดเลือกตัวแปรอิสระเข้าในตัวแบบ

ข้อมูลมิติสูง (High-Dimensional data) คือ ข้อมูลที่มีจำนวนตัวแปรอิสระจำนวนมาก และมีจำนวนมากกว่าขนาดตัวอย่าง

การถดถอยเชิงเส้นแบบพินอลไลซ์ (Penalized Linear Regression) เป็นวิธีการวิเคราะห์การถดถอยที่หาตัวประมาณพารามิเตอร์ที่ทำให้ผลรวมกำลังสองของความคลาดเคลื่อนมีค่าต่ำสุด ภายใต้ข้อจำกัดของฟังก์ชันของตัวประมาณพารามิเตอร์

ตัวแบบถดถอยเบาบาง (Sparse Model) คือ ตัวแบบการถดถอยเชิงเส้นที่มีตัวแปรอิสระจำนวนมาก และมีค่าสัมประสิทธิ์การถดถอยส่วนใหญ่เป็นศูนย์ และบางส่วนไม่เป็นศูนย์ ซึ่งอาจมาจากมีเพียงตัวแปรอิสระเพียงบางส่วนมีความสัมพันธ์กับตัวแปรตาม หรือตัวแปรอิสระเกิดความสัมพันธ์เชิงเส้นแบบพหุ

ประโยชน์ที่คาดว่าจะได้รับ

1. ทราบประสิทธิภาพของตัวประมาณสัมประสิทธิ์การถดถอยในกรณี que ศึกษาตามขอบเขตงานวิจัย
2. สามารถนำไปประยุกต์ใช้กับการวิเคราะห์การถดถอยเชิงเส้นพหุคูณกับข้อมูลจริงที่มีลักษณะเป็นข้อมูลมิติสูง

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณแบบพินอลโลซ์ 4 วิธี คือ วิธีริตจ, LASSO, Adaptive LASSO และ Relaxed LASSO ซึ่งในการเปรียบเทียบประสิทธิภาพนั้นจะพิจารณาจากความถูกต้องในการพยากรณ์ ความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอย และความถูกต้องในการคัดเลือกตัวแปรอิสระเข้าในตัวแบบ ภายใต้สถานการณ์ที่แตกต่างกัน ในบทนี้จะแบ่งออกเป็น 2 ส่วน คือ ส่วนแรก คือ ทฤษฎีที่เกี่ยวข้อง ได้แก่ การวิเคราะห์การถดถอยพหุคูณ การประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดและวิธีการประมาณค่าจากภาวะน่าจะเป็นสูงสุด วิธีการวิเคราะห์การถดถอยแบบริตจ, LASSO, Adaptive LASSO และ Relaxed LASSO ส่วนที่สอง คือ งานวิจัยที่เกี่ยวข้อง

ทฤษฎีที่เกี่ยวข้อง

1. ตัวแบบการถดถอยเชิงเส้นพหุคูณ (Multiple linear regression model)

การถดถอยเชิงเส้นพหุคูณเป็นการหาตัวแบบที่ใช้อธิบายลักษณะความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระและตัวแปรตาม ซึ่งตัวแบบถดถอยประกอบไปด้วยส่วนที่เป็นฟังก์ชันอธิบายลักษณะความสัมพันธ์ และส่วนที่เป็นความคลาดเคลื่อนสุ่ม ดังนั้น ตัวแบบการถดถอยเชิงเส้นพหุคูณ (Multiple linear regression model) สำหรับตัวแปรอิสระ p ตัว สามารถเขียนอยู่ในรูปแบบดังนี้

$$Y = \beta_0 \mathbf{1} + X\beta + \varepsilon \quad (1)$$

โดยที่ $Y = (Y_1, \dots, Y_n)'$ คือ เวกเตอร์ของค่าสังเกตของตัวแปรตาม Y_i เมื่อ $i = 1, \dots, n$

X คือ เมทริกซ์ของตัวแปรอิสระขนาด $n \times p$ โดยที่ $X = (X_1, X_2, \dots, X_n)'$ เมื่อ

$X_i = (X_{i1}, X_{i2}, \dots, X_{ip})'$ เป็นเวกเตอร์ของค่าสังเกตตัวที่ i ของตัวแปรอิสระ p ตัว

β_0 คือ ค่าของจุดตัดแกน Y (y-intercept) คือ ค่าของ Y_i เมื่อ $X_i = \mathbf{0}$

$\beta = (\beta_1, \dots, \beta_p)'$ คือ เวกเตอร์ของสัมประสิทธิ์การถดถอย ขนาด $p \times 1$

$\mathbf{1}$ คือ เวกเตอร์ที่มีสมาชิกเป็น 1 ขนาด $n \times 1$

และ ε คือ เวกเตอร์ของความคลาดเคลื่อนสุ่ม ขนาด $n \times 1$ ซึ่ง $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$

โดยที่ข้อสมมติเกี่ยวกับความคลาดเคลื่อนสุ่ม ε_i สำหรับ $i = 1, 2, \dots, n$ มีดังนี้

1) ε_i เป็นตัวแปรสุ่ม ที่มีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนคงที่เท่ากับ σ^2 นั่นคือ $E[\varepsilon_i] = 0$ และ $\text{var}[\varepsilon_i] = \sigma^2$ ทุกค่า i

2) ε_i และ ε_j ไม่มีสหสัมพันธ์กัน (Uncorrelated) นั่นคือ $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ เมื่อ $i \neq j$

3) ε_i มีการแจกแจงที่เป็นลักษณะเดียวกันและเป็นอิสระกัน (Identically Independent Distribute : i.i.d.) ที่มีค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนเท่ากับ σ^2

ดังนั้น เวกเตอร์ค่าเฉลี่ยและเมทริกซ์ความแปรปรวนร่วม (Covariance matrix) ของ $\boldsymbol{\varepsilon}$ คือ $E[\boldsymbol{\varepsilon}] = \mathbf{0}$ และ $\text{cov}[\boldsymbol{\varepsilon}] = \sigma^2 \mathbf{I}_n$

จากตัวแบบการถดถอยเชิงเส้นพหุคูณ (1) ซึ่งเป็นตัวแบบในลักษณะเชิงเส้น โดยเทอม $\beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta}$ เป็นค่าคงที่ ภายใต้ข้อสมมติของ $\boldsymbol{\varepsilon}$ ดังกล่าวข้างต้น จะได้เวกเตอร์ \mathbf{Y} มีการแจกแจงที่เป็นลักษณะเดียวกันและเป็นอิสระกัน ที่มีเวกเตอร์ค่าเฉลี่ยและเมทริกซ์ความแปรปรวนร่วม ดังนี้

$$E[\mathbf{Y}] = E[\beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}] = \beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta} + E[\boldsymbol{\varepsilon}] = \beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta}$$

$$\text{cov}[\mathbf{Y}] = \text{cov}[\beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}] = \text{cov}[\boldsymbol{\varepsilon}] = \sigma^2 \mathbf{I}_n$$

การแจกแจงของเวกเตอร์ตัวแปรตาม \mathbf{Y} ดังกล่าวนอกจากจะทำให้เราทราบรูปร่างลักษณะการกระจายตัวของข้อมูลหรือความแปรปรวนของเวกเตอร์ \mathbf{Y} แล้ว ยังทำให้สามารถหาค่าเฉลี่ยของเวกเตอร์ \mathbf{Y} เมื่อกำหนดค่า $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip})'$ นั่นคือ

$$E[Y|\mathbf{X}_i] = \beta_0 + \mathbf{X}_i\boldsymbol{\beta} \quad (2)$$

2. การถดถอยเชิงเส้นพหุคูณด้วยวิธีกำลังสองน้อยที่สุด (OLS regression)

วิธีการประมาณค่าพารามิเตอร์ $\boldsymbol{\beta}$ ที่อยู่ในตัวแบบการถดถอยเชิงเส้นพหุคูณในสมการ

(1) มีหลายวิธี ซึ่งวิธีดั้งเดิมที่รู้จักกันแพร่หลาย คือ วิธีกำลังสองน้อยที่สุด

2.1 การประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยที่สุด

จากตัวแบบการถดถอยเชิงเส้นพหุคูณ $\mathbf{Y} = \beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ กำหนดตัวอย่างสุ่ม (X_{ij}, Y_i) เมื่อ $i = 1, 2, \dots, n$, $j = 1, \dots, p$ โดยไม่เสียนัยทั่วไป (without loss of generality) เราสมมติให้ X_{ij} ถูกแปลงให้เป็นค่ามาตรฐาน (standard score) คือ $\sum_{i=1}^n X_{ij} / n = 0$ และ $\sum_{i=1}^n X_{ij}^2 / n = 1$ และ Y_i เป็นตัวแปรที่มีค่าเฉลี่ยเป็นศูนย์ (centered variable) นั่นคือ $\sum_{i=1}^n Y_i / n = 0$ ดังนั้นค่าจุดตัดแกน Y จึงไม่จำเป็นต้องใส่ไว้ในสมการถดถอย นั่นคือ $\beta_0 \mathbf{1} = \mathbf{0}$

หลักการของการประมาณค่าพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุด คือ การหาตัวประมาณของเวกเตอร์พารามิเตอร์ β ที่ทำให้เส้นถดถอยของตัวอย่างลากผ่านหรือใกล้เคียงกับค่าสังเกตมากที่สุด โดยพิจารณาจากค่า ε_i ซึ่งเป็นผลต่างระหว่างค่าสังเกตและค่าคาดหวังของตัวแปรตาม Y ณ ค่าเวกเตอร์ X_i ใดๆ นั่นคือ

$$\varepsilon_i = Y_i - E[Y|X_i] = Y_i - X_i\beta \quad (3)$$

ตัวประมาณสัมประสิทธิ์การถดถอยที่ได้จากวิธีกำลังสองน้อยที่สุด คือ ตัวประมาณที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนสุ่ม $\sum_{i=1}^n \varepsilon_i^2 = (Y - X\beta)'(Y - X\beta)$ มีค่าน้อยที่สุด (Draper, 1981) ซึ่งสามารถเขียนอยู่ในรูปผลคูณของเวกเตอร์ $\varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta)$ ดังนี้

$$\begin{aligned} \varepsilon'\varepsilon &= (Y - X\beta)'(Y - X\beta) = Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta \\ &= Y'Y - 2\beta'X'Y + \beta'X'X\beta \end{aligned} \quad (4)$$

การหาค่าประมาณ β ที่ทำให้ $(Y - X\beta)'(Y - X\beta)$ มีค่าน้อยที่สุด สามารถทำได้ โดยการหาอนุพันธ์ของ $(Y - X\beta)'(Y - X\beta)$ เทียบกับ β แล้วกำหนดให้เท่ากับศูนย์ ดังนี้

$$\begin{aligned} \frac{\partial (Y - X\beta)'(Y - X\beta)}{\partial \beta} &= 0 - 2X'Y + 2X'X\hat{\beta} = 0 \\ X'X\hat{\beta} &= X'Y \end{aligned} \quad (5)$$

เรียกสมการที่ (5) ว่า สมการปกติ (Normal Equation)

เมื่อ X เป็นเมทริกซ์ลำดับเต็ม (full rank matrix) คือ ตัวแปรอิสระทุกตัวเป็นอิสระต่อกัน จะทำให้เมทริกซ์ $X'X$ เป็นเมทริกซ์ไม่เอกฐาน (Nonsingular Matrix) นั่นคือ สามารถหาค่า $(X'X)^{-1}$ ได้ ซึ่งหาค่าประมาณ $\hat{\beta}$ จากการใช้ $(X'X)^{-1}$ คูณสมการ (5) ทั้งสองข้าง จะได้

$$\begin{aligned} (X'X)^{-1}X'X\hat{\beta} &= (X'X)^{-1}X'Y \\ \hat{\beta} &= (X'X)^{-1}X'Y \end{aligned} \quad (6)$$

ดังนั้น ตัวประมาณของสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด คือ $\hat{\beta}^{ols} = (X'X)^{-1}X'Y$

การวิเคราะห์การถดถอยเชิงเส้นพหุคูณด้วยวิธีกำลังสองน้อยที่สุดมีเงื่อนไขว่าขนาดตัวอย่างต้องมากกว่าจำนวนตัวแปรอิสระ แต่เมื่อข้อมูลที่ใช้ในการวิเคราะห์เป็นข้อมูลมิติสูง กล่าวคือ จำนวนตัวแปรอิสระมากกว่าขนาดตัวอย่าง ส่งผลให้เมทริกซ์ $X'X$ จะเป็นเมทริกซ์เอกฐาน (singular matrix) ซึ่งไม่สามารถหาตัวประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยที่สุดได้

วิธีการประมาณค่าสัมประสิทธิ์การถดถอยนอกจากวิธีกำลังสองน้อยที่สุดแล้ว ยังมีวิธีการประมาณค่าจากภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation : MLE) ซึ่งได้ตัวประมาณค่าสัมประสิทธิ์การถดถอยเดียวกับตัวประมาณที่ได้จากวิธีกำลังสองน้อยที่สุด โดยการประมาณค่าพารามิเตอร์ β ด้วยวิธีภาวะน่าจะเป็นสูงสุด จะอาศัยแนวคิดเกี่ยวกับภาวะน่าจะเป็น โดยกำหนดฟังก์ชันภาวะน่าจะเป็น (Likelihood function) ของตัวอย่างสุ่ม แล้วจึงหาค่าสูงสุดของฟังก์ชันนี้โดยการหาอนุพันธ์เทียบกับพารามิเตอร์ที่ยังไม่ทราบค่า ซึ่งในกรณีของการถดถอยพหุคูณที่ขนาดตัวอย่างเท่ากับ n และตัวแปรอิสระ p ตัว จากสมการถดถอย (1) ภายใต้ข้อสมมติเบื้องต้นเกี่ยวกับการแจกแจงของความคลาดเคลื่อนสุ่ม กล่าวคือ ε_i มีการแจกแจงปกติ ที่มีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนร่วม σ^2 และ $E(\varepsilon_i, \varepsilon_j) = 0$ เมื่อ $i \neq j$ ดังนั้น ค่าของตัวแปรตาม Y_i สำหรับ $i = 1, 2, \dots, n$ มีการแจกแจงปกติ ที่มีค่าเฉลี่ย $X_i\beta$ เมื่อ X_i คือ เวกเตอร์แถวที่ i ของเมทริกซ์ X และความแปรปรวนเท่ากัน คือ σ^2 และเนื่องจาก Y_i เป็นฟังก์ชันเชิงเส้นกับ ε_i ส่งผลให้ Y_i และ Y_j เมื่อ $i \neq j$ เป็นอิสระต่อกันด้วย ดังนั้น ฟังก์ชันภาวะน่าจะเป็นของ Y_1, Y_2, \dots, Y_n เขียนได้ดังนี้

$$\begin{aligned} L &= f(Y_1, Y_2, \dots, Y_n) \\ &= \prod_{i=1}^n f(Y_i) \\ &= f(Y_1) \cdot f(Y_2) \cdot \dots \cdot f(Y_n) \end{aligned}$$

ฟังก์ชันภาวะน่าจะเป็นนี้สามารถเขียนในรูปของฟังก์ชันลอการิทึมฐานธรรมชาติ ได้ดังนี้

$$\begin{aligned} \ln L &= \ln f(Y_1) + \ln f(Y_2) + \dots + \ln f(Y_n) \\ &= \sum_{i=1}^n \ln f(Y_i) \end{aligned} \quad (7)$$

เนื่องจาก $Y_i \sim N(X_i\beta, \sigma^2)$ ดังนั้น ฟังก์ชันความน่าจะเป็นของ Y_i คือ

$$f(Y_i) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma^2}(Y_i - X_i\beta)^2\right]$$

นำ $f(Y_i)$ แทนลงในสมการ (7) จะได้

$$\begin{aligned} \ln L &= \sum \ln \left\{ \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma^2}(Y_i - X_i\beta)^2\right] \right\} \\ &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta) \end{aligned} \quad (8)$$

สำหรับการหาตัวประมาณของ β ทำได้โดยการหาอนุพันธ์บางส่วนของสมการ (8) เทียบกับพารามิเตอร์ β แล้วกำหนดให้มีค่าเป็นศูนย์ ดังนี้

$$\begin{aligned}\frac{\partial \ln L}{\partial \beta} &= \frac{\partial}{\partial \beta} \left[-\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\beta)' (\mathbf{Y} - \mathbf{X}\beta) \right] \\ &= \frac{1}{\sigma^2} (\mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{X}\beta)\end{aligned}$$

จากนั้น กำหนด $\frac{\partial \ln L}{\partial \beta}$ ให้มีค่าเป็นศูนย์ แล้วแก้สมการเพื่อหาค่า $\tilde{\beta}$

$$\frac{1}{\sigma^2} (\mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{X}\tilde{\beta}) = \mathbf{0}$$

$$\mathbf{X}'\mathbf{X}\tilde{\beta} = \mathbf{X}'\mathbf{Y} \quad (9)$$

จะเห็นได้ว่า สมการ (9) เป็นสมการปกติในรูปแบบเดียวกับสมการ (5) ดังนั้น ตัวประมาณสัมประสิทธิ์การถดถอยสำหรับ β คือ ตัวประมาณของ β ที่ทำให้ฟังก์ชันภาวะน่าจะเป็นมีค่ามากที่สุด ซึ่งเป็นตัวประมาณเดียวกันกับตัวประมาณที่ได้จากวิธีกำลังสองน้อยที่สุด นั่นคือ $\tilde{\beta} = \hat{\beta}^{ols} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

2.2 สมบัติของตัวประมาณกำลังสองน้อยที่สุด

การหาสมบัติของตัวประมาณกำลังสองน้อยที่สุดของสัมประสิทธิ์การถดถอยสามารถหาได้จากเวกเตอร์ค่าคาดหวัง เมทริกซ์ความแปรปรวนร่วม และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squares Error : MSE) ของตัวประมาณสัมประสิทธิ์การถดถอย โดยตัวประมาณกำลังสองน้อยที่สุดเป็นไปตามทฤษฎีบทเกาส์-มาร์คอฟ (Gauss-Markov Theorem) ซึ่งกล่าวว่า ในบรรดาตัวประมาณเชิงเส้นที่ไม่เอนเอียงทั้งหลาย ตัวประมาณที่ได้จากวิธีกำลังสองน้อยที่สุดจะเป็นตัวประมาณที่มีความแปรปรวนต่ำที่สุด ดังนั้น ตัวประมาณที่ได้จากวิธีกำลังสองน้อยที่สุดจึงเป็นตัวประมาณเชิงเส้นไม่เอนเอียงดีที่สุด (Best Linear Unbiased Estimators : BLUE) ซึ่งสามารถพิสูจน์สมบัติของตัวประมาณ $\hat{\beta}^{ols}$ ได้ดังนี้

2.2.1 สมบัติเชิงเส้นตรง (Linear)

จาก $\hat{\beta}^{ols} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ เมื่อกำหนด $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ เป็นเมทริกซ์ค่าคงที่ จะได้ $\hat{\beta} = \mathbf{A}\mathbf{Y}$ ซึ่งเป็นฟังก์ชันเชิงเส้นของ \mathbf{Y} ดังนั้น $\hat{\beta}^{ols}$ จึงมีความสัมพันธ์เชิงเส้นตรงกับ \mathbf{Y}

2.2.2 สมบัติไม่เอนเอียง (Unbiased)

เวกเตอร์ค่าคาดหวังของตัวประมาณ คือ

$$\begin{aligned}E[\hat{\beta}^{ols}] &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{Y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta = \beta\end{aligned}$$

จะได้ เวกเตอร์ค่าคาดหวังของ $\hat{\beta}^{ols}$ มีค่าเท่ากับ β นั่นคือ ตัวประมาณ $\hat{\beta}^{ols}$ มีสมบัติที่เป็นตัวประมาณที่ไม่เอนเอียง (Unbiased estimator) สำหรับ β

2.2.3 สมบัติความแปรปรวนต่ำที่สุด (Minimum variance)

เมทริกซ์ความแปรปรวนร่วมของตัวประมาณ คือ

$$\begin{aligned}
 \text{cov}[\hat{\beta}^{ols}] &= \text{cov}[(X'X)^{-1}X'Y] \\
 &= ((X'X)^{-1}X')\text{cov}[Y]((X'X)^{-1}X')' \\
 &= ((X'X)^{-1}X')(\sigma^2 I_n)(X(X'X)^{-1}) \\
 &= \sigma^2 (X'X)^{-1}X'X(X'X)^{-1} \\
 &= \sigma^2 (X'X)^{-1}
 \end{aligned} \tag{10}$$

จาก (10) สมาชิกที่ตำแหน่งแนวเส้นทแยงมุม (main diagonal) ของเมทริกซ์ $\sigma^2 (X'X)^{-1}$ คือ ความแปรปรวน (variance) ของ $\hat{\beta}_j$ และสมาชิกที่ตำแหน่งนอกเส้นทแยงมุมเป็น ความแปรปรวนร่วม (covariance) ระหว่าง $\hat{\beta}_i$ และ $\hat{\beta}_j$ โดยที่ $i \neq j$ และเมทริกซ์นี้เป็นเมทริกซ์ สมมาตร (symmetric matrix) นั่นคือ $\text{cov}(\hat{\beta}_i, \hat{\beta}_j) = \text{cov}(\hat{\beta}_j, \hat{\beta}_i)$

ในบรรดาตัวประมาณเชิงเส้นที่ไม่เอนเอียงนั้น ตัวประมาณโดยวิธีกำลังสองน้อยที่สุดจะมีความแปรปรวนต่ำที่สุด ซึ่งสามารถพิสูจน์ได้ดังนี้

กำหนดให้ $\hat{\beta}^*$ แทน เวกเตอร์ของตัวประมาณเชิงเส้นของสัมประสิทธิ์การถดถอย β โดยที่ $\hat{\beta}^* = BY$ และ $B = (X'X)^{-1}X' + C$ เมื่อ C เป็นเมทริกซ์ขนาด $p \times n$ ที่ไม่เป็น เมทริกซ์ศูนย์ จะได้ เมทริกซ์ค่าคาดหวังของ $\hat{\beta}^*$ คือ

$$\begin{aligned}
 E[\hat{\beta}^*] &= E[BY] \\
 &= E[(X'X)^{-1}X' + C]y \\
 &= E[(X'X)^{-1}X' + C](X\beta + \varepsilon) \\
 &= (X'X)^{-1}X'X\beta + CX\beta + CE[\varepsilon] \\
 &= \beta + CX\beta + 0 \\
 &= \beta + CX\beta
 \end{aligned}$$

ดังนั้น ตัวประมาณ $\hat{\beta}^*$ จะเป็นตัวประมาณที่ไม่เอนเอียงของ β เมื่อ $CX = 0$ และ เมทริกซ์ความแปรปรวนร่วมของ $\hat{\beta}^*$ คือ

$$\begin{aligned}
\text{cov}[\hat{\beta}^*] &= \text{cov}[BY] \\
&= B\text{cov}[Y]B' \\
&= \sigma^2 BB' \\
&= \sigma^2[(X'X)^{-1}X' + C][(X'X)^{-1}X' + C]' \\
&= \sigma^2[(X'X)^{-1}X'X(X'X)^{-1} + CX(X'X)^{-1} + (X'X)^{-1}(CX)' + CC'] \\
&= \sigma^2[(X'X)^{-1}X'X(X'X)^{-1} + \mathbf{0}(X'X)^{-1} + (X'X)^{-1}\mathbf{0}' + CC'] \\
&= \sigma^2[(X'X)^{-1} + CC'] \\
&= \sigma^2(X'X)^{-1} + \sigma^2 CC' \\
&= \text{cov}[\hat{\beta}^{ols}] + \sigma^2 CC'
\end{aligned}$$

จะเห็นว่าเมทริกซ์ความแปรปรวนร่วมของ $\hat{\beta}^*$ มีค่าเท่ากับเมทริกซ์ความแปรปรวนร่วมของ $\hat{\beta}^{ols}$ บวกกับเทอม $\sigma^2 CC'$ เนื่องจาก $CC' \geq \mathbf{0}$ ดังนั้น $\sigma^2 CC' \geq \mathbf{0}$ นั้นหมายความว่า $\hat{\beta}^{ols}$ เป็นเวกเตอร์ตัวประมาณที่มีเมทริกซ์ความแปรปรวนร่วมต่ำที่สุด

จากข้างต้นจะได้ว่า ตัวประมาณ $\hat{\beta}^{ols}$ โดยวิธีกำลังสองน้อยที่สุด มีคุณสมบัติครบทั้ง 3 ประการที่กล่าวข้างต้น ดังนั้น ตัวประมาณโดยวิธีกำลังสองน้อยที่สุด ($\hat{\beta}^{ols}$) เป็นตัวประมาณเชิงเส้นไม่เอนเอียงดีที่สุด

2.2.4 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squares Error : MSE)

กำหนดให้ β เป็นตัวประมาณสัมประสิทธิ์การถดถอย β ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\beta}$ เขียนอยู่ในรูปแบบดังนี้

$$\begin{aligned}
\text{MSE}[\hat{\beta}] &= E[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)] \\
&= E[(\hat{\beta} - E[\hat{\beta}] + E[\hat{\beta}] - \beta)'(\hat{\beta} - E[\hat{\beta}] + E[\hat{\beta}] - \beta)] + (E[\hat{\beta}] - \beta)'(E[\hat{\beta}] - \beta) \quad (11)
\end{aligned}$$

เนื่องจาก $E[\hat{\beta}^{ols}] = \beta$ จะได้ว่า

$$\text{bias}[\hat{\beta}^{ols}] = E[\hat{\beta}^{ols}] - \beta = \mathbf{0} \quad \text{และ}$$

$$E[(\hat{\beta}^{ols} - \beta)'(\hat{\beta}^{ols} - \beta)] = \sigma^2 \text{tr}(X'X)^{-1}$$

ดังนั้น ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\beta}^{ols}$ คือ

$$\begin{aligned}
\text{MSE}[\hat{\beta}^{ols}] &= \sigma^2 \text{tr}(X'X)^{-1} + \mathbf{0} \\
&= \sigma^2 \text{tr}(X'X)^{-1} \quad (12)
\end{aligned}$$

2.3 ปัญหาการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดในกรณีข้อมูลมีสหสัมพันธ์เชิงเส้น

เมื่อตัวแปรอิสระเกิดความสัมพันธ์เชิงเส้นพหุ คือ เมทริกซ์ $X'X$ ที่เขียนอยู่ในรูปแบบของเมทริกซ์สหสัมพันธ์ (correlation matrix) ของตัวแปรอิสระ ไม่ใกล้เคียงกับเมทริกซ์เอกลักษณ์ (unit matrix) ซึ่งเรียกรูปแบบความสัมพันธ์ระหว่างตัวแปรอิสระของเมทริกซ์ $X'X$ ในลักษณะนี้ว่าไม่เป็นเชิงตั้งฉาก (non-orthogonal) โดยที่เมทริกซ์ $X'X$ ที่มีลักษณะไม่เป็นเชิงตั้งฉาก จะทำให้ $\det(X'X)$ มีค่าเข้าใกล้ศูนย์และค่าบางค่าหรือสมาชิกบางตัวของเมทริกซ์ผกผันของ $X'X$ มีค่ามาก ๆ หรือเข้าใกล้อนันต์ส่งผลให้ความแปรปรวนของตัวประมาณสัมประสิทธิ์การถดถอยมีค่าสูง ซึ่งโฮเอิร์ลและเคนนาร์ด (Hoerl and Kennard, 1970: 56) ได้กล่าวถึง ข้อดีของการถดถอยด้วยวิธีกำลังสองน้อยที่สุดสำหรับปัญหาเมทริกซ์ $X'X$ ที่มีลักษณะไม่เป็นเชิงตั้งฉาก ซึ่งอธิบายได้ด้วยสมบัติของ $\hat{\beta}^{ols}$ สองประการ คือ เมทริกซ์ความแปรปรวนร่วม $\text{cov}[\hat{\beta}^{ols}] = \sigma^2(X'X)^{-1}$ และระยะทางของ $\hat{\beta}^{ols}$ จากค่าคาดหวังของตัวประมาณ β^{ols} ที่จะบ่งชี้ให้เห็นถึงความไม่เสถียรของ $\hat{\beta}^{ols}$ ดังนี้

กำหนด L_2 คือ เวกเตอร์ของระยะทางระหว่าง $\hat{\beta}^{ols}$ และ β จะได้ว่า ผลบวกของกำลังสองของระยะทางระหว่าง $\hat{\beta}_j^{ols}$ จาก β_j คือ

$$L_2^2 = (\hat{\beta}^{ols} - \beta)'(\hat{\beta}^{ols} - \beta) \quad (13)$$

เราสามารถหาค่าคาดหวังและความแปรปรวนของ L_2^2 เขียนแทนด้วย $E[L_2^2]$ และ $V[L_2^2]$ ได้ดังนี้

$$\begin{aligned} E[L_2^2] &= E[(\hat{\beta}^{ols} - \beta)'(\hat{\beta}^{ols} - \beta)] \\ &= E[\hat{\beta}^{ols}'\hat{\beta}^{ols} - 2\beta'\hat{\beta}^{ols} - \beta'\beta] \\ &= E[\hat{\beta}^{ols}'\hat{\beta}^{ols}] - 2\beta'E[\hat{\beta}^{ols}] + \beta'\beta \\ &= E[\hat{\beta}^{ols}'\hat{\beta}^{ols}] - \beta'\beta \end{aligned}$$

เนื่องจาก $E[L_2^2] = E[(\hat{\beta}^{ols} - \beta)'(\hat{\beta}^{ols} - \beta)]$ ซึ่งก็คือค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\beta}^{ols}$ และจาก (12) $\text{MSE}[\hat{\beta}^{ols}] = \sigma^2\text{tr}(X'X)^{-1}$ จะได้ว่า

$$E[\hat{\beta}^{ols}'\hat{\beta}^{ols}] - \beta'\beta = \sigma^2\text{tr}(X'X)^{-1}$$

$$\text{ดังนั้น} \quad E[\hat{\beta}^{ols}'\hat{\beta}^{ols}] = \beta'\beta + \sigma^2\text{tr}(X'X)^{-1} \quad (14)$$

พิจารณาความแปรปรวนของ L_2^2 เมื่อเวกเตอร์ความคลาดเคลื่อน ϵ มีการแจกแจงปกติ จะได้ว่า

$$V[L_2^2] = 2\sigma^4\text{tr}(X'X)^{-1}$$

ค่าคาดหวังและความแปรปรวนของ L_2^2 สามารถเขียนอยู่ในรูปของฟังก์ชันของค่าไอเกน (eigenvalue) ของ $\mathbf{X}'\mathbf{X}$ โดยเมื่อกำหนดให้ ค่าไอเกน (eigenvalue) ของ $\mathbf{X}'\mathbf{X}$ เขียนอยู่ในรูป

$$\lambda_{max} = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p = \lambda_{min} \geq 0 \quad (15)$$

ดังนั้น ระยะเวลากำลังสองระหว่าง $\hat{\beta}^{ols}$ และ β หรือ L_1^2 มีค่าคาดหวังและความแปรปรวน เมื่อเวกเตอร์ความคลาดเคลื่อน ϵ มีการแจกแจงปกติ ดังนี้

$$E[L_1^2] = \sigma^2 \sum_{j=1}^p (1/\lambda_j)$$

$$V[L_1^2] = 2\sigma^4 \sum_{j=1}^p (1/\lambda_j)^2$$

ขอบเขตล่างของค่าคาดหวังและความแปรปรวนของ L_2^2 คือ σ^2/λ_{min} และ $2\sigma^4/\lambda_{min}^2$ ตามลำดับ นั่นคือ ถ้าค่าสังเกตของตัวแปรอิสระส่งผลให้เมทริกซ์ $\mathbf{X}'\mathbf{X}$ มีค่าไอเกนขนาดเล็กจำนวนมาก จะทำให้ระยะทางระหว่าง $\hat{\beta}^{ols}$ และ β มีค่ามาก ดังนั้นค่าสัมบูรณ์ของ $\hat{\beta}^{ols}$ จะมีขนาดใหญ่ เมื่อข้อมูลเกิดปัญหาที่ไม่เป็นเชิงตั้งฉาก

ดังนั้นในกรณีที่ตัวแปรอิสระเกิดความสัมพันธ์เชิงเส้นแบบพหุ หรือกล่าวอีกนัยหนึ่งคือ เมทริกซ์ $\mathbf{X}'\mathbf{X}$ มีลักษณะไม่เป็นเชิงตั้งฉาก ตัวประมาณของสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุดนั้นมีความแปรปรวนสูง นั่นคือ ค่าสัมประสิทธิ์การถดถอยที่ประมาณด้วยวิธีกำลังสองน้อยที่สุดนี้ไม่ใกล้เคียงกับค่าพารามิเตอร์ ซึ่งจะส่งผลให้ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของตัวประมาณสัมประสิทธิ์การถดถอยที่ได้จากวิธีกำลังสองน้อยที่สุดมีค่ามาก จึงทำให้ตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุดนี้มีความไม่เสถียร (unstable) และไม่มีประสิทธิภาพ (inefficiency) (วิธูรา, 2558: 213)

ผู้วิจัยหลายท่านได้เสนอวิธีการประมาณค่าสัมประสิทธิ์การถดถอยวิธีอื่น เพื่อแก้ปัญหาของวิธีการประมาณสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยที่สุดในกรณีเมทริกซ์ $\mathbf{X}'\mathbf{X}$ มีลักษณะไม่เป็นเชิงตั้งฉากที่กล่าวมาข้างต้น ซึ่งวิธีการหนึ่งคือ การถดถอยเชิงเส้นพหุคูณแบบพินอลไลซ์ โดยมีหลักการหาตัวประมาณของเวกเตอร์พารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดภายใต้ข้อจำกัดของพารามิเตอร์ β กล่าวคือ หาตัวประมาณของเวกเตอร์พารามิเตอร์ β ที่ทำให้ $(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$ มีค่าต่ำสุด ภายใต้ข้อจำกัดของฟังก์ชันของ β ที่เรียกว่า ฟังก์ชันพินอลตี้ (penalty function) เขียนแทนด้วย $P(\|\beta\|_q^2)$ เมื่อ $\|\beta\|_q$ เป็น L_q นอร์ม (L_q norm) ของ β นั่นคือ

$$\hat{\beta} = \operatorname{argmin}_{\beta} (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) \text{ ภายใต้ข้อจำกัด } P(\|\beta\|_q^2) < t \text{ เมื่อ } t$$
 เป็นค่าคงที่ ซึ่งสามารถเขียนอยู่ในรูปสมการที่สอดคล้องกัน คือ

$$\hat{\beta} = \operatorname{argmin}_{\beta} (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) + \lambda P(\|\beta\|_q^2)$$

เมื่อ λ คือ ตัวคูณลากรองจ์ (Lagrange multiplier)

สำหรับวิธีการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีการประมาณค่าจากภาวะน่าจะเป็นสูงสุด การถดถอยเชิงเส้นพหุคูณแบบพินอลไลซ์มีหลักการในการหาตัวประมาณซึ่งเรียกว่า วิธีการประมาณค่าจากภาวะน่าจะเป็นสูงสุดแบบพินอลไลซ์ (Penalized maximum likelihood estimation) โดยการเพิ่มฟังก์ชันพินอลตี้ในฟังก์ชันลอการิทึมของฟังก์ชันภาวะน่าจะเป็นสูงสุดจากสมการ (8) ซึ่งจะได้

$$\begin{aligned} M(\mathbf{Y}; \boldsymbol{\beta}, \sigma^2) &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) - \lambda P(\|\boldsymbol{\beta}\|_q^2) \\ &= -\left[\frac{n}{2} \ln 2\pi + \frac{n}{2} \ln \sigma^2 + \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \lambda P(\|\boldsymbol{\beta}\|_q^2) \right] \end{aligned}$$

แล้วจึงหาตัวประมาณของ $\boldsymbol{\beta}$ ที่ทำให้ $M(\mathbf{Y}; \boldsymbol{\beta}, \sigma^2)$ มีค่ามากที่สุด นั่นคือ การหาตัวประมาณของ $\boldsymbol{\beta}$ ที่ทำให้ $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \lambda P(\|\boldsymbol{\beta}\|_q^2)$ ให้มีค่าต่ำที่สุด ซึ่งสอดคล้องกับวิธีการถดถอยเชิงเส้นพหุคูณแบบพินอลไลซ์ด้วยวิธีกำลังสองน้อยที่สุด

ฟังก์ชันพินอลตี้ที่ใช้ในวิธีการถดถอยเชิงเส้นพหุคูณแบบพินอลไลซ์มีรูปแบบต่างกัน เช่น ฟังก์ชัน L_2 นอร์ม $P(\|\boldsymbol{\beta}\|_2^2) = \sum_{j=1}^p \beta_j^2$ ที่ใช้ในการถดถอยวิธีริดจ์ หรือฟังก์ชัน L_1 นอร์ม $P(\|\boldsymbol{\beta}\|_1^2) = \sum_{j=1}^p |\beta_j|$ ที่ใช้ในการถดถอยวิธีแลซโซ เป็นต้น ซึ่งจะกล่าวรายละเอียดของการถดถอยทั้งสองวิธีในหัวข้อถัดไป

3. การถดถอยเชิงเส้นพหุคูณด้วยวิธีริดจ์ (Ridge Regression)

โฮเอิร์ล ได้นำการวิเคราะห์ริดจ์ (Ridge analysis) (Hoerl, 1962: 54-59) ซึ่งเป็นผลงานวิจัยของตนเอง มาประยุกต์ใช้กับการวิเคราะห์การถดถอยเชิงเส้น และเรียกชื่อว่า การวิเคราะห์การถดถอยวิธีริดจ์ เนื่องจากวิธีการทางคณิตศาสตร์มีความคล้ายคลึงกับการวิเคราะห์ริดจ์ ต่อมาโฮเอิร์ลและเคนนาร์ด (Hoerl and Kennard, 1970: 56-67) ได้พัฒนาและนำเสนอสมบัติทางสถิติของการวิเคราะห์การถดถอยวิธีริดจ์ ซึ่งเป็นบทความวิจัยที่ทำให้การวิเคราะห์การถดถอยวิธีริดจ์เป็นที่รู้จักกันอย่างแพร่หลาย งานวิจัยของโฮเอิร์ลและเคนนาร์ด ได้กล่าวถึงข้อจำกัดของการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยที่สุดที่ใช้เทคนิคการหาค่าเหมาะสมที่สุดเชิงคณิตศาสตร์ (mathematical optimization techniques) ซึ่งให้ผลลัพธ์เป็นค่าประมาณแบบจุด และเป็นวิธีการหาตัวประมาณไม่ได้แสดงภาพให้เห็นค่าความไว (sensitivity) ของตัวประมาณ $\hat{\boldsymbol{\beta}}_{ols}$ ต่อค่าผลบวกกำลังสองของความคลาดเคลื่อน คือ $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ ดังนั้นจึงเสนอวิธีการถดถอยวิธีริดจ์ เป็นวิธีการหาค่าเหมาะสมที่สุดโดยใช้กราฟิก เนื่องจากผลบวกกำลังสองของความคลาดเคลื่อน $\boldsymbol{\epsilon}'\boldsymbol{\epsilon}$ สามารถเขียนอยู่ในรูปของฟังก์ชันกำลังสอง (quadratic function) ของ

สัมประสิทธิ์การถดถอยกับค่าผลบวกกำลังสองของความคลาดเคลื่อน $\mathbf{e}'\mathbf{e}$ และสามารถแสดงเป็นกราฟของพื้นที่ผิวตอบสนอง

จากตัวแบบการถดถอยในสมการ (1) และจากสมการ (4) ผลบวกกำลังสองของความคลาดเคลื่อน $\mathbf{e}'\mathbf{e}$ ของการถดถอยพหุคูณ สามารถเขียนอยู่ในรูป

$$\mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \quad (16)$$

ซึ่งสามารถพิจารณาเป็นรูปแบบเดียวกับฟังก์ชันพื้นที่ผิวตอบสนองควอดราติก (quadratic response surface function) $f(\mathbf{z})$ ของตัวแปร \mathbf{z} ใดๆ จำนวน p ตัว สำหรับการวิเคราะห์รีดจ์ (Hoerl, 1985: 186-187) ซึ่งเขียนอยู่ในรูป

$$f(\mathbf{z}) = a_0 + \sum_{i=1}^p a_i z_i + \sum_{i=1}^{p-1} \sum_{j=1}^p a_{ij} z_i z_j + \sum_{i=1}^p a_i^2 z_i^2$$

หรือ

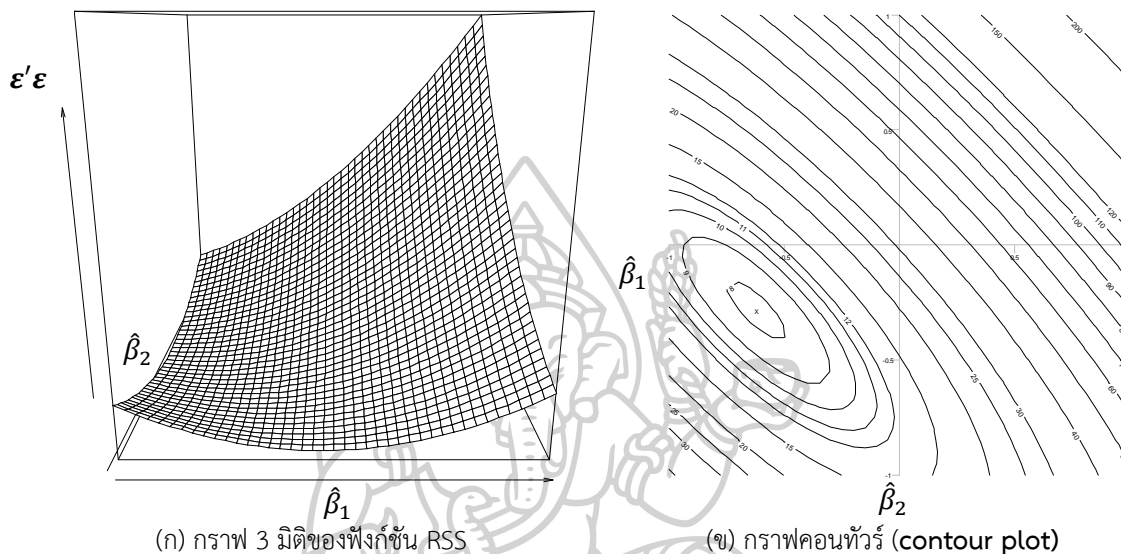
$$f(\mathbf{z}) = a_0 + \mathbf{a}'\mathbf{z} + \frac{1}{2}\mathbf{z}'\mathbf{A}\mathbf{z} \quad (17)$$

โดยที่ \mathbf{a} คือ เวกเตอร์ขนาด $p \times 1$ ของสัมประสิทธิ์เชิงเส้น, \mathbf{z} คือ เวกเตอร์ขนาด $p \times 1$ ของตัวแปรอิสระ, a_0 เป็นค่าคงที่ และ \mathbf{A} เป็นเมทริกซ์สมมาตรขนาด $p \times p$ ที่มีสมาชิกแนวทแยงเป็นค่า a_i^2 และสมาชิกนอกแนวทแยงเป็นค่าปฏิสัมพันธ์ (interaction terms) a_{ij}

จากผลบวกกำลังสองของความคลาดเคลื่อนสุ่มในสมการ (16) เทียบเคียงกับฟังก์ชันพื้นที่ผิวตอบสนองแบบควอดราติกในสมการ (17) จะได้ว่า $\boldsymbol{\beta}$ คือพิกัด (coordinate) ของ $-\mathbf{X}'\mathbf{Y}$ คือเวกเตอร์ \mathbf{a} และ $\mathbf{X}'\mathbf{X}$ คือ เมทริกซ์ \mathbf{A} ดังนั้น เราจะหาค่าต่ำสุดของผลบวกกำลังสองของความคลาดเคลื่อนสุ่มในลักษณะของฟังก์ชันของสัมประสิทธิ์การถดถอย $\hat{\boldsymbol{\beta}}$ โดยพิจารณาจากการกำหนดค่า $\boldsymbol{\beta}'\boldsymbol{\beta} = r^2$ โดยที่ r เป็นค่าคงที่ และหาค่าต่ำสุดของผลบวกกำลังสองของความคลาดเคลื่อนสุ่ม ภายใต้ข้อจำกัด $\boldsymbol{\beta}'\boldsymbol{\beta} \leq r^2$

ซีมอน (Simon, 2016: 1-9) ได้อธิบายวิธีการหาตัวประมาณรีดจ์จากการวิเคราะห์การถดถอย โดยใช้ข้อมูลจากนิตยสารเกี่ยวกับเครื่องยนต์ที่มีชื่อว่า Motor Trend US ในปี 1974 ซึ่งเป็นข้อมูลเกี่ยวกับการเผาผลาญพลังงานเชื้อเพลิง การออกแบบเครื่องยนต์ และประสิทธิภาพการทำงานของเครื่องยนต์ จำนวน 32 รุ่นที่ผลิตปี 1973-1974 ประกอบด้วยตัวแปร 11 ตัวแปร ซึ่งเป็นข้อมูลที่อ้างอิงมาจากงานวิจัยของ เฮนเดอสันและเวลลีแมน (Henderson and Velleman, 1981: 396) และอยู่ในชุดข้อมูล (datasets) ของแพ็คเกจ R ที่มีชื่อว่า mtcars ซีมอนได้แสดงตัวอย่างของการถดถอยพหุคูณวิธีรีดจ์ที่มีตัวแปรตาม คือ ระยะทางที่รถวิ่งได้ (ไมล์) ต่อน้ำมัน 1 แกลลอน และตัวแปรอิสระ 2 ตัว ได้แก่ disp คือ ความจุของเครื่องยนต์ และ hp คือ แรงม้ารวม และใช้โปรแกรม R ในการวิเคราะห์ ซึ่งเริ่มจากการแปลงค่าสังเกตของตัวแปรอิสระและตัวแปรตามให้เป็นค่ามาตรฐาน และวิเคราะห์การถดถอยวิธีกำลังสองน้อยที่สุด จากนั้นคำนวณค่าผลบวกกำลังสองของส่วนเหลือ

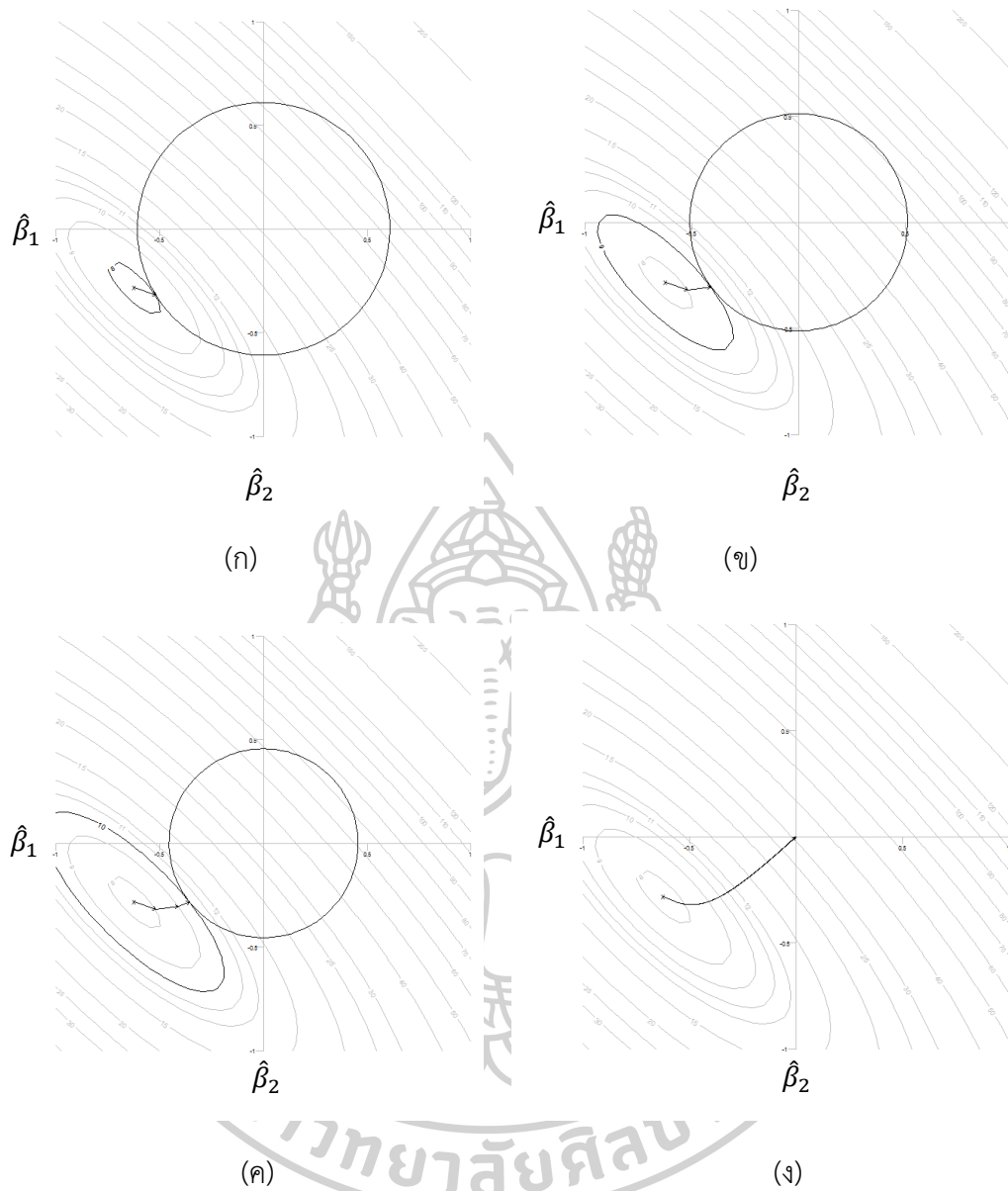
จากสมการถดถอย $\widehat{\text{mpg}} = \beta_1 \text{disp} + \beta_2 \text{hp}$ เมื่อค่าสัมประสิทธิ์การถดถอยทั้งสองตัวมีค่าต่างๆ ในช่วงที่กำหนด และแสดงภาพของพื้นที่ผิวตอบสนองและคอนทัวร์ของผลบวกกำลังสองของส่วนเหลือของพิกัด (β_1, β_2) บนระนาบพิกัดฉาก ดังภาพที่ 1 (ก) ละ (ข) ตามลำดับ



ภาพที่ 1 กราฟแสดงพื้นที่ผิวตอบสนองและคอนทัวร์ของผลบวกกำลังสองของส่วนเหลือจากสมการถดถอยพหุคูณ $\widehat{\text{mpg}} = \beta_1 \text{disp} + \beta_2 \text{hp}$ ของค่าประมาณของสัมประสิทธิ์การถดถอย β_1 และ β_2 (Simon, 2016: 3-4)

ภาพที่ 1 (ก) แสดงพื้นที่ผิวตอบสนองของผลบวกกำลังสองของส่วนเหลือของสมการถดถอยพหุคูณ $\widehat{\text{mpg}} = \beta_1 \text{disp} + \beta_2 \text{hp}$ เมื่อค่าประมาณสัมประสิทธิ์ถดถอย (β_1, β_2) มีค่าต่างๆ ในช่วงที่กำหนด ซึ่งจะได้ว่าพิกัด $(\hat{\beta}_1, \hat{\beta}_2) = (-0.62, -0.28)$ เป็นค่าที่ทำให้ผลบวกกำลังสองของส่วนเหลือมีค่าต่ำที่สุดซึ่งมีค่าเท่ากับ 7.81 ซึ่งก็คือตัวประมาณวิธีกำลังสองน้อยที่สุด จะได้ สมการถดถอย คือ $\widehat{\text{mpg}} = -0.62 \text{disp} - 0.28 \text{hp}$

ภาพที่ 1 (ข) แสดงคอนทัวร์ของพื้นที่ผิวตอบสนองของผลบวกกำลังสองของส่วนเหลือจากภาพที่ 1 (ก) โดยแกนนอน คือ ค่าของ β_1 และแกนตั้งคือค่าของ β_2 และเส้นวงรีแสดงเส้นคอนทัวร์ (Contour Line) ของพื้นที่ผิวตอบสนองของผลบวกกำลังสองของส่วนเหลือ ซึ่งแสดงค่าผลบวกกำลังสองของส่วนเหลือของพิกัด $(\hat{\beta}_1, \hat{\beta}_2)$ จากกราฟคอนทัวร์ พิกัด $(\hat{\beta}_1^{ols}, \hat{\beta}_2^{ols}) = (-0.62, -0.28)$ เป็นจุดที่ทำให้ผลบวกกำลังสองของส่วนเหลือมีค่าต่ำที่สุด ซึ่งแสดงด้วยเครื่องหมายกากบาท เส้นวงรีอยู่ถัดจากจุดกากบาทมีค่าผลบวกกำลังสองของส่วนเหลือเท่ากับ 8, 9, 10, 12, ... จะเห็นว่าวงรีที่ยิ่งไกลออกไปจากจุดกากบาท จะมีค่าให้ผลบวกกำลังสองของส่วนเหลือมากขึ้น



ภาพที่ 2 กราฟแสดงริตจต่ำสุดและเส้นทางของพิกัดของริตจต่ำสุดของผลบวกกำลังสองของส่วนเหลือ ภายใต้อัจฉกัต $\beta' \beta = \sum_{j=1}^2 \beta_j^2 = r^2$ (Simon, 2016: 6-9)

การวิเคราะห์การถดถอยวิธีริตจ์ พิจารณาค่าต่ำสุดของผลบวกกำลังสองของความคลาดเคลื่อนสุ่ม ภายใต้ข้อจำกัด $\beta' \beta \leq r^2$ สำหรับค่าคงที่ r ใดๆ ที่กำหนด จากตัวอย่างสมการถดถอย $\widehat{\text{mpg}} = \beta_1 \text{disp} + \beta_2 \text{hp}$ จะได้ว่าข้อจำกัดของ $\beta' \beta = \beta_1^2 + \beta_2^2 = r^2$ เป็นฟังก์ชันที่มีกราฟเป็นรูปวงกลม สำหรับ $0 < r^2 < t$ เมื่อ t เป็นค่าคงที่ ดังนั้นเราจะพิจารณา พิกัด $(\hat{\beta}_1, \hat{\beta}_2)$ ที่มีค่าของผลบวกกำลังสองของส่วนเหลือต่ำสุดและอยู่ภายใต้ข้อจำกัดของเส้นวงกลม ที่มีรัศมี r ต่างๆ กัน นั่นคือจุดสัมผัสของเส้นคอนทัวร์กับเส้นวงกลม เรียกว่า ริตจ์ต่ำสุด (minimum ridge) ดังแสดงในภาพที่ 2 (ก) - (ค) พิจารณาภาพที่ 2 (ก) เมื่อตรึงค่าผลบวกกำลังสองของส่วนเหลือเท่ากับ 8 แล้วหารัศมีของวงกลมที่สัมผัสกับเส้นคอนทัวร์ จะได้พิกัดของริตจ์ต่ำสุด คือ $(\hat{\beta}_1, \hat{\beta}_2) = (-0.52, -0.32)$ ภาพที่ 2 (ข) และ (ค) เมื่อตรึงค่าผลบวกกำลังสองของส่วนเหลือเท่ากับ 9 และ 10 แล้วหารัศมีของวงกลมที่สัมผัสกับเส้นคอนทัวร์จะลดลงตามลำดับ ซึ่งจะทำให้พิกัดของริตจ์ต่ำสุด $(\hat{\beta}_1, \hat{\beta}_2)$ จะเข้าใกล้จุดกำเนิด $(0,0)$ มากขึ้น เมื่อเชื่อมโยงพิกัดของริตจ์ต่ำสุด ทำให้เห็นเส้นทางของริตจ์ต่ำสุด (ridge trace) ซึ่งเป็นเส้นทางของ steepest ascent ไปยังจุดกำเนิด ดังแสดงในภาพที่ 2 (ง)

3.1 การประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีริตจ์

ตัวประมาณสัมประสิทธิ์การถดถอยด้วยวิธีริตจ์ ใช้วิธีการของการวิเคราะห์แบบริตจ์ ซึ่งสามารถอธิบายในเชิงคณิตศาสตร์ ได้ดังนี้

จุดของริตจ์ต่ำสุดสามารถหาได้จากการหาอนุพันธ์ของ $(Y - X\beta)'(Y - X\beta)$ จากสมการ (16) เทียบกับ β ภายใต้ข้อจำกัด คือ $\beta' \beta = r^2$ ดังนี้

$$\text{minimize}_{\beta} (Y - X\beta)'(Y - X\beta) \quad \text{ภายใต้ข้อจำกัด} \quad \beta' \beta = r^2 < t$$

การหาค่าเหมาะสมที่สุดภายใต้ข้อจำกัด (optimization with equality constraint) สามารถทำได้โดยใช้วิธีตัวคูณลากรองจ์ เมื่อ λ เป็นตัวคูณลากรองจ์ (Lagrange multiplier) จะได้ฟังก์ชันลากรองจ์ (Lagrangian expression) คือ

$$L(Y, \beta, \lambda) = (Y - X\beta)'(Y - X\beta) - \lambda(\beta' \beta - t)$$

อนุพันธ์บางส่วนของ $L(Y, \beta, \lambda)$ เทียบกับ β คือ

$$\begin{aligned} \frac{\partial L(Y, \beta, \lambda)}{\partial \beta} &= \frac{\partial}{\partial \beta} [(Y - X\beta)'(Y - X\beta) - \lambda(\beta' \beta - t)] \\ &= \frac{\partial}{\partial \beta} [Y'Y - 2\beta'X'Y + \beta'X'X\beta - \lambda(\beta' \beta - t)] \\ &= -2X'Y + 2X'X\beta - 2\lambda\beta \end{aligned}$$

กำหนดค่าอนุพันธ์เท่ากับศูนย์และแก้สมการหาค่าของ $\hat{\beta}$

$$-2X'Y + 2X'X\beta - 2\lambda\beta = 0$$

$$(X'X - \lambda I_p)\beta = X'Y$$

$$\text{จะได้ } \hat{\beta} = (X'X - \lambda I_p)^{-1}X'Y \quad (18)$$

ถ้า $X'X$ เป็นเมทริกซ์บวกแน่นอน (positive definite) แล้วค่าไอเกนของเมทริกซ์ $X'X$ จะมีค่าเป็นบวก ซึ่งทำให้ค่าต่ำสุดโดยรวม (overall minimum) ของฟังก์ชัน $RSS(\hat{\beta})$ หาค่าได้และอยู่ที่จุด $(X'X)^{-1}X'Y$ เราสามารถลงจุด (plot) ริตจ์ต่ำสุด (minimum ridge) ของผลบวกกำลังสองของความคลาดเคลื่อนสุมจากจุดต่ำสุดโดยรวมไปยังจุดกำเนิด ($r = 0$) โดยการกำหนดค่า λ ใน (18) ที่มีค่าน้อยกว่าค่าไอเกนต่ำสุดของเมทริกซ์ $X'X$ ซึ่งค่า λ ที่น้อยกว่าศูนย์ (หรือ $-\lambda > 0$) จะให้ค่าประมาณที่ใกล้จุดกำเนิดมากกว่าจุดต่ำสุดโดยรวม

ดังนั้น โดยทั่วไป ตัวประมาณริตจ์ นิยมเขียนอยู่ในรูปแบบของค่าคงที่ $k \geq 0$ โดยที่ $k = -\lambda$ ดังนี้

$$\hat{\beta}^{ridge} = (X'X + kI_p)^{-1}X'Y \quad (19)$$

$$= WX'Y \quad (20)$$

$$\text{เมื่อ } W = (X'X + kI_p)^{-1} \quad (21)$$

การหาตัวประมาณริตจ์ สามารถพิจารณาในแง่มุมหนึ่งซึ่งเทียบเท่ากับวิธีการข้างต้น คือ การหาของเส้นทางของพิกัดของริตจ์ต่ำสุดโดยกำหนดค่าของผลบวกกำลังสองของส่วนเหลือเท่ากับค่าใดค่าหนึ่ง และพิจารณาค่าประมาณ $\hat{\beta}$ ที่มีความยาวน้อยที่สุด ซึ่งสามารถอธิบายได้ดังนี้

ให้ $\hat{\beta}$ เป็นตัวประมาณพารามิเตอร์ β จะได้ ผลบวกกำลังสองของความคลาดเคลื่อนสุมสามารถเขียนอยู่ในรูป ดังนี้

$$\begin{aligned} \varepsilon'\varepsilon &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\ &= (Y - X\hat{\beta}^{ols})'(Y - X\hat{\beta}^{ols}) + (\hat{\beta} - \hat{\beta}^{ols})'X'X(\hat{\beta} - \hat{\beta}^{ols}) \\ &= (\varepsilon'\varepsilon)_{ols} + f(\hat{\beta}) \end{aligned} \quad (22)$$

จะเห็นว่าผลบวกกำลังสองของความคลาดเคลื่อนสุมสำหรับตัวประมาณ $\hat{\beta}$ เกิดจากผลรวมกำลังสองของความคลาดเคลื่อนสุมสำหรับตัวประมาณโดยวิธีกำลังสองน้อยที่สุดรวมกับค่าของเทอมกำลังสองของ $(\hat{\beta} - \hat{\beta}^{ols})$ เขียนอยู่ในรูปฟังก์ชัน $f(\hat{\beta})$ เมื่อกำหนดค่า $f(\hat{\beta})$ เท่ากับค่าคงที่โดย $f_0 > 0$ ซึ่งจะทำให้ $f(\hat{\beta})$ มีค่าเพิ่มขึ้น เราสามารถหาค่าต่อเนื่องของ $\hat{\beta}_0$ ที่ทำให้ $\varepsilon'\varepsilon = (\varepsilon'\varepsilon)_{ols} + f_0$

เมื่อ $f_0 > 0$ โดยการหาค่าเหมาะสมที่สุดของ $\hat{\beta}$ โดยหาค่า $\hat{\beta}$ ความยาวน้อยที่สุดภายใต้การกำหนดค่า $\epsilon'\epsilon$ ดังนี้

$$\text{minimize}_{\beta} \quad \beta'\beta$$

$$\text{ภายใต้ข้อจำกัด} \quad (\beta - \hat{\beta}^{ols})'X'X(\beta - \hat{\beta}^{ols}) = f_0$$

ซึ่งสามารถหาคำตอบได้ด้วยวิธีตัวคูณลากรองจ์ โดยกำหนด $\lambda = -1/k$ เป็นตัวคูณลากรองจ์ และจะได้ฟังก์ชันลากรองจ์ ดังนี้

$$L(Y, \beta, k) = \beta'\beta + (1/k)[(\beta - \hat{\beta}^{ols})'X'X(\beta - \hat{\beta}^{ols}) - f_0]$$

อนุพันธ์บางส่วนของ $L(Y, \beta, k)$ เทียบกับ β คือ

$$\frac{\partial}{\partial \beta} \{ \beta'\beta + (1/k)[(\beta - \hat{\beta}^{ols})'X'X(\beta - \hat{\beta}^{ols}) - f_0] \} = 0$$

$$2\beta + (1/k)[2(X'X)\beta - 2(X'X)\hat{\beta}^{ols}] = 0$$

$$2\beta + (1/k)[2(X'X)\beta - 2(X'X)(X'X)^{-1}X'Y] = 0$$

$$\beta + (1/k)[(X'X)\beta - X'Y] = 0$$

$$[(1/k)[(X'X) + kI_p]\beta = (1/k)X'Y$$

$$[(X'X) + kI_p]\beta = X'Y$$

$$\therefore \hat{\beta} = [(X'X) + kI_p]^{-1}X'Y$$

ซึ่งจะได้ตัวประมาณแบบบริดจ์ คือ $\hat{\beta}^{ridge} = (X'X + kI_p)^{-1}X'Y$ เมื่อ $k \geq 0$

เช่นเดียวกับสมการ (19)

ความสัมพันธ์ระหว่างตัวประมาณบริดจ์และตัวประมาณกำลังสองน้อยที่สุด สามารถเขียนอยู่ในรูปแบบ ดังนี้

$$\begin{aligned} \hat{\beta}^{ridge} &= [(X'X) + kI]^{-1}X'Y \\ &= [(X'X) + kI]^{-1}(X'X)(X'X)^{-1}X'Y \\ &= \{(X'X)^{-1}[(X'X) + kI]\}^{-1}\hat{\beta}^{ols} \\ &= [I_p + k(X'X)^{-1}]^{-1}\hat{\beta}^{ols} \\ &= Z\hat{\beta}^{ols} \end{aligned} \tag{23}$$

$$\text{เมื่อ } Z = [I_p + k(X'X)^{-1}]^{-1} \tag{24}$$

กำหนดเมทริกซ์ A เป็นเมทริกซ์แนวทแยงของค่าไอเกนของ $X'X$ จากสมการ (15) นั่นคือ $A = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ และเมทริกซ์ P เป็นเมทริกซ์ขนาด $p \times p$ โดยที่แต่ละคอลัมน์

ของเมทริกซ์ P คือ ไอเกนเวกเตอร์ของ $X'X$ ที่สอดคล้องกับค่าไอเกน โดยที่ $PP' = I_p$ จะได้ว่าเมทริกซ์ W และ Z จากสมการ (21) และสมการ (23) สามารถเขียนอยู่ในรูป

$$\begin{aligned} W &= [P\Lambda P' + kI_p]^{-1} \\ &= [P(\Lambda + kI_p)P']^{-1} \\ &= P(\Lambda + kI_p)^{-1}P \end{aligned} \quad (25)$$

และ

$$\begin{aligned} Z &= [I_p + k(P\Lambda P')^{-1}]^{-1} \\ &= [PP' + k(P\Lambda^{-1}P')]^{-1} \\ &= [P(I_p + k\Lambda^{-1})P']^{-1} \\ &= P(I_p + k\Lambda^{-1})^{-1}P' \end{aligned} \quad (26)$$

กำหนด lw_i และ lz_i คือ ค่าไอเกนที่ i ของ W และ Z ดังนั้น

$$lw_i = \frac{1}{\lambda_i + k} \quad (27)$$

$$lz_i = \left(\frac{1}{\lambda_i} + k\right)^{-1} = \left(\frac{\lambda_i + k}{\lambda_i}\right)^{-1} = \frac{\lambda_i}{\lambda_i + k} \quad (28)$$

เมื่อ λ_i คือ ค่าไอเกนของ $X'X$ เมื่อ $i = 1, 2, \dots, p$

ดังนั้น เมื่อ k สูงขึ้นเข้าไปล้นันต์ ส่งผลให้ lw_i ลู่เข้าสู่ศูนย์ จะได้ว่าเมทริกซ์ W ลู่เข้าสู่เมทริกซ์ศูนย์ ดังนั้นตัวประมาณสัมประสิทธิ์การถดถอยวิธีริดจ์จะลู่เข้าสู่ศูนย์ เมื่อ k เพิ่มขึ้น และ $k \neq 0$ จะได้ว่า

$$(\hat{\beta}^{ridge})'(\hat{\beta}^{ridge}) = (Z\hat{\beta}^{ols})'(Z\hat{\beta}^{ols}) = (\hat{\beta}^{ols})'Z'Z(\hat{\beta}^{ols}) < (\hat{\beta}^{ols})'\hat{\beta}^{ols}$$

3.2 สมบัติของตัวประมาณริดจ์

เราสามารถหาคุณสมบัติของตัวประมาณริดจ์ ได้แก่ ค่าคาดหวัง ความแปรปรวน และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์การถดถอยวิธีริดจ์ ($\hat{\beta}^{ridge}$) ดังนี้

3.2.1 ค่าคาดหวังของ $\hat{\beta}^{ridge}$

$$\begin{aligned} E[\hat{\beta}^{ridge}] &= E[Z\hat{\beta}^{ols}] \\ &= ZE[\hat{\beta}^{ols}] \\ &= Z\beta \end{aligned} \quad (29)$$

นั่นคือ ค่าคาดหวังของ $\hat{\beta}^{ridge}$ มีค่าเท่ากับ $Z\beta = [I_p + k(X'X)^{-1}]^{-1}\beta$ ดังนั้น ตัวประมาณ $\hat{\beta}^{ridge}$ เป็นตัวประมาณที่เอนเอียง (biased estimator) สำหรับ β

3.2.2 เมทริกซ์ความแปรปรวนร่วมของ $\hat{\beta}^{ridge}$

$$\begin{aligned} \text{cov}[\hat{\beta}^{ridge}] &= \text{cov}[Z\hat{\beta}^{ols}] \\ &= Z\text{cov}[\hat{\beta}^{ols}]Z' \\ &= Z(\sigma^2(X'X)^{-1})Z' \\ &= \sigma^2 Z(X'X)^{-1}Z' \\ &= \sigma^2 \left(\frac{\lambda_i}{\lambda_i + k} \right)^2 (X'X)^{-1} \end{aligned} \quad (30)$$

เมื่อกำหนด $\lambda_{max} = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p = \lambda_{min} \geq 0$ เป็นค่าไอเกนของเมทริกซ์ $X'X$ ดังนั้นความแปรปรวนของ $\hat{\beta}^{ridge}$ จะลดลงสู่ศูนย์ เมื่อ k เพิ่มขึ้น

3.2.3 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\beta}^{ridge}$

จากสมการ (11) ความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\beta}$ คือ

$$\text{MSE}[\hat{\beta}] = E[(\hat{\beta} - E[\hat{\beta}])(\hat{\beta} - E[\hat{\beta}])'] + (E[\hat{\beta}] - \beta)(E[\hat{\beta}] - \beta)'$$

โฮเอิร์ลและเคนนาร์ด ได้อธิบายความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\beta}^{ridge}$ (Hoerl and Kennard, 1970: 60-61) ดังนี้

$$\begin{aligned} \text{MSE}[\hat{\beta}^{ridge}] &= E[(Z\hat{\beta}^{ols} - Z\beta)(Z\hat{\beta}^{ols} - Z\beta)'] + (Z\beta - \beta)(Z\beta - \beta)' \\ &= E[(\hat{\beta}^{ols} - \beta)Z'Z(\hat{\beta}^{ols} - \beta)] + \beta'(Z - I)'(Z - I)\beta \\ &= \sigma^2 \text{tr}[(X'X)^{-1}Z'Z] + \beta'(Z - I)'(Z - I)\beta \end{aligned} \quad (31)$$

เนื่องจาก เราสามารถเขียน Z อีกรูปหนึ่ง คือ

$$\begin{aligned} Z &= (X'X + kI)^{-1}(X'X) \\ &= W(X'X) \end{aligned} \quad (32)$$

และจาก $Z = [I_p + k(X'X)^{-1}]^{-1}$

$$\begin{aligned} \text{จะได้ว่า} \quad I_p &= ZZ^{-1} = [W(X'X)][(I_p + k(X'X)^{-1})] \\ &= (W(X'X) + kW(X'X)(X'X)^{-1}) \\ &= Z + kW \end{aligned}$$

นั่นคือ $Z = I_p - kW$ (33)

ดังนั้น $MSE[\hat{\beta}^{ridge}]$ ในสมการ (31) สามารถจัดรูปของ $(X'X)^{-1}Z'Z$ และ $Z - I$ โดยใช้รูปแบบของ Z ในสมการ (32) และ (33) ดังนี้

$$\begin{aligned}(X'X)^{-1}Z'Z &= (X'X)^{-1}(WX'X)'(I_p - kW) \\ &= (X'X)^{-1}(X'X)W(I_p - kW) \\ &= W(I_p - kW) \\ &= W - kW^2\end{aligned}\tag{34}$$

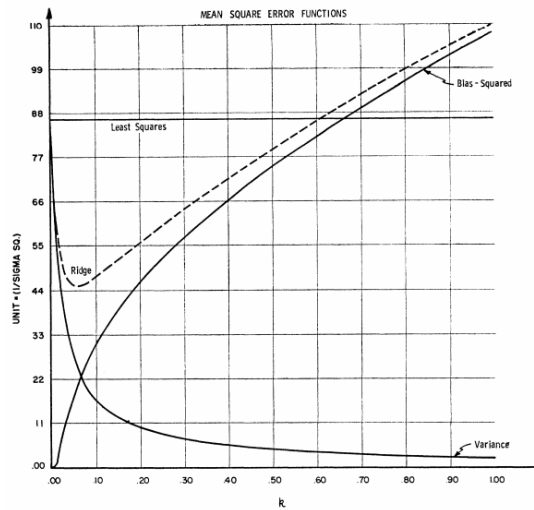
$$\begin{aligned}\text{และ } Z - I &= (I - kW) - I \\ &= -kW\end{aligned}\tag{35}$$

แทนค่า $(X'X)^{-1}Z'Z$ และ $Z - I$ จากสมการ (34) และ (35) ในสมการ (31) จะได้ว่า

$$\begin{aligned}MSE[\hat{\beta}^{ridge}] &= \sigma^2 \text{tr}[W - kW^2] + \beta'(-kW)'(-kW)\beta \\ &= \sigma^2\{\text{tr}[W] - k \text{tr}[W^2]\} + k^2\beta'W^2\beta \\ &= \sigma^2 \left\{ \sum_{i=1}^p \frac{1}{\lambda_i+k} - k \sum_{i=1}^p \frac{1}{(\lambda_i+k)^2} \right\} + k^2\beta'(X'X + kI_p)^{-2}\beta \\ &= \sigma^2 \sum_{i=1}^p \left(\frac{1}{\lambda_i+k} - \frac{k}{(\lambda_i+k)^2} \right) + k^2\beta'(X'X + kI_p)^{-2}\beta \\ &= \sigma^2 \sum_{i=1}^p \left(\frac{\lambda_i}{(\lambda_i+k)^2} \right) + k^2\beta'(X'X + kI_p)^{-2}\beta \\ &= \gamma_1(k) + \gamma_2(k)\end{aligned}$$

ดังนั้น ความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\beta}^{ridge}$ คือ ผลบวกของเทอมฟังก์ชัน $\gamma_1(k)$ และ $\gamma_2(k)$ ซึ่งความหมายของเทอม $\gamma_2(k)$ คือ ค่ากำลังสองของระยะทางระหว่าง $Z\beta$ และ β ซึ่งจะมีค่าเป็นศูนย์ ถ้า $k = 0$ เพราะว่า $Z = I_p$ ดังนั้น $\gamma_2(k)$ จึงสามารถพิจารณาเป็นค่ากำลังสองของความเอนเอียงของ $\hat{\beta}^{ridge}$ ส่วนเทอม $\gamma_1(k)$ คือ ผลรวมของค่าความแปรปรวนของตัวประมาณสัมประสิทธิ์การถดถอย $\hat{\beta}_i^{ridge}$ แต่ละตัว ซึ่งเป็นผลรวมของสมาชิกแนวทแยงของเมทริกซ์ความแปรปรวนร่วมของ $\hat{\beta}^{ridge}$ หรือ $\text{cov}[\hat{\beta}^{ridge}]$ ในสมการ (30)

ความสัมพันธ์ระหว่างความแปรปรวนร่วม ค่ากำลังสองของความเอนเอียง และค่า k สามารถแสดงได้ในภาพที่ 3 ซึ่งจะเห็นได้ว่า ผลรวมของความแปรปรวนจะลดลง เมื่อ k เพิ่มขึ้น ในขณะที่ค่ากำลังสองของความเอนเอียงจะเพิ่มขึ้นตามค่า k



ภาพที่ 3 กราฟแสดงความสัมพันธ์ระหว่างความแปรปรวนรวม
ค่ากำลังสองของความเอนเอียง และค่า k (Hoerl and Kennard, 1970: 61)

จากกราฟในภาพ 3 เส้นของกราฟที่มีลักษณะเป็นเส้นประ แสดงค่าของผลรวมของ $\gamma_1(k)$ และ $\gamma_2(k)$ ซึ่งก็คือค่า $MSE[\hat{\beta}^{ridge}]$ จะเห็นว่า มีความเป็นไปได้ที่จะหาค่า k ที่ทำให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดของตัวประมาณ $\hat{\beta}^{ridge}$ ซึ่งความเป็นไปได้นี้ สามารถอธิบายได้โดยคุณสมบัติทางคณิตศาสตร์ของ $\gamma_1(k)$ และ $\gamma_2(k)$ โดยฟังก์ชัน $\gamma_1(k)$ เป็นฟังก์ชันลดลงทางเดียว (monotonic decreasing function) ของ k และฟังก์ชัน $\gamma_2(k)$ เป็นฟังก์ชันเพิ่มขึ้นทางเดียว (monotonic increasing function) ของ k ซึ่งความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าต่ำเมื่อ k มีค่าเข้าใกล้ในบริเวณจุดกำเนิด อย่างไรก็ตามคุณลักษณะที่มียุทธศาสตร์ที่สำคัญที่สุด คือ อนุพันธ์ของแต่ละฟังก์ชันเมื่อ k มีค่าเข้าใกล้ในบริเวณจุดกำเนิด ซึ่งสามารถหาได้ดังนี้

$$\begin{aligned}\lim_{k \rightarrow 0^+} \frac{d\gamma_1}{dk} &= \lim_{k \rightarrow 0^+} \frac{d\left(\sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2}\right)}{dk} \\ &= \lim_{k \rightarrow 0^+} \left\{ \sigma^2 \sum_{i=1}^p -2\lambda_i (\lambda_i + k)^{-3} \right\} \\ &= -2\sigma^2 \sum_{i=1}^p \lambda_i^{-2}\end{aligned}$$

$$\begin{aligned}\lim_{k \rightarrow 0^+} \frac{d\gamma_2}{dk} &= \lim_{k \rightarrow 0^+} \frac{d\left(k^2 \beta'(X'X + kI_p)^{-2} \beta\right)}{dk} \\ &= \lim_{k \rightarrow 0^+} \left\{ -2k^2 \beta'(X'X + kI_p)^{-3} \beta(\beta'\beta) + 2k \beta'(X'X + kI_p)^{-2} \beta' \right\} \\ &= 0\end{aligned}$$

ดังนั้น $\gamma_1(k)$ มีค่าอนุพันธ์เป็นค่าลบ ซึ่งจะมีค่าเข้าใกล้ $-2p\sigma^2$ เมื่อ k เข้าใกล้ศูนย์ทางบวก ($k \rightarrow 0^+$) สำหรับ $X'X$ เป็นเมทริกซ์เชิงตั้งฉาก (orthogonal) และมีค่าเข้าใกล้ $-\infty$ เมื่อ $X'X$

เป็นเมทริกซ์ที่มีสถานะไม่เหมาะสม (ill-conditional) และ $\lambda_p \rightarrow 0$ ในทางตรงกันข้าม เมื่อ k เข้าใกล้ศูนย์ทางบวก ($k \rightarrow 0^+$) สมการ (28) แสดงให้เห็นว่าค่าอนุพันธ์ของ $\gamma_2(k)$ มีลักษณะแบน (flat) และมีค่าเป็นศูนย์ที่จุดกำเนิด คุณสมบัติเหล่านี้นำไปสู่ข้อสรุปได้ว่า ความเป็นไปได้ที่จะเคลื่อนที่ค่า $k > 0$ จะทำให้เกิดความเอนเอียงเล็กน้อย และลดความแปรปรวนลงได้อย่างมาก ด้วยวิธีนี้ทำให้มีการปรับแก้ไขความคลาดเคลื่อนของการประมาณค่าและการพยากรณ์ให้ดีขึ้น

หากตัวแปรอิสระทุกตัวมีความสัมพันธ์กับตัวแปรตาม เราสามารถเลือกใช้การวิเคราะห์การถดถอยวิธีริตจ์ ซึ่งจะเป็นการตั้งค่าประมาณโดยวิธีกำลังสองน้อยที่สุดลดลงไปด้านศูนย์ ทำให้ได้ค่า $\hat{\beta}_{ridge}$ ทุกตัวมีขนาดเล็กแต่ไม่เท่ากับศูนย์ ดังนั้นวิธีริตจ์จึงมีความเหมาะสมในการวิเคราะห์ข้อมูลที่มีสัมประสิทธิ์การถดถอยขนาดเล็กที่ไม่เท่ากับศูนย์จำนวนมาก ตัวประมาณนี้เป็นตัวประมาณที่เอนเอียง (biased estimator) สำหรับพารามิเตอร์ β แต่สามารถลดความแปรปรวนของตัวประมาณให้น้อยลง สมบัติเด่นของการถดถอยแบบริตจ์ คือ ตัวประมาณที่ได้จะมีความเสถียร จึงเป็นอีกวิธีที่ใช้กันมากเพื่อแก้ไขปัญหาค่าตัวแปรอิสระมีความสัมพันธ์กันสูง แม้ว่าข้อมูลจะไม่มีมิติสูงก็ตาม แต่สำหรับข้อจำกัดของวิธีนี้ คือ การขาดสมบัติในการเลือกตัวแปรอิสระทำให้การแปลผลตัวแปรในตัวแบบมีความซับซ้อนในกรณีที่ข้อมูลมีมิติสูง

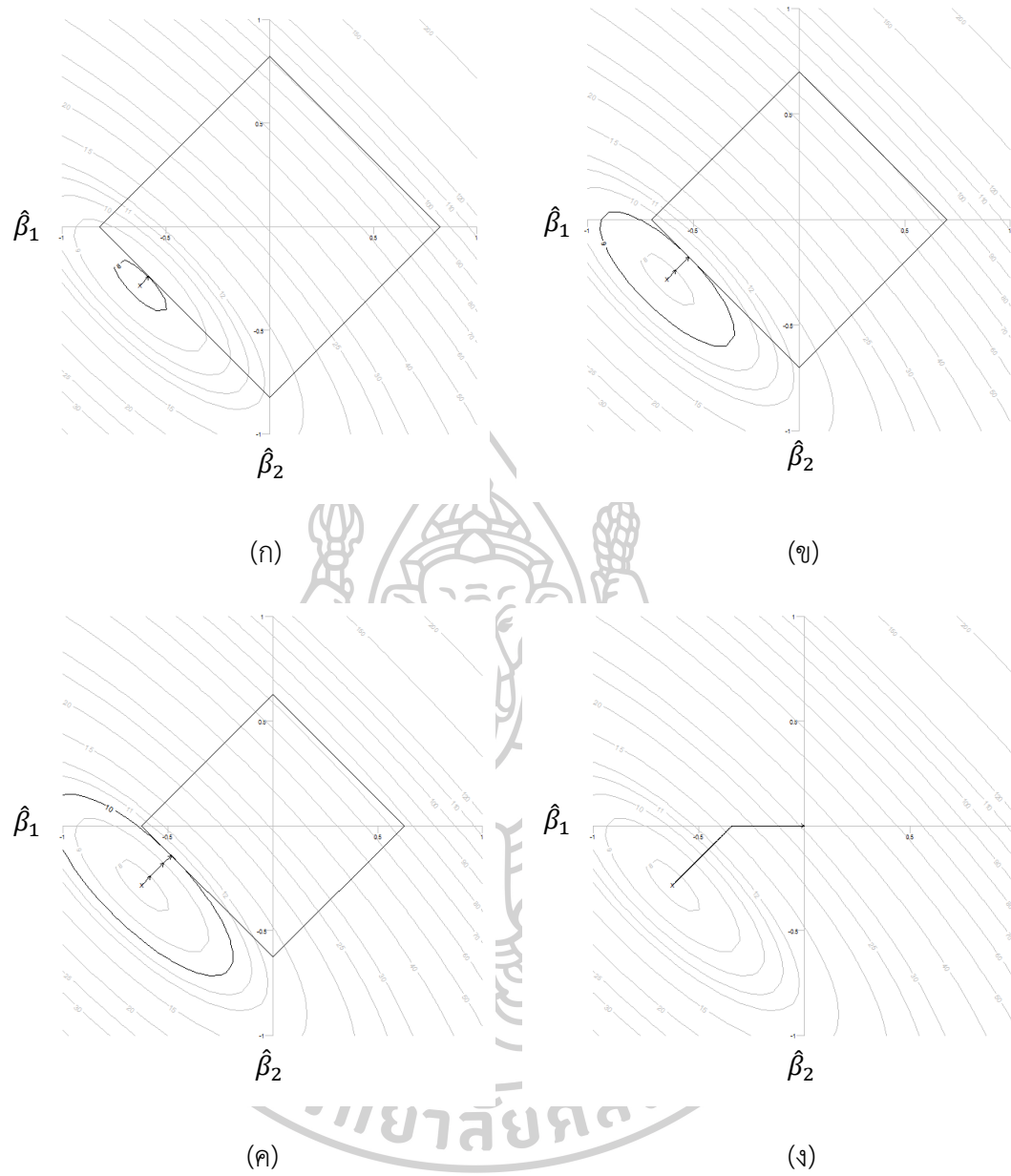
4. การถดถอยเชิงเส้นพหุคูณด้วยวิธีแลซโซ (Least absolute selection and shrinkage operator regression : LASSO regression)

ทิบชิรานี (Tibshirani, 1996: 267-268) ได้กล่าวถึงตัวประมาณกำลังสองน้อยที่สุดของการถดถอยพหุคูณใน 2 ประเด็น คือ ความถูกต้องของการพยากรณ์และความแปลความหมายของสมการถดถอย ในประเด็นแรกเกี่ยวกับความถูกต้องของการพยากรณ์ จะพบว่า การวิเคราะห์การถดถอยพหุคูณในบางครั้ง ตัวประมาณกำลังสองน้อยที่สุดที่ได้ เป็นตัวประมาณที่ไม่เอนเอียง แต่มีความแปรปรวนสูง ซึ่งทำให้การพยากรณ์ของสมการถดถอยไม่ถูกต้อง และความถูกต้องของการพยากรณ์สามารถถูกปรับปรุงให้ดีขึ้นด้วยวิธีการลดขนาดของตัวประมาณ หรือทำให้ตัวประมาณมีค่าเท่ากับศูนย์ โดยการเพิ่มความเอนเอียงของตัวประมาณ และลดความแปรปรวน ซึ่งจะทำให้การพยากรณ์มีความถูกต้องมากขึ้น ประเด็นที่สอง คือ การแปลความหมายของสมการถดถอย ถ้าตัวแปรอิสระมีจำนวนมาก นักวิจัยส่วนใหญ่จะสนใจเฉพาะเซตย่อยของตัวแปรอิสระที่มีอิทธิพลต่อตัวแปรตามมากที่สุดแทน

เทคนิคที่ช่วยในการปรับปรุงการวิเคราะห์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยที่สุดใน 2 ประเด็นที่กล่าวมาข้างต้น คือ การคัดเลือกตัวแปรอิสระ (subset selection) และการถดถอยวิธีริตจ์ เทคนิคในการคัดเลือกตัวแปรอิสระ (subset selection) ช่วยให้สมการถดถอยพหุคูณด้วยวิธีกำลังสองน้อยที่สุด มีจำนวนตัวแปรอิสระที่สามารถอธิบายความหมายได้ แต่การคัดเลือกตัวแปรอิสระ เป็นกระบวนการไม่ต่อเนื่อง (discrete process) นั่นคือ ตัวแปรอิสระที่ละตัวจะถูกพิจารณาว่าจะเข้าหรือออกจากสมการถดถอย ซึ่งถ้าข้อมูลที่นำมาวิเคราะห์มีการเปลี่ยนแปลงไปเล็กน้อย ก็จะทำให้ผลการคัดเลือกตัวแปรอิสระในสมการถดถอยพหุคูณมีความแตกต่างกันมาก และส่งผลให้ความถูกต้องของการพยากรณ์ลดลง แต่ตัวประมาณริตจ์เป็นวิธีการหาตัวประมาณที่เป็นกระบวนการต่อเนื่อง (continuous process) โดยการปรับลดขนาดของตัวประมาณ ซึ่งมีความเสถียร (stable) ของความถูกต้องของการพยากรณ์ แต่ตัวประมาณริตจ์ จะไม่มีค่าเท่ากับศูนย์ ทำให้ตัวแปรอิสระทุกตัวอยู่ในสมการถดถอย ซึ่งยากต่อการแปลความหมาย

ทิบชิรานี (Tibshirani, 1996: 268-273) จึงได้เสนอวิธีการประมาณพารามิเตอร์ที่ชื่อว่า LASSO ที่ย่อมาจากคำว่า Least Absolute Selection and Shrinkage Operator ซึ่งเป็นวิธีการปรับค่าของตัวประมาณสัมประสิทธิ์การถดถอยบางตัวให้ลดลง และให้ค่าตัวประมาณสัมประสิทธิ์การถดถอยบางตัวมีค่าเท่ากับศูนย์ โดยพยายามที่จะรักษาสสมบัติที่ดีของตัวประมาณริตจ์ และการคัดเลือกตัวแปรให้คงอยู่ในวิธีการของ LASSO ซึ่งหลักการของหาตัวประมาณสัมประสิทธิ์การถดถอยด้วยวิธี LASSO มีความคล้ายคลึงกับตัวประมาณริตจ์ คือ การหาตัวประมาณของเวกเตอร์พารามิเตอร์ β ที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนสุ่มมีค่าต่ำสุด ภายใต้ข้อจำกัดของ L_1 นอร์ม ของ β คือ $\|\beta\|^2 = \sum_{j=1}^p |\beta_j| \leq t$ เมื่อ t เป็นค่าคงที่ ซึ่งแตกต่างกับตัวประมาณริตจ์ที่มีข้อจำกัด L_2 นอร์ม ของ β

ซิมอน (Simon, 2016: 1-9) ได้อธิบายวิธีการหาตัวประมาณ LASSO จากการวิเคราะห์การถดถอยพหุคูณ โดยใช้ชุดข้อมูลที่อยู่ในแพ็คเกจ mtcars ของซอฟต์แวร์ R ซึ่งเป็นข้อมูลเกี่ยวกับการเผาผลาญพลังงานเชื้อเพลิง การออกแบบเครื่องยนต์ และประสิทธิภาพการทำงานของเครื่องยนต์จำนวน 32 รุ่นที่ผลิตปี 1973-1974 จากงานวิจัยของ เฮนเดอร์สันและเวลลีแมน (Henderson and Velleman, 1981: 396) ซึ่งเป็นชุดข้อมูลเดียวกับที่กล่าวไว้ในหัวข้อการวิเคราะห์การถดถอยวิธีริตจ์ ซิมอนได้แสดงตัวอย่างของการถดถอยพหุคูณวิธี LASSO ของสมการถดถอยพหุคูณที่มีตัวแปรอิสระ 2 ตัว คือ $\widehat{\text{mpg}} = \beta_1 \text{disp} + \beta_2 \text{hp}$ โดยที่ mpg คือ ระยะทางที่รถวิ่งได้ (ไมล์) ต่อน้ำมัน 1 แกลลอน (mpg) disp คือ ความจุของเครื่องยนต์ และ hp คือ แรงม้ารวม ภาพของพื้นที่ผิวตอบสนองและคอนทัวร์ของผลบวกกำลังสองของส่วนเหลือของพิกัด (β_1, β_2) บนระนาบพิกัดฉากเหมือนกันภาพที่ 1 (ก) ละ (ข) ตามลำดับ ซึ่งได้แสดงไว้ในหัวข้อที่แล้ว



ภาพที่ 4 กราฟแสดงวิธีการหาตัวประมาณ LASSO และเส้นทางของพิกัดของ $(\hat{\beta}_1, \hat{\beta}_2)$ ของตัวประมาณ LASSO ภายใต้ข้อจำกัด $\sum_{j=1}^2 |\beta_j| \leq t$ (Simon, 2016: 11-14)

การวิเคราะห์การถดถอยวิธี LASSO พิจารณาค่าต่ำสุดของผลบวกกำลังสองของความคลาดเคลื่อนสุ่ม ภายใต้ข้อจำกัด $\sum_{j=1}^p |\beta_j| \leq t$ สำหรับค่าคงที่ t ใดๆ ที่กำหนด จากตัวอย่างสมการถดถอย $\widehat{\text{mpg}} = \beta_1 \text{disp} + \beta_2 \text{hp}$ จะได้ว่าข้อจำกัดของ $\sum_{j=1}^2 |\beta_j| = |\beta_1| + |\beta_2| \leq t$ เป็นฟังก์ชันที่มีกราฟเป็นรูปสี่เหลี่ยมข้าวหลามตัด (diamond shape) เมื่อ t เป็นค่าคงที่ ดังนั้นเราจะพิจารณา พิกัด $(\hat{\beta}_1, \hat{\beta}_2)$ ที่มีค่าของผลบวกกำลังสองของส่วนเหลือต่ำสุดและอยู่ภายใต้ข้อจำกัดของ $|\beta_1| + |\beta_2| \leq t$ นั่นคือจุดสัมผัสของเส้นคอนทัวร์กับสี่เหลี่ยมข้าวหลามตัด ดังแสดงในภาพที่ 4 (ก) - (ค) พิจารณาภาพที่ 4 (ก) เมื่อตรงค่าผลบวกกำลังสองของส่วนเหลือเท่ากับ 8 แล้วขนาดของสี่เหลี่ยมข้าวหลามตัดที่สัมผัสกับเส้นคอนทัวร์ จุดสัมผัส คือ พิกัดของ $(\hat{\beta}_1, \hat{\beta}_2)$ ที่เป็นตัวประมาณ LASSO ภาพที่ 4 (ข) และ (ค) เมื่อตรงค่าผลบวกกำลังสองของส่วนเหลือเท่ากับ 9 และ 10 แล้วหาขนาดของสี่เหลี่ยมข้าวหลามตัดที่สัมผัสกับเส้นคอนทัวร์ ซึ่งขนาดของสี่เหลี่ยมจะลดลงเมื่อค่าผลบวกกำลังสองของส่วนเหลือมากขึ้น ซึ่งจะทำให้พิกัดของ $(\hat{\beta}_1, \hat{\beta}_2)$ จะเข้าใกล้จุดกำเนิด (0,0) มากขึ้น เมื่อเชื่อมโยงพิกัดของ $(\hat{\beta}_1, \hat{\beta}_2)$ ที่ได้จากภาพที่ 4 (ก) - (ค) ทำให้เห็นเส้นทางของผลลัพธ์ของวิธี LASSO ซึ่งค่าของตัวประมาณ LASSO จะมีค่าเข้าสู่แกนนอน $(\hat{\beta}_1, 0)$ โดยทำมุม 45 องศา จากนั้น $\hat{\beta}_1$ มีค่าเข้าสู่ศูนย์ ทำให้ $(\hat{\beta}_1, \hat{\beta}_2)$ เข้าสู่จุดกำเนิด (0,0) ดังแสดงในภาพที่ 4 (ง)

4.1 การประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีแลชโซ

ตัวประมาณสัมประสิทธิ์การถดถอยด้วยวิธี LASSO สามารถนิยามโดย

$$\text{minimize}_{\beta} \quad (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$$

$$\text{ภายใต้ข้อจำกัด} \quad \sum_{j=1}^p |\beta_j| \leq t$$

เมื่อ $t \geq 0$ เรียกว่า พารามิเตอร์การปรับแต่ง (tuning parameter) ซึ่งควบคุมขนาดของการหดตัว (shrinkage) ของตัวประมาณ $\hat{\beta}$ ถ้า $t_0 = \sum |\hat{\beta}^{ols}|$ เมื่อกำหนด ค่า $t < t_0$ จะทำให้ตัวประมาณสัมประสิทธิ์ถูกดึงค่าลดลงเข้าหาศูนย์ และตัวประมาณบางตัวจะมีค่าเท่ากับศูนย์ ซึ่งสามารถเขียนอยู่ในรูปสมการที่สอดคล้องกัน คือ

$$\hat{\beta}^{lasso} = \text{argmin}_{\beta} (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) + \lambda \sum_{j=1}^p |\beta_j|$$

เนื่องจากตัวประมาณ LASSO เป็นตัวประมาณที่เป็นคำตอบของฟังก์ชันที่ไม่เป็นเชิงเส้น และไม่สามารถหาอนุพันธ์เทียบกับ β ได้ สำหรับค่า λ ใดๆ ที่กำหนด เราจึงไม่สามารถเขียนตัวประมาณ LASSO อยู่ในสูตรที่ชัดเจนได้ในกรณีทั่วไป แต่เราสามารถศึกษาตัวประมาณ LASSO ในกรณีเฉพาะ คือ ในกรณีของเมทริกซ์ของตัวแปรอิสระ \mathbf{X} อยู่ในรูปเมทริกซ์เชิงตั้งฉากปกติ

(orthonormal design matrix) นั่นคือ $\mathbf{X}'\mathbf{X} = \mathbf{I}_p$ เพื่อแสดงให้เห็นถึงทิศทางและขนาดของการปรับลดลง (shrinkage) ของตัวประมาณ LASSO เทียบกับตัวประมาณ OLS ได้

พิจารณาในกรณีของเมทริกซ์ของตัวแปรอิสระ \mathbf{X} อยู่ในรูปเมทริกซ์เชิงตั้งฉากปกติ นั่นคือ สมมติให้ $\mathbf{X}'\mathbf{X} = \mathbf{I}_p$ ดังนั้น เราจะได้ว่า

$$\hat{\boldsymbol{\beta}}^{ols} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{Y}$$

$$\text{และ } (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{Y}'\mathbf{Y} - 2(\mathbf{X}'\mathbf{Y})'\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

$$= \mathbf{Y}'\mathbf{Y} - 2(\hat{\boldsymbol{\beta}}^{ols})'\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{\beta}$$

$$= \sum_{i=1}^n Y_{ij}^2 - 2 \sum_{j=1}^p \hat{\beta}_j^{ols} \beta_j + \sum_{j=1}^p \beta_j^2$$

ดังนั้น ในกรณีของเมทริกซ์การตั้งฉากปกติ ตัวประมาณ LASSO สามารถหาได้จาก

$$\text{minimize } \beta_j \quad \sum_{i=1}^n Y_{ij}^2 - 2 \sum_{j=1}^p \hat{\beta}_j^{ols} \beta_j + \sum_{j=1}^p \beta_j^2$$

$$\text{ภายใต้ข้อจำกัด } \sum_{j=1}^p |\beta_j| \leq t$$

ซึ่งผลเฉลยของปัญหาค่าเหมาะสมที่สุดดังกล่าวนี้ สามารถทำได้โดยใช้วิธีตัวคูณลากรองจ์ เมื่อ λ เป็นตัวคูณลากรองจ์ จะได้ฟังก์ชันลากรองจ์ คือ

$$L(\mathbf{Y}, \boldsymbol{\beta}, \lambda) = \sum_{i=1}^n Y_{ij}^2 - 2 \sum_{j=1}^p \hat{\beta}_j^{ols} \beta_j + \sum_{j=1}^p \beta_j^2 - \lambda \left(\sum_{j=1}^p |\beta_j| - t \right)$$

กำหนด $k = -\lambda$ โดยที่ $k \geq 0$ ดังนั้น

$$\begin{aligned} L(\mathbf{Y}, \boldsymbol{\beta}, \lambda) &= \sum_{i=1}^n Y_{ij}^2 - 2 \sum_{j=1}^p \hat{\beta}_j^{ols} \beta_j + \sum_{j=1}^p \beta_j^2 + k \left(\sum_{j=1}^p |\beta_j| - t \right) \\ &= \sum_{i=1}^n Y_{ij}^2 + \sum_{j=1}^p [(-2)\hat{\beta}_j^{ols} \beta_j + \beta_j^2 + k|\beta_j|] + kt \end{aligned}$$

หาอนุพันธ์บางส่วนของ $L(\mathbf{Y}, \boldsymbol{\beta}, k)$ เทียบกับ β_j

$$\begin{aligned} \frac{\partial L(\mathbf{Y}, \boldsymbol{\beta}, k)}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} \left[\sum_{i=1}^n Y_{ij}^2 + \sum_{j=1}^p [(-2)\hat{\beta}_j^{ols} \beta_j + \beta_j^2 + k|\beta_j|] + kt \right] \\ &= \frac{\partial}{\partial \beta_j} \left[\sum_{j=1}^p [(-2)\hat{\beta}_j^{ols} \beta_j + \beta_j^2 + k|\beta_j|] \right] \\ &= \frac{\partial}{\partial \beta_j} (-2\hat{\beta}_j^{ols} \beta_j + \beta_j^2 + k|\beta_j|) \end{aligned} \tag{36}$$

สมการ (36) จะมีค่าต่ำสุด เมื่อ $\hat{\beta}_j^{ols} \beta_j > 0$ นั่นคือ เครื่องหมายของ $\hat{\beta}_j^{ols}$ และ β_j จะต้องมีความหมายเดียวกัน ซึ่งอาจจะเป็นเครื่องหมายบวกทั้งคู่ หรือลบทั้งคู่ เราจึงพิจารณาเป็นสองกรณี ดังนี้

กรณี 1 : $\hat{\beta}_j^{ols} > 0$ และ $\beta_j > 0$;

$$\frac{\partial}{\partial \beta_j} (-2\hat{\beta}_j^{ols} \beta_j + \beta_j^2 + k\beta_j) = 0$$

$$-2\hat{\beta}_j^{ols} + 2\beta_j + k = 0$$

$$2\beta_j = 2\hat{\beta}_j^{ols} - k$$

$$\beta_j = \hat{\beta}_j^{ols} - \frac{k}{2}$$

เนื่องจาก $\beta_j > 0$ ทำให้ $\hat{\beta}_j^{ols} - \frac{k}{2} > 0$ ดังนั้น

$$\hat{\beta}_j = \begin{cases} \hat{\beta}_j^{ols} - \frac{k}{2} ; & \hat{\beta}_j^{ols} > \frac{k}{2} \\ 0 & ; 0 < \hat{\beta}_j^{ols} < \frac{k}{2} \end{cases}$$

กรณี 2 : $\hat{\beta}_j^{ols} \leq 0$ และ $\beta_j \leq 0$;

$$\frac{\partial}{\partial \beta_j} (-2\hat{\beta}_j^{ols} \beta_j + \beta_j^2 - k\beta_j) = 0$$

$$-2\hat{\beta}_j^{ols} + 2\beta_j - k = 0$$

$$2\beta_j = 2\hat{\beta}_j^{ols} + k$$

$$\hat{\beta}_j = \hat{\beta}_j^{ols} + \frac{k}{2}$$

เนื่องจาก $\beta_j \leq 0$ จึงทำให้ $\hat{\beta}_j^{ols} + \frac{k}{2} \leq 0$ ซึ่งสมมูลกับ $-\hat{\beta}_j^{ols} - \frac{k}{2} \geq 0$ แต่ $\hat{\beta}_j^{ols} \leq 0$ นั่นคือ $-\hat{\beta}_j^{ols} \geq 0$ ส่งผลให้ $|\hat{\beta}_j^{ols}| - \frac{k}{2} \geq 0$ ดังนั้นจะได้ว่า $|\hat{\beta}_j^{ols}| \geq \frac{k}{2}$

$$\hat{\beta}_j = \begin{cases} \hat{\beta}_j^{ols} + \frac{k}{2} ; & \hat{\beta}_j^{ols} < -\frac{k}{2} \\ 0 & ; -\frac{k}{2} < \hat{\beta}_j^{ols} < 0 \end{cases}$$

จากทั้งสองกรณีข้างต้น จะได้ตัวประมาณสัมประสิทธิ์การถดถอยด้วยวิธี LASSO คือ

$$\hat{\beta}_j^{lasso} = \begin{cases} \hat{\beta}_j^{ols} - \frac{k}{2}; & \hat{\beta}_j^{ols} > \frac{k}{2} \\ 0 & -\frac{k}{2} \leq \hat{\beta}_j^{ols} \leq \frac{k}{2} \\ \hat{\beta}_j^{ols} + \frac{k}{2}; & \hat{\beta}_j^{ols} < -\frac{k}{2} \end{cases} \quad (37)$$

เมื่อ $k \geq 0$ เป็นค่าคงที่ที่พิจารณาจากข้อจำกัด $\sum_{j=1}^p |\hat{\beta}_j| = t$

จาก (37) จะเห็นว่า $\hat{\beta}_j^{lasso}$ จะถูกปรับค่าลดลงมีขนาดเท่ากับ $|\hat{\beta}_j^{ols}| - \frac{k}{2}$ และมีเครื่องหมายเดียวกับ $\hat{\beta}_j^{ols}$ ในกรณีที่ $|\hat{\beta}_j^{ols}|$ มีค่ามากกว่า $\frac{k}{2}$ และ $\hat{\beta}_j^{lasso}$ จะมีค่าเท่ากับศูนย์ในกรณีที่ $|\hat{\beta}_j^{ols}|$ น้อยกว่า $\frac{k}{2}$ ซึ่งฟังก์ชันนี้มีรูปแบบเดียวกันกับการกำหนดเกณฑ์อย่างอ่อน (soft-thresholding operator) ที่ได้กล่าวไว้ในงานวิจัยของโดโนโล และจอห์นสโตน (Donolo and Johnstone, 1993: 8) ซึ่งสมการ (37) สามารถเขียนอยู่ในอีกรูปแบบหนึ่ง คือ

$$\hat{\beta}_j^{lasso} = \text{sign}(\hat{\beta}_j^{ols})(|\hat{\beta}_j^{ols}| - \gamma)^+$$

เมื่อ $\gamma = \frac{k}{2}$ ถูกกำหนดภายใต้ข้อจำกัด $\sum_{j=1}^p |\hat{\beta}_j| = t$ และ $(z)^+ = z I_{(z>0)}$ เมื่อ z เป็นค่าของตัวแปรในฟังก์ชัน และ I_A หมายถึง ฟังก์ชันบ่งชี้ (indicator function) โดยที่ $I_A = 1$ ถ้า A เป็นจริงและเท่ากับศูนย์ในกรณีอื่นๆ

จากรูปแบบของตัวประมาณ LASSO ในสมการ (37) จะเห็นว่า ตัวประมาณ $\hat{\beta}_j^{lasso}$ บางตัวมีค่าเป็นศูนย์ ดังนั้นจึงเป็นเหตุผลว่า การประมาณค่าสัมประสิทธิ์การถดถอยโดยวิธี LASSO นั้น มีการคัดเลือกตัวแปรเข้าสู่สมการและตัวประมาณสัมประสิทธิ์การถดถอยถูกปรับค่าลดลงเข้าหาศูนย์ในคราวเดียวกัน

4.2 การเปรียบเทียบตัวประมาณสัมประสิทธิ์การถดถอยวิธีแลซโซกับวิธีริตจ์

ดังที่ได้กล่าวมาตอนต้น ตัวประมาณ LASSO เป็นวิธีการปรับค่าของตัวประมาณสัมประสิทธิ์การถดถอยบางตัวให้ลดลง และให้ค่าตัวประมาณสัมประสิทธิ์การถดถอยบางตัวมีค่าเท่ากับศูนย์ โดยพยายามที่จะรักษาสมบัติที่ดีของตัวประมาณริตจ์ และการคัดเลือกตัวแปร ดังนั้น ถ้าพิจารณาเปรียบเทียบตัวประมาณ LASSO กับตัวประมาณวิธี best subset selection และวิธีริตจ์ ซึ่งสามารถเขียนอยู่ในรูปของฟังก์ชันของตัวประมาณ OLS เมื่อเมทริกซ์ $\mathbf{X}'\mathbf{X}$ อยู่ในรูปของเมทริกซ์เชิงตั้งฉากปกติ ได้ดังนี้ (Tibshirani, 1996: 269)

ตัวประมาณวิธี best subset selection ของตัวแบบถดถอยพหุคูณที่มีตัวแปรอิสระ p ตัว และคัดเลือกตัวประมาณสัมประสิทธิ์การถดถอยที่มีขนาดของสัมประสิทธิ์ (ค่าสัมบูรณ์) สูงที่สุด

จำนวน q ตัว และค่าของตัวประมาณสัมประสิทธิ์ถดถอยที่เหลือมีค่าเท่ากับศูนย์ สามารถเขียนอยู่ในรูปของฟังก์ชันของตัวประมาณ OLS ได้ดังนี้

สำหรับค่า γ ใดๆ ที่กำหนดเป็นเกณฑ์สำหรับการคัดเลือก และตัวประมาณสัมประสิทธิ์ถดถอยที่ไม่ถูกเลือก มีค่าเท่ากับศูนย์ นั่นคือ

$$\hat{\beta}_j = \begin{cases} \hat{\beta}_j^{ols} & ; \quad |\hat{\beta}_j^{ols}| > \gamma \\ 0 & ; \quad \text{อื่นๆ} \end{cases}$$

จากหัวข้อ 2.3 ในสมการ (21) ตัวประมาณของสัมประสิทธิ์การถดถอยวิธีริดจ์ คือ $\hat{\beta}^{ridge} = [I_p + k(X'X)^{-1}]^{-1} \hat{\beta}^{ols}$ ซึ่งในกรณีของเมทริกซ์ตั้งฉากปกติ จะได้ว่า

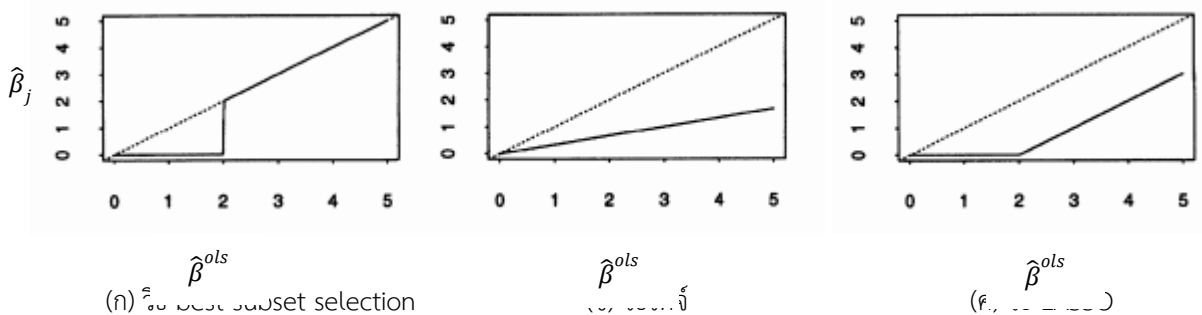
$$\begin{aligned} \hat{\beta}^{ridge} &= [I_p + kI_p]^{-1} \hat{\beta}^{ols} \\ &= (1+k)^{-1} \hat{\beta}^{ols} \end{aligned}$$

ดังนั้น

$$\hat{\beta}_j^{ridge} = \left(\frac{1}{1+k}\right) \hat{\beta}_j^{ols}$$

เมื่อ $j = 1, 2, \dots, p$ และ k ขึ้นอยู่กับค่า t

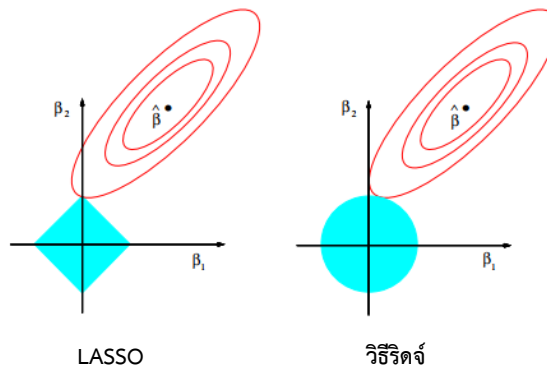
จากฟังก์ชันของตัวประมาณวิธี best subset selection ($\hat{\beta}_j$), ตัวประมาณริดจ์ ($\hat{\beta}_j^{ridge}$) และตัวประมาณ LASSO ($\hat{\beta}_j^{lasso}$) สามารถเขียนกราฟแสดงความสัมพันธ์ระหว่าง $\hat{\beta}_j$, $\hat{\beta}_j^{ridge}$ และ $\hat{\beta}_j^{lasso}$ กับตัวประมาณ $\hat{\beta}_j^{ols}$ ดังภาพที่ 5



ภาพที่ 5 แสดงความสัมพันธ์ของการสัมประสิทธิ์การถดถอยโดยวิธีต่าง ๆ กับสัมประสิทธิ์การถดถอยวิธีกำลังสองน้อยที่สุด ในกรณีของเมทริกซ์ตั้งฉากปกติ เมื่อ $\gamma = 2$ (Tibshirani, 1996 : 270)

ภาพที่ 5 แสดงความสัมพันธ์ระหว่างตัวประมาณสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยสุด และตัวประมาณวิธี Subset selection, ตัวประมาณริตจ์ และตัวประมาณ LASSO แสดงในภาพที่ 5 (ก), (ข), (ค) ตามลำดับ แกนนอนคือค่าสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยที่สุดและแกนตั้งคือค่าประมาณของสัมประสิทธิ์การถดถอยแต่ละวิธี เส้นประทำมุม 45° กับ แกนนอน คือ เส้นตรงที่ค่าสัมประสิทธิ์การถดถอยโดยวิธีต่าง ๆ เท่ากับค่าสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด ซึ่งแสดงเฉพาะจุดภาค (quadrant) ที่ 1 จากภาพที่ 5 (ก) แสดงความสัมพันธ์ระหว่างตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด และตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธี Subset selection โดย $\hat{\beta}_j = \hat{\beta}_j^{ols}$ เมื่อ $\hat{\beta}_j^{ols} > 2$ และ $\hat{\beta}_j = 0$ เมื่อ $\hat{\beta}_j^{ols} < 2$ แสดงว่า ตัวประมาณสัมประสิทธิ์การถดถอยที่มีค่ามากกว่า 2 จะถูกคัดเลือกเข้าสมการ จากภาพที่ 5 (ข) แสดงความสัมพันธ์ระหว่างตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด และตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธีริตจ์ จะเห็นว่า $\hat{\beta}_j = \left(\frac{1}{1+2}\right)\hat{\beta}_j^{ols} = \left(\frac{1}{3}\right)\hat{\beta}_j^{ols}$ นั่นคือ ตัวประมาณแบบริตจ์ จะมีค่าลดลงเป็น 1 ใน 3 ของตัวประมาณโดยวิธีกำลังสองน้อยที่สุด จากภาพที่ 5 (ค) แสดงความสัมพันธ์ระหว่างตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด และตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธี LASSO จะเห็นว่า $\hat{\beta}_j = \hat{\beta}_j^{ols} - 2$ เมื่อ $\hat{\beta}_j^{ols} > 2$ และ $\hat{\beta}_j = 0$ เมื่อ $\hat{\beta}_j^{ols} \leq 2$ ดังนั้น ตัวประมาณสัมประสิทธิ์การถดถอยที่มีค่ามากกว่า 2 จะถูกคัดเลือกในสมการ และทำให้มีขนาดลดลงโดยมีค่าน้อยกว่า $\hat{\beta}_j^{ols}$ เท่ากับ 2 หน่วย ทุกค่า

ภาพที่ 6 แสดงการเมื่อเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO และวิธีริตจ์ในกรณี $p=2$ เส้นวงรีแสดงเส้นคอนทัวร์ (Contour Line) ของฟังก์ชันภาวะน่าจะเป็นสูงสุด (Maximum likelihood function) โดยมีค่าประมาณโดยวิธีกำลังสองน้อยที่สุด ($\hat{\beta}^{ols}$) เป็นจุดศูนย์กลางวงรี พื้นที่แรเงาคือ เงื่อนไขของพารามิเตอร์แบบ LASSO ได้แก่ $|\beta_1| + |\beta_2| \leq t$ ซึ่งเป็นรูปสี่เหลี่ยมเอียง และแบบริตจ์ได้แก่ $\beta_1^2 + \beta_2^2 \leq t^2$ ตามลำดับ และจุดที่เส้นความชันสูงที่น้อยที่สุดสัมผัสกับพื้นที่เงื่อนไขบังคับ จะเป็นตัวประมาณสัมประสิทธิ์การถดถอยวิธี LASSO และวิธีริตจ์ โดยจะพบว่า วิธี LASSO นั้น บางครั้งอาจจะมีโอกาสที่เส้นความชันสูงสัมผัสกับพื้นที่เงื่อนไขบังคับที่จุดมุมของรูปสี่เหลี่ยม นั่นหมายความว่า สัมประสิทธิ์ $\beta_j = 0$ ส่งผลให้มีการคัดเลือกตัวแปรเกิดขึ้น โดยเส้นความชันสูงนี้จะต้องสัมผัสกับรูปสี่เหลี่ยมในจุดภาค (quadrant) เดียวกับ $\hat{\beta}^{ols}$ เนื่องจากข้างต้น $Q = -2\hat{\beta}_j^{ols} \beta_j + \hat{\beta}_j^{ols} \beta_j + k|\beta_j|$ โดย Q จะมีค่าต่ำสุด เมื่อ $\hat{\beta}_j^{ols} \cdot \beta_j > 0$ ดังนั้น เมื่อ $\hat{\beta}_j^{ols} > 0$ β_j จะต้องมากกว่าศูนย์ด้วย และเมื่อ $\hat{\beta}_j^{ols} \leq 0$ นั้น $\beta_j \leq 0$ ด้วย จึงส่งผลให้ $\hat{\beta}^{lasso}$ อยู่จุดภาคเดียวกันกับ $\hat{\beta}^{ols}$ แต่การประมาณแบบริตจ์นั้นไม่มีโอกาสที่เส้นความชันสูงจะสัมผัสกับพื้นที่เงื่อนไขที่จุดมุม จึงไม่มีการคัดเลือกตัวแปรเกิดขึ้นนั่นเอง



ภาพที่ 6 การประมาณค่าพารามเตอร์สมประสพการถดถอยแบบ LASSO และแบบบริดจ์
ในกรณี $p=2$ (Tibshirani, 1996: 271)

เมื่อพิจารณากรณีที่ตัวแปรอิสระ 2 ตัว นั่นคือ $p=2$ และสมมติให้ตัวประมาณ $\hat{\beta}_j^{ols}$ เมื่อ $j=1,2$ มีค่าเป็นบวกทั้งคู่ ดังนั้น จะได้ตัวประมาณแบบ LASSO คือ

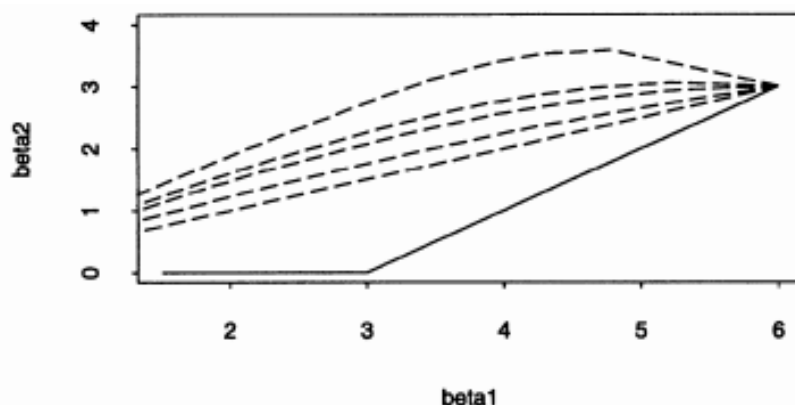
$$\hat{\beta}^{lasso} = (\hat{\beta}_j^{ols} - \gamma)^+$$

เมื่อ $t = \sum_{j=1}^2 |\hat{\beta}_j^{lasso}| = \hat{\beta}_1^{lasso} + \hat{\beta}_2^{lasso}$ และ $t_0 = \hat{\beta}_1^{ols} + \hat{\beta}_2^{ols}$ ซึ่งสูตรข้างต้นจะเป็นจริง

เมื่อ $t \leq \hat{\beta}_1^{ols} + \hat{\beta}_2^{ols}$ จะได้ผลลัพธ์สำหรับ γ คือ

$$\hat{\beta}_1^{lasso} = \left(\frac{t}{2} + \frac{\hat{\beta}_1^{ols} - \hat{\beta}_2^{ols}}{2} \right)^+ \quad \text{และ} \quad \hat{\beta}_2^{lasso} = \left(\frac{t}{2} - \frac{\hat{\beta}_1^{ols} - \hat{\beta}_2^{ols}}{2} \right)^+$$

ในทางตรงกันข้าม การปรับค่าลดลงของสัมประสิทธิ์การถดถอยวิธีบริดจ์ขึ้นอยู่กับสหสัมพันธ์ระหว่างตัวแปรอิสระ ภาพที่ 7 แสดงตัวอย่างที่เกิดจากการจำลองข้อมูล 100 ชุด จากตัวแบบถดถอย $y = 6x_1 + 3x_2$ เมื่อ x_1 และ x_2 มีการแจกแจงปรกติมาตรฐาน และสหสัมพันธ์ ρ มีค่าต่างๆ โดยแสดงค่าประมาณสัมประสิทธิ์การถดถอยวิธีบริดจ์และ LASSO ภายใต้ข้อจำกัดของ $\beta_1^2 + \beta_2^2$ และ $|\beta_1| + |\beta_2|$ ตามลำดับ โดยเส้นทึบแสดงค่าประมาณสัมประสิทธิ์การถดถอยวิธี LASSO สำหรับทุกค่า ρ และเส้นประแสดงค่าประมาณวิธีบริดจ์ซึ่งขึ้นอยู่กับค่า ρ เมื่อ $\rho = 0$ ค่าประมาณสัมประสิทธิ์การถดถอยวิธีบริดจ์จะมีปรับค่าลดลงเป็นสัดส่วน อย่างไรก็ตาม สำหรับค่า ρ ขนาดใหญ่ ค่าประมาณสัมประสิทธิ์การถดถอยวิธีบริดจ์จะปรับค่าอย่างแตกต่างกันโดยจะมีค่าลดลงเมื่อเพิ่มข้อจำกัดเพียงเล็กน้อย (Tibshirani, 1996: 272)



ภาพที่ 7 คู่อันดับ (β_1, β_2) วิธี LASSO (—) และ วิธีริดจ์(---) ภายใต้ข้อจำกัดของพารามิเตอร์ของทั้งสองวิธี สำหรับตัวแปรอิสระ 2 ตัว โดยเส้นประสัมพันธ์กับค่า $\rho = 0, 0.23, 0.45, 0.68$ และ 0.90 จากกลางขึ้นบน (Tibshirani, 1996: 272)

4.3 ขั้นตอนวิธี (Algorithm) ของการหาตัวประมาณวิธีแลซโซ

สำหรับค่าคงที่ t ใดๆ การหาตัวประมาณ LASSO ที่เป็นคำตอบของปัญหา $\text{minimize}_{\beta} (Y - X\beta)'(Y - X\beta)$ ภายใต้ข้อจำกัด $\sum_{j=1}^p |\beta_j| \leq t$ ไม่มีสูตรที่ชัดเจน ดังนั้น จำเป็นต้องใช้วิธีการหาคำตอบของปัญหาค่าเหมาะสมที่สุด โดยใช้ขั้นตอนวิธีสำหรับการคำนวณเชิงคอมพิวเตอร์เพื่อหาคำตอบปัญหาดังกล่าว ในงานวิจัยของทิบชิรานี (Tibshirani, 1996: 278) ได้เสนอวิธีการโปรแกรมกำลังสอง (quadratic programming) ในการหาตัวประมาณ LASSO ต่อมาอีฟรอน (Efron et al., 2004: 411-416) ได้นำเสนอขั้นตอนวิธีในการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยที่เรียกว่า การถดถอยมุมน้อยที่สุด (Least Angle Regression : LARS) หรือเรียกว่า LARS ซึ่งเป็นขั้นตอนวิธีที่พัฒนาจากหลักการของวิธีการคัดเลือกตัวแปรแบบไปข้างหน้า (Forward selection method) และสามารถนำมาประยุกต์ใช้กับขั้นตอนวิธีของการหาตัวประมาณ LASSO (Efron et al., 2004: 416-417) ซึ่งปัจจุบันนักวิจัยนิยมใช้วิธีการนี้อย่างแพร่หลายการหาตัวประมาณ LASSO โดยสามารถเรียกใช้ชุดคำสั่งสำเร็จรูปที่มีชื่อว่า LARS ซึ่งมีอยู่ในซอฟต์แวร์ทางสถิติ R เขียนโดย ฮาสทาย (Hastie, 2015: 2-9) สำหรับการวิเคราะห์การถดถอยด้วยวิธี LASSO

LARS เป็นวิธีการที่รวดเร็วในการคัดเลือกตัวแปรอิสระ เป็นวิธีการอย่างง่ายในการหาตัวประมาณวิธี LASSO โดยขั้นตอนแรกจะหาตัวแปรอิสระที่สัมพันธ์กับตัวแปรตามมากที่สุดและทำการหาค่าพยากรณ์ ซึ่งจะเปลี่ยนค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระนี้ไปเรื่อยๆ ถึงค่าสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุดจนกระทั่งความสัมพันธ์ระหว่างตัวแปรอิสระและส่วนเหลือ (residual) ลดลง หลังจากนั้นหาตัวแปรอิสระตัวอื่นที่มีความสัมพันธ์กับส่วนเหลือมากที่สุด ตัวแปร

อิสระตัวที่สองนี้จะเข้าสู่เซตของตัวแปรอิสระแล้วเปลี่ยนค่าสัมประสิทธิ์การถดถอยทั้งสองจนกระทั่งความสัมพันธ์ระหว่างตัวแปรอิสระและส่วนเหลือเท่าเดิมหรือลดลงแล้วหาตัวแปรอิสระตัวอื่นที่มีความสัมพันธ์กับส่วนเหลือมากที่สุด ทำกระบวนการนี้ซ้ำๆ จนกระทั่งตัวแปรอิสระทุกตัวเข้าสู่ตัวแบบซึ่งขั้นตอนวิธีของ LARS แสดงได้ดังนี้

1) แปลงค่าของตัวแปรอิสระ ให้เป็นค่ามาตรฐาน ขั้นตอนเริ่มแรก คือ สมการถดถอยพหุคูณ ไม่มีตัวแปรอิสระอยู่ในสมการ นั่นคือ $\beta_1, \beta_2, \dots, \beta_p = 0$ จะได้ว่า $\hat{Y} = \bar{Y}$ และ ส่วนเหลือคือ $\mathbf{r} = \mathbf{Y} - \bar{Y}$

2) หาตัวแปรอิสระ X_j ที่มีความสัมพันธ์กับส่วนเหลือ \mathbf{r} มากที่สุด

3) เพิ่มค่า $\hat{\beta}_j$ โดยเริ่มจากศูนย์ ไปถึงค่าสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด จนกว่าจะมีตัวแปรอิสระ X_k ที่มีความสัมพันธ์กับส่วนเหลือ $\mathbf{r} = \mathbf{Y} - \mathbf{X}\hat{\beta}$ มากกว่า X_j

4) เลือกตัวแปรอิสระ X_j และ X_k เข้าสู่ตัวแบบ แล้วเพิ่มค่า $\hat{\beta}_j$ และ $\hat{\beta}_k$ โดยเริ่มจากค่าศูนย์ไปถึงค่าสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด $\hat{\beta}_j$ และ $\hat{\beta}_k$ จนกว่าจะมีตัวแปรอิสระ X_l ที่มีความสัมพันธ์กับส่วนเหลือ เมื่อตัวแปรอิสระ X_j และ X_k อยู่ในตัวแบบ เมื่อส่วนเหลือคือ $\mathbf{r} = \mathbf{Y} - \mathbf{X}\hat{\beta}$

5) ทำซ้ำกระบวนการอย่างต่อเนื่องจนกว่าตัวแปรอิสระทั้ง p ตัว จะเข้าสู่ตัวแบบทั้งหมด หลังจาก $\min(n-1, p)$ ขั้นตอน ก็จะได้ค่าประมาณสัมประสิทธิ์การถดถอย ถ้า $p > n-1$ จะทำให้มีส่วนเหลือเป็นศูนย์หลังจาก $n-1$ ขั้นตอน (ลบ 1 เนื่องจากการแปลงข้อมูล)

4.4 สมบัติการพยากรณ์ (Oracle properties) ของตัวประมาณวิธีเลขโซ

การวิเคราะห์การถดถอยพหุคูณแบบพินอลโลซ์ ในกรณีที่ข้อมูลมีมิติสูง วิธีการหาตัวประมาณสัมประสิทธิ์การถดถอย จะต้องให้ผลลัพธ์ของสมการถดถอยที่มีความถูกต้องในการพยากรณ์สูง และคัดเลือกตัวแปรอิสระที่ใช้ทำนายได้ถูกต้อง การคัดเลือกตัวแปรอิสระเป็นสิ่งสำคัญในกรณีที่ตัวแบบถดถอยจริง มีลักษณะแบบบางเบา การบ่งชี้ตัวแปรอิสระที่ถูกต้อง ทำให้ได้สมการถดถอยที่เหมาะสมในการพยากรณ์

แฟนและลี (Fan and Li, 2001: 1348-1360) ได้กล่าวถึงสมบัติของการคัดเลือกตัวแปรในการวิเคราะห์การถดถอยพหุคูณแบบพินอลโลซ์ ซึ่งเรียกว่า สมบัติเชิงพยากรณ์ (oracle properties) ของตัวประมาณแบบพินอลโลซ์ และซู่ (Zou, 2006: 1418) ได้อธิบายถึงสมบัติเชิงพยากรณ์ (oracle properties) ดังนี้

จากตัวแบบการถดถอยเชิงเส้นพหุคูณในสมการ (1) $\mathbf{Y} = \beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ เมื่อ $\mathbf{Y} = (Y_1, \dots, Y_n)'$ และ $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip})'$ สำหรับ $i = 1, 2, \dots, n, j = 1, \dots, p$ โดย

ไม่เสียหายทั่วไป เราสมมติให้ X_{ij} ถูกแปลงให้เป็นค่ามาตรฐาน และ Y มีค่าเฉลี่ยเป็นศูนย์ ดังนั้นตัวแบบถดถอย จึงไม่จำเป็นต้องมีค่าจุดตัดแกน Y และภายใต้ข้อตกลงเบื้องต้นของความคลาดเคลื่อนสุ่ม $E(Y|X) = X\beta = X_1\beta_1 + \dots + X_p\beta_p$

กำหนดให้ $\mathcal{A} = \{j: \beta_j \neq 0\}$ และสมมติให้ขนาดของเซต \mathcal{A} หรือ $|\mathcal{A}| = p_1 < p$ นั่นคือ ตัวแบบการถดถอยเชิงเส้นพหุคูณขึ้นอยู่กับเซตย่อยของตัวแปรอิสระ ซึ่งมีจำนวน p_1 ตัว และกำหนดให้ $\hat{\beta}^\delta$ เป็นตัวประมาณสัมประสิทธิ์การถดถอยที่ได้มาจากวิธีการ δ เราสามารถกล่าวได้ว่าวิธี δ เป็นวิธีการพยากรณ์ (oracle procedure) ถ้า $\hat{\beta}^\delta$ มีสมบัติการพยากรณ์ (oracle properties) ดังต่อไปนี้

1) ความคงเส้นคงวาในการคัดเลือกตัวแปรอิสระเข้าสมการถดถอย นั่นคือ

$$\{j: \hat{\beta}_j \neq 0\} = \mathcal{A}$$

2) ตัวประมาณสัมประสิทธิ์ มีการแจกแจงเข้าสู่การแจกแจงปกติ

$$\sqrt{n}(\hat{\beta}_{\mathcal{A}}^\delta - \beta_{\mathcal{A}}) \rightarrow_d N(0, \Sigma)$$

เมื่อ Σ เป็นเมทริกซ์ความแปรปรวนของตัวแบบถดถอย

การหาตัวประมาณสัมประสิทธิ์การถดถอยด้วยวิธี LASSO เป็นเทคนิคหนึ่งที่ได้รับนิยามสำหรับการประมาณค่าพารามิเตอร์และคัดเลือกตัวแปรอิสระพร้อมๆกัน โดยใช้กระบวนการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยอย่างต่อเนื่อง ซึ่งทำให้ความถูกต้องของการพยากรณ์ของสมการถดถอยมีมากขึ้น และการคัดเลือกตัวแปรอิสระด้วยวิธี LASSO มีความคงเส้นคงวา (consistent) ถ้าสมการถดถอยสอดคล้องกับเงื่อนไขบางเงื่อนไข ซึ่งซู (Zou, 2006: 1429) ได้กล่าวถึงเงื่อนไขที่จำเป็นสำหรับการคัดเลือกตัวแปรอิสระด้วยวิธี LASSO ที่มีความคงเส้นคงวา ซึ่งมีรายละเอียดดังนี้

จากตัวแบบการถดถอยเชิงเส้นพหุคูณที่กล่าวข้างต้น ซึ่งค่าของตัวแปรอิสระ X ถูกแปลงเป็นค่ามาตรฐาน และตัวแปรตาม Y มีค่าเฉลี่ยเป็นศูนย์ กำหนดให้ความคลาดเคลื่อนสุ่ม ε_i มีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนเท่ากับ σ^2 นั่นคือ $\varepsilon_i \sim iid N(0, \sigma^2)$ และ $\frac{1}{n}X'X \rightarrow C$ เมื่อ C เป็นเมทริกซ์บวกแน่นอน

โดยไม่เสียหายทั่วไป (without loss of generality) สมมติให้ $\mathcal{A} = \{j: \beta_j \neq 0\} = \{1, 2, \dots, p_1\}$ และ $C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$ เมื่อ C_{11} เป็นเมทริกซ์ขนาด $p_1 \times p_1$

ตัวประมาณ LASSO คือ $\hat{\beta}^{lasso} = argmin_{\beta} (Y - X\beta)'(Y - X\beta) + \lambda_n \sum_{j=1}^p |\beta_j|$

โดยที่ค่า λ_n แปรผันตาม n และ ให้ $\mathcal{A}_n = \{j: \beta_j \neq 0\}$ เราจะได้ว่า ตัวประมาณ LASSO จะเป็นตัวประมาณคงเส้นคงวา ก็ต่อเมื่อ $\lim_n P(\mathcal{A}_n = \mathcal{A}) = 1$

ซู (Zou, 2006: 1429) ได้กล่าวถึงทฤษฎีบทที่เป็นเงื่อนไขจำเป็นสำหรับความคงเส้นคงวา ในการคัดเลือกตัวแปร ของตัวประมาณ LASSO ดังนี้

ทฤษฎีบท 1 ถ้า $\lim_n P(\mathcal{A}_n = \mathcal{A}) = 1$ แล้วจะมีเวกเตอร์เครื่องหมาย $\mathbf{s} = (s_1, \dots, s_{p_1})'$ $s_j = 1$ หรือ -1 ที่ทำให้

$$|\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{s}| \leq 1 \quad (38)$$

อสมการ (38) เป็นอสมการของจำนวนจริงแต่ละค่า (componentwise inequality) ที่เป็นสมาชิกของเวกเตอร์ $\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{s}$ ที่มีขนาด $(p - p_1) \times 1$ กล่าวคือ กำหนดให้ d_i เป็นสมาชิกของเวกเตอร์ $\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{s}$ เราจะได้ว่า $|d_i| \leq 1$ ก็ต่อเมื่อ ค่าสัมบูรณ์ของ d_i น้อยกว่าหรือเท่ากับ 1 หรือ $|d_{ij}| \leq 1$ สำหรับทุกค่า $i = 1, 2, \dots, p - p_1$ และ $j = 1$

ถ้าเงื่อนไข (38) ไม่เป็นจริง สามารถสรุปได้โดยทันทีว่า ตัวประมาณ LASSO เป็นตัวประมาณไม่คงเส้นคงวา

สรุปได้ว่า วิธีการประมาณพารามิเตอร์แบบ LASSO เป็นวิธีที่สามารถเลือกตัวแปรอิสระเข้าสู่ตัวแบบและประมาณค่าพารามิเตอร์ได้ในคราวเดียวกัน ตัวประมาณที่ได้จากวิธี LASSO จึงเป็นการเลือกตัวแปรอิสระได้โดยอัตโนมัติ ซึ่งวิธีนี้ถูกนำไปใช้งานจริงอย่างแพร่หลาย เนื่องจากเป็นวิธีการถดถอยเชิงเส้นพหุคูณแบบพินอลไลซ์ที่มีสมบัติของการคัดเลือกตัวแปร และปัจจุบันมีซอฟต์แวร์ทางสถิติหลายชนิดที่มีฟังก์ชันในการคำนวณตัวประมาณ LASSO เช่น R, SPSS และ SAS เป็นต้น (วิสุธา, 2558: 220) แต่ถึงแม้ว่าวิธี LASSO จะมีลักษณะเด่นตรงที่มีการคัดเลือกตัวแปร แต่ก็ยังมีข้อจำกัดที่ว่า วิธี LASSO สามารถเลือกตัวแปรอิสระเข้าสู่ตัวแบบได้มากที่สุดจำนวน n ตัว ดังนั้นหากข้อมูลมีจำนวนตัวแปรอิสระมากกว่าขนาดตัวอย่างเป็นจำนวนมาก ตัวแบบ LASSO อาจจะไม่มีความเหมาะสม และในกรณีที่ตัวแปรอิสระมีความสัมพันธ์กันสูง วิธี LASSO นี้มีแนวโน้มที่จะเลือกตัวแปรอิสระเพียงตัวแปรเดียวจากกลุ่มของตัวแปรอิสระที่มีความสัมพันธ์กันสูงนี้เข้าสู่ตัวแบบโดยไม่สนใจว่าจะเป็นตัวแปรอิสระตัวใดในกลุ่ม (Zou and Hastie, 2005: 302) จากข้อจำกัดนี้จึงทำให้วิธี LASSO เป็นวิธีที่คัดเลือกตัวแปรอิสระที่ไม่เหมาะสมในบางสถานการณ์

5. การถดถอยเชิงเส้นพหุคูณด้วยวิธีแลซโซปรับได้ (Adaptive LASSO regression)

ซู (Zou, 2006: 1418-1429) ได้กล่าวถึง สมบัติในการคัดเลือกตัวแปรอิสระและการประมาณค่าสัมประสิทธิ์การถดถอยของวิธี LASSO ว่าการคัดเลือกตัวแปรอิสระด้วยวิธี LASSO อาจจะเป็นกระบวนการที่ไม่คงเส้นคงวาได้ เนื่องจากตัวประมาณ LASSO มีการปรับค่าของตัวประมาณเท่ากันทุกตัว ภายใต้ข้อจำกัด L_1 นอร์ม ซึ่งบางทีอาจจะไม่เหมาะสม ดังนั้นจึงเสนอวิธีการประมาณแบบแลซโซปรับได้ (Adaptive LASSO) ซึ่งปรับปรุงข้อจำกัดของตัวประมาณการถดถอยแบบ LASSO โดยถ่วงน้ำหนักของค่า L_1 นอร์ม (Zou, 2006: 1420-1421)

5.1 การประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีแลซโซปรับได้

ตัวประมาณการถดถอยโดยวิธี Adaptive LASSO คือ

$$\text{minimize } (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

$$\text{ภายใต้ข้อจำกัด } \sum_{j=1}^p w_j |\beta_j| \leq t$$

ซึ่งสามารถเขียนอยู่ในรูปสมการที่สอดคล้องกัน คือ

$$\hat{\boldsymbol{\beta}}^{adp.lasso} = \text{argmin}_{\boldsymbol{\beta}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \lambda \sum_{j=1}^p w_j |\beta_j|$$

ซึ่ง $\sum_{j=1}^p w_j |\beta_j|$ คือ ฟังก์ชันพินอลตี และ w_j คือ เวกเตอร์น้ำหนัก

เช่นเดียวกับตัวประมาณ LASSO ในกรณีทั่วไป ตัวประมาณ Adaptive LASSO ไม่มีสูตรที่ชัดเจนในการคำนวณ แต่เราสามารถพิจารณาตัวประมาณ Adaptive LASSO ที่อยู่ในรูปฟังก์ชันของตัวประมาณ OLS ได้ ในกรณีที่เมทริกซ์ $\mathbf{X}'\mathbf{X} = \mathbf{I}_p$ ซึ่งเป็นเมทริกซ์เชิงตั้งฉากปกติ ดังนี้

$$\text{minimize } \sum_{i=1}^n Y_{ij}^2 - 2 \sum_{j=1}^p (\hat{\beta}_j^{ols})' \beta_j + \sum_{j=1}^p \beta_j^2$$

$$\text{ภายใต้ข้อจำกัด } \sum_{j=1}^p w_j |\beta_j| \leq t$$

ซึ่งสามารถเขียนอยู่ในรูปสมการที่สอดคล้องกัน คือ

$$\begin{aligned} L(\mathbf{Y}, \boldsymbol{\beta}, \lambda) &= \sum_{i=1}^n Y_{ij}^2 - 2 \sum_{j=1}^p \hat{\beta}_j^{ols} \beta_j + \sum_{j=1}^p \beta_j^2 + \lambda \left(\sum_{j=1}^p w_j |\beta_j| - t \right) \\ &= \sum_{i=1}^n Y_{ij}^2 + \sum_{j=1}^p [(-2) \hat{\beta}_j^{ols} \beta_j + \beta_j^2 + \lambda w_j |\beta_j|] + \lambda t \end{aligned}$$

หาอนุพันธ์บางส่วนของ $L(Y, \beta, \lambda)$ เทียบกับ β_j

$$\begin{aligned} \frac{\partial L(Y, \beta, \lambda)}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} [\sum_{i=1}^n Y_{ij}^2 + \sum_{j=1}^p [(-2)\hat{\beta}_j^{ols} \beta_j + \beta_j^2 + \lambda w_j |\beta_j|] + \lambda t] \\ &= \frac{\partial}{\partial \beta_j} [\sum_{j=1}^p ((-2)\hat{\beta}_j^{ols} \beta_j + \beta_j^2 + \lambda w_j |\beta_j|)] \\ &= \frac{\partial}{\partial \beta_j} (-2\hat{\beta}_j^{ols} \beta_j + \beta_j^2 + \lambda w_j |\beta_j|) \end{aligned} \quad (39)$$

สมการ (39) จะมีค่าต่ำสุด เมื่อ $\hat{\beta}_j^{ols} \beta_j > 0$ นั่นคือ เครื่องหมายของ $\hat{\beta}_j^{ols}$ และ β_j จะต้องมีความหมายเดียวกัน ซึ่งอาจจะเป็นเครื่องหมายบวกทั้งคู่ หรือลบทั้งคู่ เราจึงพิจารณาเป็นสองกรณีเช่นเดียวกับวิธี LASSO

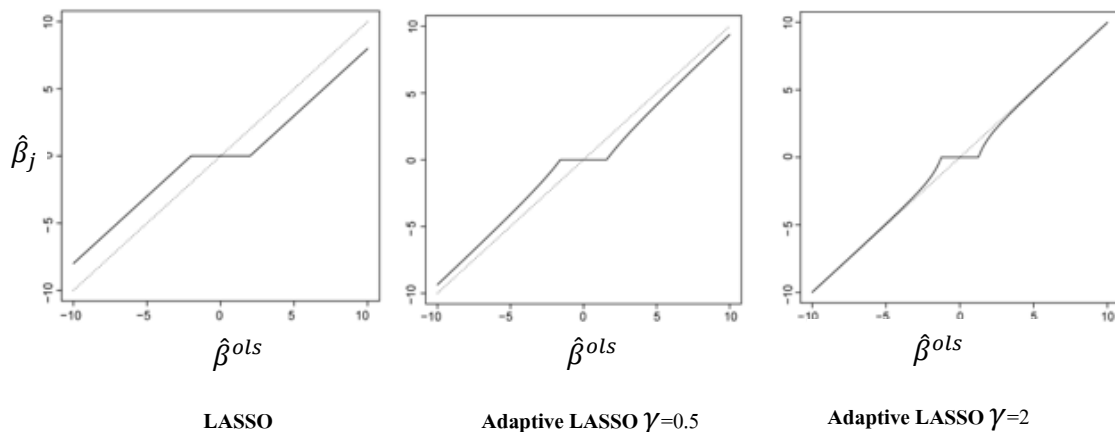
ดังนั้น จะได้ตัวประมาณสัมประสิทธิ์การถดถอยแบบ Adaptive LASSO คือ

$$\hat{\beta}_j^{adp.lasso} = \begin{cases} \hat{\beta}_j^{ols} - \frac{\lambda w_j}{2}; & \hat{\beta}_j^{ols} > \frac{\lambda w_j}{2} \\ 0 & -\frac{\lambda w_j}{2} \leq \hat{\beta}_j^{ols} \leq \frac{\lambda w_j}{2} \\ \hat{\beta}_j^{ols} + \frac{\lambda w_j}{2}; & \hat{\beta}_j^{ols} < -\frac{\lambda w_j}{2} \end{cases} \quad (40)$$

จากตัวประมาณสัมประสิทธิ์การถดถอยแบบ Adaptive LASSO ข้างต้นจะเห็นว่า ตัวประมาณสัมประสิทธิ์บางตัวมีค่าเป็นศูนย์ การประมาณค่าสัมประสิทธิ์โดยวิธี Adaptive LASSO นั้น มีการคัดเลือกตัวแปรเข้าสู่ตัวแบบและประมาณค่าพารามิเตอร์ β ได้ในคราวเดียวกัน และมีการถ่วงน้ำหนักที่สัมประสิทธิ์การถดถอยแต่ละตัว โดย $\hat{\beta}_j^{lasso}$ จะถูกทำให้หดลงมีขนาดเท่ากับ $|\hat{\beta}_j^{ols}| - \frac{\lambda w_j}{2}$ และมีเครื่องหมายเดียวกับ $\hat{\beta}_j^{ols}$ ในกรณีที่ $|\hat{\beta}_j^{ols}|$ มีค่ามากกว่า $\frac{\lambda w_j}{2}$ และ $\hat{\beta}_j^{lasso}$ จะมีค่าเท่ากับศูนย์ในกรณีที่ $|\hat{\beta}_j^{ols}|$ น้อยกว่า $\frac{\lambda w_j}{2}$

เมื่อเปรียบเทียบรูปแบบของสัมประสิทธิ์การถดถอยวิธี LASSO และ Adaptive LASSO

ดังภาพที่ 8



ภาพที่ 8 รูปแบบของการสัมประสิทธิ์การถดถอยโดยวิธีต่าง ๆ

ในกรณี orthonormal design เมื่อ $\lambda=2$ และ $w = \frac{1}{|\hat{\beta}^{ols}|^\gamma}$ (Zou, 2006: 1421)

ภาพที่ 8 แสดงความสัมพันธ์ระหว่างค่าประมาณสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด ($\hat{\beta}^{ols}$) และค่าประมาณสัมประสิทธิ์การถดถอยโดยวิธี LASSO, Adaptive LASSO โดย $\gamma=0.5$ และแบบ Adaptive LASSO โดย $\gamma=2$ ตามลำดับ ในกรณี orthonormal design เมื่อ $\lambda=2$ โดยวิธี LASSO แบบดั้งเดิมนั้น $\hat{\beta}_j = \hat{\beta}_j^{ols} - \lambda$ เมื่อ $\hat{\beta}_j^{ols} > 2$, $\hat{\beta}_j = \hat{\beta}_j^{ols} + \lambda$ เมื่อ $\hat{\beta}_j^{ols} < -2$ และ $\hat{\beta}_j = 0$ เมื่อ $|\hat{\beta}_j^{ols}| < 2$ วิธีต่อมาคือการประมาณค่าพารามิเตอร์ Adaptive LASSO $\hat{\beta}_j = \hat{\beta}_j^{ols} - \lambda w_j = \hat{\beta}_j^{ols} - \frac{\lambda}{|\hat{\beta}_j^{ols}|^\gamma}$ เมื่อ $\hat{\beta}_j^{ols} > 2$ และ $\hat{\beta}_j = \hat{\beta}_j^{ols} + \lambda w_j = \hat{\beta}_j^{ols} + \frac{\lambda}{|\hat{\beta}_j^{ols}|^\gamma}$ เมื่อ $\hat{\beta}_j^{ols} < -2$ และ $\hat{\beta}_j = 0$ เมื่อ $|\hat{\beta}_j^{ols}| \leq 2$ เมื่อ $\gamma=0.5$ และ $\gamma=2$ ตามลำดับ เมื่อ γ มีขนาดใหญ่ จะทำให้ $\lambda w_j = \frac{\lambda}{|\hat{\beta}_j^{ols}|^\gamma}$ มีขนาดเล็ก นั่นคือ $\hat{\beta}_j$ มีค่าใกล้เคียงกับ $\hat{\beta}_j^{ols}$

5.2 ขั้นตอนวิธีในการหาตัวประมาณวิธีแลซโซปรับได้

ตัวประมาณสัมประสิทธิ์การถดถอยแบบ Adaptive LASSO สามารถหาได้จากขั้นตอนวิธี LARS เช่นเดียวกับตัวประมาณ LASSO ซึ่งมีขั้นตอนวิธี ดังนี้

- (1) กำหนดให้ $\mathbf{x}_j^* = \mathbf{x}_j / \hat{w}_j, j = 1, 2, \dots, p$ ซึ่งเป็นการถ่วงน้ำหนักตัวแปรอิสระ \mathbf{x}_j ด้วย \hat{w}_j
- (2) แก้ปัญหาเพื่อหาตัวประมาณ LASSO โดยใช้ขั้นตอนวิธี LARS สำหรับทุกๆ λ , จากสมการถดถอยที่มีตัวแปรอิสระ \mathbf{x}_j^* จากขั้นตอนที่ (1) ซึ่งจะได้ตัวประมาณ

$$\hat{\beta}^{**} = \arg \min_{\beta} \{ \sum (y_i - \sum \beta_j X_{ij})^2 + \lambda \sum |\beta_j| \}$$

(3) คำนวณตัวประมาณ Adaptive LASSO จากตัวประมาณ LASSO $\hat{\beta}^{**}$ ที่ได้ในขั้นตอน (2) ดังนี้

$$\hat{\beta}^{adp.lasso} = \hat{\beta}^{**} / \hat{w}_j, j = 1, 2, \dots, p$$

5.3 สมบัติการพยากรณ์ของตัวประมาณวิธีแลซโซปรับได้

ซู (Zou, 2006: 1420) ได้กล่าวถึงทฤษฎีบทเกี่ยวกับสมบัติการพยากรณ์ของตัวประมาณ $\hat{\beta}^{adp.lasso} = \operatorname{argmin}_{\beta} (Y - X\beta)'(Y - X\beta) + \lambda_n \sum_{j=1}^p w_j |\beta_j|$ ดังนี้

กำหนดให้ $\mathcal{A} = \{j: \beta_j \neq 0\}$ และสมมติให้ขนาดของเซต \mathcal{A} หรือ $|\mathcal{A}| = p_1 < p$ นั่นคือ ตัวแบบการถดถอยเชิงเส้นพหุคูณขึ้นอยู่กับเซตย่อยของตัวแปรอิสระ ซึ่งมีจำนวน p_1 ตัว

ทฤษฎีบท 2 ถ้า $\lambda_n / \sqrt{n} \rightarrow 0$ และ $\lambda_n n^{(Y-1)/2} \rightarrow \infty$ เมื่อ $\gamma > 0$ แล้ว ตัวประมาณ Adaptive LASSO จะมีสมบัติต่อไปนี้

1) ความคงเส้นคงวาของการคัดเลือกตัวแปร:

$$\lim_n P(\mathcal{A}_n = \mathcal{A}) = 1$$

2) การแจกแจงลู่เข้าสู่การแจกแจงปกติ (Asymptotic normality):

$$\sqrt{n}(\hat{\beta}_{\mathcal{A}}^{adp.lasso} - \beta_{\mathcal{A}}) \rightarrow_d N(0, \sigma^2 \cdot C_{11}^{-1})$$

6. การถดถอยเชิงเส้นพหุคูณด้วยวิธีรีแลกซ์ แลซโซ (Relaxed LASSO regression)

นิโกลา (Nicolai, 2006: 374-393) คิดค้นประยุกต์จากการประมาณแบบ LASSO เป็นรีแลกซ์แลซโซ (Relaxed LASSO) ซึ่งเป็นกระบวนการแบบ 2 ขั้นตอน คือกระบวนการแรกจะเป็นการคัดเลือกตัวแปรอิสระเข้าสู่ตัวแบบ จากนั้นนำตัวแปรอิสระที่ได้จากกระบวนการแรกมาทำการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี penalized regression

6.1 การประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีรีแลกซ์แลซโซ

จากตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธี LASSO ในกรณีเชิงตั้งฉากปกติ คือ

$$\text{minimize } (Y - X\beta)'(Y - X\beta)$$

$$\text{ภายใต้ข้อจำกัด } \sum_{j=1}^p |\beta_j| \leq t$$

ซึ่งสามารถเขียนอยู่ในรูปสมการที่สอดคล้องกัน คือ

$$\hat{\beta}^{lasso} = \operatorname{argmin}_{\beta} (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) + \lambda \sum_{j=1}^p |\beta_j|$$

กำหนดให้ M_λ แทน เซตของตัวแปรอิสระที่ถูกเลือกโดยการประมาณแบบ LASSO โดยที่

$$M_\lambda = \{1 \leq k \leq p \mid \hat{\beta}_k^{lasso} \neq 0\}$$

สำหรับ λ ที่มีขนาดใหญ่เพียงพอ (เช่น สำหรับ $\lambda > 2 \max_k n^{-1} \sum_{i=1}^n Y_i X_i^k$) เมื่อ M_λ เป็นเซตว่าง หมายความว่า ไม่มีตัวแปรอิสระใดถูกเลือกเข้าสู่ตัวแบบ ดังนั้น ตัวประมาณสัมประสิทธิ์การถดถอยทุกตัวจะมีค่าเป็นศูนย์ และสามารถคำนวณฟังก์ชันพินอลต์ได้ ถ้าจำนวนตัวแปรอิสระ p น้อยกว่าขนาดตัวอย่าง n และในกรณีที่ $M_0 = \{1, \dots, p\}$ หมายความว่า ตัวแปรอิสระถูกเลือกเข้าสู่ตัวแบบทุกตัว

ฟังก์ชัน L_1 นอร์ม ซึ่งเป็นฟังก์ชันพินอลต์สำหรับวิธี LASSO นั้นมีผลกระทบ 2 อย่าง คือ การคัดเลือกตัวแบบและการหดลงของตัวประมาณของค่าพารามิเตอร์ โดยเมื่อเซตของสัมประสิทธิ์การถดถอยมีค่าเป็นศูนย์ ดังนั้นเซตของตัวแปรอิสระกลุ่มนั้นจะถูกคัดเลือกออกจากตัวแบบ และ M_λ เป็นเซตของตัวแปรอิสระอีกกลุ่มหนึ่งที่ถูกคัดเลือกเข้าสู่ตัวแบบนั้นมีสัมประสิทธิ์การถดถอยเข้าสู่ศูนย์เมื่อเทียบกับตัวประมาณด้วยวิธีกำลังสองน้อยที่สุด โดยผลกระทบทั้งสองนี้มีความสัมพันธ์กัน และแสดงให้เห็นถึงการหดลงของสัมประสิทธิ์การถดถอยในบริบทของเมทริกซ์เชิงตั้งฉากปกติที่มีรูปแบบเดียวกับ soft-thresholding เช่นเดียวกับวิธี LASSO อย่างไรก็ตาม ไม่เป็นที่ชัดเจนในการที่ควบคุมผลกระทบทั้งสองนี้ได้ด้วยพารามิเตอร์เพียงตัวเดียว ซึ่งอาจเป็นที่น่าพอใจในบางสถานการณ์ ในการประมาณสัมประสิทธิ์การถดถอยของตัวแปรอิสระที่ถูกคัดเลือกทั้งหมดเพียงอย่างเดียวโดยไม่ทำการหดลงของค่าประมาณ ซึ่งสอดคล้องกับ hard-thresholding

หลักการทั่วไปของ soft-thresholding และ hard-thresholding เป็นการควบคุมการคัดเลือกตัวแบบและการหดลงของตัวประมาณของค่าพารามิเตอร์โดยพารามิเตอร์ 2 ตัว คือ λ และ ϕ กับตัวประมาณสัมประสิทธิ์ถดถอยโดยวิธี Relaxed LASSO

ดังนั้น ตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธี Relaxed LASSO คือ

$$\hat{\beta}^{rex.lasso} = \operatorname{argmin}_{\beta} \sum_{i=1}^n (Y_i - X_i' \{\beta \cdot \mathbf{1}_{M_\lambda}\})^2 + \phi \lambda \sum_{j=1}^p |\beta_j|$$

เมื่อ $\lambda \in [0, \infty)$ และ $\phi \in (0, 1]$ และ $\beta \cdot \mathbf{1}_{M_\lambda}$ คือ ฟังก์ชันบ่งชี้ (indicator function) สำหรับเซตของตัวแปรอิสระ $M_\lambda \subseteq \{1, \dots, p\}$ ดังนั้น สำหรับทุก ๆ $k \in \{1, \dots, p\}$

$$\{\beta \cdot \mathbf{1}_{M_\lambda}\}_k = \begin{cases} 0 & \text{เมื่อ } k \notin M_\lambda \\ \beta_k & \text{เมื่อ } k \in M_\lambda \end{cases}$$

ตัวแปรอิสระในเซต M_λ นี้เท่านั้น ที่ถูกนำไปพิจารณาในการวิเคราะห์การถดถอยแบบ Relaxed LASSO

พารามิเตอร์ λ ควบคุมการคัดเลือกตัวแปรอิสระเข้าสู่ตัวแบบเช่นเดียวกับวิธี LASSO ส่วน ϕ จะเป็นตัวควบคุมการหดลงของตัวประมาณของค่าพารามิเตอร์ โดยถ้า $\phi = 1$ หมายความว่า ตัวประมาณจากวิธี LASSO และวิธี Relaxed LASSO เป็นตัวประมาณเดียวกัน ถ้า $\phi < 1$ หมายความว่า มีการประมาณค่าพารามิเตอร์โดยสัมประสิทธิ์การถดถอยที่ได้จากวิธีนี้จะมีค่าต่ำลงเมื่อเทียบกับจากวิธีกำลังสองน้อยที่สุด และถ้า $\phi \rightarrow 0$ หมายความว่า สัมประสิทธิ์ทุก ๆ ตัว ในเซตของตัวแปรอิสระที่ถูกเลือกเข้าสู่ตัวแบบ M_λ นั้น ถูกประมาณได้ด้วยวิธีกำลังสองน้อยที่สุด

ตัวประมาณพารามิเตอร์การถดถอยโดยวิธี Relaxed LASSO คือ

$$\begin{aligned} L(Y, \beta, \lambda) &= \sum_{i=1}^n Y_{ij}^2 - 2 \sum_{j=1}^p \hat{\beta}_j^{ols} \beta_j + \sum_{j=1}^p \beta_j^2 + \phi \lambda (\sum_{j=1}^p |\beta_j| - t) \\ &= \sum_{i=1}^n Y_{ij}^2 + \sum_{j=1}^p [(-2) \hat{\beta}_j^{ols} \beta_j + \beta_j^2 + \phi \lambda |\beta_j|] + \lambda t \end{aligned}$$

หาอนุพันธ์บางส่วนของ $L(Y, \beta, \lambda)$ เทียบกับ β_j

$$\begin{aligned} \frac{\partial L(Y, \beta, \lambda)}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} [\sum_{i=1}^n Y_{ij}^2 + \sum_{j=1}^p [(-2) \hat{\beta}_j^{ols} \beta_j + \beta_j^2 + \phi \lambda |\beta_j|] + \lambda t] \\ &= \frac{\partial}{\partial \beta_j} [\sum_{j=1}^p ((-2) \hat{\beta}_j^{ols} \beta_j + \beta_j^2 + \phi \lambda |\beta_j|)] \\ &= \frac{\partial}{\partial \beta_j} (-2 \hat{\beta}_j^{ols} \beta_j + \beta_j^2 + \phi \lambda |\beta_j|) \end{aligned} \quad (41)$$

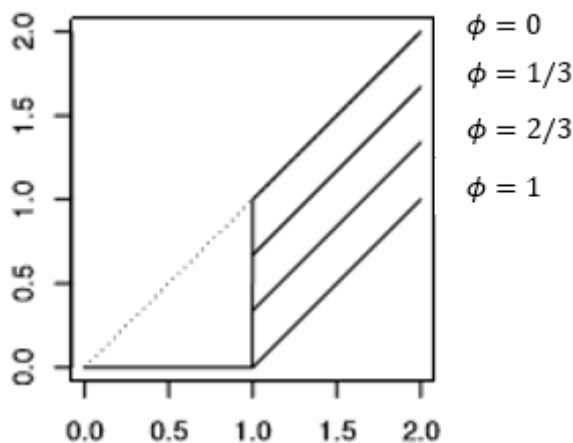
สมการ (41) จะมีค่าต่ำสุด เมื่อ $\hat{\beta}_j^{ols} \beta_j > 0$ นั่นคือ เครื่องหมายของ $\hat{\beta}_j^{ols}$ และ β_j จะต้องเหมือนกัน ซึ่งอาจจะเป็นเครื่องหมายบวกทั้งคู่ หรือลบทั้งคู่ เราจึงพิจารณาเป็นสองกรณีเช่นเดียวกับวิธี LASSO

จะได้ ตัวประมาณสัมประสิทธิ์การถดถอยแบบ Relaxed LASSO สำหรับทุก ๆ $j = 1, \dots, p$ ดังนี้

$$\hat{\beta}_j^{rex.lasso} = \begin{cases} \hat{\beta}_j^{ols} - \phi \lambda; & \hat{\beta}_j^{ols} > \lambda \\ 0 & ; -\lambda \leq \hat{\beta}_j^{ols} \leq \lambda \\ \hat{\beta}_j^{ols} + \phi \lambda; & \hat{\beta}_j^{ols} < -\lambda \end{cases} \quad (42)$$

จากข้างต้นจะเห็นว่า ตัวประมาณสัมประสิทธิ์บางตัวมีค่าเป็น 0 ดังนั้นจึงเป็นเหตุผลการประมาณค่าสัมประสิทธิ์โดยวิธี Relaxed LASSO นั้นมีการคัดเลือกตัวแปรเข้าสู่ตัวแบบและประมาณค่าพารามิเตอร์ β ได้ในคราวเดียวกัน ซึ่งจะสอดคล้องกับข้างต้นที่ว่า เมื่อ $\phi = 1$ นั้น $\hat{\beta}_j^{rex.lasso}$ จะมีค่าเท่ากับ $\hat{\beta}_j^{lasso}$ ถ้า $\phi \rightarrow 0$ นั่นคือ $\hat{\beta}_j^{rex.lasso}$ จะมีค่าใกล้เคียงกับ $\hat{\beta}_j^{ols}$ เนื่องจากการบวกเข้าหรือลบออกด้วยเทอมที่น้อยมาก ๆ ดังนั้น สัมประสิทธิ์ทุก ๆ ตัว ในเซตของ

ตัวแปรอิสระที่ถูกเลือกเข้าสู่ตัวแบบ M_λ นั้น ถูกประมาณได้ด้วยวิธีกำลังสองน้อยที่สุด และถ้า $\phi < 1$ หมายความว่า การประมาณค่าพารามิเตอร์โดยสัมประสิทธิ์การถดถอย $\hat{\beta}_j^{rex,lasso}$ จะมีค่าน้อยกว่า $\hat{\beta}_j^{ols}$



ภาพที่ 9 รูปแบบของการสัมประสิทธิ์การถดถอยโดยวิธี Relaxed LASSO

ในกรณี orthonormal design เมื่อ $\lambda=1$ และ $\phi = 0, 1/3, 2/3$ และ 1 (Nicolai, 2006: 377)

จากภาพที่ 9 เปรียบเทียบความสัมพันธ์ระหว่างค่าประมาณสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด ($\hat{\beta}^{ols}$) และค่าประมาณสัมประสิทธิ์การถดถอยโดยวิธี Relaxed LASSO ในกรณี orthonormal design เมื่อ $\lambda=1$ และ $\phi = 0, 1/3, 2/3$ และ 1 ตามลำดับ โดยแกนนอนแสดงค่าประมาณสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด ($\hat{\beta}^{ols}$) และแกนตั้งแสดงค่าประมาณสัมประสิทธิ์การถดถอยโดยวิธี Relaxed LASSO จะเห็นว่า $\hat{\beta}_j = 0$ เมื่อ $|\hat{\beta}_j^{ols}| \leq 1$ เมื่อพิจารณา $\phi = 0$ ทำให้ $\hat{\beta}_j = \hat{\beta}_j^{ols}$ และเมื่อ $\phi = 1$ ทำให้ $\hat{\beta}_j = \hat{\beta}_j^{lasso}$ และเมื่อ $\phi = 1/3$ และ $2/3$ ซึ่ง $\phi \in (0,1]$ จะมีการประมาณค่าพารามิเตอร์โดยสัมประสิทธิ์การถดถอยที่ได้จากวิธีนี้จะมีค่าต่ำลงเมื่อเทียบกับจากวิธีกำลังสองน้อยที่สุด

6.2 ขั้นตอนวิธีในการหาตัวประมาณวิธีรีแลกซ์แลซโซ

จาก ตัวประมาณสัมประสิทธิ์การถดถอยแบบ Relaxed LASSO คือ

$$\hat{\beta}^{rex,lasso} = \operatorname{argmin}_{\beta} \sum_{i=1}^n (Y_i - X_i' \{\beta \cdot \mathbf{1}_{M_\lambda}\})^2 + \phi \lambda \sum_{j=1}^p |\beta_j|$$

การคำนวณหาตัวประมาณสัมประสิทธิ์การถดถอยเหมือนกับการวิธี LASSO โดยมีขั้นตอนวิธีอย่างง่าย (Simple Algorithm) (Nicolai, 2006: 377-378) ดังนี้

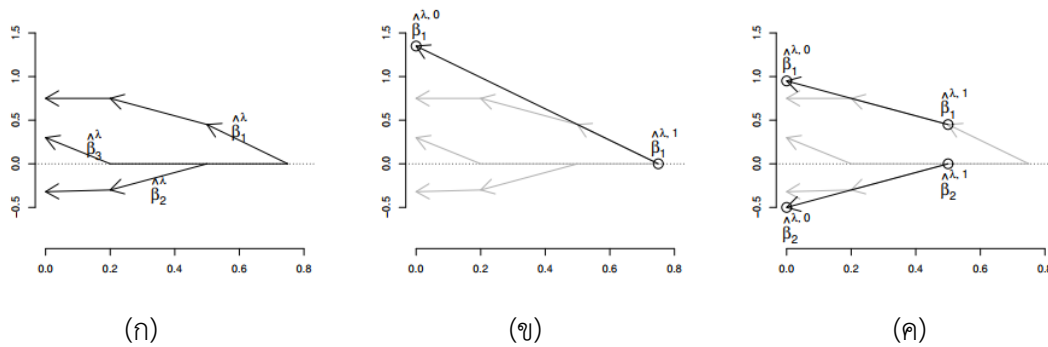
(1) คำนวณหาตัวประมาณสัมประสิทธิ์การถดถอยวิธี LASSO โดยกำหนดให้ M_1, \dots, M_m เป็นเซตของตัวแปรอิสระที่ถูกคัดเลือกเข้าสู่ตัวแบบ จำนวน m ตัวแบบ และให้ $\lambda_1 > \dots > \lambda_m = 0$ เป็นลำดับของค่าพินอลตีพารามิเตอร์ ดังนั้น $M_\lambda = M_k$ ก็ต่อเมื่อ $\lambda \in (\lambda_k, \lambda_{k-1}]$ เมื่อ $\lambda_0 := \infty$

(2) สำหรับ $k = 1, \dots, m$ คำนวณหาตัวประมาณสัมประสิทธิ์การถดถอยวิธี LASSO บนเซตของตัวแปรอิสระ M_k โดยพินอลตีพารามิเตอร์เปลี่ยนแปลงระหว่างศูนย์ถึง λ_k ซึ่งผลลัพธ์ที่ได้จะเป็นเซตของตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธี Relaxed LASSO ($\beta^{\lambda, \phi}$) สำหรับ $\lambda \in \Lambda_k$ ซึ่งผลลัพธ์ของตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธี Relaxed LASSO สำหรับทุกๆ พินอลตีพารามิเตอร์การรวมกันของเซตเหล่านี้

ขั้นตอนวิธีดังกล่าวเป็นขั้นตอนวิธีของ Relaxed LASSO สำหรับทุกๆ ค่าของพินอลตีพารามิเตอร์ $\phi \in [0, 1]$ และ $\lambda > 0$ ซึ่งความซับซ้อนในการคำนวณของขั้นตอนวิธีนี้ โดยมีการคำนวณซ้ำในขั้นตอนที่ (2) ของขั้นตอนวิธีอย่างง่าย

อย่างไรก็ตาม ขั้นตอนวิธีนี้ไม่เหมาะสมในกรณีทั่วไป การคำนวณของวิธี LASSO นั้นมีข้อมูลที่สามารถใช้ประโยชน์ได้ในขั้นตอนที่ 2 เมื่อหาผลลัพธ์ของวิธี LASSO สำหรับทุกๆ $M_k, k = 1, \dots, m$ ตัวอย่างเช่น ในภาพที่ 10 แสดงทิศทาง (direction) ของค่าประมาณวิธี Relaxed LASSO มีทิศทางเดียวกับค่าประมาณวิธี LASSO ซึ่งการที่มีทิศทางไปในทางเดียวกันนี้ทำให้ไม่ต้องคำนวณใหม่อีกครั้ง การประมาณค่าวิธี Relaxed LASSO ด้วยวิธี LASSO สามารถทำได้นั้นก็ต่อเมื่อ ค่าประมาณของวิธี LASSO ไม่มีค่าเป็นศูนย์ ในตัวอย่างนี้ตัวประมาณวิธี LASSO นั้นเพิ่มขึ้นทางเดียวเมื่อ พารามิเตอร์พินอลตี λ มีค่าลดลง แต่อย่างไรก็ตาม ถ้าค่าประมาณของวิธี LASSO สำหรับเซตของ M_k มีค่าเป็นศูนย์ แล้วต้องคำนวณค่าประมาณของวิธี LASSO อีกครั้งเพื่อให้แน่นชัดสำหรับเซตนี้

ความสัมพันธ์ระหว่างพารามิเตอร์พินอลตี λ และตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธี LASSO และ Relaxed LASSO แสดงในภาพที่ 10



ภาพที่ 10 ความสัมพันธ์ระหว่างพารามิเตอร์พินอลตี้ λ และตัวประมาณ LASSO และ Relaxed LASSO สำหรับข้อมูลที่มีตัวแปรอิสระ 3 ตัว (Nicolai, 2006: 378)

จากภาพที่ 10 (ก) แสดงเส้นทางของตัวประมาณ LASSO และ Relaxed LASSO สำหรับข้อมูลที่มีตัวแปรอิสระ 3 ตัว เมื่อค่าของ λ มีขนาดใหญ่ ตัวประมาณทุกตัวมีค่าเป็นศูนย์ ในช่วง $\lambda \in (0.45, 0.75]$ มีเพียงตัวแปรอิสระตัวแรกเท่านั้นที่มีค่าประมาณไม่เท่ากับศูนย์ ภาพที่ 10 (ข) แสดงค่าประมาณวิธี Relaxed LASSO ถ้า ในช่วง $\lambda \in (0.45, 0.75]$ ตัวประมาณวิธี Relaxed LASSO มีทิศทางเดียวกับตัวประมาณวิธี LASSO ตัวประมาณวิธี Relaxed LASSO สำหรับ $\emptyset = 0$ นั้นมีค่าสอดคล้องกับตัวประมาณโดยวิธีกำลังสองน้อยที่สุด ในทำนองเดียวกัน ภาพที่ 10 (ค) แสดงค่าประมาณวิธี Relaxed LASSO ถ้า ในช่วง $\lambda \in (0.2, 0.45]$ ตัวประมาณวิธี Relaxed LASSO มีทิศทางเดียวกับตัวประมาณวิธี LASSO ตัวประมาณวิธี Relaxed LASSO สำหรับ $\emptyset = 0$ นั้นมีค่าสอดคล้องกับตัวประมาณโดยวิธีกำลังสองน้อยที่สุดสำหรับตัวแปรอิสระ 2 ตัวที่ถูกเลือกโดยวิธี LASSO

จากเหตุผลข้างต้นจึงทำให้ นิโคลาส (Nicolai, 2006: 378) เสนอขั้นตอนวิธีในการหาค่าประมาณวิธี LASSO ให้มีการคำนวณให้น้อยลง ดังนี้

(1) คำนวณหาตัวประมาณสัมประสิทธิ์การถดถอยวิธี LASSO เช่นเดียวกับขั้นตอนที่ 1 ในขั้นตอนวิธีอย่างง่าย

(2) สำหรับทุกๆ $k = 1, \dots, m$ กำหนดให้ $\delta(k) = (\beta^{\lambda_k} - \beta^{\lambda_{k-1}})/(\lambda_{k-1} - \lambda_k)$ และ $\tilde{\beta} = \beta^{\lambda_k} + \lambda_k \delta(k)$ ถ้ามี l อย่างน้อย 1 ตัวที่ทำให้เครื่องหมายของ $\tilde{\beta}_l$ และ $\beta_l^{\lambda_k}$ ไม่เหมือนกัน นั่นคือ $sign(\tilde{\beta}_l) \neq sign(\beta_l^{\lambda_k})$ แล้วการหาตัวประมาณสัมประสิทธิ์การถดถอยวิธี Relaxed LASSO สำหรับ นั้นคำนวณจากขั้นตอนที่ 2 ในขั้นตอนวิธีอย่างง่าย ถ้าไม่สอดคล้องกับเงื่อนไขดังกล่าวนี้ ตัวประมาณโดยวิธี Relaxed LASSO สำหรับ $\lambda \in \Lambda_k$ และ $\phi \in [0, 1]$ คำนวณโดยการประมาณค่าในช่วงเชิงเส้น (linear interpolation) ระหว่าง $\beta^{\lambda_{k-1}}$ (ซึ่งสอดคล้องกับ $\emptyset = 1$) และ $\tilde{\beta}$ (ซึ่งสอดคล้องกับ $\emptyset = 0$)

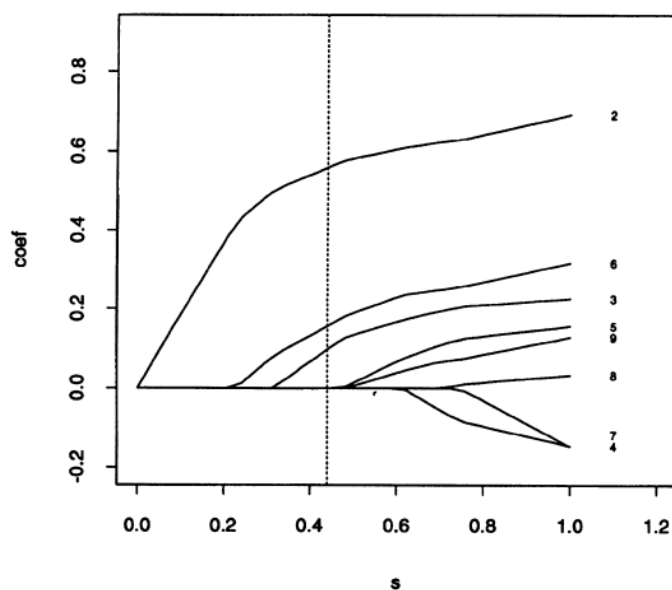
7. การหาพารามิเตอร์การปรับแต่ง (Tuning parameter) โดยวิธี Cross-validation

วิธี Cross validation เป็นวิธีที่นิยมในการทำวิจัยเพื่อใช้ในการวัดประสิทธิภาพของตัวแบบโดยใช้ค่าความถูกต้อง เนื่องจากผลที่ได้มีความน่าเชื่อถือ โดยจะมีการแบ่งข้อมูลออกเป็น 2 ส่วน คือ ชุดข้อมูลฝึกฝน (Training set) และชุดข้อมูลทดสอบ (Test set) (Zhang, 2011: 40-43) โดย cross-validation จะกระทำกับชุดข้อมูลฝึกฝน ซึ่งมีหลายวิธี เช่น generalized cross-validation, v-fold Cross validation และ Least-one-out cross validation เป็นต้น

งานวิจัยนี้ใช้วิธี v-fold Cross validation ในการหาค่าประมาณของพารามิเตอร์ปรับแต่ง ซึ่งมีรายละเอียดดังนี้

วิธี v-fold Cross validation จะนำชุดข้อมูลฝึกฝน มาแบ่งข้อมูลออกเป็น v ส่วนเท่า ๆ กัน ในการทดลองครั้งแรกจะให้ข้อมูลชุดที่ 1 เป็นชุดข้อมูลทดสอบย่อย (sub-test set) และชุดข้อมูล V-1 ที่เหลือเป็นชุดข้อมูลฝึกฝนย่อย (sub-training set) จากนั้นประมาณค่าพารามิเตอร์ที่สนใจจากชุดข้อมูลฝึกฝนย่อยแล้วนำค่าที่ได้ไปพยากรณ์ในชุดข้อมูลทดสอบย่อย แล้วคำนวณหาค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Prediction mean square error : PMSE) ของข้อมูลชุดแรก $PMSE_1$ โดย $PMSE_v = \sum_{i=1}^{n_t} (Y_i - \hat{Y}_i)^2 / n_t$ เมื่อ n_t คือ ขนาดตัวอย่างของชุดข้อมูลฝึกฝนย่อย ในการทดลองครั้งที่สองจะให้ข้อมูลชุดที่ 2 เป็นชุดข้อมูลทดสอบย่อยและชุดข้อมูล V-1 ที่เหลือเป็นชุดข้อมูลฝึกฝนย่อย จากนั้นประมาณค่าพารามิเตอร์ที่สนใจจากชุดข้อมูลฝึกฝนย่อย แล้วนำค่าที่ได้ไปพยากรณ์ในชุดข้อมูลทดสอบย่อย แล้วคำนวณหาค่า $PMSE_2$ ทำแบบนี้จนกระทั่งข้อมูลทั้ง V ชุดถูกนำมาเป็นชุดข้อมูลทดสอบย่อย แล้วหาค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (APMSE) โดยพารามิเตอร์ปรับแต่ง คือ พารามิเตอร์ ณ ตำแหน่งที่ทำให้ค่า APMSE ต่ำที่สุด ข้อดีของการเลือกสุ่มข้อมูลแบบ V-fold Cross validation คือ ข้อมูลทุกตัวจะถูกนำมาเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบ แต่ข้อจำกัดคือ ใช้เวลามากเนื่องจากต้องทดลองข้อมูลทั้งหมด V ครั้ง โดยส่วนใหญ่นิยมกำหนดค่า V เป็น 5 หรือ 10

ตัวอย่างการวิเคราะห์ถดถอยพหุคูณด้วยวิธี LASSO โดยใช้ข้อมูลมะเร็งต่อมลูกหมาก และการหาพารามิเตอร์การปรับแต่งโดยใช้วิธีการ generalized cross-validation (Tibshirani, 1996: 273-274) แสดงในภาพที่ 11



ภาพที่ 11 แสดงการปรับค่าสัมประสิทธิ์การถดถอยวิธี LASSO ในตัวอย่างข้อมูลมะเร็งต่อมลูกหมาก (Stamey et al., 1989) โดยแต่ละเส้นแสดงความสัมพันธ์ระหว่างสัมประสิทธิ์การถดถอยและพารามิเตอร์ของวิธี LASSO คือ $s = t / \sum |\beta_j^{OLS}|$ เส้นประแสดงตัวแบบสำหรับ $\hat{s} = 0.44$ ซึ่งถูกเลือกโดยวิธี generalized cross-validation (Tibshirani, 1996: 273)

ภาพที่ 11 แสดงตัวประมาณแบบ LASSO กับฟังก์ชันของค่า t โดย $s = t / \sum |\beta_j^{OLS}|$ จะเห็นว่า ค่าสัมบูรณ์ของสัมประสิทธิ์การถดถอยแต่ละตัวเข้าสู่ศูนย์เมื่อ s เข้าใกล้ศูนย์ ในตัวอย่างนี้ เส้นของสัมประสิทธิ์การถดถอยลดลงทางเดียวไปสู่ศูนย์ แต่ก็ไม่จำเป็นที่จะเกิดขึ้นเสมอไปในกรณีต่างๆ ไป เส้นประแสดงสมการถดถอยสำหรับ $\hat{s} = 0.44$ ซึ่งถูกเลือกโดยวิธี generalized cross-validation จะมีเพียงตัวแปรอิสระที่อยู่ครึ่งแรกของเส้นประเท่านั้นที่ถูกคัดเลือกเข้าสู่สมการถดถอย ซึ่งในตัวอย่างนี้คือตัวแปรอิสระตัวที่ 2, 6 และ 3 เท่านั้น (Tibshirani, 1996: 274)

งานวิจัยที่เกี่ยวข้อง

ในปี 1970 Hoerl และ Kennard (Hoerl and Kennard, 1970: 55-67) ได้พัฒนาและนำเสนอคุณสมบัติทางสถิติของวิธีการประมาณสัมประสิทธิ์การถดถอยเรียกว่า การถดถอยแบบบริดจ์ ได้กล่าวถึงความไม่เพียงพอของการวิเคราะห์การถดถอยด้วยวิธีกำลังสองน้อยที่สุดสำหรับปัญหาในลักษณะที่ไม่เป็นเชิงตั้งฉาก (non-orthogonal) ได้กล่าวถึงการประมาณค่าสัมประสิทธิ์ถดถอยด้วยวิธีกำลังสองน้อยที่สุดเผชิญกับข้อบกพร่องของเทคนิคการหาค่าเหมาะสมที่สุดเชิงคณิตศาสตร์ ซึ่งให้

ผลลัพธ์เป็นค่าของตัวประมาณแบบจุด วิธีการหาตัวประมาณโดยวิธีกำลังสองน้อยที่สุดไม่ได้แสดงภาพให้เห็นค่าความไว ของตัวประมาณต่อเกณฑ์การหาค่าเหมาะสมที่สุด และได้เสนอวิธีการวิเคราะห์แบบบริดจ์ ซึ่งเป็นวิธีที่แสดงภาพเชิงกราฟิกของลักษณะพื้นที่ผิวตอบสนองของผลบวกกำลังสองของความคลาดเคลื่อนสุ่ม และหาตำแหน่งของบริเวณที่มีค่าต่ำสุดโดยรวม (global minimum) และค่าต่ำสุดเฉพาะที่ (local minimum) ตัวประมาณสัมประสิทธิ์การถดถอยที่ได้เป็นตัวประมาณที่มีความเอนเอียง (Bias estimator) แต่ความแปรปรวนและค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของตัวประมาณน้อยกว่าตัวประมาณด้วยวิธีกำลังสองน้อยที่สุด

ในปี 1996 Tibshirani (Tibshirani, 1996: 267-288) ได้เสนอวิธีการประมาณสัมประสิทธิ์การถดถอยที่ชื่อว่า LASSO ซึ่งเป็นวิธีการประมาณสัมประสิทธิ์การถดถอยที่นำคุณลักษณะที่ดีของวิธีริดจ์และวิธี subset selection มารวมกัน และได้ทำการเปรียบเทียบวิธีการประมาณสัมประสิทธิ์การถดถอยแบบ LASSO เทียบกับวิธีอื่น ๆ คือ วิธีกำลังสองน้อยที่สุด, non-negative garotte, best subset selection และวิธีริดจ์ พบว่า เมื่อตัวแปรอิสระที่มีอิทธิพลต่อตัวแปรตามสูงมีจำนวนน้อย วิธี best subset selection เหมาะสมที่สุด ส่วนวิธี LASSO และ วิธีริดจ์ไม่ค่อยเหมาะสม แต่เมื่อตัวแปรอิสระที่มีอิทธิพลต่อตัวแปรตามปานกลางมีจำนวนน้อยถึงปานกลางนั้น วิธี LASSO เหมาะสมที่สุด รองลงมาคือวิธีริดจ์และวิธี best subset selection และเมื่อจำนวนตัวแปรอิสระที่มีอิทธิพลต่อตัวแปรตามน้อยมีจำนวนมากนั้น วิธีริดจ์ดีที่สุด รองลงมาคือ LASSO และ best subset selection ตามลำดับ

ในปี 2006 Zou (Zou, 2006: 1418-1429) ได้ศึกษาคุณสมบัติในการคัดเลือกตัวแปรอิสระและการประมาณค่าสัมประสิทธิ์การถดถอยของวิธี LASSO ว่าโดยทั่วไปแล้วการคัดเลือกตัวแปรอิสระโดยวิธี LASSO ไม่คงเส้นคงวา และได้เสนอเงื่อนไขบางเงื่อนไขที่ทำให้ตัวประมาณวิธี LASSO มีความคงเส้นคงวา และได้เสนอตัวประมาณที่ชื่อว่า Adaptive LASSO ซึ่งเป็นตัวประมาณที่สอดคล้องกับคุณสมบัติของ Oracle ทำให้ตัวประมาณที่ได้มีความคงเส้นคงวา และได้ทำการเปรียบเทียบประสิทธิภาพของวิธีการประมาณแบบ Adaptive LASSO เทียบกับวิธีอื่นๆ คือ LASSO, SCAD และ non-negative garotte โดยกำหนดขนาดตัวอย่างเท่ากับ 20,40,60,80 และจำนวนตัวแปรอิสระเท่ากับ 8 และสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระคือ $\rho^{|k-j|}$ เมื่อ $\rho = 0.5$ ภายใต้สถานการณ์ที่สัมประสิทธิ์การถดถอยขนาดใหญ่มีจำนวนน้อย และสัมประสิทธิ์การถดถอยขนาดเล็กมีจำนวนมาก ซึ่งผลจากการศึกษาการจำลองข้อมูลพบว่า โดยภาพรวมแล้ว วิธี Adaptive LASSO มีประสิทธิภาพมากที่สุดในบรรดาวิธีที่เปรียบเทียบทั้ง 4 ที่กล่าวข้างต้น

ในปี 2006 Nicolai (Nicolai, 2006: 374-393) ได้เสนอวิธีการประมาณสัมประสิทธิ์การถดถอยที่ชื่อว่า Relaxed LASSO ซึ่งเป็นวิธีการประมาณสัมประสิทธิ์การถดถอยที่ควบคุมผลกระทบ 2 อย่าง คือ การคัดเลือกตัวแปรอิสระ และการหดลงของค่าประมาณสัมประสิทธิ์การถดถอยด้วย

พารามิเตอร์ 2 ตัว และเสนอขั้นตอนวิธีในการหาตัวประมาณวิธี Relaxed LASSO นอกจากนี้ได้ทำการเปรียบเทียบวิธี Relaxed LASSO กับวิธี LASSO โดยกำหนดขนาดตัวอย่างเท่ากับ 50, 100, 200 จำนวนตัวแปรอิสระเท่ากับ 50, 100, 200, 400, 800 กำหนดจำนวนตัวแปรอิสระที่อยู่ในตัวแบบจริง $p_1 = 5, 15, 25, 50$ ตัว สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระคือ $\rho^{|k-j|}$ เมื่อ $\rho = 0, 0.3$ พบว่า วิธี Relaxed LASSO มีจำนวนตัวแปรอิสระที่ถูกคัดเลือกน้อยกว่าวิธี LASSO และเมื่อตัวแปรอิสระที่เกี่ยวข้อง p_1 มีขนาดเล็ก คือ $p_1 = 5, 15$ วิธี Relaxed LASSO ให้ผลดีกว่าวิธี LASSO ในทอมของประสิทธิภาพการพยากรณ์

ในปี 2008 Huang, Ma และ Zhang (Huang, Ma and Zhang, 2008: 1603-1618) ได้ทำการเปรียบเทียบประสิทธิภาพของวิธีการประมาณแบบ LASSO และ Adaptive LASSO กรณีเมทริกซ์ของตัวแปรอิสระ เป็นเมทริกซ์เชิงตั้งฉาก (orthogonal matrix) ไม่เป็นเมทริกซ์เชิงตั้งฉาก (non-orthogonal matrix) และแบบแบ่งกลุ่ม (grouping structure) โดยกำหนดขนาดตัวอย่างเท่ากับ 100 จำนวนตัวแปรอิสระเท่ากับ 200 และ 400 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.5 มีสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระ คือ $\rho^{|k-j|}$ เมื่อ $\rho = 0.5, 0.95$ กรณีเป็นตัวแบบถดถอยเบาบาง (Sparse model) และกำหนดค่าสัมประสิทธิ์การถดถอย $\beta_1, \dots, \beta_5 = 2.5, \beta_6, \dots, \beta_{10} = 1.5, \beta_{11}, \dots, \beta_{15} = 0.5$ และ $\beta_{16}, \dots, \beta_p = 0$ เพื่อกำหนดค่าพารามิเตอร์ให้สอดคล้องกับลักษณะของตัวแบบการถดถอยเบาบางเบา (sparse model) พบว่า วิธี Adaptive LASSO มีประสิทธิภาพในการพยากรณ์สูงกว่าวิธี LASSO และมีจำนวนตัวแปรอิสระเข้าสู่ตัวแบบน้อยกว่าวิธี LASSO แต่อย่างไรก็ตาม วิธี Adaptive LASSO มีจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยมากกว่าที่กำหนด

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยนี้เป็นการวิจัยเชิงทดลองโดยวิธีการจำลอง (Simulation study) ซึ่งมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณแบบ penalized regression 4 วิธี คือ วิธีริคต์, LASSO, Adaptive LASSO และ Relaxed LASSO ซึ่งในการเปรียบเทียบประสิทธิภาพนั้นจะพิจารณาจากความถูกต้องในการพยากรณ์ ความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอย และความถูกต้องในการคัดเลือกตัวแปรอิสระเข้าในตัวแบบ โดยการวิจัยครั้งนี้ได้จำลอง (Simulation) ข้อมูลด้วยโปรแกรม R เพื่อสร้างข้อมูลของตัวแปรอิสระและตัวแปรตาม และกำหนดสถานการณ์ต่าง ๆ ตามขอบเขตของงานวิจัย โดยในแต่ละสถานการณ์ทำซ้ำ 500 ครั้ง ซึ่งมีขอบเขตของงานวิจัยและขั้นตอนในการวิจัย ดังนี้

ขอบเขตของการวิจัยและการทดลอง

ผู้วิจัยจำลองข้อมูลที่ใช้ในการทดลองภายใต้สถานการณ์ต่าง ๆ ดังนี้

1. ขนาดตัวอย่าง $n = 50, 100$ และ 200
2. จำนวนตัวแปรอิสระ $p = 100, 200, 400$ และ 800
3. เมทริกซ์ของตัวแปรอิสระ (X) ขนาด $n \times p$ มีการแจกแจงปกติหลายตัวแปร ที่มีเวกเตอร์ค่าเฉลี่ยเท่ากับเวกเตอร์ศูนย์ และเมทริกซ์ความแปรปรวนร่วม $\Sigma_{p \times p}$
4. ความคลาดเคลื่อนสุ่ม ε มีการแจกแจงปกติ ที่มีค่าเฉลี่ยเป็นศูนย์ และความแปรปรวนคงที่ โดยกำหนดความแปรปรวน $\sigma^2 = 1.5^2$
5. เวกเตอร์ตัวแปรตาม (Y) สร้างจากตัวแบบการถดถอยเชิงเส้นพหุคูณ $Y = X\beta + \varepsilon$
6. กำหนดตัวแบบการถดถอยพหุคูณของการทดลองสำหรับแต่ละค่า n และ p โดยกำหนดรูปแบบความสัมพันธ์ของตัวแปรอิสระ X และค่าพารามิเตอร์ β ในแต่ละกรณี ดังนี้

กรณีที่ 1 ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระเป็นเชิงตั้งฉากบางส่วน (partial orthogonal sparse model)

ความสัมพันธ์ของตัวแปรอิสระ X เป็นแบบเมทริกซ์เชิงตั้งฉากบางส่วน (partial orthogonal matrix) คือ ตัวแปรอิสระที่สัมพันธ์กันแบ่งออกเป็น 2 กลุ่มที่อิสระต่อกัน โดยกลุ่มแรกคือตัวแปรอิสระจำนวน 15 ตัวที่สัมพันธ์กัน $(x_{i1}, \dots, x_{i(15)})$ และกลุ่มที่สองคือตัวแปรอิสระตัวที่

เหลือจำนวน $p - 15$ ตัวที่สัมพันธ์กัน $(x_{i(16)}, \dots, x_{i(p)})$ โดยตัวแปรอิสระกลุ่มที่ 1 และกลุ่มที่ 2 เป็นอิสระกัน ซึ่งค่าสหสัมพันธ์ (pairwise correlation) ระหว่างตัวแปรอิสระตัวที่ k และ j ของ $(x_{i1}, \dots, x_{i(15)})$ คือ $\rho^{|k-j|}$ เมื่อ $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$, $k, j = 1, \dots, 15$ และค่าสหสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ k และ j ของ $(x_{i(16)}, \dots, x_{i(p)})$ คือ $\rho^{|k-j|}$ เมื่อ $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$, และ $k, j = 16, \dots, p$ ซึ่ง $x_{i1}, \dots, x_{ip} \sim N_p(\mathbf{0}, \Sigma)$ โดยที่ $x_{i1}, \dots, x_{i(15)} \sim N_{15}(\mathbf{0}, \Sigma_1)$ และ $x_{i(16)}, \dots, x_{i(p)} \sim N_{p-15}(\mathbf{0}, \Sigma_2)$

เมื่อทำการแปลง x_{ij} ให้เป็นค่ามาตรฐาน จะได้ว่าเมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระสามารถหาได้จากเมทริกซ์สหสัมพันธ์ระหว่างตัวแปรอิสระ โดย $\Sigma = \rho$

$$\text{นั่นคือ } \Sigma_1 = \rho_1 = \begin{bmatrix} 1 & (\rho)^1 & \dots & (\rho)^{14} \\ (\rho)^1 & 1 & & \vdots \\ \vdots & & & 1 \\ (\rho)^{14} & \dots & & \end{bmatrix}_{15 \times 15}$$

$$\text{และ } \Sigma_2 = \rho_2 = \begin{bmatrix} 1 & (\rho)^1 & \dots & (\rho)^{p-16} \\ (\rho)^1 & 1 & & \vdots \\ \vdots & & & 1 \\ (\rho)^{p-16} & \dots & & \end{bmatrix}_{(p-15) \times (p-15)}$$

ดังนั้น เมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระ \mathbf{X} คือ

$$\Sigma = \rho = \begin{bmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \Sigma_2 \end{bmatrix}_{p \times p}$$

และกำหนดค่าสัมประสิทธิ์การถดถอย $\beta_1, \dots, \beta_5 = 2.5$, $\beta_6, \dots, \beta_{10} = 1.5$, $\beta_{11}, \dots, \beta_{15} = 0.5$ และ $\beta_{16}, \dots, \beta_p = 0$ เพื่อกำหนดค่าพารามิเตอร์ให้สอดคล้องกับลักษณะของตัวแบบการถดถอยแบบบางเบา (sparse model) คือ ตัวแปรอิสระส่วนน้อยอยู่ในตัวแบบ

กรณีที่ 2 ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก (non-orthogonal sparse model)

ความสัมพันธ์ของตัวแปรอิสระ \mathbf{X} ไม่เป็นเมทริกซ์เชิงตั้งฉาก (non-orthogonal matrix) คือ ตัวแปรอิสระ p ตัว สัมพันธ์กันทั้งหมด โดยที่ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ k และ j ของ $(x_{i1}, \dots, x_{i(p)})$ คือ $\rho^{|k-j|}$ เมื่อ $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$, $k, j = 1, \dots, p$ ซึ่ง $x_{i1}, \dots, x_{ip} \sim N_p(\mathbf{0}, \Sigma)$

$$\text{โดยที่ } \Sigma = \rho = \begin{bmatrix} 1 & (\rho)^1 & \dots & (\rho)^{p-1} \\ (\rho)^1 & 1 & & \vdots \\ \vdots & & & 1 \\ (\rho)^{p-1} & \dots & & \end{bmatrix}_{p \times p}$$

และกำหนด $\beta_1, \dots, \beta_5 = 2.5$, $\beta_6, \dots, \beta_{10} = 1.5$, $\beta_{11}, \dots, \beta_{15} = 0.5$ และ $\beta_{16}, \dots, \beta_p = 0$ เพื่อกำหนดค่าพารามิเตอร์ให้สอดคล้องกับลักษณะของตัวแบบการถดถอยแบบบางเบาเช่นเดียวกับรูปแบบแรก

กรณีที่ 3 ตัวแบบถดถอยบางเบาที่ความสัมพันธ์ของตัวแปรอิสระมีโครงสร้างแบบแบ่งกลุ่ม (grouping structure sparse model)

ความสัมพันธ์ของตัวแปรอิสระ \mathbf{X} มีโครงสร้างแบบแบ่งกลุ่ม (grouping structure) (Zou และ Hastie, 2005) โดยตัวแปรอิสระ p ตัว แบ่งออกเป็น 4 กลุ่ม ตัวแปรอิสระในกลุ่มเดียวกันมีความสัมพันธ์กันสูงและตัวแปรอิสระแต่ละกลุ่มอิสระต่อกัน ซึ่งตัวแปรอิสระถูกสร้างขึ้นจากตัวแปรสุ่ม Z_{ij} และ e_{ij} ดังนี้

$$X_{ij} = Z_{i1} + e_{ij}, j = 1, \dots, 5, i = 1, \dots, n$$

$$X_{ij} = Z_{i2} + e_{ij}, j = 6, \dots, 10, i = 1, \dots, n$$

$$X_{ij} = Z_{i3} + e_{ij}, j = 11, \dots, 15, i = 1, \dots, n$$

และ $X_{ij} = Z_{ij}, j = 16, \dots, p, i = 1, \dots, n$

เมื่อ $Z_{ij} \sim N(0,1)$ และ $e_{ij} \sim N(0,1/100)$

โดยตัวแปรอิสระกลุ่มแรก มีค่าคาดหวัง ความแปรปรวน ความแปรปรวนร่วมของตัวแปรอิสระ ดังนี้

$$E[X_{ij}] = E[Z_{i1} + e_{ij}] = E[Z_{i1}] + E[e_{ij}] = 0$$

$$V[X_{ij}] = V[Z_{i1} + e_{ij}] = V[Z_{i1}] + V[e_{ij}] + \text{cov}[Z_{i1}, e_{ij}] = 1 + \frac{1}{100} + 0 = 1.01$$

$$\text{cov}[X_{ij}, X_{ij'}] = E[(X_{ij} - E[X_{ij}])(X_{ij'} - E[X_{ij'}])]$$

$$= E[(X_{ij})(X_{ij'})]$$

$$= E[(Z_{i1} + e_{ij})(Z_{i1} + e'_{ij})]$$

$$= E[Z_{i1}^2 + e_{ij}Z_{i1} + e'_{ij}Z_{i1} + e_{ij}e'_{ij}]$$

$$= V[Z_{i1}] + \text{cov}[e_{ij}, Z_{i1}] + \text{cov}[e'_{ij}, Z_{i1}] + \text{cov}[e_{ij}, e'_{ij}]$$

$$= 1$$

ในทำนองเดียวกัน มีค่าคาดหวัง ความแปรปรวน ความแปรปรวนร่วมของตัวแปรอิสระกลุ่มที่ 2 และ 3 คือ $E[X_{ij}] = 0$, $V[X_{ij}] = 1.01$ และ $\text{cov}[X_{ij}, X_{ij'}] = 1$

กำหนด $\beta_1, \dots, \beta_{15} = 1.5$ และที่เหลือ $\beta_{16}, \dots, \beta_p = 0$ เพื่อกำหนดค่าพารามิเตอร์ของทั้งสามกลุ่มมีค่าเท่ากัน นั่นคือ ให้ความสำคัญของกลุ่มทั้งสามเท่ากัน และตัวแปรอิสระในกลุ่มเดียวกันมีความสัมพันธ์กันสูง

กรณีที่ 4 ตัวแบบถดถอยไม่บางเบาและเมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก (non-orthogonal and non-sparse model)

ความสัมพันธ์ของตัวแปรอิสระ X ไม่เป็นเมทริกซ์เชิงตั้งฉากเช่นเดียวกับกรณีที่ 2 แต่กำหนด $\beta_1, \dots, \beta_p = 0.5$ เพื่อกำหนดค่าพารามิเตอร์ให้สอดคล้องกับลักษณะของตัวแบบการถดถอยแบบไม่บางเบา และกำหนดให้ค่าพารามิเตอร์ทุกตัวไม่เท่ากับศูนย์ มีขนาดเล็กและมีค่าเท่ากัน เพื่อให้ให้ความสำคัญของตัวแปรอิสระทุกตัวเท่ากัน

เกณฑ์ที่ใช้ในการพิจารณา

เกณฑ์ที่ใช้ในการพิจารณาเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยวิธีรีดจ์, LASSO, Adaptive LASSO และ Relaxed LASSO ได้แก่

1. ประสิทธิภาพของการพยากรณ์โดยวัดจากค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Median of Prediction Mean Square Error : Med_PMSE) ในการจำลองข้อมูล 500 ครั้ง

$$\text{Med_PMSE} = \text{Median}(\text{PMSE}_1, \dots, \text{PMSE}_{500})$$

โดยที่ $\text{PMSE}_r = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$ เมื่อ n คือ ขนาดตัวอย่าง และ $r = 1, \dots, 500$

2. ประสิทธิภาพของการประมาณโดยวัดจากค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์การถดถอย (Median of Mean Square Error : Med_MSE) ในการจำลองข้อมูล 500 ครั้ง มีค่าน้อยที่สุด

$$\text{Med_MSE} = \text{Median}(\text{MSE}_1, \dots, \text{MSE}_{500})$$

โดยที่ $\text{MSE}_r = \frac{\sum_{j=1}^p (\hat{\beta}_j - \beta_j)^2}{p}$ เมื่อ p คือ จำนวนตัวแปรอิสระ

3. ความความผิดพลาดของการเลือกตัวประมาณ เมื่อกำหนดค่าพารามิเตอร์ β ตามสถานการณ์ต่าง ๆ ซึ่งส่วนใหญ่มีค่าเป็นศูนย์และบางส่วนไม่ได้มีค่าเป็นศูนย์ ทำการนับจำนวนตัวแปรที่ถูกคัดเลือกผิดพลาดไปจากตัวแบบจริง โดยความผิดพลาดในการคัดเลือกตัวแปรมี 2 แบบ คือ

กรณีที่ค่าพารามิเตอร์ไม่เท่ากับศูนย์ แต่ตัวแปรอิสระไม่ถูกคัดเลือกเข้าสู่ตัวแบบ (IC1) และกรณีที่ค่าพารามิเตอร์เท่ากับศูนย์ แต่ตัวแปรอิสระถูกคัดเลือกเข้าสู่ตัวแบบ (IC2) แล้วจึงพิจารณาค่าเฉลี่ยของจำนวนของตัวแปรอิสระที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอิสระ (Mean of IC : Mean_IC) จากการจำลอง 500 รอบ, ค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรอิสระเมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (Percent of IC : %IC) และจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ย (Mean of number of covariates : Mean_Covariate) จากการจำลอง 500 รอบ โดยมีวิธีการคำนวณ ดังนี้

3.1 ค่าเฉลี่ยของจำนวนของตัวแปรอิสระที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอิสระ จากการจำลอง 500 รอบ คือ

$$\text{Mean_IC1} = \frac{\sum_{r=1}^{500} \text{IC1}_r}{500} \quad \text{และ} \quad \text{Mean_IC2} = \frac{\sum_{r=1}^{500} \text{IC2}_r}{500}$$

โดยที่ จำนวนของตัวแปรอิสระที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอิสระ คือ

$$\text{IC1}_r = \#\{j : \beta_j \neq 0, \hat{\beta} = 0\}$$

$$\text{และ} \quad \text{IC2}_r = \#\{j : \beta_j = 0, \hat{\beta} \neq 0\}$$

3.2 ค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรอิสระเมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ คือ

$$\% \text{IC1} = \frac{\text{Mean_IC1}}{15} \times 100 \quad \text{และ} \quad \% \text{IC2} = \frac{\text{Mean_IC2}}{p-15} \times 100$$

3.3 จำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ย คือ

$$\text{Mean_Covariate} = 15 - \text{Mean_IC1} + \text{Mean_IC2}$$

ขั้นตอนการวิจัย

ขั้นตอนการวิจัยของวิธีการหาตัวประมาณพารามิเตอร์ด้วยวิธีริตจ

1. ทำการสร้าง (generate) ข้อมูลค่าสังเกต (Y_i, X_{ij}) , $i = 1, \dots, n$ ของชุดข้อมูลฝึกฝน (Training set) และชุดข้อมูลทดสอบ (Test set) อย่างละ n ค่า โดยกำหนดให้ T และ T^C คือ ชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบ ตามลำดับ โดยมีขั้นตอนดังนี้

1.1) สร้างเมทริกซ์ของตัวแปรอิสระ (\mathbf{X}) ขนาด $n \times p$ โดย \mathbf{X}_i มีการแจกแจงปรกติหลายตัวแปร (Multivariate normal distribution) ที่มีเวกเตอร์ค่าเฉลี่ยเท่ากับเวกเตอร์ศูนย์ และเมทริกซ์ความแปรปรวนร่วม $\Sigma_{p \times p}$ นั่นคือ $\mathbf{X}_i \sim N_p(\mathbf{0}, \Sigma)$, $i = 1, \dots, n$ ซึ่งเมทริกซ์ความแปรปรวนร่วม Σ จะถูกกำหนดแตกต่างกันตามสถานการณ์ต่าง ๆ ตามขอบเขตของการวิจัย

1.2) สร้างเวกเตอร์ความคลาดเคลื่อนสุ่ม $\boldsymbol{\varepsilon} = (\varepsilon_i)$ โดยที่ ε_i มีการแจกแจงปรกติที่มีค่าเฉลี่ยเป็นศูนย์และความแปรปรวนคงที่ (σ^2) สำหรับ $i = 1, \dots, n$ โดยกำหนดส่วนเบี่ยงเบนมาตรฐาน (σ) มีค่าเท่ากับ 1.5

1.3) สร้างเวกเตอร์ของตัวแปรตาม \mathbf{Y} จากตัวแบบ $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ เมื่อ \mathbf{X} และ $\boldsymbol{\varepsilon}$ ได้มาจากข้อ 1.1) และ 1.2) และกำหนด $\boldsymbol{\beta}$ ต่างกันตามสถานการณ์ต่าง ๆ ตามขอบเขตของการวิจัย

2. นำเมทริกซ์ \mathbf{X} และ \mathbf{Y} จากชุดข้อมูลฝึกฝน T มาหาพารามิเตอร์ปรับแต่ง (tuning parameter : t) โดยวิธี V-fold cross-validation โดยมีขั้นตอน ดังนี้

2.1) แบ่งข้อมูล \mathbf{X}_i และ \mathbf{Y}_i , $i = 1, \dots, n$ ใน T ออกเป็น 5 ส่วน ($V=5$) จะได้ T_1, T_2, T_3, T_4, T_5 ให้หนึ่งส่วนเป็นชุดข้อมูลทดสอบย่อย (sub-test set) T_v และอีก 4 ส่วนที่เหลือเป็นชุดข้อมูลฝึกฝนย่อย (sub-training set) $T_{n_v} = T - T_v$ โดยที่ $n_v = \frac{n}{5}$, $v = 1, \dots, 5$

2.2) ทำการแปลงตัวอย่างสุ่ม \mathbf{X}_j จากชุดข้อมูลฝึกฝนย่อยแต่ละเซต $T - T_v$ ให้เป็นค่ามาตรฐาน และแปลงค่าตัวอย่างสุ่ม \mathbf{Y} ให้มีค่าเฉลี่ยเท่ากับศูนย์ ดังนี้

$$X_j^* = \frac{X_j - \bar{X}}{S_j} \quad \text{เมื่อ } i \in T - T_v$$

$$Y_i^* = Y_i - \bar{Y}$$

โดยที่ S_j คือ ส่วนเบี่ยงเบนมาตรฐานตัวอย่าง (sample standard deviation)

\bar{X} คือ เวกเตอร์ค่าเฉลี่ยตัวอย่างของเวกเตอร์ค่าสังเกต \mathbf{X} ในแต่ละชุดข้อมูลฝึกฝน $T - T_v$

\bar{Y} คือ เวกเตอร์ค่าเฉลี่ยตัวอย่างของเวกเตอร์ค่าสังเกต \mathbf{Y} ในแต่ละชุดข้อมูลฝึกฝน $T - T_v$

2.3) กำหนด $t \geq 0$ เมื่อ t เป็นค่าคงที่ใด ๆ ณ จุดค่าใดค่าหนึ่งของค่า k หา $\hat{\boldsymbol{\beta}}_v^{ridge}$ ที่สอดคล้องกับเงื่อนไขบังคับและทำให้ $(\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{\beta})'(\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{\beta})$ มีค่าต่ำที่สุด ทำเช่นนี้ในทุก ๆ ชุดข้อมูลฝึกฝนย่อย $T - T_v$ เมื่อ $v = 1, \dots, 5$

2.4) จากนั้นนำค่าประมาณ $\hat{\boldsymbol{\beta}}_v^{ridge}$ ที่ได้จากแต่ละชุดข้อมูลฝึกฝนย่อยไปหาค่าพยากรณ์ของ \mathbf{Y} ในชุดข้อมูลทดสอบย่อย T_v เมื่อ $v = 1, \dots, 5$ โดยใช้สมการถดถอย $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}_v^{ridge}$ แล้วคำนวณค่าความคลาดเคลื่อนกำลังสองของค่าพยากรณ์ $PMSE_v = \sum_{i=1}^{n_t} (Y_i^* - \hat{Y}_i^*)^2 / n_t$ โดยที่ n_t เป็นขนาดตัวอย่างของ T_v จาก $Y_i \in T_v$

2.5) หาค่าเฉลี่ย $APMSE(J) = \frac{1}{5} \sum_{v=1}^5 PMSE_v$ ที่ t จุดนั้น

2.6) เลือกค่า k ที่ให้ค่าเฉลี่ยของ PMSE ต่ำที่สุด ซึ่งค่า k ค่านั้น คือ ค่าพารามิเตอร์ปรับแต่ง (tuning parameter)

3. หาตัวประมาณสัมประสิทธิ์การถดถอยวิธีริดจ์ ($\hat{\beta}^{*ridge}$) จากชุดข้อมูลฝึกฝน T โดยใช้ค่า t ที่ได้จากข้อ 2.6) แล้วแปลงค่า $\hat{\beta}_j^{*ridge}$ ให้เป็น $\hat{\beta}_j^{ridge} = \hat{\beta}_j^{*ridge} / S_j$

4. นำค่าประมาณ $\hat{\beta}^{ridge}$ ที่ได้จากข้อ 3. ไปหาค่าพยากรณ์ของ Y ที่อยู่ในชุดข้อมูลทดสอบ T^C จากสมการถดถอย $\hat{Y}_j = \beta_0 + \sum \hat{\beta}_j^{ridge} X_{ij}$ หรือ $\hat{Y} = \beta_0 \mathbf{1} + \mathbf{X} \hat{\beta}^{ridge}$ เมื่อ $\beta_0 = \bar{Y} - \sum_{j=1}^p \hat{\beta}_j^{ridge} \bar{X}_j$ แล้วคำนวณค่า PMSE = $\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 / n$ จากชุดข้อมูลทดสอบ T^C

5. คำนวณค่า $MSE = \sum_{j=1}^p (\hat{\beta}_j - \beta_j)^2 / p$

6. นับจำนวนตัวประมาณสัมประสิทธิ์การถดถอย $\hat{\beta}_j^{ridge}$ ในเวกเตอร์ $\hat{\beta}^{ridge}$ ที่มีค่าเท่ากับศูนย์ แต่ค่าพารามิเตอร์ β_j ไม่เท่ากับศูนย์ (IC1) และนับจำนวนตัวประมาณสัมประสิทธิ์การถดถอย $\hat{\beta}_j^{ridge}$ ในเวกเตอร์ $\hat{\beta}^{ridge}$ ที่มีค่าไม่เท่ากับศูนย์ แต่ค่าพารามิเตอร์ β_j เท่ากับศูนย์ (IC2) สำหรับ $j = 1, \dots, p$

7. ทำซ้ำขั้นตอน 1.-6. จำนวน 500 ครั้ง

8. หาค่ามัธยฐานของค่า PMSE, MSE, ค่าเฉลี่ยของจำนวนของตัวแปรอิสระที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอิสระ, ค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรอิสระเมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ และจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ย จากผลการทดลองทั้งหมด 500 ครั้ง

ขั้นตอนการวิจัยของวิธีการหาตัวประมาณพารามิเตอร์ด้วยวิธีแลซโซ

เมื่อดำเนินการวิจัยของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีริดจ์เสร็จสิ้นแล้ว ผู้วิจัยนำชุดข้อมูลฝึกฝน และชุดข้อมูลทดสอบ ที่สร้างขึ้นจากข้อที่ 1 ของขั้นตอนวิธีริดจ์มาหาพารามิเตอร์ปรับแต่งและตัวประมาณสัมประสิทธิ์การถดถอยแบบ LASSO ดังรายละเอียดต่อไปนี้

1. นำ \mathbf{X} และ \mathbf{Y} จากชุดข้อมูลฝึกฝน T มาหาพารามิเตอร์ปรับแต่ง โดยวิธี V-fold cross-validation โดยมีขั้นตอน ดังนี้

1.1) แบ่งข้อมูล $(\mathbf{X}_i, \mathbf{Y}_i)$ ใน T ออกเป็น 5 ส่วน ($V=5$) จะได้ T_1, T_2, T_3, T_4, T_5 ให้หนึ่งส่วนเป็นชุดข้อมูลทดสอบย่อย T_v และอีก 4 ส่วนที่เหลือเป็นชุดข้อมูลฝึกฝนย่อย $T - T_v$

1.2) กำหนดค่าพารามิเตอร์ปรับแต่ง $t \geq 0$ เมื่อ t เป็นคงที่ใด ๆ โดยที่ $0 \leq s = \frac{t}{\sum |\hat{\beta}_{ols}|} \leq 1$ ณ จุดค่าใดค่าหนึ่งของค่า t หา $\hat{\beta}_v^{lasso}$ ที่สอดคล้องกับเงื่อนไขบังคับและ

ทำให้ $(Y^* - X^*\beta)'(Y^* - X^*\beta)$ มีค่าต่ำที่สุด ทำเช่นนี้ในทุก ๆ ชุดข้อมูลฝึกฝนย่อย $T - T_v$ เมื่อ $v = 1, \dots, 5$ ตามขั้นตอนวิธี LARS ที่ได้กล่าวในหัวข้อ 4.3 ขั้นตอนวิธีในการหาตัวประมาณวิธี LASSO

1.3) จากนั้นนำค่าประมาณ $\hat{\beta}_v^{lasso^*}$ ที่ได้จากแต่ละชุดข้อมูลฝึกฝนย่อยไปหาค่าพยากรณ์ของ Y ในชุดข้อมูลทดสอบย่อย T_v เมื่อ $v = 1, \dots, 5$ โดยใช้สมการถดถอย $\hat{Y} = X^*\hat{\beta}_v^{lasso^*}$ แล้วคำนวณค่าความคลาดเคลื่อนกำลังสองของค่าพยากรณ์ $PMSE_v = \sum_{i=1}^{n_v} (Y_i^* - \hat{Y}_i^*)^2 / n_v$ โดยที่ n_v เป็นขนาดตัวอย่างของ T_v^* จาก $Y_i^* \in T_v$

1.4) หาค่าเฉลี่ย $APMSE(J) = \frac{1}{5} \sum_{v=1}^5 PMSE_v$ ที่ t จุดนั้น

1.5) เลือกค่า t ที่ให้ค่าเฉลี่ยของ PMSE ต่ำที่สุด ซึ่งค่า t ค่านั้น คือ ค่าพารามิเตอร์ปรับแต่ง

2. หาตัวประมาณสัมประสิทธิ์การถดถอยวิธี LASSO ($\hat{\beta}^{lasso^*}$) จากจากชุดข้อมูลฝึกฝน T โดยใช้ค่า t ที่ได้จากข้อ 1.5) แล้วแปลงค่า $\hat{\beta}_j^{lasso^*}$ ให้เป็น $\hat{\beta}_j^{lasso} = \hat{\beta}_j^{lasso^*} / S_j$

3. ทำเหมือนขั้นตอนที่ 4.-8. ของวิธีรีดจ์ โดยใช้ตัวประมาณที่ได้จากวิธี LASSO ในข้อ 2.

ขั้นตอนการวิจัยของวิธีการหาตัวประมาณพารามิเตอร์ด้วยวิธีแลชโซปรับได้

เมื่อดำเนินการวิจัยของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีรีดจ์และ LASSO เสร็จสิ้นแล้ว ผู้วิจัยนำชุดข้อมูลฝึกฝน และชุดข้อมูลทดสอบ ที่สร้างขึ้นจากข้อที่ 1 ของขั้นตอนวิธีรีดจ์มาหาพารามิเตอร์ปรับแต่งและตัวประมาณสัมประสิทธิ์การถดถอยแบบ Adaptive LASSO ดังรายละเอียดต่อไปนี้

1. นำ X และ Y จากชุดข้อมูลฝึกฝน T มาหาพารามิเตอร์ปรับแต่ง (tuning parameter : t) โดยวิธี V-fold cross-validation โดยมีขั้นตอน ดังนี้

1.1) แบ่งข้อมูล X_i และ Y_i , $i = 1, \dots, n$ ใน T ออกเป็น 5 ส่วน ($V=5$) จะได้ T_1, T_2, T_3, T_4, T_5 ให้หนึ่งส่วนเป็นชุดข้อมูลทดสอบย่อย T_v และอีก 4 ส่วนที่เหลือเป็นชุดข้อมูลฝึกฝนย่อย $T - T_v$

1.2) นำชุดข้อมูลฝึกฝนย่อย $T - T_v$ มาทำการแปลงตัวอย่างสุ่ม X_j ให้เป็นค่ามาตรฐาน และแปลงค่าตัวอย่างสุ่ม Y ให้มีค่าเฉลี่ยเท่ากับศูนย์ ดังนี้

$$X_{jv}^* = (X_{jv} - \bar{X}_j) / S_j \quad \text{และ} \quad Y_{iv}^* = Y_{iv} - \bar{Y}$$

เมื่อ $i \in T - T_v$, $j = 1, \dots, p$ โดยกำหนดให้ $(T - T_v)^*$ คือ ชุดข้อมูลฝึกฝนที่ได้จากการแปลงข้อมูลค่าสังเกตจาก $T - T_v$

1.3) สำหรับแต่ละชุดข้อมูลฝึกฝนย่อย v นำข้อมูล X_{jv}^* และ Y_{jv}^* มาหาเวกเตอร์ถ่วงน้ำหนัก จาก $w_j = |\tilde{\beta}_j|^{-1}$ เมื่อ $j = 1, \dots, p$ โดยที่ $\tilde{\beta}_j = \frac{(X_{jv}^*)' Y_{jv}^*}{n_T - T_v}$

1.4) ทำการถ่วงน้ำหนักตัวอย่างสุ่ม X_{jv}^* จากชุดข้อมูลฝึกฝนย่อย $T - T_v$ โดย $X_{jv}^{**} = X_{jv}^* / w_{jv}$

1.5) นำข้อมูล X_{jv}^{**} และ Y_{jv}^* มาหาพารามิเตอร์ปรับแต่งด้วยวิธีการของวิธี LASSO ดำเนินตามขั้นตอนที่ 1.2)-1.5) ของวิธี LASSO

2. หาตัวประมาณสัมประสิทธิ์การถดถอย $\hat{\beta}$ จากจากชุดข้อมูลฝึกฝน T โดยมีขั้นตอนดังนี้

2.1) ทำการแปลงตัวอย่างสุ่ม X_i จากชุดข้อมูลฝึกฝน T ให้เป็นค่ามาตรฐาน และแปลงค่าตัวอย่างสุ่ม Y ให้มีค่าเฉลี่ยเท่ากับศูนย์ ดังนี้

$$X_j^* = \frac{X_j - \bar{X}}{s_j} \quad \text{เมื่อ } s_j \text{ คือ } \dots \text{ (sample standard deviation)}$$

และ $Y_i^* = Y_i - \bar{Y}$

2.2) นำชุดข้อมูลฝึกฝน T มาหาเวกเตอร์ถ่วงน้ำหนัก จาก $w_j = |\tilde{\beta}_j|^{-1}$ โดย $\tilde{\beta}_j$ คือ สัมประสิทธิ์การถดถอยเชิงเส้นตัวแปรเดียว (univariate linear regression coefficient) เมื่อ $j = 1, \dots, p$ โดยที่ $\tilde{\beta}_j = \frac{(X_j^*)' Y^*}{n}$

2.3) ทำการแปลงตัวอย่างสุ่ม X_j^* จากชุดข้อมูลฝึกฝน (T^*) โดย $X_j^{**} = X_j^* / w_j$

2.4) ทำการแปลงสัมประสิทธิ์การถดถอยกลับเป็น $\hat{\beta}^* = \hat{\beta} / w$

3. ทำเหมือนขั้นตอนที่ 4.-8. ของวิธีริตจ โดยใช้ตัวประมาณที่ได้จากวิธี Adaptive LASSO ในข้อ 2.4

ขั้นตอนการวิจัยของวิธีการหาตัวประมาณพารามิเตอร์ด้วยวิธีรีแลกซ์แลชโซ

เมื่อดำเนินการวิจัยของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีริตจ, LASSO และ Adaptive LASSO เสร็จสิ้นแล้ว ผู้วิจัยนำชุดข้อมูลฝึกฝน และชุดข้อมูลทดสอบ ที่สร้างขึ้นจากข้อที่ 1 ของขั้นตอนวิธีริตจมาหาพารามิเตอร์ปรับแต่งและตัวประมาณสัมประสิทธิ์การถดถอยแบบ Relaxed LASSO ดังรายละเอียดต่อไปนี้

1. นำ X และ Y จากชุดข้อมูลฝึกฝน T มาหาพารามิเตอร์ปรับแต่ง ϕ โดยวิธี V-fold cross-validation โดยมีขั้นตอน ดังนี้

1.1) แบ่งข้อมูล X_i และ Y_i , $i = 1, \dots, n$ ใน T ออกเป็น 5 ส่วน ($V=5$) จะได้ T_1, T_2, T_3, T_4, T_5 ให้หนึ่งส่วนเป็นชุดข้อมูลทดสอบย่อย T_v และอีก 4 ส่วนที่เหลือเป็นชุดข้อมูลฝึกฝนย่อย $T - T_v$

1.2) นำ $\hat{\beta}^{lasso}$ ที่ได้จากข้อ 2.2) ในวิธี LASSO มาคำนวณหาค่า λ จาก $\lambda = \{ |(Y - X\hat{\beta})(M_{11(v)})| \text{ และ } \max|X'Y| \}$ โดยที่ $M_{11(v)}$ คือ ค่าของสมาชิกในแถวที่ 1 คอลัมน์ที่ 1 ของเมทริกซ์ตัวแปรอิสระที่ถูกคัดเลือกเข้าสู่ตัวแบบ เมื่อ $v = 1, \dots, 5$

1.3) กำหนดค่าพารามิเตอร์ปรับแต่ง ϕ เป็นค่าที่ใด ๆ โดยที่ $0 \leq \phi \leq 1$ ณ ทุกจุดของค่า λ และ ϕ นำข้อมูล X_{jv}^* และ Y_v^* มาหาพารามิเตอร์ปรับแต่งด้วยวิธีการของวิธี LASSO ดำเนินตามขั้นตอนที่ 1.2)-1.4) ของวิธี LASSO

1.4) เลือกค่า ϕ และ λ ที่ให้ค่าเฉลี่ยของ PMSE ต่ำที่สุด ซึ่งค่า ϕ และ λ ค่านั้น คือ ค่าพารามิเตอร์ปรับแต่ง

2. หาตัวประมาณสัมประสิทธิ์การถดถอย $\hat{\beta}^{rex.lasso}$ จากชุดข้อมูลฝึกฝน T ทุก จุด ϕ และ λ

3. เลือกตัวประมาณสัมประสิทธิ์การถดถอย $\hat{\beta}^{rex.lasso}$ จากชุดข้อมูลฝึกฝน T ณ จุด λ และ ϕ ที่ให้ค่าเฉลี่ยของ PMSE ต่ำที่สุด

4. ทำเหมือนขั้นตอนที่ 4.-8. ของวิธีรีดจ์ โดยใช้ตัวประมาณที่ได้จากวิธี Relaxed LASSO ในข้อ 3.

เมื่อดำเนินการวิจัยของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 4 วิธี คือ วิธีรีดจ์, LASSO, Adaptive LASSO และ Relaxed LASSO เสร็จสิ้นแล้ว จากนั้นผู้วิจัยทำการเปรียบเทียบค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์การถดถอย ค่าเฉลี่ยของจำนวนของตัวแปรอิสระที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอิสระ ค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรอิสระเมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ และจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ย

บทที่ 4

ผลการวิเคราะห์ข้อมูล

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสัมประสิทธิ์การถดถอย 4 วิธี ในกรณีที่ข้อมูลมีมิติสูง ได้แก่ วิธีริคต์, LASSO, Adaptive LASSO และ Relaxed LASSO โดยพิจารณาจากความถูกต้องในการพยากรณ์ ความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอย และความถูกต้องในการคัดเลือกตัวแปรอิสระเข้าในตัวแบบ

ในการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสัมประสิทธิ์การถดถอย ทำการศึกษา 4 กรณี คือ กรณีที่ 1 คือ ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระเป็นเชิงตั้งฉากบางส่วน (partial orthogonal sparse model) ตัวแปรอิสระ X_k และ X_j ที่มีความสัมพันธ์กัน มีค่าสัมประสิทธิ์สหสัมพันธ์ คือ $\rho^{|k-j|}$ เมื่อกำหนด 5 กรณี คือ $\rho = 0.1, 0.3, 0.5, 0.7$ และ 0.9 และกำหนดค่าพารามิเตอร์ให้สอดคล้องกับลักษณะของตัวแบบการถดถอยแบบบางเบา กรณีที่ 2 คือ ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก (non-orthogonal sparse model) กำหนดค่าสัมประสิทธิ์สหสัมพันธ์และค่าพารามิเตอร์ให้สอดคล้องกับลักษณะของตัวแบบการถดถอยแบบบางเบาเช่นเดียวกับกรณีแรก กรณีที่ 3 คือ ตัวแบบถดถอยบางเบาที่ความสัมพันธ์ของตัวแปรอิสระมีโครงสร้างแบบแบ่งกลุ่ม (grouping structure sparse model) นั่นคือ ตัวแปรอิสระในกลุ่มเดียวกันมีความสัมพันธ์กันสูงและตัวแปรอิสระแต่ละกลุ่มอิสระต่อกัน ซึ่งกำหนดค่าพารามิเตอร์ของทั้งสามกลุ่มมีค่าเท่ากัน เพื่อให้ความสำคัญของกลุ่มทั้งสามเท่ากัน และกำหนดค่าพารามิเตอร์ให้สอดคล้องกับลักษณะของตัวแบบการถดถอยแบบบางเบา และกรณีที่ 4 คือ ตัวแบบถดถอยไม่บางเบาและเมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก (non-orthogonal and non-sparse model) กำหนดรูปแบบความสัมพันธ์ของตัวแปรอิสระเช่นเดียวกับกรณีที่ 2 แต่กำหนดค่าพารามิเตอร์ทุกตัวให้มีขนาดเล็กแต่ไม่เท่ากับศูนย์และมีค่าเท่ากัน โดยแต่ละกรณีจะทำการจำลองข้อมูลที่มีขนาดตัวอย่าง (n) เท่ากับ 50, 100 และ 200 และจำนวนตัวแปรอิสระ (p) เป็น 2 เท่า และ 4 เท่า ของขนาดตัวอย่าง

โดยเกณฑ์ที่ใช้ในการพิจารณาเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 4 วิธี ได้แก่

1. ประสิทธิภาพของการพยากรณ์โดยวัดจากค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Med_PMSE) ในการจำลองข้อมูล 500 รอบ

2. ประสิทธิภาพของการประมาณโดยวัดจากค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์การถดถอย (Med_MSE) ในการจำลองข้อมูล 500 รอบ

3. ความผิดพลาดของการเลือกตัวประมาณโดยวัดจาก

3.1) ค่าเฉลี่ยของจำนวนของตัวแปรอิสระที่เกิดความผิดพลาดของการคัดเลือกตัวแปรอิสระ (Mean_IC) จากการจำลอง 500 รอบ

3.2) ค่าร้อยละของความผิดพลาดของการคัดเลือกตัวแปรอิสระเมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (%IC)

3.3) จำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ย (Mean_Covariate) จากการจำลอง 500 รอบ

งานวิจัยนี้ทำการจำลองข้อมูลโดยโปรแกรม R เวอร์ชัน 3.2.2 โดยแต่ละสถานการณ์มีการทำซ้ำ 500 รอบ ซึ่งผลการวิจัยมีรายละเอียด ดังนี้

ความถูกต้องของการพยากรณ์ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริตจ, LASSO, Adaptive LASSO และ Relaxed LASSO

ความถูกต้องของการพยากรณ์ของตัวประมาณวิธีริตจ, LASSO, Adaptive LASSO และ Relaxed LASSO โดยวัดจากค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Med_PMSE) ในการจำลองข้อมูล 500 รอบ ซึ่งผลการวิจัยแบ่งออกเป็น 4 กรณี คือ 1) ตัวแบบถดถอยพหุคูณ partial orthogonal sparse model 2) ตัวแบบถดถอยพหุคูณ non-orthogonal sparse model 3) ตัวแบบถดถอยพหุคูณ grouping structure sparse model และ 4) ตัวแบบถดถอยพหุคูณ non-orthogonal and non-sparse model ซึ่งมีรายละเอียดดังนี้

กรณีที่ 1 ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระเป็นเชิงตั้งฉากบางส่วน

ในกรณีตัวแบบถดถอยพหุคูณ partial orthogonal sparse model กล่าวคือ ตัวแปรอิสระ X_1, \dots, X_{15} มีความสัมพันธ์กันโดยที่สัมประสิทธิ์สหสัมพันธ์ระหว่าง X_k และ X_j คือ $\rho^{|k-j|}$ และตัวแปรอิสระ X_{16}, \dots, X_p มีความสัมพันธ์กันโดยที่สัมประสิทธิ์สหสัมพันธ์ระหว่าง X_k และ X_j คือ $\rho^{|k-j|}$ แต่ตัวแปรอิสระกลุ่มที่ 1 คือ X_1, \dots, X_{15} และตัวแปรอิสระกลุ่มที่ 2 คือ X_{16}, \dots, X_p ไม่มีความสัมพันธ์กัน และกำหนดค่าพารามิเตอร์ให้สอดคล้องกับลักษณะของตัวแบบการถดถอยแบบบางเบา

ความถูกต้องของการพยากรณ์ของตัวประมาณวิธีริตจ, LASSO, Adaptive LASSO และ Relaxed LASSO ในกรณีตัวแบบถดถอยพหุคูณ partial orthogonal sparse model แสดงในตารางที่ 1



ตารางที่ 1 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Med_PMSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริดจ์, LASSO, Adaptive LASSO และ Relaxed LASSO กรณี partial orthogonal sparse model

ρ	n	p	Med_PMSE			
			Ridge Regression	LASSO	Adaptive LASSO	Relaxed LASSO
0.1	50	100	27.2358	8.1755	9.6714	8.0078
		200	39.8373	13.9576	15.6093	13.8015
	100	200	26.8089	4.5830	5.0076	4.1168
		400	39.1016	5.3221	5.7751	4.6574
	200	400	26.8768	3.3814	3.5492	2.8734
800	39.4978	3.5542	3.6911	3.0731		
0.3	50	100	23.8201	6.6440	7.1333	6.0180
		200	45.3991	9.1994	10.0434	8.6079
	100	200	24.2603	4.0748	4.3612	3.5205
		400	44.2122	4.4717	4.6015	3.8820
	200	400	23.6613	3.2000	3.2733	2.6601
800	43.6643	3.1998	3.2533	2.6594		
0.5	50	100	18.3839	5.3710	5.8036	4.8517
		200	46.3906	6.2364	6.4275	5.5830
	100	200	17.4131	3.7306	3.9032	3.1004
		400	45.6925	3.8454	3.8510	3.1792
	200	400	17.7979	3.1225	3.1176	2.5742
800	44.6804	3.0054	2.9881	2.5403		
0.7	50	100	10.7791	4.7863	5.1542	4.0411
		200	36.1065	4.8963	4.8177	4.2306
	100	200	10.6624	3.6175	3.5303	2.8900
		400	35.5068	3.4454	3.3995	2.9069
	200	400	10.6850	3.1526	2.9855	2.5196
800	35.0857	2.8847	2.7735	2.4660		
0.9	50	100	5.8069	4.6682	4.4853	3.5452
		200	11.3853	4.2349	4.1440	3.5264
	100	200	5.9723	3.7001	3.3315	2.7566
		400	11.0559	3.4644	3.1948	2.7953
	200	400	6.1983	3.1901	2.8210	2.5051
800	11.0967	2.9644	2.7099	2.4732		

จากตารางที่ 1 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ partial orthogonal sparse model ในทุกกรณีของค่า ρ , n และ p พบว่า ค่ามัธยฐานของค่า PMSE ของวิธี Relaxed LASSO มีค่าต่ำที่สุด และค่ามัธยฐานของ PMSE ของวิธีริตจมีค่าสูงสุด เมื่อเปรียบเทียบวิธี LASSO และ Adaptive LASSO พบว่า ในกรณีที่ $\rho = 0.1, 0.3,$ และ 0.5 ค่ามัธยฐานของ PMSE ของวิธี LASSO ต่ำกว่าวิธี Adaptive LASSO ในทุกๆ ค่า n และ p ยกเว้น กรณีที่ $n = 200$ ส่วนวิธี Adaptive LASSO มีค่ามัธยฐานของค่า PMSE ต่ำกว่าวิธี LASSO ในกรณีที่ $\rho = 0.7$ และ 0.9 ในทุกกรณีของค่า n และ p

กรณีที่ 2 ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก

ในกรณีตัวแบบถดถอยพหุคูณ non-orthogonal sparse model กล่าวคือ ตัวแปรอิสระ X_1, \dots, X_p มีความสัมพันธ์กันโดยที่สัมประสิทธิ์สหสัมพันธ์ระหว่าง X_k และ X_j คือ $\rho^{|k-j|}$ แต่ตัวแปรกลุ่มที่ 1 คือ X_1, \dots, X_{15} และตัวแปรอิสระกลุ่มที่ 2 คือ X_{16}, \dots, X_p มีความสัมพันธ์กัน และกำหนดค่าพารามิเตอร์ให้สอดคล้องกับลักษณะของตัวแบบการถดถอยแบบบางเบา

ความถูกต้องของการพยากรณ์ของตัวประมาณวิธีริตจ, LASSO, Adaptive LASSO และ Relaxed LASSO ในกรณีตัวแบบถดถอยพหุคูณ non-orthogonal sparse model แสดงในตารางที่ 2



ตารางที่ 2 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Med_PMSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริดจ์, LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal sparse model

ρ	n	p	Med_PMSE			
			Ridge	LASSO	Adaptive	Relaxed
0.1	50	100	27.5173	8.5301	10.1093	8.2879
		200	39.1090	12.8956	14.3263	13.4529
	100	200	26.6689	4.6824	5.0065	4.1616
		400	39.4489	5.4150	5.7989	4.6714
	200	400	26.6649	3.4315	3.5567	2.9022
		800	39.1583	3.5668	3.7296	3.0881
0.3	50	100	24.7933	6.4828	7.3312	5.9970
		200	45.4593	8.6545	9.9742	8.3358
	100	200	23.9852	4.1047	4.2676	3.5427
		400	44.2997	4.4720	4.6377	3.8164
	200	400	23.5136	3.2092	3.2733	2.6379
		800	43.9608	3.2344	3.2592	2.6748
0.5	50	100	18.1360	5.6117	5.8861	4.8800
		200	45.6011	6.3273	6.3972	5.9148
	100	200	17.7304	3.7665	3.9389	3.0689
		400	45.5766	3.7831	3.8600	3.1515
	200	400	17.7592	3.1349	3.0531	2.5347
		800	45.1763	3.0216	3.0115	2.5443
0.7	50	100	11.0951	4.8947	5.0632	3.9808
		200	36.3771	5.0157	4.9724	4.3485
	100	200	10.5289	3.6507	3.5418	2.8771
		400	35.4822	3.4949	3.4461	2.9496
	200	400	10.4778	3.1176	2.9758	2.5013
		800	34.6760	2.8982	2.7858	2.4973
0.9	50	100	5.6919	4.6336	4.4450	3.5253
		200	11.8775	4.1888	4.2261	3.4946
	100	200	6.0281	3.6045	3.3396	2.7768
		400	11.3465	3.4357	3.1950	2.7587
	200	400	6.1322	3.1585	2.9133	2.4887
		800	11.2196	3.0049	2.7289	2.4983

จากตารางที่ 2 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ non-orthogonal sparse model ในทุกกรณีของค่า ρ , n และ p พบว่า ค่ามัธยฐานของค่า PMSE ของวิธี Relaxed LASSO มีค่าต่ำที่สุด ยกเว้นกรณีที่ $\rho = 0.1$, $n = 50$ และ $p = 200$ และค่ามัธยฐานของ PMSE ของวิธีรีดจ์ มีค่าสูงสุด เมื่อเปรียบเทียบวิธี LASSO และ Adaptive LASSO พบว่า ในกรณีที่ $\rho = 0.1$, 0.3 , และ 0.5 ค่ามัธยฐานของ PMSE ของวิธี LASSO ต่ำกว่าวิธี Adaptive LASSO ในทุกๆ ค่า n และ p ยกเว้น กรณีที่ $\rho = 0.5$ และ $n = 200$ ส่วนวิธี Adaptive LASSO มีค่ามัธยฐานของค่า PMSE ต่ำกว่าวิธี LASSO ในกรณีที่ $\rho = 0.7$ และ 0.9 ในทุกกรณีของค่า n และ p ยกเว้นกรณีที่ $\rho = 0.7$ และ 0.9 , $n = 50$ และ $p = 100$

กรณีที่ 3 ตัวแบบถดถอยบางเบาที่ความสัมพันธ์ของตัวแปรอิสระมีโครงสร้างแบบแบ่งกลุ่ม

ในกรณีตัวแบบถดถอยพหุคูณ grouping structure sparse model กล่าวคือ ตัวแปรอิสระจำนวน p ตัว แบ่งออกเป็น 4 กลุ่ม ซึ่งถูกสร้างขึ้นจากตัวแปรสุ่ม Z_{ij} และ e_{ij} ดังนี้

$$X_{ij} = Z_{i1} + e_{ij}, j = 1, \dots, 5,$$

$$X_{ij} = Z_{i2} + e_{ij}, j = 6, \dots, 10,$$

$$X_{ij} = Z_{i3} + e_{ij}, j = 11, \dots, 15 \text{ และ}$$

$$X_{ij} = Z_{ij}, j = 16, \dots, p, i = 1, \dots, n$$

โดยตัวแปรอิสระแต่ละกลุ่มอิสระต่อกัน และกำหนดค่าพารามิเตอร์ให้สอดคล้องกับลักษณะของตัวแบบการถดถอยแบบบางเบา

ความถูกต้องของการพยากรณ์ของตัวประมาณวิธีรีดจ์, LASSO, Adaptive LASSO และ Relaxed LASSO ในกรณีตัวแบบถดถอยพหุคูณ grouping structure sparse model แสดงในตารางที่ 3

ตารางที่ 3 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Med_PMSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริดจ์, LASSO, Adaptive LASSO และ Relaxed LASSO กรณี grouping structure sparse model

n	p	Med_PMSE			
		Ridge regression	LASSO	Adaptive LASSO	Relaxed LASSO
50	100	13.4745	3.4091	3.7128	2.5780
	200	34.6734	3.5171	3.3431	2.6232
100	200	13.3672	2.9297	2.9990	2.3950
	400	35.2044	2.9566	2.8699	2.4190
200	400	13.6068	2.7380	2.6785	2.3395
	800	34.2631	2.6570	2.5930	2.3491

จากตารางที่ 3 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ grouping structure sparse model ในทุกกรณีของค่า n และ p พบว่า ค่ามัธยฐานของค่า PMSE ของวิธี Relaxed LASSO มีค่าต่ำที่สุดในทุกๆ กรณี วิธี Adaptive LASSO มีค่ามัธยฐานของค่า PMSE ต่ำกว่าวิธี LASSO ยกเว้นกรณีที่ n = 50, p = 100 และ n = 100, p = 200 และวิธีริดจ์มีค่ามัธยฐานของค่า PMSE สูงที่สุด

กรณีที่ 4 ตัวแบบถดถอยไม่บางเบาและเมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก

ในกรณีตัวแบบถดถอยพหุคูณ non-orthogonal and non-sparse model กล่าวคือ ตัวแปรอิสระ X_1, \dots, X_p มีความสัมพันธ์กันโดยที่สัมประสิทธิ์สหสัมพันธ์ระหว่าง X_k และ X_j คือ $\rho^{|k-j|}$ แต่ตัวแปรกลุ่มที่ 1 คือ X_1, \dots, X_{15} และตัวแปรอิสระกลุ่มที่ 2 คือ X_{16}, \dots, X_p มีความสัมพันธ์กัน แต่กำหนดให้ค่าพารามิเตอร์ทุกตัวไม่เท่ากับศูนย์ มีขนาดเล็กและมีค่าเท่ากัน (non-sparse model)

ความถูกต้องของการพยากรณ์ของตัวประมาณวิธีริดจ์, LASSO, Adaptive LASSO และ Relaxed LASSO ในกรณีตัวแบบถดถอยพหุคูณ non-orthogonal and non-sparse model แสดงในตารางที่ 4

ตารางที่ 4 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Med_PMSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริดจ์, LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal and non-sparse model

ρ	n	p	Med_PMSE			
			Ridge	LASSO	Adaptive	Relaxed
0.1	50	100	17.2754	30.3788	30.7969	31.3444
		200	45.2894	63.4621	65.7713	65.8134
	100	200	29.9547	55.8951	58.2164	57.5619
		400	86.5258	86.5258	125.5908	126.1321
	200	400	54.1215	105.1692	109.3690	107.2480
		800	170.0667	244.9744	251.4116	248.5342
0.3	50	100	15.9371	35.8241	37.6409	37.6984
		200	51.6886	90.3328	94.5037	95.0314
	100	200	26.0349	64.3920	70.3231	67.5479
		400	99.6378	178.0761	182.1035	182.3775
	200	400	47.7469	118.5768	130.1534	123.8643
		800	194.0940	344.5003	354.2469	352.0305
0.5	50	100	11.9829	34.7794	40.4338	41.2252
		200	51.6730	119.7407	122.0156	130.3565
	100	200	18.8660	63.9535	74.2398	72.1619
		400	99.5920	229.9371	238.5948	242.8630
	200	400	31.3373	120.0695	144.6126	132.5610
		800	195.2576	442.7951	461.7521	460.9673
0.7	50	100	7.6771	28.6733	35.4871	38.7441
		200	35.5480	140.0892	147.1222	162.2275
	100	200	9.8452	53.7429	65.2040	67.3050
		400	67.0588	271.7855	293.4571	304.9877
	200	400	14.1049	105.0322	126.2674	125.8682
		800	128.8989	533.0429	570.9250	582.2999
0.9	50	100	5.0374	13.3519	14.2029	19.6038
		200	7.4647	109.9255	113.6145	147.6962
	100	200	5.4674	25.0061	25.0892	33.9368
		400	10.3003	10.3003	10.3003	268.1004
	200	400	5.9783	48.6608	47.9993	62.2899
		800	16.0924	445.0594	445.0263	511.7671

จากตารางที่ 4 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ non-orthogonal and non-sparse model ในทุกกรณีของค่า ρ , n และ p พบว่า ค่ามัธยฐานของค่า PMSE ของวิธีรีดจ์มีค่าต่ำที่สุด ต่อมาคือวิธี LASSO ยกเว้นกรณีที่ $\rho = 0.9$ และ $n = 200$ ทุกค่า p เมื่อเปรียบเทียบวิธี Adaptive LASSO และ Relaxed LASSO พบว่า ในกรณีที่ $\rho = 0.1, 0.3,$ และ 0.5 เมื่อ $n = 50$ ทุกค่า p และ $n = 100, p = 400$ พบว่า ค่ามัธยฐานของ PMSE ของวิธี Adaptive LASSO ต่ำกว่าวิธี Relaxed LASSO และเมื่อ $n = 100, p = 200$ และ $n = 200$ ทุกค่า p พบว่า วิธี Relaxed LASSO มีค่ามัธยฐานของค่า PMSE ต่ำกว่าวิธี Adaptive LASSO ส่วนในกรณีที่ $\rho = 0.7$ และ 0.9 ในทุกกรณีของค่า n และ p ยกเว้นกรณีที่ $\rho = 0.7, n = 200$ และ $p = 400$ วิธี Adaptive LASSO มีค่ามัธยฐานของค่า PMSE ต่ำกว่าวิธี Relaxed LASSO

ความถูกต้องของการประมาณค่าพารามิเตอร์ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีรีดจ์, LASSO, Adaptive LASSO และ Relaxed LASSO

ความถูกต้องของการประมาณค่าพารามิเตอร์ของตัวประมาณวิธีรีดจ์, LASSO, Adaptive LASSO และ Relaxed LASSO โดยวัดจากค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์การถดถอย (Med_MSE) ในการจำลองข้อมูล 500 รอบ ซึ่งผลการวิจัยแบ่งออกเป็น 4 กรณี คือ 1) ตัวแบบถดถอยพหุคูณ partial orthogonal sparse model 2) ตัวแบบถดถอยพหุคูณ non-orthogonal sparse model 3) ตัวแบบถดถอยพหุคูณ grouping structure sparse model และ 4) ตัวแบบถดถอยพหุคูณ non-orthogonal and non-sparse model ซึ่งมีรายละเอียดดังนี้

กรณีที่ 1 ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระเป็นเชิงตั้งฉากบางส่วน
 ตารางที่ 5 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย
 (Med_MSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริดจ์, LASSO, Adaptive LASSO และ
 Relaxed LASSO กรณี partial orthogonal sparse model

ρ	n	p	Med_MSE			
			Ridge	LASSO	Adaptive	Relaxed
0.1	50	100	0.2290	0.0559	0.0732	0.0550
		200	0.1646	0.0518	0.0603	0.0544
	100	200	0.1128	0.0112	0.0127	0.0088
		400	0.0807	0.0070	0.0080	0.0056
	200	400	0.0563	0.0027	0.0031	0.0015
		800	0.0403	0.0015	0.0017	0.0010
0.3	50	100	0.1806	0.0426	0.0486	0.0385
		200	0.1466	0.0286	0.0331	0.0283
	100	200	0.0898	0.0086	0.0100	0.0065
		400	0.0717	0.0048	0.0054	0.0041
	200	400	0.0446	0.0022	0.0026	0.0011
		800	0.0354	0.0011	0.0012	0.0006
0.5	50	100	0.1355	0.0352	0.0433	0.0322
		200	0.1193	0.0207	0.0224	0.0203
	100	200	0.0657	0.0082	0.0098	0.0058
		400	0.0583	0.0039	0.0044	0.0034
	200	400	0.0327	0.0022	0.0024	0.0010
		800	0.0287	0.0009	0.0011	0.0005
0.7	50	100	0.0999	0.0464	0.0526	0.0373
		200	0.0813	0.0205	0.0224	0.0203
	100	200	0.0493	0.0107	0.0114	0.0073
		400	0.0401	0.0044	0.0048	0.0039
	200	400	0.0244	0.0030	0.0027	0.0015
		800	0.0198	0.0011	0.0011	0.0008
0.9	50	100	0.1093	0.1119	0.1193	0.0804
		200	0.0441	0.0447	0.0477	0.0429
	100	200	0.0614	0.0274	0.0244	0.0175
		400	0.0213	0.0110	0.0106	0.0089
	200	400	0.0349	0.0076	0.0060	0.0041
		800	0.0106	0.0028	0.0023	0.0022

จากตารางที่ 5 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ partial orthogonal sparse model ในทุกกรณีของค่า ρ , n และ p พบว่า ค่ามัธยฐานของค่า MSE ของวิธี Relaxed LASSO มีค่าต่ำที่สุด ยกเว้นกรณีที่ $\rho = 0.1$, $n = 50$ และ $p = 200$ และค่ามัธยฐานของ MSE ของวิธีริตจมีค่าสูงสุด ยกเว้นกรณีที่ $\rho = 0.9$ และ $n = 50$ เมื่อเปรียบเทียบวิธี LASSO และ Adaptive LASSO พบว่า ในกรณีที่ $\rho = 0.1, 0.3, 0.5$ และ 0.7 วิธี LASSO มีค่ามัธยฐานของ MSE ต่ำกว่าวิธี Adaptive LASSO ในทุกๆ ค่า n และ p ยกเว้น กรณีที่ $\rho = 0.7$ และ $n = 200$ ส่วนวิธี Adaptive LASSO มีค่ามัธยฐานของค่า MSE ต่ำกว่าวิธี LASSO ในกรณีที่ $\rho = 0.9$ และ $n = 100$ และ 200 ในทุกค่า p



กรณีที่ 2 ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก

ตารางที่ 6 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย (Med_MSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริคต์, LASSO, Adaptive LASSO และ Relaxed LASSO กรณี partial non-orthogonal sparse model

ρ	n	p	Med_MSE			
			Ridge	LASSO	Adaptive	Relaxed
0.1	50	100	0.2296	0.0577	0.0746	0.0575
		200	0.1645	0.0492	0.0587	0.0509
	100	200	0.1130	0.0112	0.0129	0.0087
		400	0.0809	0.0072	0.0083	0.0057
	200	400	0.0558	0.0027	0.0031	0.0015
		800	0.0403	0.0015	0.0017	0.0010
0.3	50	100	0.1841	0.0397	0.0504	0.0379
		200	0.1452	0.0280	0.0329	0.0283
	100	200	0.0907	0.0088	0.0103	0.0068
		400	0.0718	0.0048	0.0052	0.0039
	200	400	0.0444	0.0022	0.0025	0.0009
		800	0.0353	0.0011	0.0012	0.0006
0.5	50	100	0.1354	0.0375	0.0448	0.0328
		200	0.1187	0.0207	0.0222	0.0204
	100	200	0.0660	0.0084	0.0102	0.0060
		400	0.0582	0.0039	0.0046	0.0032
	200	400	0.0329	0.0023	0.0024	0.0011
		800	0.0289	0.0010	0.0010	0.0006
0.7	50	100	0.0996	0.0471	0.0521	0.0376
		200	0.0827	0.0215	0.0219	0.0201
	100	200	0.0494	0.0110	0.0117	0.0074
		400	0.0400	0.0046	0.0049	0.0039
	200	400	0.0246	0.0030	0.0028	0.0015
		800	0.0197	0.0011	0.0011	0.0008
0.9	50	100	0.1142	0.1133	0.1256	0.0828
		200	0.0455	0.0444	0.0468	0.0418
	100	200	0.0630	0.0282	0.0269	0.0171
		400	0.0218	0.0108	0.0106	0.0088
	200	400	0.0352	0.0077	0.0070	0.0043
		800	0.0107	0.0029	0.0026	0.0022

จากตารางที่ 6 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ non-orthogonal sparse model ในทุกกรณีของค่า ρ , n และ p พบว่า ค่ามัธยฐานของค่า MSE ของวิธี Relaxed LASSO มีค่าต่ำที่สุด ยกเว้นกรณีที่ $\rho = 0.1$ และ 0.3 , $n = 50$ และ $p = 200$ และค่ามัธยฐานของ MSE ของวิธีริตจมีค่าสูงสุด ยกเว้นกรณีที่ $\rho = 0.9$ และ $n = 50$ เมื่อเปรียบเทียบวิธี LASSO และ Adaptive LASSO พบว่า ในกรณีที่ $\rho = 0.1, 0.3, 0.5$ และ 0.7 วิธี LASSO มีค่ามัธยฐานของ MSE ต่ำกว่าวิธี Adaptive LASSO ในทุกๆ ค่า n และ p ยกเว้น กรณีที่ $\rho = 0.7$ และ $n = 200$ ส่วนวิธี Adaptive LASSO มีค่ามัธยฐานของค่า MSE ต่ำกว่าวิธี LASSO ในกรณีที่ $\rho = 0.9$ และ $n = 100$ และ 200 ในทุกค่า p

กรณีที่ 3 ตัวแบบถดถอยบางเบาที่ความสัมพันธ์ของตัวแปรอิสระมีโครงสร้างแบบแบ่งกลุ่ม

ตารางที่ 7 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย (Med_MSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริตจ, LASSO, Adaptive LASSO และ Relaxed LASSO กรณี grouping structure sparse model

n	p	Med_MSE			
		Ridge regression	LASSO	Adaptive LASSO	Relaxed LASSO
50	100	0.0791	1.1218	1.1870	1.3917
	200	0.0677	0.5421	0.5646	0.6960
100	200	0.0389	0.5292	0.5380	0.7018
	400	0.0339	0.2593	0.2657	0.3519
200	400	0.0196	0.2458	0.2479	0.3401
	800	0.0168	0.1190	0.1209	0.1698

จากตารางที่ 7 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ grouping structure sparse model ในทุกกรณีของค่า n และ p พบว่า ค่ามัธยฐานของค่า MSE ของวิธีริตจมีค่าต่ำที่สุดในทุกกรณี วิธี LASSO มีค่ามัธยฐานของค่า PMSE ต่ำกว่าวิธี Adaptive LASSO และวิธี Relaxed LASSO มีค่ามัธยฐานของค่า PMSE สูงที่สุด

กรณีที่ 4 ตัวแบบถดถอยไม่บางเบาและเมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก

ตารางที่ 8 ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย (Med_MSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธีริจด์, LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal and non-sparse model

ρ	n	p	Med_MSE			
			Ridge	LASSO	Adaptive	Relaxed
0.1	50	100	0.1424	0.2426	0.2500	0.2500
		200	0.1859	0.2500	0.2611	0.2619
	100	200	0.1256	0.2349	0.2474	0.2416
		400	0.1819	0.2500	0.2573	0.2567
	200	400	0.1191	0.2293	0.2410	0.2324
		800	0.1800	0.2500	0.2568	0.2527
0.3	50	100	0.1119	0.2511	0.2680	0.2681
		200	0.1611	0.2707	0.2939	0.2920
	100	200	0.0965	0.2473	0.2637	0.2531
		400	0.1576	0.2707	0.2943	0.2889
	200	400	0.0905	0.2435	0.2585	0.2467
		800	0.1547	0.2733	0.2934	0.2784
0.5	50	100	0.0833	0.2881	0.3202	0.3084
		200	0.1253	0.3384	0.3633	0.3665
	100	200	0.0668	0.2815	0.3134	0.2946
		400	0.1209	0.3490	0.3661	0.3615
	200	400	0.0576	0.2758	0.3178	0.2836
		800	0.1184	0.3549	0.3719	0.3622
0.7	50	100	0.0649	0.3480	0.4122	0.4043
		200	0.0761	0.4947	0.5143	0.5529
	100	200	0.0439	0.3378	0.4006	0.3736
		400	0.0696	0.4934	0.5277	0.5357
	200	400	0.0318	0.3315	0.3990	0.3564
		800	0.0663	0.4937	0.5313	0.5160
0.9	50	100	0.0892	0.4669	0.4889	0.6216
		200	0.0264	0.8923	0.9089	1.1629
	100	200	0.0554	0.4487	0.4532	0.5667
		400	0.0180	0.8768	0.8944	1.0891
	200	400	0.0331	0.4372	0.4405	0.5260
		800	0.0138	0.8693	0.8860	1.0146

จากตารางที่ 8 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ non-orthogonal and non-sparse model ในทุกกรณีของค่า ρ , n และ p พบว่า ค่ามัธยฐานของค่า MSE ของวิธีริตจมีค่าต่ำที่สุด ต่อมาคือวิธี LASSO เมื่อเปรียบเทียบกับวิธี Adaptive LASSO และ Relaxed LASSO ในกรณีที่ $\rho = 0.1, 0.3,$ และ 0.5 วิธี Relaxed LASSO มีค่ามัธยฐานของค่า MSE ต่ำกว่าวิธี Adaptive LASSO ทุกกรณี ยกเว้นกรณีที่ $\rho = 0.1, n = 50$ และ $p = 200, \rho = 0.3, n = 50$ และ $p = 100$ และ $\rho = 0.5, n = 50$ และ $p = 200$ ส่วนในกรณีที่ $\rho = 0.9$ ในทุกกรณีของค่า n และ p วิธี Adaptive LASSO มีค่ามัธยฐานของค่า MSE ต่ำกว่าวิธี Relaxed LASSO

ความถูกต้องในการคัดเลือกตัวแปรอิสระ

ความถูกต้องในการคัดเลือกตัวแปรอิสระ วัดจากจำนวนของตัวแปรอิสระที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอิสระวิธี LASSO, Adaptive LASSO และ Relaxed LASSO โดยเฉลี่ยจากการจำลอง 500 รอบ, ค่าร้อยละของ IC1 และ IC2 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ และจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยจากการจำลอง 500 รอบ ซึ่งผลการวิจัยแบ่งออกเป็น 4 กรณี คือ 1) ตัวแบบถดถอยพหุคูณ partial orthogonal sparse model 2) ตัวแบบถดถอยพหุคูณ non-orthogonal sparse model 3) ตัวแบบถดถอยพหุคูณ grouping structure sparse model และ 4) ตัวแบบถดถอยพหุคูณ non-orthogonal and non-sparse model ซึ่งมีรายละเอียดดังนี้

กรณีที่ 1 ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระเป็นเชิงตั้งฉากบางส่วน
 ตารางที่ 9 ค่าเฉลี่ยของ IC1 (Mean_IC1) และค่าร้อยละของ IC1 เมื่อเทียบกับจำนวนตัวแปรอิสระ
 ในตัวแบบ (%IC1) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ
 Relaxed LASSO กรณี partial orthogonal sparse model

ρ	n	p	LASSO		Adaptive LASSO		Relaxed LASSO	
			Mean_IC1	%IC1	Mean_IC1	%IC1	Mean_IC1	%IC1
0.1	50	100	2.000	13.333	2.992	19.947	3.262	211.747
		200	3.876	25.840	4.990	33.267	5.442	36.280
	100	200	0.492	3.280	1.394	9.293	2.370	15.800
		400	1.074	7.160	2.368	15.787	3.032	20.213
	200	400	0.014	0.093	0.714	4.760	0.914	6.093
800	0.054	0.360	1.162	7.747	1.442	9.613		
0.3	50	100	1.334	8.893	2.090	13.933	2.552	17.013
		200	2.422	16.147	3.332	22.213	3.766	25.107
	100	200	0.266	1.773	0.980	6.533	1.542	10.280
		400	0.508	3.387	1.632	10.880	2.254	15.027
	200	400	0.012	0.080	0.456	3.040	0.428	2.853
800	0.020	0.133	0.656	4.373	0.642	4.280		
0.5	50	100	1.050	7.000	1.624	10.827	2.142	14.280
		200	1.504	10.027	2.286	15.240	2.714	18.093
	100	200	0.218	1.453	0.682	4.547	1.178	7.853
		400	0.330	2.200	1.044	6.960	1.490	9.933
	200	400	0.010	0.067	0.240	1.600	0.358	2.387
800	0.028	0.187	0.386	2.573	0.438	2.920		
0.7	50	100	1.138	7.587	1.272	8.480	1.998	13.320
		200	1.224	8.160	1.534	10.227	2.270	15.133
	100	200	0.336	2.240	0.420	2.800	1.068	7.120
		400	0.382	2.547	0.534	3.560	1.278	8.520
	200	400	0.050	0.333	0.046	0.307	0.342	2.280
800	0.062	0.413	0.104	0.693	0.444	2.960		
0.9	50	100	2.000	13.333	1.184	7.893	2.436	16.240
		200	1.890	12.600	1.602	10.680	2.610	17.400
	100	200	1.030	6.867	0.578	3.853	1.520	10.133
		400	1.068	7.120	0.844	5.627	1.586	10.573
	200	400	0.452	3.013	0.276	1.840	0.814	5.427
800	0.484	3.227	0.370	2.467	0.926	6.173		

จากตารางที่ 9 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ partial orthogonal sparse model ในทุกกรณีของค่า ρ , n และ p พบว่า จำนวนเฉลี่ยของตัวแปรอิสระที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอิสระ ซึ่งเมื่อกำหนดให้พารามิเตอร์มีค่าไม่เท่ากับศูนย์ แต่ตัวประมาณสัมประสิทธิ์การถดถอยที่ได้มีค่าเป็นศูนย์ (IC1) และค่าร้อยละของ IC1 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบของวิธี LASSO มีค่าต่ำที่สุด ยกเว้นกรณีที่ $\rho = 0.7$, $n = 200$ และ $p = 400$ และ $\rho = 0.9$ และจำนวนเฉลี่ยของตัวแปรอิสระและค่าร้อยละของการเกิด IC1 ของวิธี Relaxed LASSO มีค่าสูงสุดทุกกรณี



ตารางที่ 10 ค่าเฉลี่ยของ IC2 (Mean_IC2) และค่าร้อยละของ IC2 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (%IC2) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี partial orthogonal sparse model

ρ	n	p	LASSO		Adaptive LASSO		Relaxed LASSO	
			Mean_IC2	%IC2	Mean_IC2	%IC2	Mean_IC2	%IC2
0.1	50	100	25.644	30.169	23.708	27.892	12.960	15.247
		200	26.768	14.469	22.384	12.099	13.652	7.379
	100	200	48.414	26.170	44.900	24.270	7.576	4.095
		400	51.144	13.284	42.866	11.134	9.566	2.485
	200	400	81.462	21.159	67.662	17.575	4.406	1.144
		800	83.436	10.629	58.188	7.412	4.618	0.588
0.3	50	100	24.474	28.793	23.280	27.388	9.072	10.673
		200	25.896	13.998	23.340	12.616	10.896	5.890
	100	200	45.180	24.422	40.964	22.143	4.684	2.532
		400	47.634	12.372	39.034	10.139	4.754	1.235
	200	400	76.536	19.879	62.936	16.347	1.798	0.467
		800	72.882	9.284	54.992	7.005	1.704	0.217
0.5	50	100	22.608	26.598	21.778	25.621	5.456	6.419
		200	23.704	12.813	21.064	11.386	6.484	3.505
	100	200	42.760	23.114	38.278	20.691	2.344	1.267
		400	40.900	10.623	35.150	9.130	2.370	0.616
	200	400	74.742	19.414	60.152	15.624	0.534	0.139
		800	66.658	8.491	51.172	6.519	0.610	0.078
0.7	50	100	21.574	25.381	20.654	24.299	2.806	3.301
		200	20.768	11.226	19.018	10.280	2.710	1.465
	100	200	42.256	22.841	35.564	19.224	1.016	0.549
		400	36.792	9.556	31.620	8.213	0.986	0.256
	200	400	77.508	20.132	56.832	14.762	0.452	0.117
		800	64.480	8.214	46.612	5.938	0.552	0.070
0.9	50	100	21.826	25.678	18.308	21.539	2.126	2.501
		200	20.062	10.844	17.966	9.711	1.724	0.932
	100	200	42.356	22.895	29.256	15.814	1.318	0.712
		400	39.260	10.197	28.200	7.325	1.392	0.362
	200	400	77.262	20.068	45.206	11.742	0.840	0.218
		800	71.434	9.100	39.416	5.021	0.956	0.122

จากตารางที่ 10 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ partial orthogonal sparse model ในทุกกรณีของค่า ρ , n และ p พบว่า จำนวนเฉลี่ยของตัวแปรอิสระที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอิสระ ซึ่งเมื่อกำหนดให้พารามิเตอร์มีค่าเท่ากับศูนย์ แต่ตัวประมาณสัมประสิทธิ์การถดถอยที่ได้มีค่าไม่เป็นศูนย์ (IC2) และค่าร้อยละของ IC2 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบของวิธี Relaxed LASSO มีค่าต่ำที่สุดในทุกๆ กรณี และจำนวนเฉลี่ยของตัวแปรอิสระและค่าร้อยละของการเกิด IC2 ของวิธี LASSO มีค่าสูงสุดทุกกรณี



ตารางที่ 11 ค่าเฉลี่ยของจำนวนตัวแปรอิสระในสมการถดถอย (Mean_covariate) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี partial orthogonal sparse model

ρ	n	p	Mean_Covariate		
			LASSO	Adaptive	Relaxed
0.1	50	100	38.644	35.716	24.698
		200	37.892	32.394	23.210
	100	200	62.922	58.506	20.206
		400	65.070	55.498	21.534
	200	400	96.448	81.948	18.492
		800	98.382	72.026	18.176
0.3	50	100	38.140	36.190	21.520
		200	38.474	35.008	22.130
	100	200	59.914	54.984	18.142
		400	62.126	52.402	17.500
	200	400	91.524	77.480	16.370
		800	87.862	69.336	16.062
0.5	50	100	36.558	35.154	18.314
		200	37.200	33.778	18.770
	100	200	57.542	52.596	16.166
		400	55.570	49.106	15.880
	200	400	89.732	74.912	15.176
		800	81.630	65.786	15.172
0.7	50	100	35.436	34.382	15.808
		200	34.544	32.484	15.440
	100	200	56.920	50.144	14.948
		400	51.410	46.086	14.708
	200	400	92.458	71.786	15.110
		800	79.418	61.508	15.108
0.9	50	100	34.826	32.124	14.690
		200	33.172	31.364	14.114
	100	200	56.326	43.678	14.798
		400	53.192	42.356	14.806
	200	400	91.810	59.930	15.026
		800	85.950	54.046	15.030

จากตารางที่ 11 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ partial orthogonal sparse model ในทุกกรณีของค่า ρ , n และ p พบว่า จำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยของวิธี Relaxed LASSO มีค่าน้อยที่สุดซึ่งใกล้เคียงกับจำนวนตัวแปรอิสระที่กำหนดในตัวแบบจริงเท่ากับ 15 ตัว เมื่อเปรียบเทียบวิธี LASSO และวิธี Adaptive LASSO พบว่าวิธี LASSO มีจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยมากกว่าวิธี Adaptive LASSO และทั้งสองวิธีมีจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยจำนวนมาก



กรณีที่ 2 ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก

ตารางที่ 12 ค่าเฉลี่ยของ IC1 (Mean_IC1) และค่าร้อยละของ IC1 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (%IC1) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal sparse model

ρ	n	p	LASSO		Adaptive LASSO		Relaxed LASSO	
			Mean_IC1	%IC1	Mean_IC1	%IC1	Mean_IC1	%IC1
0.1	50	100	2.084	13.893	3.184	21.227	3.514	23.427
		200	3.932	26.213	4.884	32.560	5.510	36.733
	100	200	0.470	3.133	1.434	9.560	2.422	16.147
		400	1.036	6.907	2.388	15.920	3.078	20.520
	200	400	0.006	0.040	0.744	4.960	0.872	5.813
		800	0.050	0.333	1.158	7.720	1.352	9.013
0.3	50	100	1.406	9.373	2.106	14.040	2.640	17.600
		200	2.358	15.720	3.294	21.960	3.812	25.413
	100	200	0.234	1.560	0.976	6.507	1.552	10.347
		400	0.508	3.387	1.522	10.147	2.092	13.947
	200	400	0.008	0.053	0.420	2.800	0.390	2.600
		800	0.008	0.053	0.672	4.480	0.622	4.147
0.5	50	100	1.082	7.213	1.560	10.400	2.134	14.227
		200	1.554	10.360	2.298	15.320	2.756	18.373
	100	200	0.252	1.680	0.760	5.067	1.224	8.160
		400	0.328	2.187	1.010	6.733	1.426	9.507
	200	400	0.026	0.173	0.214	1.427	0.330	2.200
		800	0.018	0.120	0.354	2.360	0.454	3.027
0.7	50	100	1.140	7.600	1.310	8.733	2.022	13.480
		200	1.366	9.107	1.670	11.133	2.384	15.893
	100	200	0.438	2.920	0.508	3.387	1.216	8.107
		400	0.456	3.040	0.658	4.387	1.312	8.747
	200	400	0.060	0.400	0.058	0.387	0.338	2.253
		800	0.062	0.413	0.136	0.907	0.410	2.733
0.9	50	100	2.248	14.987	1.352	9.013	2.572	17.147
		200	2.092	13.947	1.714	11.427	2.752	18.347
	100	200	1.252	8.347	0.708	4.720	1.552	10.347
		400	1.156	7.707	0.960	6.400	1.656	11.040
	200	400	0.612	4.080	0.344	2.293	0.864	5.760
		800	0.542	3.613	0.388	2.587	0.898	5.987

จากตารางที่ 12 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ non-orthogonal sparse model ในทุกกรณีของค่า ρ , n และ p พบว่า จำนวนเฉลี่ยของตัวแปรอิสระที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอิสระ ซึ่งเมื่อกำหนดให้พารามิเตอร์มีค่าไม่เท่ากับศูนย์ แต่ตัวประมาณสัมประสิทธิ์การถดถอยที่ได้มีค่าเป็นศูนย์ (IC1) และค่าร้อยละของ IC1 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบของวิธี LASSO มีค่าต่ำที่สุด ยกเว้นกรณีที่ $\rho = 0.7$, $n = 200$ และ $p = 400$ และ $\rho = 0.9$ และจำนวนเฉลี่ยของตัวแปรอิสระและค่าร้อยละของการเกิด IC1 ของวิธี Relaxed LASSO มีค่าสูงสุดทุกกรณี



ตารางที่ 13 ค่าเฉลี่ยของ IC2 (Mean_IC2) และค่าร้อยละของ IC2 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (%IC2) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal sparse model

ρ	n	p	LASSO		Adaptive LASSO		Relaxed LASSO	
			Mean_IC2	%IC2	Mean_IC2	%IC2	Mean_IC2	%IC2
0.1	50	100	25.584	30.099	22.638	26.633	12.350	14.529
		200	25.716	13.901	22.640	12.238	12.976	7.014
	100	200	48.824	26.391	44.704	24.164	7.718	4.172
		400	51.272	13.317	42.542	11.050	10.046	2.609
	200	400	82.166	21.342	66.174	17.188	4.556	1.183
		800	81.686	10.406	59.290	7.553	5.624	0.716
0.3	50	100	24.390	28.694	23.650	27.824	9.128	10.739
		200	25.762	13.925	22.670	12.254	10.496	5.674
	100	200	45.150	24.405	40.850	22.081	4.678	2.529
		400	46.814	12.159	37.852	9.832	5.610	1.457
	200	400	76.504	19.871	62.580	16.255	1.862	0.484
		800	73.610	9.377	54.402	6.930	2.036	0.259
0.5	50	100	22.666	26.666	22.020	25.906	5.552	6.532
		200	23.432	12.666	20.960	11.330	6.582	3.558
	100	200	42.358	22.896	39.326	21.257	2.234	1.208
		400	41.170	10.694	34.290	8.906	2.512	0.652
	200	400	75.552	19.624	59.542	15.465	0.696	0.181
		800	65.344	8.324	50.488	6.432	0.536	0.068
0.7	50	100	21.968	25.845	20.770	24.435	2.914	3.428
		200	20.630	11.151	19.288	10.426	3.514	1.899
	100	200	42.604	23.029	35.372	19.120	1.418	0.766
		400	37.662	9.782	32.136	8.347	1.356	0.352
	200	400	77.552	20.143	56.996	14.804	0.576	0.150
		800	64.054	8.160	46.680	5.946	0.884	0.113
0.9	50	100	22.380	26.329	18.666	21.960	2.472	2.908
		200	20.216	10.928	18.032	9.747	2.186	1.182
	100	200	42.498	22.972	31.062	16.790	1.534	0.829
		400	39.118	10.161	28.792	7.478	1.584	0.411
	200	400	76.818	19.953	0.344	0.089	1.380	0.358
		800	71.756	9.141	41.030	5.227	1.066	0.136

จากตารางที่ 13 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ non-orthogonal sparse model ในทุกกรณีของค่า ρ , n และ p พบว่า จำนวนเฉลี่ยของตัวแปรอิสระที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอิสระ ซึ่งเมื่อกำหนดให้พารามิเตอร์มีค่าเท่ากับศูนย์ แต่ตัวประมาณสัมประสิทธิ์การถดถอยที่ได้มีค่าไม่เป็นศูนย์ (IC2) และค่าร้อยละของ IC2 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบของวิธี Relaxed LASSO มีค่าต่ำที่สุดยกเว้นกรณีที่ $\rho = 0.9$, $n = 200$ และ $p = 400$ และจำนวนเฉลี่ยของตัวแปรอิสระและค่าร้อยละของการเกิด IC2 ของวิธี LASSO มีค่าสูงสุดทุกกรณี



ตารางที่ 14 ค่าเฉลี่ยของจำนวนตัวแปรอิสระในสมการถดถอย (Mean_Covariate) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal sparse model

ρ	n	p	Mean_Covariate		
			LASSO	Adaptive	Relaxed
0.1	50	100	38.500	34.454	23.836
		200	36.784	32.756	22.466
	100	200	63.354	58.270	20.296
		400	65.236	55.154	21.968
	200	400	97.160	80.430	18.684
		800	96.636	73.132	19.272
0.3	50	100	37.984	36.544	21.488
		200	38.404	34.376	21.684
	100	200	59.916	54.874	18.126
		400	61.306	51.330	18.518
	200	400	91.496	77.160	16.472
		800	88.602	68.730	16.414
0.5	50	100	36.584	35.460	18.418
		200	36.878	33.662	18.826
	100	200	57.106	53.566	16.010
		400	55.842	48.280	16.086
	200	400	90.526	74.328	15.366
		800	80.326	65.134	15.082
0.7	50	100	35.828	34.460	15.892
		200	34.264	32.618	16.130
	100	200	57.166	49.864	15.202
		400	52.206	46.478	15.044
	200	400	92.492	71.938	15.238
		800	78.992	61.544	15.474
0.9	50	100	35.132	32.314	14.900
		200	33.124	31.318	14.434
	100	200	56.246	45.354	14.982
		400	52.962	42.832	14.928
	200	400	91.206	15.000	15.516
		800	86.214	55.642	15.168

จากตารางที่ 14 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ non-orthogonal sparse model ในทุกกรณีของค่า p , n และ p พบว่า จำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยของวิธี Relaxed LASSO มีค่าน้อยที่สุดซึ่งใกล้เคียงกับ 15 มากที่สุด เมื่อเปรียบเทียบกับวิธี LASSO และวิธี Adaptive LASSO พบว่าวิธี LASSO มีจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยมากกว่าวิธี Adaptive LASSO และทั้งสองวิธีมีจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยจำนวนมาก

กรณีที่ 3 ตัวแบบถดถอยบางเบาที่ความสัมพันธ์ของตัวแปรอิสระมีโครงสร้างแบบแบ่งกลุ่ม

ตารางที่ 15 ค่าเฉลี่ยของ IC1 (Mean_IC1) และค่าร้อยละของ IC1 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (%IC1) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี grouping structure sparse model

n	p	LASSO		Adaptive LASSO		Relaxed LASSO	
		Mean_IC1	%IC1	Mean_IC1	%IC1	Mean_IC1	%IC1
50	100	11.072	73.813	11.116	74.107	11.264	75.093
	200	11.066	73.773	11.078	73.853	11.294	75.293
100	200	10.556	70.373	10.562	70.413	11.016	73.440
	400	10.542	70.280	10.46	69.733	11.002	73.347
200	400	9.834	65.560	9.792	65.280	10.7	71.333
	800	9.762	65.080	9.758	65.053	10.752	71.680

จากตารางที่ 15 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ grouping structure sparse model เมื่อ $n = 50$ ในทุกๆ p และ $n = 100$, $p = 200$ พบว่า จำนวนเฉลี่ยของตัวแปรอิสระที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอิสระ ซึ่งเมื่อกำหนดให้พารามิเตอร์มีค่าไม่เท่ากับศูนย์ แต่ตัวประมาณสัมประสิทธิ์การถดถอยที่ได้มีค่าเป็นศูนย์ (IC1) และค่าร้อยละของ IC1 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบของวิธี LASSO มีค่าต่ำที่สุด แต่เมื่อ $n = 100$, $p = 400$ และ $n = 200$ ทุก p พบว่า วิธี Adaptive LASSO มีจำนวนเฉลี่ยของตัวแปรอิสระและค่าร้อยละของการเกิด IC1 มีค่าต่ำที่สุด และในทุกกรณีของค่า n และ p พบว่า จำนวนเฉลี่ยของตัวแปรอิสระและค่าร้อยละของการเกิด IC1 ของวิธี Relaxed LASSO มีค่าสูงสุด

ตารางที่ 16 ค่าเฉลี่ยของ IC2 (Mean_IC2) และค่าร้อยละของ IC2 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (%IC2) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี grouping structure sparse model

n	p	LASSO		Adaptive LASSO		Relaxed LASSO	
		Mean_IC2	%IC2	Mean_IC2	%IC2	Mean_IC2	%IC2
50	100	24.504	28.828	25.926	30.501	1.028	1.209
	200	22.884	12.370	21.704	11.732	0.894	0.483
100	200	41.214	22.278	38.858	21.004	0.554	0.299
	400	36.562	9.497	32.584	8.463	0.56	0.145
200	400	69.52	18.057	53.35	13.857	0.576	0.150
	800	58.61	7.466	46.754	5.956	0.392	0.050

จากตารางที่ 16 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ grouping structure sparse model ในทุกกรณีของค่า n และ p พบว่า จำนวนเฉลี่ยของตัวแปรอิสระที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอิสระ ซึ่งเมื่อกำหนดให้พารามิเตอร์มีค่าเท่ากับศูนย์ แต่ตัวประมาณสัมประสิทธิ์การถดถอยที่ได้มีค่าไม่เป็นศูนย์ (IC2) และค่าร้อยละของ IC2 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบของวิธี Relaxed LASSO มีค่าต่ำที่สุดทุกกรณี และจำนวนเฉลี่ยของตัวแปรอิสระและค่าร้อยละของการเกิด IC2 ของวิธี LASSO มีค่าสูงสุด ยกเว้นกรณีที่ n = 50 และ p = 100

ตารางที่ 17 ค่าเฉลี่ยของจำนวนตัวแปรอิสระในสมการถดถอย (Mean_Covariate) ของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยแบบ LASSO, Adaptive LASSO และ Relaxed LASSO กรณี grouping structure sparse model

n	p	Mean_Covariate		
		LASSO	Adaptive LASSO	Relaxed LASSO
50	100	28.432	29.810	4.764
	200	26.818	25.626	4.600
100	200	45.658	43.296	4.538
	400	41.020	37.124	4.558
200	400	74.686	58.558	4.876
	800	63.848	51.996	4.640

จากตารางที่ 17 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ grouping structure sparse model ในทุกกรณีของค่า n และ p พบว่า จำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยของวิธี Relaxed LASSO มีค่าน้อยที่สุดซึ่งน้อยกว่าจำนวนตัวแปรอิสระในตัวแบบจริงที่กำหนดเท่ากับ 15 ตัว ส่วนวิธี LASSO และวิธี Adaptive LASSO มีจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยมากกว่าที่กำหนด



กรณีศึกษาที่ 4 ตัวแบบถดถอยพหุคูณ non-orthogonal and non-sparse model

ตารางที่ 18 ค่าเฉลี่ยของ IC1 (Mean_IC1) และค่าร้อยละของ IC1 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (%IC1) ของการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal and non-sparse model

ρ	n	p	LASSO		Adaptive LASSO		Relaxed LASSO	
			Mean_IC1	%IC1	Mean_IC1	%IC1	Mean_IC1	%IC1
0.1	50	100	73.842	73.842	76.404	76.404	80.922	80.922
		200	182.076	91.038	181.492	90.746	188.076	94.038
	100	200	138.178	69.089	146.090	73.045	152.786	76.393
		400	368.618	92.155	371.566	92.892	378.500	94.625
	200	400	269.506	67.377	288.808	72.202	295.072	73.768
		800	0.250	0.031	744.140	93.018	757.756	94.720
0.3	50	100	64.426	64.426	67.262	67.262	73.238	73.238
		200	175.514	87.757	175.472	87.736	181.964	90.982
	100	200	118.004	59.002	127.516	63.758	134.328	67.164
		400	347.668	86.917	347.622	86.906	359.648	89.912
	200	400	221.014	55.254	244.392	61.098	246.386	61.597
		800	689.374	86.172	697.210	87.151	711.786	88.973
0.5	50	100	55.132	55.132	58.764	58.764	65.976	65.976
		200	164.860	82.430	164.978	82.489	173.824	86.912
	100	200	105.620	52.810	113.214	56.607	120.768	60.384
		400	322.554	80.639	324.950	81.238	339.292	84.823
	200	400	204.520	51.130	219.980	54.995	227.904	56.976
		800	629.478	78.685	635.122	79.390	657.240	82.155
0.7	50	100	51.592	51.592	53.882	53.882	61.566	61.566
		200	154.384	77.192	156.514	78.257	165.792	82.896
	100	200	101.294	50.647	104.250	52.125	115.258	57.629
		400	303.888	75.972	308.170	77.043	320.242	80.061
	200	400	201.072	50.268	205.694	51.424	222.118	55.530
		800	603.518	75.440	610.180	76.273	628.294	78.537
0.9	50	100	51.098	51.098	51.300	51.300	59.640	59.640
		200	151.106	75.553	151.544	75.772	161.186	80.593
	100	200	101.000	50.500	101.064	50.532	114.530	57.265
		400	301.000	75.250	301.230	75.308	315.588	78.897
	200	400	201.000	50.250	201.000	50.250	221.608	55.402
		800	601.000	75.125	601.090	75.136	622.524	77.816

จากตารางที่ 18 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ non-orthogonal and non-sparse model ในทุกกรณีของค่า ρ , n และ p พบว่า จำนวนเฉลี่ยของตัวแปรอิสระที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอิสระ ซึ่งเมื่อกำหนดให้พารามิเตอร์มีค่าไม่เท่ากับศูนย์ แต่ตัวประมาณสัมประสิทธิ์การถดถอยที่ได้มีค่าเป็นศูนย์ (IC1) และค่าร้อยละของ IC1 เมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบของวิธี LASSO มีค่าต่ำที่สุด ยกเว้นกรณีที่ $\rho = 0.7$, $n = 50$, $p = 200$ และกรณีที่ $\rho = 0.9$, $n = 50$, $p = 200$ และ $n = 100$, $p = 400$ และจำนวนเฉลี่ยของตัวแปรอิสระและค่าร้อยละของการเกิด IC1 ของวิธี Relaxed LASSO มีค่าสูงสุดทุกกรณี



ตารางที่ 19 ค่าเฉลี่ยของจำนวนตัวแปรอิสระในสมการถดถอย (Mean_Covariate) ของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยแบบ LASSO, Adaptive LASSO และ Relaxed LASSO กรณี non-orthogonal and non-sparse model

ρ	n	p	Mean_Covariate		
			LASSO	Adaptive	Relaxed
0.1	50	100	26.158	23.596	19.078
		200	17.924	18.508	11.924
	100	200	61.822	53.910	47.214
		400	31.382	28.434	21.500
	200	400	130.494	111.192	104.928
		800	799.750	55.860	42.244
0.3	50	100	35.574	32.738	26.762
		200	24.486	24.528	18.036
	100	200	81.996	72.484	65.672
		400	52.332	52.378	40.352
	200	400	178.986	155.608	153.614
		800	110.626	102.790	88.214
0.5	50	100	44.868	41.236	34.024
		200	35.140	35.022	26.176
	100	200	94.380	86.786	79.232
		400	77.446	75.050	60.708
	200	400	195.480	180.020	172.096
		800	170.522	164.878	142.760
0.7	50	100	48.408	46.118	38.434
		200	45.616	43.486	34.208
	100	200	98.706	95.750	84.742
		400	96.112	91.830	79.758
	200	400	198.928	194.306	177.882
		800	196.482	189.820	171.706
0.9	50	100	48.902	48.700	40.360
		200	48.894	48.456	38.814
	100	200	99.000	98.936	85.470
		400	99.000	98.770	84.412
	200	400	199.000	199.000	178.392
		800	199.000	198.910	177.476

จากตารางที่ 19 จะเห็นได้ว่า เมื่อตัวแบบถดถอยพหุคูณ non-orthogonal and non-sparse model เมื่อกำหนดให้ตัวแปรอิสระทุกตัวอยู่ในตัวแบบ ในทุกกรณีของค่า ρ , n และ p พบว่า วิธี Relaxed LASSO มีจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยน้อยที่สุด ต่อมาคือวิธี LASSO และวิธี Adaptive LASSO มีจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยมากที่สุด ซึ่งทั้งสามวิธีมีจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยน้อยกว่าจำนวนตัวแปรอิสระที่กำหนด



บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่องานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสัมประสิทธิ์การถดถอย 4 วิธี ในกรณีที่ข้อมูลมีมิติสูง ได้แก่ วิธีริดจ์, LASSO, Adaptive LASSO และ Relaxed LASSO โดยพิจารณาจากความถูกต้องในการพยากรณ์ ความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอย และความถูกต้องในการคัดเลือกตัวแปรอิสระเข้าในตัวแบบ ภายใต้ขอบเขตงานวิจัย โดยแบ่งเป็น 4 กรณี คือ กรณีที่ 1 คือ ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระเป็นเชิงตั้งฉากบางส่วน และกำหนดให้ตัวแปรอิสระอยู่ในตัวแบบจำนวน 15 ตัว กรณีที่ 2 คือ ตัวแบบถดถอยบางเบาที่เมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก และกำหนดให้ตัวแปรอิสระอยู่ในตัวแบบจำนวน 15 ตัว กรณีที่ 3 คือ ตัวแบบถดถอยบางเบาที่ความสัมพันธ์ของตัวแปรอิสระมีโครงสร้างแบบแบ่งกลุ่ม และกำหนดให้ตัวแปรอิสระอยู่ในตัวแบบจำนวน 15 ตัว และกรณีที่ 4 คือ ตัวแบบถดถอยไม่บางเบาและเมทริกซ์สหสัมพันธ์ของตัวแปรอิสระไม่เป็นเชิงตั้งฉาก และกำหนดให้ตัวแปรอิสระทุกตัวอยู่ในตัวแบบ โดยแต่ละกรณีจะทำการจำลองข้อมูลที่มีขนาดตัวอย่าง เท่ากับ 50, 100 และ 200 และจำนวนตัวแปรอิสระเป็น 2 เท่า และ 4 เท่าของขนาดตัวอย่าง

สรุปผลการวิจัย

จากผลการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 4 วิธี ภายใต้ขอบเขตของงานวิจัย สามารถสรุปผลการวิจัยแบ่งเป็น 3 ประเด็น ดังนี้

1. ประสิทธิภาพของการพยากรณ์ โดยใช้ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Med_PMSE) เป็นเกณฑ์ในการเปรียบเทียบ สามารถสรุปแยกตามรูปแบบความสัมพันธ์ของตัวแปรอิสระ ได้ดังนี้

กรณีตัวแบบถดถอยพหุคูณ partial orthogonal sparse model, non-orthogonal sparse model และ grouping structure sparse model พบว่า วิธี Relaxed LASSO มีประสิทธิภาพของการพยากรณ์สูงสุด และวิธีริดจ์ต่ำที่สุด เมื่อเปรียบเทียบวิธี LASSO และ Adaptive LASSO พบว่า วิธี LASSO มีประสิทธิภาพของการพยากรณ์สูงกว่าวิธี Adaptive LASSO เมื่อตัวแปรอิสระมีความสัมพันธ์กันต่ำ ส่วนวิธี Adaptive LASSO มีประสิทธิภาพของการพยากรณ์สูงกว่าวิธี LASSO เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูง เมื่อพิจารณา ความสัมพันธ์ระหว่างขนาดตัวอย่าง, จำนวนตัวแปรอิสระ, และค่า p กับค่ามัธยฐานของค่า PMSE พบว่า เมื่อขนาดตัวอย่าง

มากขึ้นและกำหนดให้จำนวนตัวแปรอิสระและค่า ρ คงที่ พบว่า ค่ามัธยฐานของ PMSE ของทุกวิธีลดลง และเมื่อค่า ρ มีค่าสูงขึ้นและกำหนดให้ขนาดตัวอย่างและจำนวนตัวแปรอิสระคงที่ พบว่า ค่ามัธยฐานของ PMSE ของทุกวิธีลดลง

กรณีทีตัวแบบถดถอยพหุคูณ non-orthogonal and non-sparse model พบว่า วิธีริตจมีประสิทธิภาพของการพยากรณ์สูงสุด รองลงมาคือวิธี LASSO เมื่อเปรียบเทียบวิธี Adaptive LASSO และ Relaxed LASSO พบว่า วิธี Adaptive LASSO มีประสิทธิภาพของการพยากรณ์สูงกว่าวิธี Relaxed LASSO เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูง เมื่อพิจารณา ความสัมพันธ์ระหว่างขนาดตัวอย่าง, จำนวนตัวแปรอิสระ, และค่า ρ กับค่ามัธยฐานของค่า PMSE พบว่า เมื่อขนาดตัวอย่างสูงขึ้นและกำหนดจำนวนตัวแปรอิสระและค่า ρ คงที่และเมื่อจำนวนตัวแปรอิสระมากขึ้นและค่า ρ คงที่ พบว่า ค่ามัธยฐานของ PMSE ของทุกวิธีเพิ่มขึ้น

2. ประสิทธิภาพของการประมาณ โดยใช้ค่ามัธยฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์การถดถอย (Median MSE) เป็นเกณฑ์ในการเปรียบเทียบ สามารถสรุปแยกตามรูปแบบความสัมพันธ์ของตัวแปรอิสระ ได้ดังนี้

กรณีตัวแบบถดถอยพหุคูณ partial orthogonal sparse model และ non-orthogonal sparse model พบว่า วิธี Relaxed LASSO มีประสิทธิภาพของการประมาณสูงสุด ตามด้วยวิธี LASSO, Adaptive LASSO และวิธีริตจมีประสิทธิภาพของการประมาณต่ำที่สุด เมื่อพิจารณา ความสัมพันธ์ระหว่างขนาดตัวอย่าง, จำนวนตัวแปรอิสระ, และค่า ρ กับค่ามัธยฐานของค่า MSE พบว่า เมื่อขนาดตัวอย่างสูงขึ้นและกำหนดจำนวนตัวแปรอิสระค่า ρ คงที่ และเมื่อจำนวนตัวแปรอิสระเพิ่มมากขึ้นและกำหนดขนาดตัวอย่างและค่า ρ คงที่ พบว่า ค่ามัธยฐานของ MSE ของทุกวิธีลดลง และเมื่อค่า ρ มีค่าสูงขึ้นและขนาดตัวอย่างและจำนวนตัวแปรอิสระคงที่ ค่ามัธยฐานของ MSE ของทุกวิธีลดลง ยกเว้นตัวแปรอิสระมีความสัมพันธ์กันสูง

กรณีตัวแบบถดถอยพหุคูณ grouping structure sparse model และ non-orthogonal and non-sparse model พบว่า วิธีริตจมีประสิทธิภาพของการประมาณสูงสุด ตามด้วยวิธี LASSO เมื่อเปรียบเทียบวิธี Adaptive LASSO และ Relaxed LASSO พบว่า วิธี Relaxed LASSO มีประสิทธิภาพของการประมาณสูงกว่าวิธี Adaptive LASSO เมื่อตัวแปรอิสระมีความสัมพันธ์กันต่ำ ส่วนวิธี Adaptive LASSO มีประสิทธิภาพของการประมาณสูงกว่าวิธี Relaxed LASSO เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูง เมื่อพิจารณา ความสัมพันธ์ระหว่างขนาดตัวอย่าง, จำนวนตัวแปรอิสระ, และค่า ρ กับค่ามัธยฐานของค่า MSE พบว่า เมื่อขนาดตัวอย่างเพิ่มมากขึ้นและกำหนดจำนวนตัวแปรอิสระและค่า ρ คงที่ ค่ามัธยฐานของ MSE ของทุกวิธีลดลง แต่เมื่อจำนวนตัวแปรอิสระเพิ่มขึ้นและกำหนดขนาดตัวอย่างและค่า ρ คงที่ พบว่า ค่ามัธยฐานของ MSE

ของทุกวิธีเพิ่มขึ้น และเมื่อค่า ρ มีค่าสูงขึ้น ค่ามัธยฐานของ MSE ของทุกวิธีลดลง ยกเว้นที่ตัวแปรอิสระมีความสัมพันธ์กันสูง

3. ความผิดพลาดของการคัดเลือกตัวแปรอิสระเข้าสู่ตัวแบบ โดยใช้ค่าเฉลี่ยของจำนวนของตัวแปรอิสระที่เกิดความผิดพลาดของการคัดเลือกตัวแปรอิสระ (Mean_IC) ค่าร้อยละของความผิดพลาดของการคัดเลือกตัวแปรอิสระเมื่อเทียบกับจำนวนตัวแปรอิสระในตัวแบบ (%IC) และจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ย (Mean_Covariate) เป็นเกณฑ์ในการเปรียบเทียบสามารถสรุปแยกตามรูปแบบความสัมพันธ์ของตัวแปรอิสระ ได้ดังนี้

กรณีตัวแบบถดถอยพหุคูณ partial orthogonal sparse model, non-orthogonal sparse model และ grouping structure sparse model เนื่องจาก ตัวประมาณสัมประสิทธิ์การถดถอยวิธีริตจเป็นตัวประมาณที่ให้ค่าประมาณทุกตัวมีค่าเข้าใกล้ศูนย์แต่ไม่เท่ากับศูนย์ จึงทำให้ความผิดพลาดของการคัดเลือกตัวแปรอิสระแบบที่ 1 คือ เมื่อกำหนดให้ตัวแปรอิสระอยู่ในตัวแบบ แต่ตัวแปรอิสระนั้นไม่ถูกคัดเลือกเข้าสู่ตัวแบบ (IC1) มีความผิดพลาด 0% หรือไม่มีความผิดพลาดของการคัดเลือกตัวแปรอิสระ แต่เกิดความผิดพลาดของการคัดเลือกตัวแปรอิสระแบบที่ 2 คือ กำหนดให้ตัวแปรอิสระไม่อยู่ในตัวแบบ แต่ตัวแปรอิสระนั้นถูกคัดเลือกเข้าสู่ตัวแบบ (IC2) มีความผิดพลาด 100% ดังนั้น จึงพิจารณาความผิดพลาดของการคัดเลือกตัวแปรอิสระเฉพาะวิธี LASSO, Adaptive LASSO และ Relaxed LASSO เท่านั้น เมื่อพิจารณาความผิดพลาดแบบ IC1 พบว่า ทั้งสามวิธี คือ LASSO, Adaptive LASSO และ Relaxed LASSO มีความผิดพลาดแบบ IC1 ใกล้เคียงกัน แต่เมื่อพิจารณาความผิดพลาดแบบ IC2 พบว่า วิธี Relaxed LASSO มีความผิดพลาดน้อยที่สุด แต่อย่างไรก็ตาม เมื่อพิจารณาจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ย พบว่า วิธี Relaxed LASSO มีจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยใกล้เคียงกับตัวแบบจริงมากที่สุด เมื่อเปรียบเทียบวิธี LASSO และวิธี Adaptive LASSO พบว่าวิธี LASSO มีจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยมากกว่าวิธี Adaptive LASSO และทั้งสองวิธีมีจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยจำนวนมาก

กรณีตัวแบบถดถอยพหุคูณ non-orthogonal and non-sparse model เนื่องจาก กำหนดให้ ตัวแปรอิสระทุกตัวอยู่ในตัวแบบ จึงพิจารณาเฉพาะความผิดพลาดของการคัดเลือกตัวแปรอิสระแบบ IC1 เท่านั้น และในทำนองเดียวกัน วิธีริตจไม่เกิดความผิดพลาดในการคัดเลือกตัวแปรอิสระ เมื่อเปรียบเทียบความผิดพลาดของการคัดเลือกตัวแปรอิสระวิธี LASSO, Adaptive LASSO และ Relaxed LASSO พบว่า วิธี LASSO มีความผิดพลาดของการคัดเลือกตัวแปรอิสระแบบ IC1 น้อยที่สุด และวิธี Relaxed LASSO มีความผิดพลาดของการคัดเลือกตัวแปรอิสระแบบ IC1 มากที่สุด แต่อย่างไรก็ตาม เมื่อพิจารณาจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ย เมื่อกำหนดให้ตัวแปร

อิสระทุกตัวอยู่ในตัวแบบ พบว่า วิธี Relaxed LASSO มีจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยน้อยที่สุด ต่อมาคือวิธี LASSO และวิธี Adaptive LASSO มีจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยมากที่สุด ซึ่งทั้งสามวิธีมีจำนวนตัวแปรอิสระในสมการถดถอยโดยเฉลี่ยน้อยกว่าจำนวนตัวแปรอิสระที่กำหนด

จากข้างต้น สามารถสรุปประสิทธิภาพของตัวประมาณสัมประสิทธิ์การถดถอยวิธีรีดจ์, LASSO, Adaptive LASSO และ Relaxed LASSO ภายใต้ขอบเขตงานวิจัย ดังนี้

ตารางที่ 20 เปรียบเทียบประสิทธิภาพของวิธีการประมาณ จำแนกตามตัวแบบการถดถอยพหุคูณ

กรณี	ประสิทธิภาพของการพยากรณ์	ประสิทธิภาพของการประมาณ	ความผิดพลาดในการคัดเลือกตัวแปรอิสระ
partial orthogonal sparse model	- วิธี Relaxed LASSO มีประสิทธิภาพสูงที่สุด	วิธี Relaxed LASSO มีประสิทธิภาพสูงที่สุด	- วิธี LASSO มีความผิดพลาดแบบ IC1 น้อยที่สุด แต่มีความผิดพลาดแบบ IC2 มากที่สุด
non-orthogonal sparse model	- วิธีรีดจ์มีประสิทธิภาพต่ำที่สุด - วิธี Adaptive LASSO มีประสิทธิภาพสูงกว่าวิธี LASSO เมื่อตัวแปรอิสระมีความสัมพันธ์สูง	ต่อมาคือ วิธี LASSO, Adaptive LASSO และวิธีรีดจ์	- วิธี Relaxed LASSO มีความผิดพลาดแบบ IC2 น้อยที่สุด แต่มีความผิดพลาดแบบ IC1 มากที่สุด
grouping structure sparse model	- วิธีรีดจ์มีประสิทธิภาพสูงที่สุด ต่อมาคือวิธี LASSO	- วิธีรีดจ์มีประสิทธิภาพสูงที่สุดต่อมาคือวิธี LASSO	- วิธี LASSO มีความผิดพลาดแบบ IC1 น้อยที่สุด
non-orthogonal and non-sparse model	- ส่วนวิธี Adaptive LASSO มีประสิทธิภาพสูงกว่าวิธี Relaxed LASSO เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูง	- ส่วนวิธี Adaptive LASSO มีประสิทธิภาพสูงกว่าวิธี Relaxed LASSO เมื่อตัวแปรอิสระมีความสัมพันธ์สูง	- วิธี Relaxed LASSO มีความผิดพลาดแบบ IC1 มากที่สุด

อภิปรายผลการวิจัย

ในกรณีข้อมูลมิติสูงแบบบางเบา (high-dimensional sparse data) พบว่า วิธี Relaxed LASSO มีประสิทธิภาพของการพยากรณ์และประสิทธิภาพของการประมาณสูงที่สุด เนื่องจากวิธี Relaxed LASSO แบ่งการควบคุมอิทธิพลของการคัดเลือกตัวแบบและการหดลงของตัวประมาณของค่าพารามิเตอร์ด้วยพารามิเตอร์ 2 ตัว และวิธีริดจ์มีประสิทธิภาพต่ำที่สุดเนื่องจาก ตัวประมาณสัมประสิทธิ์การถดถอยวิธีริดจ์เป็นตัวประมาณที่ให้ค่าประมาณสัมประสิทธิ์การถดถอยทุกตัว มีค่าเข้าใกล้ศูนย์แต่ไม่เท่ากับศูนย์ และเมื่อกำหนดให้ค่าพารามิเตอร์เป็นศูนย์จำนวนมาก จึงทำให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์การถดถอยและค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณแตกต่างกันมาก ส่งผลให้วิธีริดจ์มีประสิทธิภาพของการพยากรณ์และประสิทธิภาพของการประมาณต่ำ เมื่อเปรียบเทียบกับวิธี LASSO และ Adaptive LASSO พบว่า วิธี LASSO มีประสิทธิภาพสูงกว่าวิธี Adaptive LASSO เมื่อตัวแปรอิสระมีความสัมพันธ์กันต่ำ ส่วนวิธี Adaptive LASSO มีประสิทธิภาพสูงกว่าวิธี LASSO เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูง เนื่องจากเมื่อตัวแปรอิสระมีความสัมพันธ์สูงส่งผลให้เงื่อนไข (38) จากทฤษฎีบท 1 ไม่เป็นจริง จึงทำให้ตัวประมาณสัมประสิทธิ์การถดถอยวิธี LASSO เป็นตัวประมาณไม่คงเส้นคงวา ดังนั้น วิธี Adaptive LASSO มีประสิทธิภาพสูงกว่าวิธี Adaptive LASSO เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูง เมื่อพิจารณาความผิดพลาดของการคัดเลือกตัวแปรอิสระ พบว่า ทั้งสามวิธี คือ LASSO, Adaptive LASSO และ Relaxed LASSO มีความผิดพลาดแบบ IC1 ใกล้เคียงกัน แต่เมื่อพิจารณาความผิดพลาดแบบ IC2 พบว่า วิธี Relaxed LASSO มีความผิดพลาดน้อยที่สุด

ในกรณีข้อมูลมิติสูงแบบไม่บางเบา (high-dimensional non-sparse data) พบว่า วิธีริดจ์มีประสิทธิภาพของการพยากรณ์และประสิทธิภาพของการประมาณสูงที่สุด เนื่องจาก ตัวประมาณสัมประสิทธิ์การถดถอยวิธีริดจ์เป็นตัวประมาณที่ให้ค่าประมาณสัมประสิทธิ์การถดถอยทุกตัว มีค่าเข้าใกล้ศูนย์แต่ไม่เท่ากับศูนย์ และเมื่อกำหนดให้ค่าพารามิเตอร์ทุกตัวมีขนาดเล็กแต่ไม่เท่ากับศูนย์ จึงทำให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์การถดถอยและค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณแตกต่างกันน้อย รองลงมาคือวิธี LASSO เนื่องจากวิธี Relaxed LASSO เป็นกระบวนการหาตัวประมาณสัมประสิทธิ์การถดถอย 2 ขั้นตอน ซึ่งมีการคัดเลือกตัวแปรอิสระจากกระบวนการแรกแล้วนำเซตของตัวแปรอิสระนั้นมาทำการประมาณสัมประสิทธิ์การถดถอยอีกรอบหนึ่ง จึงทำให้วิธี Relaxed LASSO มีตัวแปรอิสระที่อยู่ในตัวแบบน้อยกว่าวิธี LASSO แต่สถานการณ์นี้กำหนดให้ตัวแปรอิสระทุกตัวอยู่ในตัวแบบ ดังนั้น วิธี LASSO จึงมีประสิทธิภาพสูงกว่าวิธี Relaxed LASSO เมื่อเปรียบเทียบกับวิธี Adaptive LASSO และ Relaxed LASSO พบว่า วิธี Adaptive LASSO มีประสิทธิภาพสูงกว่าวิธี Relaxed LASSO เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูง และเมื่อพิจารณาประสิทธิภาพของการคัดเลือกตัวแปรอิสระ พบว่า วิธี LASSO มี

ประสิทธิภาพสูงสุด และวิธี Relaxed LASSO มีประสิทธิภาพต่ำที่สุด แต่อย่างไรก็ตาม วิธี Relaxed LASSO มีจำนวนตัวแปรอิสระในตัวแบบน้อยที่สุด

ข้อเสนอแนะ

1. เมื่อข้อมูลที่ใช้ในการวิเคราะห์เป็นข้อมูลมีมิติสูงแบบบางเบา มีตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามจำนวนน้อยควรเลือกใช้วิธี Relaxed LASSO ในการประมาณค่าสัมประสิทธิ์การถดถอย
2. เมื่อข้อมูลที่ใช้ในการวิเคราะห์เป็นแบบไม่บางเบาควรเลือกใช้วิธีริตจ์ในการประมาณค่าสัมประสิทธิ์การถดถอย

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. ในงานวิจัยนี้ศึกษากรณีที่กำหนดตัวแปรอิสระอยู่ในตัวแบบจริงจำนวน 15 ตัว ในงานวิจัยครั้งต่อไปควรกำหนดจำนวนตัวแปรอิสระในตัวแบบจำนวนที่หลากหลาย
2. เนื่องจากในงานวิจัยนี้กำหนดให้เวกเตอร์ถ่วงน้ำหนักในวิธี Adaptive LASSO คือ $w_j = |\beta_j|^{-1}$ เมื่อ $j = 1, \dots, p$ โดยที่ $\beta_j = \frac{X_j'Y}{n}$ ในงานวิจัยครั้งต่อไปควรกำหนดเวกเตอร์ถ่วงน้ำหนักในวิธี Adaptive LASSO ในรูปแบบอื่น
3. ในงานวิจัยนี้ศึกษาในกรณีที่กำหนดความคลาดเคลื่อนสุ่มมีการแจกแจงปกติ ในงานวิจัยครั้งต่อไปควรศึกษาเพิ่มเติมในกรณีที่ความคลาดเคลื่อนสุ่มมีการแจกแจงอื่น



ภาคผนวก

โปรแกรมที่ใช้ในงานวิจัย

```

library(lars)
library(MASS)
library(relaxo)
nrep<-500          #จำนวนการทำซ้ำ 500 รอบ
r=0.1              #กำหนดค่า  $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$ 
n=50               #ขนาดตัวอย่าง n = 50, 100, 200
p=100              #จำนวนตัวแปรอิสระ p = 100, 200, 400, 800
p1=15              #กำหนดจำนวนตัวแปรอิสระในตัวแบบจริง p1=15, p
p2=p-p1           #จำนวนตัวแปรอิสระที่ไม่อยู่ในตัวแบบจริง
nb=p1/3
nt=n
v=5
nv=nt/v

#กำหนดเมทริกซ์ความแปรปรวนร่วมกรณีความสัมพันธ์ของตัวแปรอิสระอยู่ในรูปแบบ partial
orthogonal model
sigma1=matrix(0,p1,p1)
for(a in 1:p1)
  for(b in 1:p1)
    sigma1[a,b]=r^(abs(a-b))
sigma0=matrix(0,p1,p2)
for(a in 1:p1)
  for(b in 1:p2)
    sigma0[a,b]=0
sigma00=matrix(0,p2,p1)
for(a in 1:p2)
  for(b in 1:p1)
    sigma00[a,b]=0
sigma2=matrix(0,p2,p2)
for(a in 1:p2)

```

```

for(b in 1:p2)
  sigma2[a,b]=r^(abs(a-b))
sigmaa<-cbind(sigma1,sigma0)
sigmaaa<-cbind(sigma00,sigma2)
sigma<-rbind(sigmaa,sigmaaa)
sigmae<-diag(rep(1.5^2,p))

#กำหนดเมทริกซ์ความแปรปรวนร่วมกรณีความสัมพันธ์ของตัวแปรอิสระอยู่ในรูปแบบ partial non-
orthogonal model
sigma=matrix(0,p,p)
for(a in 1:p)
  for(b in 1:p)
    sigma[a,b]=r^(abs(a-b))

b<-c(rep(2.5,nb),rep(1.5,nb),rep(0.5,nb),rep(0,p2)) #กำหนดค่าสัมประสิทธิ์การถดถอย

set.seed(1234)

pmse.ridge<-rep(0,nrep)
pmse.lasso<-rep(0,nrep)
pmse.adplasso<-rep(0,nrep)
pmse.rexlasso<-rep(0,nrep)
mse.ridge<-rep(0,nrep)
mse.lasso<-rep(0,nrep)
mse.adplasso<-rep(0,nrep)
mse.rexlasso<-rep(0,nrep)
ic1.ridge<-rep(0,nrep)
ic2.ridge<-rep(0,nrep)
ic1.lasso<-rep(0,nrep)
ic2.lasso<-rep(0,nrep)
ic1.adplasso<-rep(0,nrep)

```



```

ic2.adplasso<-rep(0,nrep)
ic1.rexlasso<-rep(0,nrep)
ic2.rexlasso<-rep(0,nrep)
  for(k in 1:nrep)
  {
#สร้างข้อมูลค่าสังเกตของชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบ
กรณีความสัมพันธ์ของตัวแปรอิสระอยู่ในรูปแบบ partial orthogonal model, partial non-orthogonal model และ non-sparse model
x.train<-mvrnorm(n,rep(0,p),sigma)
e.train<-rnorm(n,0,1.5)
y.train <- x.train%*%b+e.train
data.train<-data.frame(y=y.train,x=l(x.train))

x.test<-mvrnorm(n,rep(0,p),sigma)
e.test<-rnorm(n,0,1.5)
y.test <- x.test%*%b+e.test
data.test<-data.frame(y=y.test,x=l(x.test))

กรณีความสัมพันธ์ของตัวแปรอิสระอยู่ในรูปแบบ grouping structure
z1.tr<-rnorm(n,0,1)
z2.tr<-rnorm(n,0,1)
z3.tr<-rnorm(n,0,1)
e1.tr=matrix(0,n,nb)
for(a in 1:nb)
  e1.tr[,a]=rnorm(n,0,1)
e2.tr=matrix(0,n,nb)
for(a in 1:nb)
  e2.tr[,a]=rnorm(n,0,1)
e3.tr=matrix(0,n,nb)
for(a in 1:nb)
  e3.tr[,a]=rnorm(n,0,1)

```

```

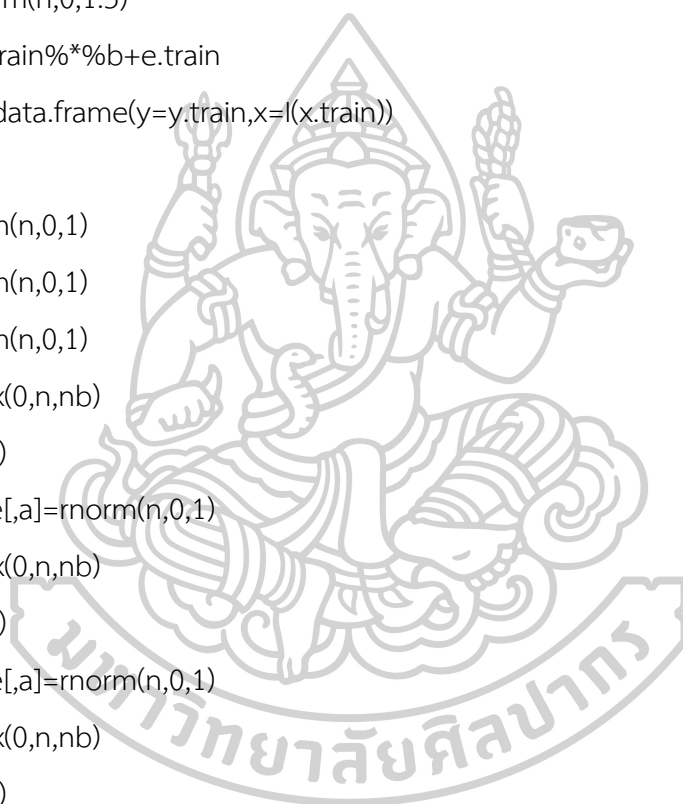
x1.tr<-z1.tr+e1.tr
x2.tr<-z2.tr+e2.tr
x3.tr<-z3.tr+e3.tr
x4.tr<-matrix(0,n,p2)
for(a in 1:p2)
  x4.tr[,a]=rnorm(n,0,1)
x.train<-cbind(x1.tr,x2.tr,x3.tr,x4.tr)
e.train<-rnorm(n,0,1.5)
y.train <- x.train%*%b+e.train
data.train<-data.frame(y=y.train,x=l(x.train))

```

```

z1.te<-rnorm(n,0,1)
z2.te<-rnorm(n,0,1)
z3.te<-rnorm(n,0,1)
e1.te=matrix(0,n,nb)
for(a in 1:nb)
  e1.te[,a]=rnorm(n,0,1)
e2.te=matrix(0,n,nb)
for(a in 1:nb)
  e2.te[,a]=rnorm(n,0,1)
e3.te=matrix(0,n,nb)
for(a in 1:nb)
  e3.te[,a]=rnorm(n,0,1)
x1.te<-z1.te+e1.te
x2.te<-z2.te+e2.te
x3.te<-z3.te+e3.te
x4.te<-matrix(0,n,p2)
for(a in 1:p2)
  x4.te[,a]=rnorm(n,0,1)
x.test<-cbind(x1.te,x2.te,x3.te,x4.te)
e.test<-rnorm(n,0,1.5)

```



```

y.test <- x.test%*%b+e.test
data.test<-data.frame(y=y.test,x=l(x.test))
ind<-list(v)
for(i in 1:(v-1))
{
  ind[[i]]<-((i-1)*nv+1):(i*nv)
}
ind[[v]]<-((v-1)*nv+1):(nt)

s.vec<-seq(0,1,0.05)
#การวิเคราะห์การถดถอยวิธีริดจ์
pmse.cv.ridge<-matrix(0,nv,length(s.vec))
pmse.lambda.ridge<-matrix(0,v,length(s.vec))
beta.ridge.hat.cv<-matrix(0,v,(p+1))
for(i in 1:v)
{
  vdata.train<-data.train[-ind[[i]],]
  vdata.test<-data.train[ind[[i]],]
  lsridge<-lm.ridge(y~x,data=vdata.train,lambda=s.vec)
  beta.hat.v.ridge<-coef(lsridge)
  yhat.v.ridge<-cbind(1,vdata.test$x)%*%t(beta.hat.v.ridge)
  for(j in 1:length(s.vec))
  pmse.cv.ridge[j]<-(vdata.test$y-yhat.v.ridge[,j])^2
  pmse.lambda.ridge[i,j]<-colSums(pmse.cv.ridge)/nv
  min.pmse.cv.ridge<-min(pmse.lambda.ridge[i,])
  tuning.cv.ridge<-s.vec[which(pmse.lambda.ridge[i,]==min.pmse.cv.ridge)]
  beta.ridge.hat.cv[i,j]<-beta.hat.v.ridge[which(s.vec==tuning.cv.ridge),]
}
apmse.ridge<-colSums(pmse.lambda.ridge)/v
apmse.min.ridge<-min(apmse.ridge)
tuning.ridge<-s.vec[which(apmse.ridge==apmse.min.ridge)]

```

```

ridge.test<-lm.ridge(y~x,data=data.train,lambda=tuning.ridge)
coef.ridge<-coef(ridge.test)
beta.hat.ridge<-coef.ridge[-1]
yhat.ridge<-cbind(1,data.test$x)%*%coef.ridge
pmse.ridge[k]<-sum((data.test$y-yhat.ridge)^2)/nt
mse.ridge[k]<-sum((beta.hat.ridge-b)^2)/p
ic1.ridge[k] = sum(b!=0 & beta.hat.ridge==0)
ic2.ridge[k] = sum(b==0 & beta.hat.ridge!=0)

#การวิเคราะห์การถดถอยวิธี LASSO
pmse.cv.lasso<-matrix(0,nv,length(s.vec))
pmse.lambda.lasso<-matrix(0,v,length(s.vec))
beta.lasso.hat.cv<-matrix(0,v,p)
for(i in 1:v)
{
vdata.train<-data.train[-ind[[i]],]
vdata.test<-data.train[ind[[i]],]
out.v.lasso<-lars(x=vdata.train$x,y=vdata.train$y,type="lasso")
beta.hat.v.lasso<-predict(out.v.lasso,s=s.vec,type="coef",mode="fraction")$coef
xbar.v.lasso<-colMeans(vdata.train$x)
ybar.v.lasso<-mean(vdata.train$y)
b0.v.lasso<-ybar.v.lasso-beta.hat.v.lasso%*%xbar.v.lasso
b0.test.lasso<-matrix(rep(b0.v.lasso,nv),nv,length(s.vec),byrow=TRUE)
yhat.v.lasso<-vdata.test$x%*%t(beta.hat.v.lasso)+b0.test.lasso
vdata.test.y.lasso<-matrix(vdata.test$y,nv,length(s.vec))
pmse.cv.lasso<-(vdata.test.y.lasso-yhat.v.lasso)^2
pmse.lambda.lasso[i,]<-colSums(pmse.cv.lasso)/nv
min.pmse.cv.lasso<-min(pmse.lambda.lasso[i,])
tuning.cv.lasso<-s.vec[which(pmse.lambda.lasso[i,]==min.pmse.cv.lasso)]
beta.lasso.hat.cv[i,]<-beta.hat.v.lasso[which(s.vec==tuning.cv.lasso),]
}

```

```

apmse.lasso<-colSums(pmse.lambda.lasso)/v
apmse.min.lasso<-min(apmse.lasso)
tuning.lasso<-s.vec[which(apmse.lasso==apmse.min.lasso)]
out.lasso<-lars(x=data.train$x,y=data.train$y,type="lasso")
beta.hat.lasso<-predict(out.lasso,s=tuning.lasso,type="coef",mode="fraction")$coef
xbar<-colMeans(data.train$x)
ybar<-mean(data.train$y)
b0.lasso<-ybar-beta.hat.lasso%*%xbar
yhat.lasso<-data.test$x%*%beta.hat.lasso+rep(b0.lasso,nt)
pmse.lasso[k]<-sum((data.test$y-yhat.lasso)^2)/nt
mse.lasso[k]<-sum((beta.hat.lasso-b)^2)/p
ic1.lasso[k] = sum(b!=0 & beta.hat.lasso==0)
ic2.lasso[k] = sum(b==0 & beta.hat.lasso!=0)

#การวิเคราะห์การถดถอยวิธี Adaptive LASSO
tran_data_train.x<-scale(data.train$x,center=TRUE,scale=TRUE)
tran_data_train.y<-scale(data.train$y,center=TRUE,scale=FALSE)
betatilde<-(t(tran_data_train.x)%*%tran_data_train.y)/n
w <- 1/abs(betatilde)
data_train.x<-scale(tran_data_train.x,center=FALSE,scale=w)
data_train.y<-(tran_data_train.y)
data_train<-data.frame(y=data_train.y,x=l(data_train.x))
sd.tran_data_train.x<-apply(data.train$x,2,sd)
pmse.cv.adplasso<-matrix(0,nv,length(s.vec))
pmse.lambda.adplasso<-matrix(0,v,length(s.vec))
beta.adplasso.hat.cv<-matrix(0,v,p)
for(i in 1:v)
{
vdata.train.adplasso<-data.train[-ind[[i]],]
vdata.test.adplasso<-data.train[ind[[i]],]
tran.vdata.train.adp.x<-scale(vdata.train.adplasso$x,center=TRUE,scale=TRUE)

```

```

tran.vdata.train.adp.y<-scale(vdata.train.adplasso$y,center=TRUE,scale=FALSE)
betatilde.cv<-(t(tran.vdata.train.adp.x)%*%tran.vdata.train.adp.y)/nv
w.cv<-1/abs(betatilde.cv)
vdata_train.x.adplasso<-scale(tran.vdata.train.adp.x,center=FALSE,scale=w.cv)
vdata_train.y.adplasso<-(tran.vdata.train.adp.y)
out.v.adplasso<-
lars(x=vdata_train.x.adplasso,y=vdata_train.y.adplasso,type="lasso",normalize=FALSE)
beta.hat.v.adplasso<-predict(out.v.adplasso,s=s.vec,type="coef",mode="fraction")$coef
beta.hatstar.v.adplasso<-beta.hat.v.adplasso/t(matrix(w.cv,length(w.cv),length(s.vec)))
xbar.v.adplasso<-colMeans(tran.vdata.train.adp.x)
ybar.v.adplasso<-mean(tran.vdata.train.adp.y)

b0.v.adplasso<-ybar.v.adplasso-beta.hatstar.v.adplasso%*%xbar.v.adplasso
b0.test.adplasso<-matrix(rep(b0.v.adplasso,nv),nv,length(s.vec),byrow=TRUE)
yhat.v.adplasso<-vdata.test.adplasso$x%*%t(beta.hatstar.v.adplasso)+b0.test.adplasso
for(j in 1:length(s.vec))
pmse.cv.adplasso[,j]<-(vdata.test.adplasso$y-yhat.v.adplasso[,j])^2
pmse.lambda.adplasso[i,j]<-colSums(pmse.cv.adplasso)/nv
}
apmse.adplasso<-colSums(pmse.lambda.adplasso)/v
apmse.min.adplasso<-min(apmse.adplasso)
tuning.adplasso<-s.vec[which(apmse.adplasso==apmse.min.adplasso)]
out.adplasso<-lars(x=data_train$x,y=data_train$y,type="lasso",normalize=FALSE)
beta.hat.star<-
predict(out.adplasso,s=tuning.adplasso,type="coef",mode="fraction")$coef
beta.hat.adplasso<-beta.hat.star/(w*sd.tran_data_train.x)
xbar.adplasso<-colMeans(data.train$x)
ybar.adplasso<-mean(data.train$y)
b0.adplasso<-ybar.adplasso-t(beta.hat.adplasso)%*%xbar.adplasso
yhat.adplasso<-data.test$x%*%beta.hat.adplasso+rep(b0.adplasso,nt)
pmse.adplasso[k]<-sum((data.test$y-yhat.adplasso)^2)/nt

```

```

mse.adplasso[k]<-sum((beta.hat.adplasso-b)^2)/p
ic1.adplasso[k] = sum(b!=0 & beta.hat.adplasso==0)
ic2.adplasso[k] = sum(b==0 & beta.hat.adplasso!=0)

#การวิเคราะห์การถดถอยวิธี Relaxed LASSO
phi<-sort(seq(0,1,0.05),decreasing=TRUE)
max.steps=min(2*length(data.train$y),2*ncol(data.train$x))
fast=TRUE
keep.data=TRUE
pmse.cv.relaxo<-rep(0,length=length(phi)*(max.steps-1))
for(i in 1:v)
{
vdata.train<-data.train[-ind[[i]],]
vdata.test<-data.train[ind[[i]],]

out.v.rexlasso<-relaxo(vdata.train$x,vdata.train$y,phi=phi,
                      fast=fast,keep.data=FALSE,warn=FALSE,max.step=max.steps)
beta.hat.v.rexlasso<-out.v.rexlasso$beta
yhat.v.rexlasso<-vdata.test$x%*%t(beta.hat.v.rexlasso)
pmse.cv.rexlasso<-(vdata.test$y-yhat.v.rexlasso)^2
pmse.lambda.rexlasso<-colSums(pmse.cv.rexlasso)/nv
pmse.cv.relaxo[1:ncol(yhat.v.rexlasso)]<-pmse.cv.relaxo[1:ncol(yhat.v.rexlasso)] +
colSums((vdata.test$y-yhat.v.rexlasso)^2)/v
if(length(pmse.cv.relaxo)>ncol(yhat.v.rexlasso))
  pmse.cv.relaxo[(ncol(yhat.v.rexlasso)+1):length(pmse.cv.relaxo)]<-Inf
}
select<-which.min(pmse.cv.relaxo)
out.rexlasso<-relaxo(data.train$x,data.train$y,phi=phi,fast=fast,
                    keep.data=keep.data,warn=FALSE)
beta.hat.rexlasso<-out.rexlasso$beta[select,,drop=FALSE]
xbar<-colMeans(data.train$x)

```

```

ybar<-mean(data.train$y)
b0.rexlasso<-ybar-beta.hat.rexlasso%*%xbar
yhat.rexlasso<-data.test$x%*%t(beta.hat.rexlasso)+rep(b0.rexlasso,nt)
pmse.rexlasso[k]<-sum((data.test$y-yhat.rexlasso)^2)/nt
mse.rexlasso[k]<-sum((beta.hat.rexlasso-b)^2)/p

ic1.rexlasso[k] = sum(b!=0 & beta.hat.rexlasso==0)
ic2.rexlasso[k] = sum(b==0 & beta.hat.rexlasso!=0)
      }

med.pmse.ridge<-median(pmse.ridge)
med.pmse.lasso<-median(pmse.lasso)
med.pmse.adplasso<-median(pmse.adplasso)
med.pmse.rexlasso<-median(pmse.rexlasso)
med.mse.ridge<-median(mse.ridge)
med.mse.lasso<-median(mse.lasso)
med.mse.adplasso<-median(mse.adplasso)
med.mse.rexlasso<-median(mse.rexlasso)
meanic1.ridge<-mean(ic1.ridge)
meanic2.ridge<-mean(ic2.ridge)
meanic1.lasso<-mean(ic1.lasso)
meanic2.lasso<-mean(ic2.lasso)
meanic1.adplasso<-mean(ic1.adplasso)
meanic2.adplasso<-mean(ic2.adplasso)
meanic1.rexlasso<-mean(ic1.rexlasso)
meanic2.rexlasso<-mean(ic2.rexlasso)
med.pmse.ridge
med.pmse.lasso
med.pmse.adplasso
med.pmse.rexlasso
med.mse.ridge

```


med.mse.lasso
med.mse.adplasso
med.mse.rexlasso
meanic1.ridge
meanic2.ridge
meanic1.lasso
meanic2.lasso
meanic1.adplasso
meanic2.adplasso
meanic1.rexlasso
meanic2.rexlasso



รายการอ้างอิง

- วิฐุรา พึ่งพาพงศ์. (2558). “บทวิเคราะห์วิธีวิเคราะห์การถดถอยเชิงเส้นสำหรับข้อมูลที่มีมิติสูง.” **วารสารวิทยาศาสตร์และเทคโนโลยี** 23, 2 (เมษายน-มิถุนายน): 212-223.
- Donoho, David L. and Iain M. Johnstone. (1994). “Ideal Spatial Adaptation by Wavelet Shrinkage.” Department of Statistics, Stanford University, Stanford, U.S.A.
- Draper, N and Smith, H. (1981). **Applied regression analysis**. New York: John Wiley.
- Efron, Bradley et al. (2004), “Least Angle Regression,” **The Annals of Statistics** 32, 2: 407–499.
- Fan, Jianging, and Runze Li. (2001), “Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties.” **Journal of the American Statistical Association** 96, 456 (December): 1348–1360.
- Hastie, Trevor (2015). **Package ‘lars’**. Accessed April 30. Available from <https://cran.r-project.org/web/packages/lars/lars.pdf>
- Henderson, Harold V. and Paul F. Velleman. (1981). “Building Multiple Regression Models Interactively.” **International Biometric Society** 37, 2 (January): 391-411.
- Hoerl, Arthur E. (1962). “Application of ridge analysis to regression problems.” **Chemical Engineering Progress** 58: 54-59.
- Hoerl, Arthur E. and Robert W. Kennard. (1970). “Ridge regression: biased estimation for nonorthogonal problems.” **American Statistical Association** 12, 1 (February): 56-67.
- Hoerl, Roger W. (1985). “Ridge Analysis 25 Years Later.” **The American Statistician** 39, 3: 183-192.
- Huang, Jian, Shuangge Ma and Cun-Hui Zhang. (2008). “Adaptive Lasso for sparse high-dimensional regression models.” **Institute of Statistical Science, Academia Sinica** 18, 4 (October): 1603-1618.
- Nicolai, Meinshausen. (2006). “Relaxed Lasso” **Computational Statistics & Data Analysis**, 52: 374 – 393.
- Simon S. (2016). “Understand how lasso regression work” 1-29.

Tibshirani, Robert. (1996), “Regression Shrinkage and Selection via the Lasso,”

Journal of the Royal Statistical Society 58, 1 (January): 267–288.

Zhang, Feng. (2011). “Cross-validation and regression analysis in high-dimensional sparse linear models.” Ph.D. dissertation, Department of Statistics and the committee on graduate studies, University of Sheffield.

Zou, Hui. (2006). “The Adaptive Lasso and Its Oracle Properties.” **Journal of the**

American Statistical Association 101, 476 (December): 1418-1429.

Zou, Hui and Trevor Hastie. (2005). “Regularization and variable selection via the elastic net.” **J. R. Statist** 67: 301-320.



รายการอ้างอิง





ประวัติผู้เขียน

ชื่อ-สกุล	พัชราภรณ์ พรดำเนินสวัสดิ์
วัน เดือน ปี เกิด	16 เมษายน 2536
สถานที่เกิด	นครปฐม ประเทศไทย
วุฒิการศึกษา	วท.บ. (สถิติ) มหาวิทยาลัยศิลปากร
ที่อยู่ปัจจุบัน	บ้านเลขที่ 124 หมู่ที่ 6 ต.คลองใหม่ อ.สามพราน จ.นครปฐม 73110
ผลงานตีพิมพ์	Porndumnernsawat, P. and N. Meejun (2015). A comparison of products of coefficients methods to test mediation. The 4th ICADA 2015-SSIS, Bangkok, Thailand. Porndumnernsawat, P. (2016). A Comparison of Product of Coefficients Methods to Test Mediation Effect. ICAS2016, Phuket, Thailand.



