การเปรียบเทียบวิธีการแก้ไขพหุสัมพันธ์ในการถดถอยเชิงเส้นพหุคูณ

นางสาวเปรมวดี ชูไสว

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2548

ISBN 974-53-2835-9

A COMPARISION OF MULTICOLLINEARITY CORRECTION METHODS

IN MULTIPLE LINEAR REGRESSIONS

Miss Premvadee Chuesawai

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2005

ISBN 974-53-2835-9

หัวข้อวิทยานิพนธ์	การเปรียบเทียบวิธีการแก้ไขพหุสัมพันธ์ในการถดถอยเชิงเส้นพหุคูณ
โดย	นางสาวเปรมวดี ชูไสว
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

ดนุชา คุณพนิชกิจ ....คณบดีคณะพาณิชยศาสตร์และการบัญชี

(ผู้ช่วยศาสตราจารย์ ดร.ดนุชา คุณพนิชกิจ)

คณะกรรมการสอบวิทยานิพนธ์

.... ประธานกรรมการ

(รองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา)

.... อาจารย์ที่ปรึกษา

(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

.... กรรมการ

(รองศาสตราจารย์ วัชราภรณ์ สุริยาภิวัฒน์)

เปรมวดี ชูไลว : การเปรียบเทียบวิธีการแก้ไขพหุสัมพันธ์ในการถดถอยเชิงเส้นพหุคูณ (COMPARISION OF MULTICOLLINEARITY CORRECTION METHOD IN MULTIPLE LINEAR REGRESSIONS) อ.ที่ปรึกษา : รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา 116 หน้า ISBN 974-53-2835-9

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ เมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ วิธีการเปรียบเทียบที่นำมาพิจารณาคือ วิธีกำลังสองน้อยสุด วิธีริดจ์รีเกรสชั่น วิธีการถดถอยองค์ประกอบหลัก โดยเกณฑ์เปรียบเทียบคือ อัตราส่วนของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอย โดยจำนวนตัวแปรอิสระที่ใช้เท่ากับ 3,6 และ 9 มีส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนเท่ากับ 1,5 และ 10 ขนาดตัวอย่างเท่ากับ 5p,10p,15p,20p,25p และ 30p เมื่อ p คือจำนวนตัวแปรอิสระ โดยแบ่งระดับความสัมพันธ์ระหว่างตัวแปรอิสระ เป็น น้อย (0.15-0.30) ปานกลาง (0.31-0.5, 0.51-0.65) และมาก (0.66-0.85,0.86-0.99)

ผลที่ได้จากการวิจัย พบว่า มากกว่า 99% ของสถานการณ์ สามารถสรุปได้ว่าในระดับความสัมพันธ์มาก ช่วง 0.86-0.99 นั้นจะทำให้เกิดปัญหาพหุสัมพันธ์สูงจนส่งผลกระทบต่อการประมาณค่าด้วยตัวประมาณแบบวิธีกำลังสองน้อยสุด ดังนั้นวิธีการถดถอยองค์ประกอบหลักจึงเป็นวิธีที่มีประสิทธิภาพที่สุด ในทุกกรณี ในส่วนระดับความสัมพันธ์อื่น นั้น จะต้องพิจารณาค่าส่วนเบี่ยงเบนของค่าคลาดเคลื่อนประกอบด้วย ถ้าส่วนเบี่ยงเบนของค่าคลาดเคลื่อน เท่ากับ 1 วิธีกำลังสองน้อยสุดจะมีประสิทธิภาพมากสุด และในส่วนของ การเบี่ยงเบนค่าคลาดเคลื่อนเท่ากับ 5 และ 10 วิธีกำลังสองน้อยสุด หลังจากที่ทำการแก้ไขให้ x เป็นแกนตั้งฉากซึ่งกันและกันแล้ว จะเป็นวิธีที่ดีที่สุด

ปัจจัยที่มีผลต่อความคลาดเคลื่อนนั้น พบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้น AMSE ที่ได้จะให้ค่าลดลง ตรงกันข้าม ถ้าระดับความสัมพันธ์ระหว่างตัวแปรอิสระ และ ส่วนเบี่ยงเบนของค่าคลาดเคลื่อนเพิ่มขึ้น AMSE ที่ได้จะมีค่าเพิ่มขึ้น ยกเว้นกรณีวิธีถดถอยองค์ประกอบหลัก ที่ เมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระช่วง 0.86-0.99 จะน้อยกว่า ค่า AMSE ในระดับอื่น

ภาควิชา.........สถิติ............ ลายมือชื่อนิสิต.......เปรมวดี ชูไลว.......

สาขาวิชา.......สถิติ............ ลายมือชื่ออาจารย์ที่ปรึกษา..........

#4682317126 : MAJOR STATISTICS
KEY WORDS : Multivariate Regression / Multicollinearity / ordinary least square / ridge regression / principal component regression / Gram–Schmidt orthogonalization

PREMVADEE CHUESAWAI : A COMPARISION OF MULTICOLLINEARITY CORRECTION METHOD IN MULTIPLE LINEAR REGRESSIONS. THESIS ADVISOR : ASSOC.PROF. SUPOL DURONGWATANA., Ph.D. , 116 pp. ISBN 974-53-2835-9

The objective of this study is to compare the parameter estimations of multiple regression coefficient when multicollinearity is occurred. The method under consideration are : ordinary least square method (OLS), ridge regression method (RID), principal component regression method (PCR) and ordinary least square method of transforming X by Gram–Schmidt orthogonalization (OLS – ORT). The criterion of comparison is a ratio of average mean square error of coefficient.

In this study, there are four factors that should affect AMSE in coefficient estimation which are the number of independent variable, sample size, level of multicollinearity and standard error. Numbers of independent variables are 3, 6 and 9. Sample sizes of each independent variable have 5p, 10p, 15p, 20p, 25p and 30p when p is number of independent variables. Error in the model are random variables sampling from normal distribution with zero mean and various standard deviation, ie. 1,5,10. Level of multicollinearity split into 3 levels (low, medium and high)

The results of this study is that OLS is the method, having the best efficiency for prediction in regression model if it has low level of multicollinearity and standard error equal to 1. Generally, the efficiency of PCR and OLS – ORT are not different in all cases except case that have high level of multicollinearity. OLS - ORT are the efficient methods if it has standard error equal to 5 or 10 and has medium multicollinearity. PCR is the best method if it has high multicollinearity in all level of standard error.

The factors that affected AMSE in coefficient estimation which are ranked from the most are level of multicollinearity, standard error, the number of independent variables and sample size. The value of AMSE of 4 methods is increased when standard error increased or level of multicollinearity increased. Contradictory, the value of AMSE of all methods is decreased when sample size increased.

Department.........Statistics.........
Field of Study......Statistics.........
Academic Year.......2005............

Student's signature.......Premvadee Chuesawai
Advisor's signature..........................

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความช่วยเหลืออย่างดียิ่งจาก รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา อาจารย์ที่ปรึกษาวิทยานิพนธ์ รองศาสตราจารย์ ดร. กัลยา วานิชย์บัญชา และ รองศาสตราจารย์วัชราภรณ์ สุริยาภิวัฒน์ อาจารย์กรรมการ ที่กรุณาให้คำปรึกษา แนะนำ และแก้ไขข้อบกพร่องต่าง ๆ เป็นอย่างดีมาโดยตลอด ผู้วิจัยใคร่ขอกราบขอบพระคุณเป็นอย่างสูง

นอกจากนี้ ขอกราบขอบพระคุณ ผู้ใหญ่และผู้มีพระคุณทุกท่านที่ให้การสนับสนุนเสมอมา และ ขอขอบคุณพี่ๆ เพื่อนๆ น้องๆที่ช่วยเหลือเป็นอย่างดีมาก และเป็นกำลังใจซึ่งกันเสมอ

สารบัญ

สารบัญตาราง

สารบัญรูป

รูปที่		หน้า

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในการศึกษาการเปลี่ยนแปลงของตัวแปรตาม (dependent variable) ว่ามีผลมาจากตัวแปรอิสระ (independent variables) ชุดหนึ่งอย่างไรนั้น วิธีหนึ่งของการวิเคราะห์ความสำคัญของตัวแปรเหล่านั้นก็คือ การวิเคราะห์ความถดถอยพหุคูณ (multiple regression analysis) ซึ่งเป็นกรณีหนึ่งของการวิเคราะห์ความถดถอยเชิงเส้น

ในการวิเคราะห์ความถดถอยพหุคูณนั้นเป็นการใช้ตัวแปรอิสระที่เหมาะสมมากกว่า 1 ตัว โดยทั่ว ๆ ไปย่อมทำให้ผลการประมาณค่าตัวแปรตามมีความถูกต้องมากกว่าการใช้ตัวแปรอิสระเพียงเดียว สมการตัวแบบทั่วไป (General model) ที่แสดงความสัมพันธ์ระหว่างตัวแปรอิสระ และตัวแปรตามแบบเชิงเส้น (linear relationship) จะมีลักษณะดังนี้

(1.1) $$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{y}$ แทนเวกเตอร์ของตัวแปรตามขนาด $n \times 1$

โดย n แทนขนาดตัวอย่างของตัวแปรแต่ละตัว

X แทนเมตริกซ์ตัวแปรอิสระขนาด $n \times (p+1)$

p แทนจำนวนตัวแปรอิสระ

$\underset{\sim}{\beta}$ แทนเวกเตอร์ของพารามิเตอร์ของตัวแปรอิสระขนาด $(p+1) \times 1$

และ $\underset{\sim}{\varepsilon}$ แทนเวกเตอร์ของความคลาดเคลื่อนขนาด $n \times 1$

โดยที่ $\underset{\sim}{\varepsilon} \sim N(\underset{\sim}{0}, \sigma_\varepsilon^2 I_n)$, $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$, $COV(\underset{\sim}{\varepsilon}) = \sigma_\varepsilon^2 I_n$

จุดประสงค์สำคัญอย่างหนึ่งของการวิเคราะห์ความถดถอยคือ การพยากรณ์หรือศึกษาเกี่ยวกับแนวโน้มของตัวแปรตาม ดังนั้นจึงจำเป็นที่จะต้องทราบถึงอิทธิพลของตัวแปรอิสระที่มีผลต่อตัวแปรตาม นั้นคือการประมาณค่าพารามิเตอร์หรือเวกเตอร์สัมประสิทธิ์ความถดถอย $\underset{\sim}{\beta}$

การประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณจากตัวแบบดังกล่าวนี้ วิธีที่นิยมใช้มากสุด คือ วิธีกำลังสองน้อยสุดสามัญ (Ordinary Least Square Method) ซึ่งตัวประมาณสัมประสิทธิ์ความถดถอยพหุคูณอยู่ในรูปของ

$$\hat{\underset{\sim}{\beta}}_{OLS} = (X'X)^{-1}X'\underset{\sim}{y} \qquad (1.2)$$

โดยที่ X มีค่าลำดับชั้นเต็ม (full column rank) = p + 1

ตัวประมาณในสมการที่ (1.2) มีคุณสมบัติเป็นตัวประมาณไม่เอนเอียงและให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (mean square error) ต่ำสุดในบรรดาตัวประมาณไม่ เอนเอียงเชิงเส้น แต่การประมาณค่า $\underset{\sim}{\beta}$ ด้วยวิธีกำลังสองน้อยสุดมีข้อสมมุติฐานสำคัญข้อหนึ่งคือตัวแปรอิสระแต่ละตัวต้องไม่มีความสัมพันธ์กับตัวแปรอิสระตัวอื่นซึ่งในทางปฏิบัติจะเป็นไปได้น้อยมาก

เมื่อตัวแปรอิสระมีพหุสัมพันธ์กันสูงจะทำให้เมตริกซ์ $X'X$ เกิดเงื่อนไขที่ไม่ดี (ill – condition) คือทำให้ $|X'X|$ มีค่าเล็กลงเข้าใกล้ศูนย์ เนื่องจากเมตริกซ์ความแปรปรวนร่วมของค่าประมาณสัมประสิทธิ์ความถดถอยพหุคูณอยู่ในรูปของ $Cov(\hat{\underset{\sim}{\beta}}) = \sigma^2(X'X)^{-1}$ จึงส่งผลให้ความแปรปรวนของค่าประมาณสัมประสิทธิ์ความถดถอยมีค่ามาก นั่นคือ ค่าประมาณสัมประสิทธิ์ความถดถอยพหุคูณที่ได้ขาดความแม่นยำ (accuracy) ดังนั้นถ้าตัวแปรอิสระมีความสัมพันธ์กันสูงเราอาจแก้ไขโดยตัดตัวแปรอิสระบางตัวออกจากตัวแบบ แต่ในบางกรณีความสัมพันธ์ระหว่างตัวแปรอิสระไม่ชัดเจน จึงไม่สามารถตัดตัวแปรอิสระตัวใดตัวหนึ่งออกได้เพราะถือว่าตัวแปรอิสระทุกตัวมีผลในการอธิบายตัวแปรตามได้พอสมควร

ดังนั้นเพื่อทำการแก้ปัญหาดังกล่าวที่เกิดขึ้นนี้ จึงมีแนวคิดในการปรับตัวประมาณของพารามิเตอร์โดยการเลือกที่จะใช้ตัวประมาณที่สามารถแก้ปัญหาความสัมพันธ์ของตัวแปรอิสระที่เกิดขึ้น ค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองของตัวแบบเชิงเส้นที่ได้จึงมีความคลาดเคลื่อนลดลงเนื่องจากความสัมพันธ์ของตัวแปรอิสระได้ถูกปรับแก้ แต่ตัวประมาณที่ได้นั้นจะได้เป็นตัวประมาณที่มีความเอนเอียง เพราะตัวแปรต่าง ๆ ที่นำมาศึกษาอาจมีความสัมพันธ์ (Multicollinearity) กัน

Hoerl and Kennard (1970 : 69 – 82) ได้ศึกษาวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่า วิธีกำลังสองน้อยที่สุด โดยให้ชื่อว่า ริดจ์ รีเกรสชั่น (Ridge regression) ซึ่งวิธีนี้ ไม่ต้องตัดตัวแปรอิสระออกจากตัวแบบ ถึงแม้จะเกิดพหุสัมพันธ์ในระหว่างตัวแปรอิสระก็ตาม แต่ค่าประมาณสัมประสิทธิ์ความถดถอยพหุคูณที่ได้จากริดจ์รีเกรสชั่นนี้จะเอนเอียง (Bias)

เนื่องจาก ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของวิธีกำลังสองน้อยที่สุด เป็นฟังก์ชั่นของ $(X'X)^{-1}$ ฉะนั้น การที่จะพยายามลดค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ให้ต่ำลง จึงต้องพยายามลดค่า $(X'X)^{-1}$ ให้ต่ำลง ซึ่ง ริดจ์รีเกรสชั่น อาจทำได้ โดยการบวกค่าคงที่มากกว่าศูนย์กับสมาชิกทุกตัวบนเส้นทแยงมุมของเมตริกซ์ $(X'X)$ ได้ตัวประมาณค่าสัมประสิทธิ์ ความถดถอยพหุคูณโดยวิธี ริดจ์ รีเกรสชั่นดังนี้

$$\hat{\beta}_R = (X'X + kI)^{-1} XY \qquad ; \qquad k > 0$$

นอกจากนี้ Hoerl and Kennard ยังได้กล่าวว่าตัวประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณที่ได้จากวิธี ริดจ์รีเกรสชั่นนี้จะมีลักษณะค่อนข้างคงที่ ค่าสมบูรณ์ของตัวประมาณมีค่าสมเหตุสมผล และเครื่องหมายของค่าประมาณสัมประสิทธิ์ความถดถอยพหุคูณ จะถูกต้อง

ในปี 1973 Conniffe and Stone (1973 : 181 – 187) ได้วิจารณ์เกี่ยวกับการใช้ ริดจ์ รีเกรสชั่น ว่า การเปลี่ยนแปลงของเครื่องหมายของค่าประมาณสัมประสิทธิ์ความถดถอยพหุคูณนี้ เมื่อ $(k = 0)$ แล้ว $\hat{\beta}_R$ มีเครื่องหมายตรงกันข้ามกับค่าประมาณสัมประสิทธิ์ความถดถอยพหุคูณที่ได้จากวิธีกำลังสองน้อยสุดสามัญ

Gunst and Mason (1977 : 616 – 628) มีแนวคิดที่ว่า σ^2 (ความแปรปรวนของความคลาดเคลื่อน) มีค่าน้อย ตัวประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณที่ได้จากวิธีกำลังสองน้อยที่สุดจะเป็นตัวประมาณที่ดีถึงแม้ว่าตัวแปรอิสระจะมีพหุสัมพันธ์กันก็ตาม จึงได้เสนอวิธีความถดถอยองค์ประกอบหลัก (Principal Component Regression : PCR) มาใช้ในการคำนวณค่าสัมประสิทธิ์ความถดถอยพหุคูณเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ โดยการจัด รูป แบบ ตัวแปรอิสระเสียใหม่ก่อนที่จะนำไปหาค่าสัมประสิทธิ์ความถดถอยพหุคูณโดยให้ตัวแปรใหม่เป็นผลบวกเชิงเส้นของตัวแปรเดิม ซึ่งเป็นการลดผลกระทบของการเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ มีผลทำให้ตัวประมาณสัมประสิทธิ์ความถดถอยพหุคูณมีความถูกต้องมากขึ้นจึงช่วยให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์ความถดถอยพหุคูณมีค่าต่ำกว่าวิธีกำลังสองน้อยที่สุด ถึงแม้ว่าตัวประมาณที่ได้จากวิธีนี้มีคุณสมบัติเป็นตัวประมาณที่เอนเอียง (biased estimator)

จากคุณสมบัติในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณทั้ง 3 วิธีข้างต้น เป็นสิ่งที่น่าสนใจว่า เมื่อตัวแปรอิสระเกิดพหุสัมพันธ์กัน การประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณจากวิธีใด ที่จะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยที่สุด และมีประสิทธิภาพมากที่สุด ผู้วิจัยจึงสนใจศึกษาถึงการเปรียบเทียบการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ โดยจะเน้นในการศึกษาการประมาณค่าจากวิธีกำลังสองน้อยที่สุด

หลังจากแก้ไข X ให้ไม่มีความสัมพันธ์กันด้วยวิธี Orthogonalization ซึ่งเป็นวิธีพื้นฐานที่จะได้ตัวประมาณที่ไม่เอนเอียง แล้วนำไปเปรียบเทียบกับวิธีพื้นฐานอื่น ซึ่งได้แก่ วิธีริดจ์รีเกรสชั่น และวิธีความถดถอยองค์ประกอบหลัก

1.2 วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้วัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระซึ่งในที่นี้จะมี 4 วิธีคือ

1. วิธีกำลังสองน้อยที่สุด (Ordinary Least Square Method)
2. วิธีกำลังสองน้อยที่สุด หลังจากแก้ไขด้วยวิธีทำให้เป็นแนวตั้งเชิงตั้งฉาก (Orthogonal Least Square Method)
3. วิธีริดจ์รีเกรสชั่น (Ridge Regression Method)
4. วิธีความถดถอยองค์ประกอบหลัก (principle Component Regression Method)

1.3 สมมติฐานการวิจัย

เมื่อตัวแปรอิสระมีพหุสัมพันธ์กันมากวิธีการประมาณค่าพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุด (Ordinary Least Square Method (OLS)) หลังจากที่ได้ทำการแก้ไขพหุสัมพันธ์ระหว่างตัวแปรอิสระด้วยวิธีแนวตั้งเชิงตั้งฉาก แล้วน่าจะให้ค่าประมาณที่มีความถูกต้องเชื่อถือได้มากกว่าวิธีอื่น ๆ ไม่ว่าขนาดตัวอย่างหรือความแปรปรวนมีค่ามากหรือน้อยก็ตาม

1.4 ข้อตกลงเบื้องต้น

1. รูปแบบทั่วไปของสมความถดถอยมีรูปแบบดังสมการที่ (1.1)

2. ตัวแปรอิสระ $\underset{\sim}{X} = (X_1, X_2, \ldots, X_m)'$ มีการแจกแจงร่วม คือ การแจกแจงปกติของหลายตัวแปร (multivariate normal distribution) เขียนเป็นสัญลักษณ์ $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ โดยที่ $\underset{\sim}{\mu} = E(\underset{\sim}{X})$ และ $\Sigma = \mathrm{cov}(\underset{\sim}{X})$

3. ตัวแปรอิสระ X และตัวแปรตาม y เป็นตัวแปรเชิงปริมาณ

4. ในการวิจัยครั้งนี้ใช้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเป็นตัวตัดสินใจในการเลือกวิธีที่จะใช้ในการประมาณสัมประสิทธิ์ความถดถอยเชิงเส้นพหุคูณเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

1.5 ขอบเขตการวิจัย

1. ตัวแปรอิสระ (p) ที่ใช้ในการวิจัยมี 3 จำนวน คือ 3, 6 และ 9 ตัวแปร
2. ขนาดตัวอย่าง (n) ที่ศึกษาคือ 5p , 10p , 15p , 20p , 25p และ 30p
3. ความคลาดเคลื่อนมีการแจกแจงแบบปกติ (Normal Distribution) เมื่อ $\mu = 0$, $\sigma = 1, 5$ และ 10
4. กำหนดค่าสัมประสิทธิ์ความถดถอยพหุคูณ $(\underset{\sim}{\beta})$ ให้เป็นค่าคงที่ใด ๆ ชุดหนึ่งเพื่อสร้างค่า $\underset{\sim}{y}$ ขึ้นจากตัวแบบ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$
5. ระดับพหุสัมพันธ์ของตัวแปรอิสระแบ่งเป็น 3 ระดับ คือ

ระดับต่ำ	ค่า ρ มีค่าอยู่ในช่วง 0.15 ถึง 0.30
ระดับปานกลาง	ค่า ρ มีค่าอยู่ในช่วง 0.31 ถึง 0.50
	ค่า ρ มีค่าอยู่ในช่วง 0.50 ถึง 0.65
ระดับสูง	ค่า ρ มีค่าอยู่ในช่วง 0.61 ถึง 0.85
	ค่า ρ มีค่าอยู่ในช่วง 0.85 ถึง 0.99

โดยจะสร้างข้อมูลให้มีลักษณะสัมพันธ์กันเป็นไปตามโครงสร้างที่กำหนดด้วย Covariance Matrix อย่างเช่นในกรณี x = 3 จะได้โครงสร้างเมตริกซ์ตั้งต้นดังนี้

$$\text{Covariance Matrix} = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

ซึ่งความสัมพันธ์ระหว่าง x แต่ละตัวนั้น จะอยู่ในช่วงระดับเดียวกัน แต่อาจจะมีค่าแตกต่างกันในตัวแปรแต่ละคู่ แล้วแต่จะสุ่มในช่วงระดับนั้นได้

6. ข้อมูลที่ใช้ในการศึกษาวิจัยครั้งนี้ใช้โปรแกรมคอมพิวเตอร์จำลองด้วยเทคนิคมอนติคาร์โล กระทำซ้ำ n รอบ โดย n จะขึ้นอยู่กับความเหมาะสมในแต่ละสถานการณ์ โดยเงื่อนไขหยุดคือ ค่า AMSE ที่ได้ในแต่ละรอบความแตกต่างกันไม่เกิน 5% ของค่า AMSE เดิม

1.6 เกณฑ์การตัดสินใจ

เกณฑ์ที่ใช้ในการตัดสินใจว่า วิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณวิธีการใดให้ค่าพารามิเตอร์ใกล้เคียงกับค่าจริงมากที่สุดนั้นจะพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (average mean squares error ($AMSE$)) และเกณฑ์ที่ใช้ในการประกอบการตัดสินใจจะใช้ค่าเปอร์เซ็นต์ของอัตราส่วนผลต่างค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ratio of different average mean squares error ($Diff$)) มีสูตรดังนี้

การหาค่าความคลาดเคลื่อนของค่าประมาณสำหรับสัมประสิทธิ์ความถดถอยในแต่ละวิธี

$$MSE_p = (\hat{\beta}_p - \beta_p)^2$$

เมื่อ p คือ ตัวแปรอิสระ

β_p แทนตัวประมาณสัมประสิทธิ์ความถดถอยพหุคูณตัวที่ p ที่ประมาณได้ในครั้งที่ i

$$MSE_m = \frac{\sum_{p=1}^{p+1} MSE_p}{p+1}$$

$$MSEm = \sqrt{MSEm}$$

เมื่อ m คือ วิธีการประมาณค่าสัมประสิทธิ์ความถดถอย

การหาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์ความถดถอยจากการทำซ้ำ

$$AMSE_m = \frac{\sum_{i=1}^{loops} MSE_i}{loops}$$

โดย $loops$ คือจำนวนรอบที่ทำซ้ำ

$AMSE_m$ แทนค่าเฉลี่ยจากค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากสัมประสิทธิ์ความถดถอยพหุคูณเฉลี่ยแต่ละวิธี

การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจากการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธีต่าง ๆ

$$DIFF = \frac{(AMSE_m - AMSE_{\min})}{AMSE_{\min}} \times 100\ \% \qquad ; \quad m = 1,\ 2,\ 3,\ 4$$

โดย $DIFF$ แทนเปอร์เซ็นต์ของอัตราส่วนผลต่างค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจากการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณดังวิธีต่าง ๆ

1.7 วิธีดำเนินการวิจัย

1. สร้างข้อมูลของความคลาดเคลื่อนตามที่กำหนดไว้ในขอบเขตการวิจัย

2. สร้างข้อมูลตัวแปรอิสระ (X) ให้มีระดับพหุสัมพันธ์ที่กำหนดไว้ในขอบเขตการวิจัย และสร้างข้อมูลของตัวแปรตาม (y) จากรูปแบบความสัมพันธ์ $\underset{\sim}{y} = X \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

3. ทำการแปลงข้อมูลที่มีพหุสัมพันธ์โดยการแปลงเมตริกซ์ X เพื่อให้เป็นแนวตั้งเชิงตั้งฉาก ด้วยกระบวนการของแกรมสมิทต์ (Gram-Schmidt Orthogonalization Process) เพื่อมาใช้ประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยสุดสามัญ (OLS) เท่านั้น

4. ประมาณค่าพารามิเตอร์จากข้อมูลที่ไม่ได้ทำการแปลงด้วยวิธีการแกรมสมิทต์ โดยใช้วิธี

   4.1 วิธีริดจ์รีเกรสชั่น (Ridge Regression Method : RID)

   4.2 วิธีความถดถอยองค์ประกอบหลัก (Principle Component Regression Method : PCR)

5 คำนวณหาค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ($AMSE$) ของแต่ละวิธี

6 เปรียบเทียบเปอร์เซ็นต์ของอัตราส่วนผลต่างค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ($Diff$) กับแต่ละวิธี

7 สรุปผลการวิจัยและข้อเสนอแนะ

1.8 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางในการประมาณค่าสัมประสิทธิ์ของตัวแบบถดถอยพหุนามกรณีที่ใช้ตัวประมาณค่าสัมประสิทธิ์ที่สามารถแก้ปัญหาพหุสัมพันธ์ (multicollinearity) และทำให้ได้สมการที่ใช้พยากรณ์ค่าของตัวแปรตามที่มีความคลาดเคลื่อนลดลง

บทที่ 2

ทฤษฎีและสถิติที่ใช้ในการวิจัย

การวิเคราะห์ความถดถอยพหุคูณ คือ การวิเคราะห์ รูปแบบหรือสมการของความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม โดยแสดงให้เห็นว่า ตัวแปรอิสระแต่ละตัวมีความสัมพันธ์กับตัวแปรตามหรือไม่มากน้อยเพียงใด และความสัมพันธ์นั้นเป็นไปในเชิงบวก หรือเป็นไปในเชิงลบ เทคนิคการวิเคราะห์จะทดสอบว่าตัวแปรอิสระแต่ละตัวมีผลต่อตัวแปรตามหรือไม่และตัวแปรอิสระทุกตัวมีผลต่อตัวแปรตามมากน้อยเพียงใด ในการใช้เทคนิคดังกล่าวผู้วิจัยจะต้องมีข้อสมมติว่า ตัวแปรอิสระแต่ละตัวเป็นอิสระซึ่งกันและกัน

เนื่องจาก การวิเคราะห์ความถดถอยพหุคูณมีเป้าหมายที่จะศึกษาอิทธิพลของตัวแปรอิสระแต่ละตัวที่มีต่อตัวแปรตาม เมื่อพิจารณาร่วมกับตัวแปรอิสระอื่น ๆ ที่นำมาวิเคราะห์ โดยดูจากค่าสัมประสิทธิ์ความถดถอยพหุคูณ และเพื่อใช้ในการประมาณค่าของตัวแปรตามเมื่อกำหนดให้ตัวแปรอิสระเปลี่ยนแปลงไป ความไม่คลาดเคลื่อนและความแม่นยำของค่าประมาณสัมประสิทธิ์ความถดถอยพหุคูณเมื่อใช้วิธีกำลังสองน้อยที่สุด จึงเป็นสิ่งที่มีความสำคัญมาก

เมื่อตัวแปรอิสระมีพหุสัมพันธ์กันสูง จะทำให้ค่าประมาณสัมประสิทธิ์ความถดถอยที่ได้จากวิธีกำลังสองน้อยที่สุด ขาดความแม่นตรง การหลีกเลี่ยงปัญหาการเกิดพหุสัมพันธ์ในตัวแปรอิสระโดยการตัดตัวแปรใดตัวแปรหนึ่งออกไป อาจทำให้ค่าประมาณสัมประสิทธิ์ความถดถอยพหุคูณที่ได้คลาดเคลื่อนไปและความคลาดเคลื่อนนี้จะมากหากตัวแปรที่ถูกตัดทิ้งไปมีความสัมพันธ์สูงกับตัวแปรอื่น ๆ ที่มีอยู่ในสมการ

โดยสรุปเทคนิคการวิเคราะห์ความถดถอยพหุคูณ เป็นเทคนิคที่มีประโยชน์ในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระหลาย ๆ ตัวโดยค่าสัมประสิทธิ์ความถดถอยพหุคูณที่ได้จะบอกให้ทราบว่าตัวแปรอิสระแต่ละตัวมีผลต่อตัวแปรตามมากน้อยเพียงใด ทั้งนี้ต้องคำนึงถึงตัวแปรอิสระตัวอื่น ๆ ด้วย และค่าความผิดพลาดมาตรฐานของค่าสัมประสิทธิ์ความถดถอย จะบอกให้ทราบว่าค่าสัมประสิทธิ์ความถดถอยที่ได้นั้นมีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญจริงหรือไม่

ในบทนี้จะกล่าวถึงการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณของตัวแปรอิสระโดยใช้วิธีกำลังสองน้อยสุดสามัญ (ordinary least square method(OLS)) วิธีริดจ์รีเกรสชั่น (Ridge regression method (RID)) และวิธีความถดถอยองค์ประกอบหลัก (principal component regression (PCR))

2.1 ตัวแบบทั่วไป

ตัวแบบทั่วไปของการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ สามารถเขียนได้ดังนี้

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

เมื่อ $\underset{\sim}{y}$ เป็นเวกเตอร์ของตัวแปรตามขนาด $n \times 1$

X เป็นเมตริกซ์ของตัวแปรอิสระขนาด $n \times (p+1)$

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของสัมประสิทธิ์ความถดถอยพหุคูณขนาด $(p+1) \times 1$

$\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของความคลาดเคลื่อนขนาด $n \times 1$

n เป็นขนาดตัวอย่าง

และ p เป็นจำนวนตัวแปรอิสระ

โดยที่ $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ และ $Cov(\underset{\sim}{\varepsilon}) = \sigma^2 I_n$

2.2 **วิธีกำลังสองน้อยสุดสามัญ** (Ordinary Least Squares Method)

การประมาณค่าพารามิเตอร์ด้วยตัวประมาณกำลังสองน้อยสุดใช้หลักการประมาณค่าพารามิเตอร์ที่ทำให้ผลรวมของความคลาดเคลื่อนกำลังสอง (Sum Squares Error: SSE) มีค่าน้อยที่สุด

ให้ $\hat{\beta}_{\sim OLS}$ เป็น เวกเตอร์ของตัวประมาณแบบกำลังสองน้อยสุดของค่าพารามิเตอร์ $\underset{\sim}{\beta}$ และ $\underset{\sim}{e}$ เป็น เวกเตอร์ของส่วนเหลือที่เป็นตัวประมาณของความคลาดเคลื่อน $\underset{\sim}{\varepsilon}$ เมื่อแทนที่ $\underset{\sim}{\beta}$ ด้วย $\hat{\beta}_{\sim OLS}$ และแทนที่ $\underset{\sim}{\varepsilon}$ ด้วย $\underset{\sim}{e}$ ในตัวแบบเชิงเส้น จะได้

(2.2.1) $$\underset{\sim}{y} = X\hat{\beta}_{\sim OLS} + \underset{\sim}{e}$$

และ $$\underset{\sim}{e} = \underset{\sim}{\hat{\varepsilon}} = \underset{\sim}{y} - X\hat{\beta}_{\sim OLS}$$

เมื่อพิจารณา ผลบวกของความคลาดเคลื่อนกำลังสอง

$$SSE = \left(\underset{\sim}{e}'\underset{\sim}{e}\right)$$

$$= \left(\underset{\sim}{y} - X\hat{\beta}_{\sim OLS}\right)'\left(\underset{\sim}{y} - X\hat{\beta}_{\sim OLS}\right)$$

(2.2.2) $$= \underset{\sim}{y}'\underset{\sim}{y} - 2\hat{\beta}_{\sim OLS}'X'\underset{\sim}{y} + \hat{\beta}_{\sim OLS}'X'X\hat{\beta}_{\sim OLS}$$

หาค่าของ $\hat{\beta}_{\sim OLS}$ ที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนมีค่าต่ำสุด โดยการหาอนุพันธ์อันดับที่ 1 ของ SSE เทียบกับ $\hat{\beta}_{\sim OLS}$ แล้วให้เท่ากับ $\underset{\sim}{0}$ ได้ว่า

$$\frac{\partial}{\partial \hat{\beta}_{\sim OLS}} SSE = \underset{\sim}{0}$$

$$-2X'\underset{\sim}{y} + 2X'X\hat{\beta}_{\sim OLS} = 0$$

$$X'X\hat{\beta}_{\sim OLS} = X'\underset{\sim}{y}$$

(2.2.3) $$\hat{\beta}_{\sim OLS} = (X'X)^{-1}X'\underset{\sim}{y}$$

ตัวประมาณที่ได้ในสมการที่ (2.2.3) นี้มีคุณสมบัติเป็นตัวประมาณไม่เอนเอียงกล่าวคือ

$$E(\hat{\underset{\sim}{\beta}}_{OLS}) = E[(X'X)^{-1}X'\underset{\sim}{y}]$$

$$= E[(X'X)^{-1}X'(X\underset{\sim}{\beta}+\underset{\sim}{\varepsilon}]$$

$$= E[\beta]+(X'X)^{-1}X'E(\underset{\sim}{\varepsilon})$$

$$= \underset{\sim}{\beta}_{OLS}$$

ตัวประมาณ $\underset{\sim}{\beta}_{OLS}$ จะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุดในบรรดาตัวประมาณไม่เอนเอียงเชิงเส้น แต่ในการประมาณค่า $\underset{\sim}{\beta}$ ด้วยวิธีกำลังสองน้อยสุดมีข้อสมมุติฐานสำคัญข้อหนึ่งคือตัวแปรอิสระแต่ละตัวต้องไม่มีความสัมพันธ์กับตัวแปรอิสระตัวอื่นซึ่งในทางปฏิบัติจะเป็นไปได้น้อยมาก เมื่อตัวแปรอิสระมีพหุสัมพันธ์กันสูงจะทำให้เมตริกซ์ $X'X$ เกิดเงื่อนไขที่ไม่ดี (ill – condition) อาจมีผลทำให้การประมาณ $\underset{\sim}{\beta}$ ด้วยวิธีกำลังสองน้อยสุดไม่ได้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด ดังนั้นจึงควรพิจารณาคุณสมบัติของตัวประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธีกำลังสองน้อยที่สุด 2 ประการคือ

1. เมตริกซ์ความแปรปรวนของตัวประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธีกำลังสองน้อยที่สุด

2. ค่าเฉลี่ยกำลังสองของความแตกต่าง $\underset{\sim}{\beta}$ และ $\hat{\underset{\sim}{\beta}}$ (ระยะทางจาก $\underset{\sim}{\beta}$ ไปยัง $\hat{\underset{\sim}{\beta}}$) ซึ่งสามารถเขียนให้อยู่ในรูปฟังก์ชันของ $X'X$ และ σ^2

(2.2.4) $$Cov(\hat{\underset{\sim}{\beta}}) = \sigma^2(X'X)^{-1}$$

ให้ L_1 คือระยะทางจาก $\hat{\underset{\sim}{\beta}}$ ไปยัง $\underset{\sim}{\beta}$ ดังนั้น

(2.2.5) $$L_1^2 = (\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})$$

ดังนั้นจะได้ค่าเฉลี่ยของกำลังสองระยะทางจาก $\hat{\underset{\sim}{\beta}}$ ไปยัง $\underset{\sim}{\beta}$ ในรูปของ

(2.2.6) $$E(L_1^2) = \sigma^2 trace(X'X)^{-1}$$

$$E(L_1^2) = E[(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})]$$

$$= E[\hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}}]-\underset{\sim}{\beta}'\underset{\sim}{\beta}$$

(2.2.7) $$E(\hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}}) = \underset{\sim}{\beta}'\underset{\sim}{\beta}+\sigma^2 trace(X'X)^{-1}$$

เมื่อ $\underset{\sim}{\varepsilon}$ มีการแจกแจงแบบปกติจะได้ว่า

(2.2.8) $$Var(L_1^2) = 2\sigma^4 trace(X'X)^{-2}$$

จากสมการที่ (2.2.4) , (2.2.6) และ (2.2.8) พบว่า $Cov(\underset{\sim}{\hat{\beta}})$, $E[L_1^2]$ และ $Var(L_1^2)$ ต่างก็เป็นฟังก์ชันของเมตริกซ์ $X'X$ ดังนั้นเพื่อความสะดวกในการทำความเข้าใจเราจึงแปลงเมตริกซ์ $X'X$ ให้อยู่ในรูปของค่าเฉพาะ (eigenvalue) ของเมตริกซ์ $X'X$ โดยใช้ทฤษฎีที่สำคัญข้อหนึ่งคือ ถ้า λ_i เป็นค่าเฉพาะของเมตริกซ์ $X'X$ แล้ว $\sum_{i=1}^{p}\lambda_i = trace(X'X)$; $i = 1, 2, ..., p$ เมื่อ p เป็นจำนวนตัวแปรอิสระ

กำหนดให้ค่าเฉพาะของเมตริกซ์ $X'X$ มีค่าเป็น

$$(\lambda_{\max} = \lambda_1) \geq \lambda_2 \geq ... \geq (\lambda_{\min} = \lambda_p) \qquad ; \ \lambda_1 \ , \lambda_2 \ , ... , \lambda_p \ > 0$$

จากสมการที่ (2.2.8) เราสามารถเขียนค่าเฉลี่ยของกำลังสองระยะทางจาก $\underset{\sim}{\hat{\beta}}$ ไปยัง $\underset{\sim}{\beta}$ ในรูปฟังก์ชันของค่าเฉพาะได้ดังนี้

(2.2.9)
$$E(L_1^2) = \sigma^2 \sum_{i=1}^{k}\left(\frac{1}{\lambda_i}\right)$$

และจากสมการที่ (2.2.9) เราสามารถเขียนค่าความแปรปรวนของกำลังสองระยะทางจาก $\underset{\sim}{\hat{\beta}}$ ไปยัง $\underset{\sim}{\beta}$ อยู่ในรูปฟังก์ชันของค่าเฉพาะได้ดังนี้

(2.2.10)
$$Var(L_1^2) = 2\sigma^4 \sum_{i=1}^{k}\left(\frac{1}{\lambda_i}\right)^2$$

ในกรณีที่ตัวแปรอิสระมีสภาพไม่เหมาะสม กล่าวคือเกิดความสัมพันธ์ระหว่างตัวแปรอิสระในอัตราที่สูงจะทำให้ $|X'X|$ มีค่าเล็กลงเข้าใกล้ศูนย์ เนื่องจาก $|X'X|$ มีค่าเท่ากับผลคูณของค่าเฉพาะของเมตริกซ์ $X'X$ จึงส่งผลให้ค่าเฉพาะบางค่าต่ำมาก ดังนั้นจากสมการที่ (2.2.9) และ (2.2.10) เราจะเห็นได้ว่า $E(L_1^2)$ และ $E(L_1^2)$ จึงมีค่าสูงขึ้นตามไปด้วย นอกจากนี้การเกิดความสัมพันธ์ระหว่างตัวแปรอิสระส่งผลให้ความแปรปรวนของค่าประมาณสัมประสิทธิ์ความถดถอยมีค่ามากและเกิดความสัมพันธ์กันสูง ระหว่างสัมประสิทธิ์ความถดถอยที่ใช้ประมาณค่า [1]

2.3 วิธีริดจ์รีเกรสชั่น

2.3.1 ทฤษฎีและแนวคิด

ในกรณีที่ตัวแปรอิสระในสมความถดถอยมีสหสัมพันธ์กันอย่างสูงนั้น ตัวประมาณกำลังสองน้อยสุดจะให้ค่าประมาณที่ขาดความแม่นยำ (imprecise) ลักษณะของสหสัมพันธ์อาจจะเป็นคู่ของตัวแปร หรือ อาจเป็นลักษณะที่ตัวแปรอิสระตัวหนึ่ง เป็นผลรวมเชิงเส้น (linear combination) ของตัวแปรอิสระตัวอื่น ๆ ในสมการความถดถอย ซึ่งในสถานการณ์เช่นนี้ จะมีตัวแปรอิสระอย่างน้อย 1 ตัวที่มีค่าความคลาดเคลื่อนมาตรฐานสูง ซึ่งจะส่งผลให้การทดสอบสมมติฐานโดยใช้การทดสอบที (t - test) ของสัมประสิทธิ์ความถดถอยของตัวแปรดังกล่าว มีโอกาสผิดพลาดสูง

ในปี ค.ศ. 1970 โฮเอิล (Hoerl) และ เคนนาร์ด (Kennard) ได้ศึกษาหาตัวประมาณค่าพารามิเตอร์ในตัวแบบความถดถอยพหุคูณเชิงเส้น ที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าตัวประมาณกำลังสองน้อยสุด เมื่อข้อมูลเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ โดยตัวประมาณนี้สร้างจากหลักการนำค่าคงที่ค่าหนึ่งที่มากกว่าศูนย์ (k) มาบวกกับสมาชิกทุกตัวในแนวทแยงมุมของเมตริกซ์ $X'X$ จะทำให้เมตริกซ์ $X'X$ ลดสภาพความเป็นเมตริกซ์ที่ไม่ดี (ill - condition) ลงได้ ก่อนจะหาเมตริกซ์ผกผันของ $X'X$ ดังกล่าว

สมการปกติ (Normal Equation) ของตัวประมาณความถดถอยริดจ์ สามารถเขียนได้ในรูปแบบดังนี้

(2.3.1.1) $$(X'X + kI_n)\underset{\sim}{\hat{\beta}}_{RID} = X'\underset{\sim}{y}$$

ดังนั้น [2]ตัวประมาณความถดถอยริดจ์ (Ridge Regression Estimater) เขียนได้ในรูป

(2.3.1.2) $$\underset{\sim}{\hat{\beta}}_{RID} = (X'X + kI_n)^{-1}X'\underset{\sim}{y} \quad , \quad k > 0$$

เมื่อ k เป็นค่าคงที่ หรือเรียกว่าเป็น พารามิเตอร์ที่เอนเอียง (biasing parameter)

และ I เป็นเมตริกซ์เอกลักษณ์ (Identity Matrix) ที่มีขนาด $(q \times q)$

ตัวประมาณความถดถอยริดจ์เป็นตัวประมาณที่เอนเอียง แต่จะมีค่าความแปรปรวนน้อยกว่าตัวประมาณกำลังสองน้อยสุด

Hoerl and Kennard ได้ใช้วิธีการทางคณิตศาสตร์มาควบคุมค่าความแปรปรวนของตัวประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณ โดยวิธีกำลังสองน้อยที่สุดให้มีค่าต่ำสุดได้ดังนี้

(2.3.1.3) $$\hat{\beta}_R = (X'X + kI)^{-1} X'Y$$

ให้ $W = (X'X + kI)^{-1}$

(2.3.1.4) $$\therefore \quad \hat{\beta}_R = WX'Y$$

ตัวประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณ โดยวิธีริดจ์ รีเกรสชั่น สามารถเขียนให้อยู่ในรูปฟังก์ชั่นของตัวประมาณ ค่าสัมประสิทธิ์ความถดถอยพหุคูณ โดยวิธีกำลังสองน้อยที่สุด เพื่อความสะดวกในการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณที่คำนวณได้จากสองวิธีนี้

จากสมการ 2.3.1.4 อาจเขียนได้ว่า

$$\hat{\beta}_R = \left[I + k(X'X)^{-1}\right]^{-1} \hat{\beta} = z\hat{\beta}$$

โดยที่ $z = \left[I + k(X'X)^{-1}\right]^{-1}$ และ $W = (X'X + kI)^{-1}$

ได้ $E_i(W)$ และ $e_i(z)$ เป็น eigenvalue ของ W และ Z ตามลำดับ ซึ่งได้จากการแก้สมการ Characteristic equation

$$|W - E_i I| = 0 \quad \text{และ} \quad |Z - E_i I| = 0$$

$$E_i(Z) = \frac{1}{(\lambda_i + k)}$$

$$E_i(W) = \frac{1}{(\lambda_i + k)}$$

อาจเขียน Z ให้อยู่ในรูปฟังก์ชันของ kW ได้ดังนี้

$$Z = I - k(X'X + kI)^{-1}$$

$$= I - kW$$

ค่าประมาณ $\hat{\beta}_R$ จะมีค่าน้อยกว่า β เมื่อ $k > 0$

$$\hat{\beta}'_R \beta_R < \hat{\beta}'\beta$$

<u>พิสูจน์จากนิยาม</u> : $\hat{\beta}_R = z\hat{\beta}$; $X'X$ และ Z มีคุณสมบัติเป็น Symmetric positive definite ฉะนั้นจะได้

$$\hat{\beta}_R'\beta_R = (Z\hat{\beta})'(Z\hat{\beta})$$
$$= \sum_{i=1}^{p} E_i^2(Z)\hat{\beta}'\hat{\beta}$$
$$< E_{i\ (man)}^2(Z)\hat{\beta}'\beta$$

$E_{\max}(z) = \dfrac{\lambda_l}{(\lambda_l + k)}$; λ_l เป็น eigenvalue ที่มีค่ามากที่สุดของเมตริกซ์ $X'X$ ดังนั้น $\hat{\beta}_R'\hat{\beta}_R < \hat{\beta}'\hat{\beta}$

ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณ โดยวิธีริดจ์ รีเกรสชั่นจะมีค่าเป็น

$$MSE(\hat{\beta}_R) = E\left[(\hat{\beta}_R - \beta)^2\right]$$
(2.3.1.5) $$= Var(\hat{\beta}_R) + \left[E(\hat{\beta}_R) - \beta\right]^2$$

การที่ยอมให้เกิดความเอนเอียงขึ้นใน $\hat{\beta}_R$ จะทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ $\hat{\beta}_R$ มีค่าน้อยกว่าความแปรปรวนของตัวประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณที่ไม่เอนเอียง สิ่งซึ่งส่งผลตามมาก็คือช่วงความเชื่อมั่นของตัวประมาณ $\hat{\beta}_R$ จะแคบกว่า ช่วงความเชื่อมั่นของตัวประมาณ $\hat{\beta}$ แสดงว่า $\hat{\beta}_R$ เป็นตัวประมาณที่ค่อนข้างคงที่ (Stable) มากกว่า $\hat{\beta}$ จากสมการ 2.3.1.5 อาจเขียนได้ว่า

$$MSE(\hat{\beta}_R) = \sigma^2 trace(X'X + kI)X'X(X'X + kI)^{-1} + k^2\beta'(X'X + kI)^{-2}\beta$$
$$= \sigma^2\sum_{j=1}^{p}\frac{\lambda_j}{(\lambda_j + k)^2} + k^2\beta'(X'X + kI)^{-2}\beta$$

ในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วย วิธีริดจ์ รีเกรสชั่น จะต้องเลือกค่า k ซึ่งให้ค่าความแปรรวนของตัวประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณลดลงมากกว่าการเพิ่มขึ้นของความเอนเอียงยกกำลังสอง ถ้าทำได้เช่นนี้จะทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณโดยวิธี ริดจ์ รีเกรสชั่น มีค่าน้อยกว่าความแปรปรวนตัวประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณ โดยวิธีกำลังสองน้อยที่สุด

ค่าความคลาดเคลื่อนกำลังสองของตัวประมาณค่าตัวแปรตาม จะมีค่าเป็น

$$SSE = (Y - X\hat{\beta}_R)'(Y - X\hat{\beta}_R)$$
$$= (Y - X\hat{\beta}_R)'(Y - X\hat{\beta}_R) + (\hat{\beta}_R - \hat{\beta})'X'X(\hat{\beta}_R - \hat{\beta})$$

2.3.2 การประมาณค่าพารามิเตอร์ k

ในปี ค.ศ. 1970 Hoerl and Kennard ได้เสนอการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธี ริดจ์ รีเกรสชั่น ยังไม่สามารถหาค่า k ที่แน่นอนได้ นอกจากจะต้องมีการทดลองให้ k มีค่าเพิ่มขึ้น จากศูนย์ทีละน้อยและแทนค่า k ลงในสมการ 2.3.1.1 จนกว่าจะได้ค่า k ที่เหมาะสม และทุกครั้งที่กำหนดค่า k เพื่อคำนวณค่า $\hat{\beta}_R$ จะนำค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณที่ได้จากวิธี ริดจ์ รีเกรสชั่น และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณที่ได้จากวิธีกำลังสองน้อยที่สุดมาเปรียบเทียบกัน ซึ่งจะทำเช่นนี้ไปเรื่อย ๆ จนกว่าค่า k ใดที่จะทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณ $\hat{\beta}_R$ มีค่าน้อยกว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณ β และจะถือว่าค่า k นั้นเริ่มจะเป็นค่า k ที่เหมาะสมจากนั้นจะค่อย ๆ เพิ่มค่า k ไปทีละน้อยอีกจนกว่าจะได้ k ที่เหมาะสมที่สุด ซึ่งจะได้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์ความถดถอยพหุคูณมีค่าต่ำสุด จะเห็นว่าจุดอ่อนของริดจ์ รีเกรสชั่น ก็คือ ไม่สามารถกำหนดค่า k ที่แน่นอนได้ แต่ต่อมาได้มีผู้พยายามจะประมาณค่า k ให้เหมาะสมยิ่งขึ้น ซึ่งวิธีนี้ก็ต้องเอาค่า k ไปปรับอีกเช่นกันเพียงแต่ต่างกันที่ค่าเริ่มต้นของ k เท่านั้น ซึ่งมีผู้เสนอวิธีการหาไว้หลายวิธี แต่ในการวิจัยเรื่องนี้ จะทำการศึกษา ริดจ์ รีเกรสชั่นโดยการกำหนดค่าของ k ด้วยวิธีของ Hoerl, Kennard (1970 : 105 - 123) ซึ่งกล่าวไว้ว่า ค่า k ที่เหมาะสมจะเป็น

(2.3.2.1)
$$k = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$$

โดยที่ p คือจำนวนตัวแปรอิสระ $\hat{\beta}$ และ $\hat{\sigma}^2$ เป็นค่าประมาณซึ่งได้จากวิธีกำลังสองน้อยที่สุด เขาได้แสดงให้เห็นว่าค่าเฉลี่ยคลาดเคลื่อนกำลังสองของ วิธี ริดจ์ รีเกรสชั่น ที่ลดลงจากวิธีกำลังสองน้อยที่สุด จะมีนัยสำคัญ

นอกจากจะใช้วิธี ริดจ์ รีเกรสชั่นในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณ เมื่อตัวแปรอิสระมีพหุสัมพันธ์กันแล้ว ก็อาจใช้วิธีเกรสชั่นพริ้นซีเปิ้ลคอมโพเน้นท์ (Regression principal component) ซึ่งจะได้ตัวประมาณที่เอนเอียง และให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าวิธีกำลังสองน้อยที่สุด

2.4 วิธีการถดถอยองค์ประกอบหลัก

วิธีนี้ได้นำการวิเคราะห์องค์ประกอบหลัก (Principal component analysis) มาช่วยในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณ เมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระด้วยการจัดรูปแบบตัวแปรอิสระเสียใหม่ก่อนที่จะนำไปหาค่าสัมประสิทธิ์ความถดถอยพหุคูณโดยให้ตัวแปรใหม่เป็นผลบวกเชิงเส้นของตัวแปรเดิม

2.4.1 แนวคิดพื้นฐาน

ความถดถอยองค์ประกอบหลักเป็นเทคนิคการประมาณที่เอนเอียงวิธีหนึ่งที่ใช้แก้ปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระ ในวิธีนี้จะใช้การประมาณด้วยวิธีกำลังสองน้อยสุดกับกลุ่มของตัวแปรที่เรียกว่า องค์ประกอบหลัก (principle component) ของเมตริกซ์สหสัมพันธ์ โดยมีการจำกัดจำนวนองค์ประกอบหลักที่แน่นอนซึ่งมีผลต่อการลดความแปรปรวนลง แต่จะเป็นตัวประมาณที่เอนเอียง โดยในกรณีที่มีพหุสัมพันธ์ระหว่างตัวแปรอิสระ วิธีนี้จะให้ผลลัพธ์ในการประมาณค่าและการพยากรณ์ค่าได้ดีกว่าวิธีกำลังสองน้อยสุด

ในกรณีตัวแปร X และ Y มีความสัมพันธ์กันสูง และเป็นความสัมพันธ์เชิงเส้นตรงในทางบวก ถ้าสามารถกำหนดความลาดชันของเส้นตรงได้ ก็จะสามารถหาค่าของ X เมื่อรู้ค่าของ Y และสามารถหาค่าของ X เมื่อรู้ค่าของ Y ได้ ซึ่งเส้นตรงที่ได้นี้จะเรียกว่า แกนหลัก (Principal axis) ถ้าจุดต่าง ๆ อยู่บนเส้นตรงแกนหลักทั้งหมด แกนหลักก็สามารถที่จะให้ข้อมูลเกี่ยวกับ X และ Y ได้อย่างถูกต้องสมบูรณ์ทุกค่า แต่ถ้าจุดแสดงค่า x และ y กระจายออกไปมาก ก็ต้องอาศัยแกนเพิ่มอีกหนึ่งแกน ซึ่งแกนที่เพิ่มขึ้นนี้จะต้องมีจุดเริ่มต้นตั้งฉากกับแกนหลัก แกนหลักจะลากผ่านจุดต่าง ๆ ที่ทำให้ระยะทางระหว่างจุดกันแกนหลัก (โดยการลากเส้นจากจุดมาตั้งฉากกับแกนหลัก) สั้นที่สุด และทำให้ผลรวมของระยะทางยกกำลังสองมีค่าต่ำสุด

การหาค่าต่ำสุดของแกนหลักนี้ แตกต่างจากวิธีกำลังสองน้อยที่สุด ซึ่งพยายามที่จะหาเส้นตรง $y = a + bx$ และพยายามที่จะหาค่า $\hat{Y}$ ที่ประมาณได้ต่างจาก Y เดิมให้น้อยที่สุด กล่าวคือ $\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2$ มีค่าน้อยที่สุด การลากเส้นระยะทางตามวิธีกำลังสองน้อยที่สุดเป็นการลากเส้นขนานกับแกน แต่การลากเส้นระยะทางตามวิธีการวิเคราะห์องค์ประกอบหลักจะลากเส้นตั้งฉากกับแกนหลัก

ถ้ามีจำนวนตัวแปรเพิ่มขึ้น จำนวนมิติของกราฟก็จะเพิ่มขึ้น เช่นถ้ามี 3 ตัวแปร ก็ต้องเพิ่มเส้นแสดงมิติเพิ่มอีกหนึ่งเส้นและในการลงจุดก็ต้องคำนึงถึงค่าของตัวแปร 3 ตัวพร้อม ๆ กัน ในการหาแกนหลักก็ต้องหาแกนซึ่งสามารถอธิบายความผันแปรของตัวแปรทั้ง 3 ตัวให้ได้มากที่สุดและแกนต่อ ๆ ไป ก็จะต้องอธิบายความผันแปรที่เหลือให้ได้มากที่สุด

วิธีการแยกแกนตามลำดับจากแกนหลักไปยังแกนรอง หรือการลดตัวแปรตามวิธีการวิเคราะห์องค์ประกอบหลักนี้ อาศัยสมการที่เรียกว่า Eigen – equation ซึ่งสามารถเขียนในรูปเมตริกซ์ได้ดังนี้

(2.4.1.1) $$R\lambda = V$$

โดยที่ R คือ เมตริกซ์ที่ต้องการหา

V คือ เมตริกซ์ของ eigenvector

λ คือ eigenvalue

การแก้สมการ 2.4.1.1 จะได้ eigenvalue และ eigenvector ที่สัมพันธ์กับเมตริกซ์ข้อมูล ซึ่งวิธีการแก้สมการนี้ อาศัยสมการที่เรียกว่า Determinant equation

(2.4.1.2) $$Det(R - I\lambda) = 0$$

ถ้าเมตริกซ์ประกอบด้วย 2 ตัวแปร จะได้

(2.4.1.3) $$Det\begin{bmatrix} 1-\lambda & r_{12} \\ r_{21} & 1-\lambda \end{bmatrix} = 0$$

ซึ่งสามารถกระจายเป็นสมการ ได้ดังนี้

(2.4.1.4) $$(1-\lambda)(1-\lambda) - r_{12}r_{21} = 0$$

(2.4.1.5) $$\lambda^2 + 2\lambda + 1 - r_{12}^2 = 0$$

ในการแก้สมการ 2.4.1.5 ก็ทำได้ในทำนองเดียวกันการแก้สมการ $ax^2 + bx + c = 0$ ได้ eigenvalues ของเมตริกซ์ ความสัมพันธ์ระหว่างตัวแปร 2 ตัวคือ $\lambda = 1 + r_{12}$ และ $\lambda = 1 - r_{12}$ ถ้าตัวแปร 2 ตัวมีความสัมพันธ์กันแบบสมบูรณ์ $(r_{12} = 1)$ eigenvalue ค่าที่หนึ่ง (λ_1) จะมีค่าเป็นสอง และ eigenvalue ค่าที่สอง (λ_2) มีค่าเป็นศูนย์ แต่ถ้าตัวแปรอิสระไม่มีความสัมพันธ์กันเลย $(r_{12} = 0)$ eigenvalue ทั้งสองค่า (λ_1 และ λ_2) มีค่าเท่ากับหนึ่ง ซึ่งจะพบว่าผลรวมของ λ_1 และ λ_2 จะเท่ากับจำนวนตัวแปรอิสระ ส่วนผลคูณของ λ_1 และ λ_2 จะเท่ากับ $(1 - r_{12}^2)$ ก็คือค่า Determinant ของเมตริกซ์ความสัมพันธ์

แกนหลักคือ แกนที่ eigenvaule มีค่ามากที่สุด ส่วนแกนของ eigenvalue จะมีค่ารองลงมาตามลำดับ ผลรวมของ eigenvalue ทั้งหมดจะเท่ากับจำนวนตัวแปร ถ้าหาร eigenvalue ค่าแรกด้วยจำนวนตัวแปรจะได้ค่าของความผันแปรของตัวแปรที่อธิบายได้โดยแกนหลักและค่าผลหารต่อ ๆ ไปคือค่าของความผันแปรที่อธิบายได้โดยแกนรองตามลำดับ อาจสรุปได้ดังนี้

สัดส่วนของความผันแปรของแกนแต่ละแกนเท่ากับ eigenvalue ของแต่ละแกนหารด้วยจำนวนตัวแปร น้ำหนักของตัวแปรที่มีต่อองค์ประกอบ (Principal component loading) คือ อัตราความผันแปรของตัวแปรที่อธิบายได้โดยแกนหลักและแกนรอง ซึ่งเท่ากับผลคูณระหว่าง eigenvectors และรากที่สองของ eigenvalue ของแกน

ตัวประมาณที่เอนเอียงของค่าสัมประสิทธิ์ความถดถอยพหุคูณ อาจหาได้จากวิธีรีเกรสชั่นพริ้นซีเปิ้ลคอมโพเน้นท์ จากตัวแบบทั่วไปของการวิเคราะห์ความถดถอยพหุคูณ

ตัวแบบและวิธีการวิเคราะห์ความถดถอยองค์ประกอบหลัก

พิจารณาเมตริกซ์ของเวกเตอร์เฉพาะ (eigenvectors) ที่สอดคล้องกับค่าเฉพาะ (eigenvalue) $\lambda_1, \lambda_2, \ldots, \lambda_p$ ของ $X'X$ โดยที่ $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ และทราบว่า $T'T = I$ โดยที่ T เป็นเมตริกซ์เชิงตั้งฉาก (orthogonal matrix) ดังนั้นจึงสามารถเขียนตัวแบบความถดถอยจากตัวแบบเดิม คือ

(2.4.1.6) $$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

ไปเป็นตัวแบบ

(2.4.1.7) $$\underset{\sim}{y} = Z\underset{\sim}{\alpha} + \underset{\sim}{\varepsilon}$$

$$Z = XT \qquad \text{และ} \qquad \underset{\sim}{\alpha} = T\underset{\sim}{\beta}$$

$$Z'Z = T'X'XT = \Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_p)$$

เมื่อ $\underset{\sim}{y}$ เป็นเวกเตอร์ซึ่งมีสมาชิกอยู่ในเทอมค่ามาตรฐาน มีขนาด $n \times 1$

X เป็นเมตริกซ์ซึ่งมีสมาชิกอยู่ในเทอมค่ามาตรฐาน มีขนาด $n \times p$

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของสัมประสิทธิ์ความถดถอยพหุคูณ มีขนาด $p \times 1$

T เป็นเมตริกซ์ตั้งฉาก ขนาด $p \times p$ โดยที่แต่ละคอลัมน์ของเมตริกซ์เป็นเวกเตอร์เฉพาะที่สอดคล้องกับค่าเฉพาะ λ_1 , λ_2 , ... , λ_p ตามลำดับ

Z เป็นเมตริกซ์ซึ่งแต่ละสดมภ์คือตัวแปรอิสระชุดใหม่ซึ่งเกิดจากผลรวมเชิงเส้น (linear combination) ของตัวแปรอิสระอาจกล่าวได้ว่าแต่ละสดมภ์ประกอบไปด้วย

$\underset{\sim}{Z_1}$ คือ องค์ประกอบหลักที่ 1

$\underset{\sim}{Z_2}$ คือ องค์ประกอบหลักที่ 2

$\vdots$

$\underset{\sim}{Z_p}$ คือ องค์ประกอบหลักที่ p

Λ เป็นเมตริกซ์ในแนวทแยงมุม (diagonal matrix) ที่มีขนาด $p \times p$ มีสมาชิกในแนวเส้นทแยงมุมเป็นค่าเฉพาะของเมตริกซ์ $X'X$

จากสมการที่ (2.4.1.7) เราสามารถคำนวณค่า $\underset{\sim}{\hat{\alpha}}$ โดยใช้วิธีกำลังสองน้อยที่สุดประมาณค่าดังนี้

(2.4.1.8) $$\underset{\sim}{\hat{\alpha}} = (Z'Z)^{-1}Z'\underset{\sim}{y}$$

เมื่อ $\underset{\sim}{\hat{\alpha}}$ เป็นเวกเตอร์ของสัมประสิทธิ์ตัวใหม่ มีขนาด $p \times 1$

และจากสมการที่ 2.4.1.8 ถ้าความถดถอยอยู่ในรูปแบบของ Z เราสามารถหาความแปรปรวนของสัมประสิทธิ์ (สมาชิกในแนวทแยงของ $(Z'Z)^{-1}$ หารด้วย σ^2) เป็นส่วนกลับของค่าเฉพาะดังนี้

$$\frac{Var(\hat{\alpha}_j)}{\sigma^2} = \frac{1}{\lambda_j} \qquad ; \quad j = 1,2,\ldots,p$$

เราจะสังเกตได้ว่า ถ้าทุกองค์ประกอบหลักยังคงอยู่ในตัวแบบถดถอยแล้ว การแปลงตัวแปรเป็นองค์ประกอบหลักเหล่านั้นเป็นเพียงการหมุนแกนของตัวแปรถดถอยนั่นเอง อย่างไรก็ตามตัวแปรใหม่เหล่านี้ตั้งฉากกัน ขนาดของความแปรปรวนยังคงเท่าเดิม แต่เนื่องจากความแปรปรวนทั้งหมดถูกแจกแจงใหม่ ถ้าตัวแปรอิสระมีพหุสัมพันธ์กันจะมีอย่างน้อย 1 องค์ประกอบที่มีค่าเฉพาะน้อยที่สุด การตัดองค์ประกอบหลักออกไปอย่างน้อย 1 องค์ประกอบอาจจะทำให้ลดความแปรปรวนทั้งหมดในตัวแบบให้น้อยลงและสามารถปรับปรุงสมการที่ใช้พยากรณ์ให้เหมาะสมยิ่งขึ้นได้

2.4.2 เกณฑ์ที่ใช้ในการพิจารณาจำนวนองค์ประกอบหลักและการประมาณสัมประสิทธิ์ความถดถอยองค์ประกอบหลัก

ในการวิเคราะห์องค์ประกอบหลัก เพื่อนำมาประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณเมื่อมีพหุสัมพันธ์ระหว่างตัวแปรอิสระจะนำองค์ประกอบมาใช้เพียงบางองค์ประกอบเท่านั้น เนื่องจากหลักและความคิดพื้นฐานของความถดถอยองค์ประกอบหลักเหมือนกับหลักการของกำลังสองน้อยที่สุด คือ เป็นการประมาณค่าด้วยวิธีกำลังสองน้อยที่สุดที่ดำเนินการบนองค์ประกอบหลักดังนั้นถ้าองค์ประกอบหนึ่งถูกตัดทิ้ง ผลลัพธ์ของตัวประมาณของสัมประสิทธิ์ความถดถอยของตัวแปรเดิม (X) จะเป็นตัวประมาณที่เอนเอียง ความแปรปรวนที่เกิดจากการสร้างองค์ประกอบนั้น ๆ จะถูกตัดทิ้งด้วย ส่งผลให้ความแปรปรวนมีค่าลดลง การพิจารณาจำนวนองค์ประกอบหลักที่จะนำมาเกณฑ์ด้วยกัน โดยในงานวิจัยที่เราได้ใช้ Kaiser's Criterion เป็นเกณฑ์นั้นคือ หากพบว่า λ ตัวใดมีค่าต่ำกว่า 1 จะตัดออก

โดยจำนวนองค์ประกอบหลักที่เราจะตัดออกจะมีค่าเท่ากับจำนวนของค่าเฉพาะที่มีค่าน้อยกว่าก็ให้ตัดทิ้งไปซึ่งมีผลเหมือนกำหนดให้ λ ตัวนั้นมีค่าเป็น 0 จำนวนองค์ประกอบที่เหลือจะนำมาสร้างตัวแบบถดถอยด้วยวิธีกำลังสองน้อยสุด

สมมติให้มีค่าเฉพาะจำนวน s ค่าสุดท้าย $(\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_s \geq \lambda_p > 0)$ ที่มีค่าน้อยกว่าค่าเฉพาะน้อยกว่า 1 เราจะได้ $\underset{\sim}{\hat{\alpha}}$ ซึ่งอยู่ในรูปของ $\underset{\sim}{\hat{\alpha}}_{pca}$ ดังนี้

$$\underset{\sim}{\hat{\alpha}}_{pca} = \underset{\sim}{\beta}\underset{\sim}{\hat{\alpha}} \quad \text{โดยที่} \quad \underset{\sim}{\beta} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}$$

$$b_1 = b_2 = \ldots = b_{p-s} = 1 \quad \text{และ} \quad b_{p-s+1} = b_{p-s+2} = \ldots = b_p = 0$$

$$\underset{\sim}{\hat{\alpha}}_{pcz} = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \vdots \\ \hat{\alpha}_{p-s} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$= \Lambda_{p-s}^{-1} T'_{p-s} X' \underset{\sim}{y}$$

จากนั้นจึงใช้วิธีกำลังสองน้อยที่สุดประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณได้ตัวประมาณสัมประสิทธิ์ความถดถอยพหุคูณ โดยวิธีความถดถอยองค์ประกอบหลัก ดังนี้

$$\begin{aligned}\underset{\sim}{\hat{\beta}}_{pca} &= T\underset{\sim}{\hat{\alpha}}_{pca} \\ &= T_{p-s}\left(\Lambda_{p-s}^{-1}T'_{p-s}X'\underset{\sim}{y}\right) \\ &= \sum_{j=1}^{p-s}\lambda_j^{-1}t'_jX'\underset{\sim}{y}t_j\end{aligned}$$

$$\begin{aligned}Var(\underset{\sim}{\hat{\beta}}_{pca}) &= \sigma^2T\Lambda^{-1}T' \\ &= \sigma^2\sum_{j=1}^{p}\lambda_j^{-1}\underset{\sim}{t}'_j\underset{\sim}{t}_j\end{aligned}$$

เมื่อ $\underset{\sim}{t}_j$ แทนเวกเตอร์ของเมตริกซ์เชิงตั้งฉาก สดมภ์ที่ j

โดย $\underset{\sim}{\hat{\beta}}_{pca}$ เป็นตัวประมาณที่เป็นมาตรฐาน (Standardize estimator) สมาชิกแต่ละตัวของ $\underset{\sim}{\hat{\beta}}_{pca}$ เรียกว่าสัมประสิทธิ์ความถดถอยที่เป็นมาตรฐาน (standardize regression coefficients) นอกจากนี้ $\underset{\sim}{\hat{\beta}}_{pca}$ ยังมีคุณสมบัติเป็นตัวประมาณที่เอนเอียง ซึ่งไม่มีการตัด $\underset{\sim}{X}_j$ ใด ๆ ออกจากสมการ แต่จะมีการตัด $\underset{\sim}{Z}_j$ บางตัวออกแทน

2.4.3 **ความเอนเอียงของสัมประสิทธิ์ความถดถอยองค์ประกอบหลัก**

เมื่อพิจารณาองค์ประกอบหลักที่ถูกตัดทิ้งไป s องค์ประกอบ และเหลืออยู่ p-s องค์ประกอบ โดยมีเมตริกซ์ของเวกเตอร์เฉพาะ $T = \left[\underset{\sim}{T}_1, \underset{\sim}{T}_2, \underset{\sim}{T}_3, \ldots, \underset{\sim}{T}_p \right]$ ซึ่งสามารถแบ่งส่วนออกเป็น

$$T = \left[\underset{\sim}{T}_x : \underset{\sim}{T}_{p-s} \right]$$

ในทำนองเดียวกัน พิจารณา Λ ซึ่งเป็นเมตริกซ์ในแนวทแยงมุมที่มีสมาชิกในแนวเส้นทแยงมุมเป็นค่าเฉพาะที่สัมพันธ์กับองค์ประกอบที่ตัดทิ้ง และค่าเฉพาะที่สัมพันธ์กับองค์ประกอบที่เลือกไว้ตามลำดับ เราสามารถเขียนสมการแสดงความสัมพันธ์ในการประมาณค่า $\underset{\sim}{\hat{\alpha}}$ ด้วยวิธีกำลังสองน้อยที่สุด ได้ดังนี้

(2.4.3.1)
$$\underset{\sim}{\hat{\alpha}} = (Z'Z)^{-1} Z' \underset{\sim}{y}$$

และแสดงตัวประมาณ $\underset{\sim}{\hat{\alpha}}$ ที่อยู่ในตัวแบบ คือ

$$\underset{\sim}{\hat{\alpha}}_{p-s} = (Z'Z)^{-1} Z' \underset{\sim}{y} = \Lambda_{p-s}^{-1} T'_{p-s} X' \underset{\sim}{y}$$

และจะได้ว่า
$$Cov(\underset{\sim}{\hat{\alpha}}) = \sigma^2 (Z'Z)^{-1} = \sigma^2 \Lambda^{-1}$$

เนื่องจากหลักการในการสร้างตัวแบบด้วยวิธีความถดถอยองค์ประกอบหลักเป็นเหมือนกับวิธีกำลังสองน้อยที่สุด และองค์ประกอบเหล่านี้ตั้งฉากกัน เราสามารถแสดงได้ว่า $\underset{\sim}{\hat{\alpha}}_{p-s}$ เป็นตัวประมาณที่ไม่เอนเอียงของ $\underset{\sim}{\alpha}_{p-s}$ และสามารถแปลงองค์ประกอบหลักในตัวแบบกลับเป็นตัวแปรดั้งเดิมได้ดังนี้

จาก
$$\underset{\sim}{\alpha} = T' \underset{\sim}{\beta}$$

ไปเป็น
$$\underset{\sim}{\beta} = T \underset{\sim}{\alpha}$$

เมื่อเราตัดองค์ประกอบ s องค์ประกอบสุดท้ายทิ้งไป เราสามารถเขียนตัวประมาณกำลังสองน้อยที่สุดของสัมประสิทธิ์ความถดถอยสำหรับพารามิเตอร์ p ตัวได้เป็น

(2.4.3.2) $$\hat{\underset{\sim}{\beta}}_{pca} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} = \begin{bmatrix} \underset{\sim}{T}_1 & \underset{\sim}{T}_s & \cdots & \underset{\sim}{T}_{p-s} \end{bmatrix} \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_p \end{bmatrix}$$

หรือ $$\hat{\underset{\sim}{\beta}}_{pca} = T_{p-s}\hat{\alpha}_{p-s}$$

ดังนั้น $$E\left[\hat{\underset{\sim}{\beta}}_{pca}\right] = T_{p-s}\,\underset{\sim}{\alpha}_{p-s} = T_{p-s}T'_{p-s}\,\underset{\sim}{\beta}$$

เนื่องจาก $TT' = I = T_sT'_s + T_{p-s}T'_{p-s}$

จะได้ว่า $$\begin{aligned} E\left[\hat{\underset{\sim}{\beta}}_{pca}\right] &= [I - T_sT'_s]\underset{\sim}{\beta} \\ &= \underset{\sim}{\beta} - T_sT'_x\,\underset{\sim}{\beta} \\ &= \underset{\sim}{\beta} - T_s\,\underset{\sim}{\alpha}_s \end{aligned}$$

ดังนั้น ตัวประมาณสัมประสิทธิ์ความถดถอยมีความเอนเอียงเท่ากับ $T_s\underset{\sim}{\alpha}_s$ โดยที่ $\underset{\sim}{\alpha}_s$ เป็นเมตริกซ์ขององค์ประกอบหลักที่ถูกตัดทิ้งไป

2.4.4 ความแปรปรวนของสัมประสิทธิ์ความถดถอยองค์ประกอบหลัก

การตัดองค์ประกอบเป็นผลให้ความแปรปรวนของสัมประสิทธิ์ความถดถอย $\hat{\underset{\sim}{\beta}}_{pca}$ มีค่าลดลงโดยสามารถพิจารณาได้จากสมการต่อไปนี้

$$\frac{Var(\hat{\underset{\sim}{\beta}})}{\sigma^2} = (X'X)^{-1}$$

$$= T\Lambda^{-1}T'$$

(2.4.4.1)
$$= T_s\Lambda_s^{-1}T_s' + T_{p-s}\Lambda_{p-s}^{-1}T_{p-s}'$$

เนื่องจาก $Var(\hat{\alpha}_{p-s}) = \Lambda_{p-s}^{-1}$ ดังนั้นเราจะได้เมตริกซ์ความแปรปรวนร่วม คือ

(2.4.4.2)
$$\frac{Var(\hat{\underset{\sim}{\beta}}_{pca})}{\sigma^2} = T_{p-s}\Lambda_{p-s}^{-1}T_{p-s}'$$

ดังนั้นเมื่อพิจารณาจากสมการที่ 2.4.4.1 และ 2.4.4.2 จะสังเกตได้ว่า ความแตกต่างของเมตริกซ์ความแปรปรวนและความแปรปรวนร่วมสำหรับตัวประมาณกำลังสองน้อยที่สุดและตัวประมาณองค์ประกอบหลัก คือ $T_s\Lambda_s^{-1}T_s'$ โดยเมื่อตัวองค์ประกอบที่ไม่ต้องการออก จะทำให้ได้ความแปรปรวนที่มีค่าลดลงดังสมการที่ (2.4.4.2)

2.5 **วิธีแปลงเมตริกซ์ X เพื่อให้เป็นแนวตั้งเชิงตั้งฉาก** (Orthogonalization Process)

เนื่องจากตัวแบบความถดถอยเมตริกซ์ X จะเกิดความสัมพันธ์ร่วมเชิงพหุ และ $\det(X'X)=0$ ดังนั้นจึงจำเป็นที่จะต้องปรับปรุงแก้ไขด้วยวิธีการแปลงเมตริกซ์ X เพื่อให้เป็นแนวตั้งเชิงตั้งฉาก เพื่อให้สอดคล้องกับข้อข้อสมมุติฐานสำคัญข้อหนึ่งคือตัวแปรอิสระแต่ละตัวต้องไม่มีความสัมพันธ์กับตัวแปรอิสระตัวอื่นในการประมาณค่าสัมประสิทธิ์ด้วยวิธีกำลังสองน้อยสุด

ในการประมาณอิทธิพลบางตัวแปรในตัวแบบเชิงเส้น มีความจำเป็นต้องตรวจสอบค่าลำดับชั้นของ X ถ้า X มีค่าลำดับชั้นเท่ากับจำนวนแนวตั้งของมันก็จะทำให้การหาค่าประมาณอิทธิพลของตัวแปรอิสระง่ายขึ้นเพราะทำให้สามารถหา $(X'X)^{-1}$ ได้ ซึ่งในงานวิจัยนี้ได้เลือกใช้วิธีแปลงด้วยกระบวนการของแกรม-สมิตซ์ (Gram-Schmidt) รวมทั้งการสร้างฐานหลักเชิงตั้งฉากปรกติ (Orthogonal basis) สำหรับ n มิติ ซึ่งเป็นผลคูณภายในปริภูมิ V โดยจะทำการเริ่มจากการพิจารณาฐานหลัก (basis) $x_1, x_2, ..., x_n$ โดยมีกระบวนการดังนี้

$$
\begin{aligned}
Z_1 &= x_1 \\
Z_2 &= x_2 - \frac{x_2 Z_1}{Z_1 Z_1} Z_1 \\
Z_3 &= x_3 - \frac{x_3 Z_2}{Z_2 Z_2} Z_2 - \frac{x_3 Z_1}{Z_1 Z_1} Z_1 \\
&\vdots \\
Z_r &= x_r - \frac{x_r Z_{r-1}}{Z_{r-1} Z_{r-1}} Z_{r-1} - \frac{x_r Z_{r-2}}{Z_{r-2} Z_{r-2}} Z_{r-2} - \ldots - \frac{x_r Z_1}{Z_1 Z_1} Z_1
\end{aligned}
$$

เมื่อ $Z_1, Z_2, \ldots, Z_r$ คือ ฐานหลักใหม่ (new basis) ซึ่ง $Z_i \ (i = 1,2,\ldots,r)$ จะตั้งฉากกัน (mutually orthogonal)

2.6 ดัชนีพหุสัมพันธ์ (Degree of multicollinearity)

ดัชนีพหุสัมพันธ์คือ ค่าซึ่งบอกให้ทราบว่าตัวแปรอิสระมีความสัมพันธ์กันมากหรือน้อยเพียงใดมีผู้เสนอค่าต่าง ๆ ซึ่งสามารถใช้เป็นดัชนีพหุสัมพันธ์ไว้หลายคน

2.6.1 สหสัมพันธ์อย่างง่ายระหว่างตัวแปรอิสระ (Simple correlation among the regression variales)

พิจารณาเมตริกซ์สหสัมพันธ์ของ $X^{*'}X^{*}$ โดยที่สมาชิกแถวที่ i และแนวตั้งที่ j ของเมตริกซ์ $X^{*'}X^{*}$ คือ $X_{ij}^{*} = X_{ij} - \bar{x}$ สมาชิกแต่ละตัวของเมตริกซ์สหสัมพันธ์ $X^{*'}X^{*}$ แสดงถึงสหสัมพันธ์อย่างง่ายของตัวแปรอิสระ X_i และ X_j ถ้าสหสัมพันธ์อย่างง่ายของตัวแปรอิสระ X_i และ X_j ใด ๆ มีค่าสูงมาก แสดงว่าตัวแปรอิสระมีแนวโน้มที่จะมีพหุสัมพันธ์ อย่างไรก็ตามการพิจารณาด้วยวิธีนี้ไม่ได้กำหนดค่าสหสัมพันธ์อย่างง่ายของตัวแปรอิสระแต่ละคู่ว่าควรเป็นเท่าไร จึงจะเกิดพหุสัมพันธ์

2.6.2 Variance Inflation Factors (VIF)

2.6.2.1 Marquart (1970 ; 299) ได้เสนอให้ใช้ค่าสมาชิกในแนวเส้นทะแยงมุมของอินเวอร์สเมตริกซ์ความสัมพันธ์ ซึ่งให้ชื่อว่า Variance inflation factor (VIF) เป็นดัชนีพหุสัมพันธ์โดยที่

$$VIF = \frac{1}{(1 - R_i^2)}$$

โดยที่ R_j^2 คือค่าสัมประสิทธิ์การตัดสินใจของตัวแปร X_j ซึ่งถดถอยกับตัวแปรอิสระที่เหลือ

ถ้า x เกือบจะเป็นอิสระกับตัวแปรอื่น ๆ R^2 จะมีค่าน้อย และ VIF จะมีค่าเข้าใกล้หนึ่ง แต่ถ้า x มีความสัมพันธ์เชิงเส้นกับตัวแปรที่เหลือบางตัว R_j^2 จะมีค่าใกล้เคียงกับหนึ่ง และ VIF จะมีค่ามาก ถ้าข้อมูลมี VIF บางค่าซึ่งมีค่ามากจะบอกได้ ว่าตัวแปรอิสระมีพหุสัมพันธ์กัน โดยทั่วไปในทางปฏิบัติถ้า VIF ค่าใดมีค่ามากกว่า 5 หรือ 10 แสดงว่าค่าประมาณสัมประสิทธิ์ความถดถอยพหุคูณซึ่งสอดคล้องกับ VIF ค่านั้นเป็นตัวประมาณที่ไม่ดีเนื่องจากมีพหุสัมพันธ์ระหว่างตัวแปรอิสระ

2.6.2.2 Chaterjee and Price (1977 , 1983) ได้กล่าวว่า VIF อาจใช้หาค่าคาดหวังของระยะห่างกำลังสองของค่าประมาณ โดยวิธีกำลังสองน้อยที่สุด จากค่าจริงซึ่งจะใช้วัดความแม่นยำของค่าประมาณจากวิธีกำลังสองน้อยที่สุด ถ้าระยะห่างมีค่าน้อยลงตัวประมาณจะยิ่งมีความแม่นยำมากขึ้น

ให้ L^2 คือ ระยะห่างกำลังสอง

$$\therefore\ L^2 = \sigma^2 \sum_{i=1}^{p} VIF_i \quad ; \quad p = \text{จำนวนตัวแปรอิสระ}$$

ถ้าตัวแปรอิสระไม่มีพหุสัมพันธ์กัน VIF จะมีค่าเท่ากับหนึ่งทุกค่า ซึ่งจะทำให้

$$L^2 = p\sigma^2 \quad , \quad R_L = \frac{\sigma^2 \sum_{i=1}^{p} VIF_i}{p\sigma^2} = \frac{\sum_{i=1}^{p} VIF_i}{p}$$

R_L จะวัดความคลาดเคลื่อนกำลังสองของวิธีกำลังสองน้อยที่สุด เมื่อเปรียบเทียบกับขนาดของความคลาดเคลื่อนของข้อมูลนั้น ถ้าตัวแปรอิสระไม่มีพหุสัมพันธ์กัน ดังนั้นค่า R_L จึงอาจใช้เป็นดัชนีพหุสัมพันธ์ได้ ตัวอย่างเช่น ถ้า $R_L = 124$ หมายความว่า ความคลาดเคลื่อนกำลังสองของตัวประมาณวิธีกำลังสองน้อยที่สุด จะเป็น 124 เท่าของความคลาดเคลื่อนกำลังสองของตัวประมาณกำลังสองน้อยที่สุด ถ้าตัวแปรอิสระไม่มีพหุสัมพันธ์กัน

2.6.3 ค่าเจาะจงของ $X'X$

Montgomery (1982 : 301) ถ้า $\lambda_1 , \lambda_2 , \ldots , \lambda_p$ เป็น eigenvalue ของเมตริกซ์ความสัมพันธ์ให้ $m = \dfrac{\lambda_{\max}}{\lambda_{\min}}$ อาจใช้ค่า m เป็นดัชนีพหุสัมพันธ์ได้ ค่า m จะวัดการกระจายของ eigenvalue โดยทั่วไปถ้า $m < 100$ การที่ตัวแปรอิสระมีพหุสัมพันธ์กันจะไม่ก่อให้เกิดปัญหาในการประมาณมากหนัก แต่ถ้า $m > 100$ การที่ตัวแปรอิสระมีพหุสัมพันธ์กันจะก่อให้เกิดปัญหาในการประมาณมาก

บทที่ 3

วิธีดำเนินการวิจัย

ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากการจำลองด้วยเทคนิคมอนติคาร์โล (Monte Carlo) โดยใช้โปรแกรม S plus 2000 ซึ่งมีขั้นตอนของแผนการทดลองและโปรแกรมที่ใช้ในการศึกษาดังต่อไปนี้

3.1 แผนการทดลอง

ในการวิจัยครั้งนี้กำหนดสถานการณ์ต่าง ๆ ที่ต้องการศึกษาดังนี้

1. เลือกกลุ่มตัวอย่างจากประชากรที่มีการแจกแจงแบบปกติ โดยมีพารามิเตอร์ μ เท่ากับ 0 และ σ เท่ากับ 1.0 , 5.0 และ 10.0 ตามลำดับ
2. จำนวนตัวแปรอิสระที่ใช้ในการศึกษาเท่ากับ 3 , 6 และ 9
3. ขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ 5p,10p,15p,15p,20p,25p และ 30p เมื่อ p คือจำนวนตัวแปรอิสระ

เนื่องจากการวิจัยครั้งนี้มีวัตถุประสงค์ที่ จะเปรียบเทียบตัวประมาณสัมประสิทธิ์ความถดถอยพหุคูณเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปร ผู้วิจัยจึงกำหนดระดับความสัมพันธ์ระหว่างตัวแปรเป็น 3 ระดับ คือ

ระดับต่ำ	ค่า ρ มีค่าอยู่ในช่วง 0.15 ถึง 0.30
ระดับปานกลาง	ค่า ρ มีค่าอยู่ในช่วง 0.31 ถึง 0.50
	ค่า ρ มีค่าอยู่ในช่วง 0.50 ถึง 0.65
ระดับสูง	ค่า ρ มีค่าอยู่ในช่วง 0.61 ถึง 0.85
	ค่า ρ มีค่าอยู่ในช่วง 0.85 ถึง 0.99

3.2 ขั้นตอนการวิจัย

ขั้นตอนในการวิจัยมีดังนี้

3.2.1 สร้างข้อมูลของตัวแปรอิสระ (X) ให้มีระดับความสัมพันธ์ตามที่กำหนด และสร้างข้อมูลของตัวแปรตาม $(\underset{\sim}{y})$ จากรูปแบบความสัมพันธ์ $\underset{\sim}{y} = \underset{\sim}{X}\,\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

3.2.2 สร้างข้อมูลของความคลาดเคลื่อนให้มีการแจกแจงแบบปกติ

3.2.3 แปลงข้อมูล X โดยการใช้ Gramsmihdt Orthogonalization

3.2.4 ประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธีที่สนใจศึกษา 4 วิธี คือ วิธีกำลังสองน้อยที่สุด, วิธีริดจ์รีเกรสชั่น, วิธีถดถอยองค์ประกอบหลัก วิธีรีเกรสชั่นเชิงตั้งฉาก

3.2.5 หาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยของความถดถอยแบบพหุคูณพร้อมทั้งเปรียบเทียบค่าเฉลี่ยและสรุปผลที่ได้

สำหรับรายละเอียดของแต่ละขั้นตอนมีดังนี้

จากตัวแบบ $Y = X\beta + \varepsilon$; $E(\varepsilon) = 0$; $E(\varepsilon'\varepsilon) = \sigma^2 I_n$ ในการวิจัยครั้งนี้จึงต้องสร้างความคลาดเคลื่อนของข้อมูลแต่ละชุดให้มีการแจกแจงแบบปกติ มีค่าเฉลี่ยเป็นศูนย์ และความแปรปรวนของความคลาดเคลื่อนมีค่าเป็น σ^2 โดยกำหนดให้ $\sigma = 0.1, 1, 5$ และ 10 ค่าของตัวแปรตาม Y_i จะได้จากการแทนค่าในสมการ

$$Y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + \ldots + x_{ip}\beta_p + \varepsilon_i$$

ในที่สุดก็จะได้ข้อมูลชุดหนึ่งซึ่งประกอบไปด้วยตัวแปรอิสระและตัวแปรตาม โดยที่ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามเป็นแบบเชิงเส้น ข้อมูลที่ได้จากตัวแบบนี้เป็นข้อมูลมาตรฐาน ซึ่งในการใช้ข้อมูลมาตรฐานนี้ Gunst and Mason (1977 : 616 - 628) ได้กล่าวว่า เป็นการลดความคลาดเคลื่อนเนื่องจากการคำนวณลงเท่านั้น ไม่ได้ทำให้ความคลาดเคลื่อนเนื่องจากเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระลดลงแต่อย่างใด

3.2.1 การสร้างตัวแปรอิสระ X ให้มีความสัมพันธ์ในระดับต่าง ๆ

ในการวิจัยครั้งนี้ สร้างข้อมูลของตัวแปรอิสระ X ที่มีการแจกแจงปกติของหลายตัวแปร (multivariate normal distribution) ที่มีค่าพารามิเตอร์ $\underset{\sim}{\mu}$ เท่ากับ $\underset{\sim}{0}$ และ $\underset{\sim}{V}$ เป็นเมตริกซ์ความแปรปรวนร่วม (covariance matrix) ที่แสดงความสัมพันธ์ระหว่างตัวแปรอิสระใน X สามารถเขียนได้เป็น $X \sim N_p(\underset{\sim}{\mu}, \underset{\sim}{V})$

ปี ค.ศ. 1972 Barr และ Slesak ได้เสนอวิธีการสร้างข้อมูลที่มีการแจกแจงปกติของหลายตัวแปรดังนี้

กำหนดให้ $X = (X_1, X_2, ..., X_p)'$ เป็นเวกเตอร์ของตัวแปรที่มีการแจกแจงแบบปกติที่มีความสัมพันธ์โดยมีเวกเตอร์ค่าเฉลี่ย $\underset{\sim}{\mu} = (\mu_1, \mu_2, ..., \mu_p)'$ สามารถเขียนเมตริกซ์ความแปรปรวนร่วม

$$\underset{\sim}{V} = E\left[\left(\underset{\sim}{X} - \underset{\sim}{\mu}\right)\left(\underset{\sim}{X} - \underset{\sim}{\mu}\right)'\right] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \dots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

เมื่อ $\sigma_{ij} = Cov(X_i, X_j) = \sigma_{ji} = Cov(X_j, X_i)$ สำหรับ $i \neq j$ และ $\sigma_{ii} = Var(X_i)$, $i = 1,2,..., p \quad j = 1,2,..., p$

ขั้นตอนในการสร้างตัวแปร X ที่มีการแจกแจงแบบปกติหลายตัวแปร มีดังนี้

1. คำนวณหาเมตริกซ์ C ซึ่งเป็นเมตริกซ์เชิงสามเหลี่ยนล่าง (lower triangular matrix) ซึ่งทำให้ได้ $V = C'C$

$$C = \begin{bmatrix} c_{11} & 0 & \dots & 0 \\ c_{21} & c_{22} & \cdots & 0 \\ \vdots & \vdots & \dots & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix}$$

2. สร้าง $\underset{\sim}{Z}$ ที่ประกอบด้วยตัวแปรที่มีการแจกแจงแบบปกติมาตรฐานที่เป็นอิสระซึ่งกันและกัน p ตัว

3. คำนวณ $\underset{\sim}{X} = \underset{\sim}{\mu} + \underset{\sim}{C}\underset{\sim}{Z}$ เนื่องจาก V เป็นเมตริกซ์สมมาตรเรา จึงใช้ Cholesky Factorization ในการคำนวณหาเมตริกซ์ C ได้ดังนี้

3.1) กำหนดให้ $a = \sqrt{\sigma_{11}}$

$$c_{i1} = \frac{\sigma_{i1}}{z} \qquad ; \; i = 1, 2, 3, \ldots , p$$

3.2) $$c_{ii} = \left(\sigma_{ii} - \sum_{j=1}^{i-1} c_{ij}^2\right)^{\frac{1}{2}}$$

เมื่อ $$\sigma_{ii} = \sum_{j=1}^{i-1} c_{ij}^2 \qquad ; \; i = 1, 2, 3, \ldots , p$$

3.3) $$c_{ij} = \frac{\left(\sigma_{ij} - \sum_{k=1}^{j-1} c_{ik} c_{jk}\right)}{c_{jj}}$$

เมื่อ $$\sigma_{ij} = \sum_{k=1}^{j-1} c_{ik} c_{jk} \qquad ; \; i = 1, 2, 3, \ldots , p \, , \; j = 1, 2, 3, \ldots , i-1$$

จากสมการ

(3.2.1.1) $$X = \underset{\sim}{\mu} + \underset{\sim}{C}\underset{\sim}{Z}$$

ในที่นี้ให้ $\underset{\sim}{\mu}$ แทน $\underset{\sim}{0}$ และ $Z = (Z_1, Z_2, \ldots, Z_p)'$ เมื่อ $\underset{\sim}{Z} \sim N(\underset{\sim}{0}, \underset{\sim}{1})$ จากสมการที่ (3.2.1.1) สามารถจำลองข้อมูลของตัวแปรอิสระ $\underset{\sim}{X}$ ให้มีความสัมพันธ์ในระดับต่าง ๆ ตามที่ต้องการได้

สำหรับในการวิจัยครั้งนี้ (การวิจัยครั้งนี้ทำการทดลอง 1,000 รอบในแต่ละสถานการณ์) ผู้วิจัยได้กำหนดให้ $\underset{\sim}{C}$ เป็นเมตริกซ์ที่คงที่ หลังจากนั้นจึงจำลอง $\underset{\sim}{Z}$ ซึ่งมีการแจกแจงแบบปกติ แล้วนำมาคูณกับ $\underset{\sim}{C}$ ในแต่ละรอบของการทดลอง โดยผลลัพธ์ที่ได้คือ $\underset{\sim}{X}$ ที่ได้จะมีระดับความสัมพันธ์ตามที่ต้องการ

รายละเอียดของขั้นตอนของโปรแกรมการคำนวณในการสร้างตัวแปรอิสระ $\underset{\sim}{X}$ ที่มีการแจกแจงแบบปกติหลายตัวแปร แสดงไว้ในภาคผนวก

ดังนั้น จะต้องกำหนดค่า $\underset{\sim}{\beta}$ เพื่อจะสร้างค่า Y ขึ้นจากตัวแบบดังกล่าว การกำหนดค่า $\underset{\sim}{\beta}$ จะกำหนดให้เป็นค่าคงที่ เท่ากับ 1

3.2.2 การสร้างความคลาดเคลื่อนให้มีการแจกแจงตามที่ต้องการศึกษา

การสร้างค่าความคลาดเคลื่อนให้มีลักษณะการแจกแจงตามที่ต้องการศึกษานั้นใช้โปรแกรม S – plus 2000 โดยสามารถสร้างลักษณะการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ μ และค่าความแปรปรวนเท่ากับ σ^2 โดยทฤษฎีในการสร้างนี้จะเริ่มจากการสร้างเลขสุ่ม(random number) ซึ่งมีการแจกแจงแบบสม่ำเสมอ (uniform distribution) ในช่วง (0,1) แล้วมาปรับเป็น N(μ ,σ^2)

การผลิตเลขสุ่มที่มีการแจกแจงแบบปกติ

การผลิตเลขสุ่มที่มีการแจกแจงแบบปกติใช้วิธีของ George E.P Box และ Mervin E.Muller (ค.ศ.1958) โดยใช้การแปลงตัวแปรสุ่ม คือ จากตัวแปรสุ่มมาตรฐานอิสระกัน Z_1 และ Z_2 ได้จุดบนระนาบในระบบพิกัดคาร์ทีเซียน (Cartesian coordinates) แปลงตัวแปรสุ่มในระบบพิกัดคาร์ทีเซียนเป็นตัวแปรสุ่มในระบบพิกัดเชิงขั้ว (Polar coordinates) เป็นจุด (P,Θ) โดยที่

$$Z_1 = P\cos\theta$$
$$Z_1 = P\sin\theta$$

การแปลง $z_1 = p\cos\theta$ และ $z_2 = p\sin\theta$ เป็นการแปลงจากปริภูมิ

$R_{z_1,z_2} = \{(z_1,z_2) : -\infty < z_1 < \infty \text{ , } -\infty < z_1 < \infty\}$ ไปยังปริภูมิ

$R_{P,\theta} = \{(p,\theta) : 0 \le p < \infty \text{ , } 0 \le \theta \le 2\pi\}$ โดยใช้จาโคเบียน (Jacobian) ของการแปลง J ดังนี้

$$J = \begin{vmatrix} \dfrac{\partial z_1}{\partial p} & \dfrac{\partial z_1}{\partial \theta} \\ \dfrac{\partial z_2}{\partial p} & \dfrac{\partial z_2}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -p\sin\theta \\ \sin\theta & p\cos\theta \end{vmatrix}$$

$$= p(\cos^2\theta + \sin^2\theta) = p$$

จากเทคนิคการแปลงในทฤษฎีความน่าจะเป็นได้ว่า P และ Θ มีฟังก์ชันความหนาแน่นร่วม (joint density function) คือ

$$f_{p,\Theta}(p,\theta) = f_{z_1,z_2}(p\cos\theta, p\sin\theta)|J|$$

เนื่องจาก Z_1 และ Z_2 มีฟังก์ชันความหนาแน่นร่วม คือ

$$
\begin{aligned}
f_{z_1,z_2}(z_1,z_2) &= f_{z_1}(z_1)f_{z_2}(z_2) \\
&= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z_1^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z_2^2} \\
&= \frac{1}{2\pi}e^{-\frac{1}{2}(z_1^2+z_2^2)}
\end{aligned}
$$

ดังนั้นโดยการ แทนค่าจะได้ผลลัพธ์

$$
\begin{aligned}
f_{P,\Theta}(p,\theta) &= \frac{1}{2\pi}e^{-\frac{1}{2}p^2}p \\
&= \frac{1}{2\pi}pe^{-\frac{1}{2}p^2} \qquad ;\ 0 \le \theta \le 2\pi\ ,\ 0 \le p < \infty \\
&= f_{\Theta}(\theta)f_p(p)
\end{aligned}
$$

โดยที่ $f_\theta(\theta) = \dfrac{1}{2\pi}\ ;\ 0 \le \theta \le 2\pi$ เป็นฟังก์ชันของ θ เท่านั้น ไม่ขึ้นอยู่กับ p และ $f_p(p) = pe^{-\frac{1}{2}p^2}\ ;\ p \ge 0$ เป็นฟังก์ชันของ p เท่านั้น ไม่ขึ้นอยู่กับ θ เพราะฉะนั้นจาก คุณสมบัติของตัวแปรอิสระ ได้ว่า P และ Θ เป็นอิสระกัน(เชิงสถิติ) ในการจำลอง Z_1 และ Z_2 เราจะจำลอง P และ Θ อย่างอิสระกัน โดยจำลอง P จาก $f_P(p) = pe^{-\frac{1}{2}p^2}$ ซึ่งด้วยวิธีการแปลผกผันได้ตัวแบบจำลอง $P = \sqrt{-2\ln R_1}\ ;\ R_1 \sim U(0,1)$ และจำลอง Θ จากการแจกแจง $U(0,2\pi)$ ได้ $\Theta = 2\pi R_2\ ;\ R_2 \sim U(0,1)$ ดังนั้นเราจะได้ตัวแบบจำลองตัวแปรสุ่ม $Z_1 \sim N(0,1)$ และ $Z_2 \sim N(0,1)$ ซึ่งเป็นอิสระกัน คือ

$$
\begin{aligned}
Z_1 &= (-2\ln R_1)^{\frac{1}{2}}\cos(2\pi R_2) \\
Z_2 &= (-2\ln R_1)^{\frac{1}{2}}\sin(2\pi R_2)
\end{aligned}
$$

เมื่อเราได้เลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานแล้ว จากนั้นก็มีการปรับให้มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ μ และค่าความแปรปรวนเท่ากับ σ^2 (NORMAL ~ N (μ,σ^2))

โดย

$$\text{normal} = \mu + \sigma Z_1$$

$$\text{normal} = \mu + \sigma Z_2$$

ซึ่งสำหรับโปรแกรม S – plus 2000 ได้มีการให้ใช้คำสั่งเพื่อ สร้าง error ขึ้นโดยใช้

3.2.3 <u>การแปลงข้อมูล X โดยการใช้ Gramsmihdt Orthogonalization</u>

ได้ใช้วิธีการแปลงด้วย วิธีแปลงด้วยกระบวนการของแกรม-สมิตซ์ (Gram-Schmidt) รวมทั้งการสร้างฐานหลักเชิงตั้งฉากปรกติ (Orthogonal basis) สำหรับ n มิติ ซึ่งเป็นผลคูณภายในปริภูมิ V โดยจะทำการเริ่มจากการพิจารณาฐานหลัก (basis) $x_1, x_2, ..., x_n$ โดยมีกระบวนการดังนี้

$$\begin{aligned}
Z_1 &= x_1 \\
Z_2 &= x_2 - \frac{x_2 Z_1}{Z_1 Z_1} Z_1 \\
Z_3 &= x_3 - \frac{x_3 Z_2}{Z_2 Z_2} Z_2 - \frac{x_3 Z_1}{Z_1 Z_1} Z_1 \\
&\vdots \\
Z_r &= x_r - \frac{x_r Z_{r-1}}{Z_{r-1} Z_{r-1}} Z_{r-1} - \frac{x_r Z_{r-2}}{Z_{r-2} Z_{r-2}} Z_{r-2} - \dots - \frac{x_r Z_1}{Z_1 Z_1} Z_1
\end{aligned}$$

เมื่อ $Z_1, Z_2, \dots, Z_r$ คือ ฐานหลักใหม่ (new basis) ซึ่ง $Z_i \ (i = 1,2,\dots,r)$ จะตั้งฉากกัน (mutually orthogonal)

3.2.4 <u>การประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธีความถดถอยองค์ประกอบหลัก วิธีกำลังสองน้อยสุด วิธีริดจ์รีเกรสชั่น วิธีกำลังสองน้อยสุดหลังจากแปลงค่า X แล้ว</u>

จากหัวข้อ 3.2.1) ถึง 3.2.3) เราสามารถสร้างค่าความคลาดเคลื่อนตามการแจกแจงที่ต้องการศึกษา และสามารถสร้างตัวแปรอิสระ (X) ให้มีความสัมพันธ์เชิงเส้นตรงตามที่กำหนด จึงทำให้สามารถสร้างตัวแปรตาม $(\underset{\sim}{y})$ ได้ จากนั้นจึงทำการประมาณค่าสัมประสิทธิ์ความถดถอยพหุของวิธีต่าง ๆ ซึ่งมีขั้นตอนดังนี้

1) ทำการปรับตัวแปรอิสระ (X) และตัวแปรตาม $(\underset{\sim}{y})$ ให้เป็นมาตรฐาน (standardization)

2) ทำการแปลง (X) เป็น orthogonal ด้วย Gram Smith

3) คำนวณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุดสามัญ และวิธีริดจ์ รีเกรสชั่น วิธีการถดถอยองค์ประกอบหลัก

4) คำนวณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุดสามัญด้วย (X) ที่แปลงเป็นแนวแกนเชิงตั้งฉากแล้ว

3.2.5 <u>การหาค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของสัมประสิทธิ์ความถดถอยพหุคูณ</u>

การหาค่าความคลาดเคลื่อนของค่าประมาณสำหรับสัมประสิทธิ์ความถดถอยในแต่ละวิธี

$$MSE_p = (\hat{\beta}_p - \beta_p)^2$$

เมื่อ p คือ ตัวแปรอิสระ

β_p แทน ตัวประมาณสัมประสิทธิ์ความถดถอยพหุคูณตัวที่ p ที่ประมาณได้ในครั้งที่ i

$$MSE_m = \frac{\sum_{p=1}^{p+1} MSE_p}{p+1}$$

$$MSEm = \sqrt{MSEm}$$

เมื่อ m คือ วิธีการประมาณค่าสัมประสิทธิ์ความถดถอย

การหาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์ความถดถอยจากการทำซ้ำ

$$AMSE_m = \frac{\sum_{i=1}^{loops} MSE_i}{loops}$$

โดย $loops$ คือ จำนวนรอบที่ทำซ้ำ

$AMSE_m$ แทนค่าเฉลี่ยจากค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากสัมประสิทธิ์ความถดถอยพหุคูณเฉลี่ยแต่ละวิธี

การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจากการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธีต่าง ๆ

$$DIFF = \frac{(AMSE_m - AMSE_{\min})}{AMSE_{\min}} \times 100\ \% \qquad ; \quad m = 1, 2, 3, 4$$

โดย $DIFF$ แทนเปอร์เซ็นต์ของอัตราส่วนผลต่างค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจากการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณดังวิธีต่าง ๆ

บทที่ 4

ผลการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อหาข้อสรุปในการเลือกตัวประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ โดยผู้วิจัยจะทำการเปรียบเทียบประสิทธิภาพของตัวประมาณจาก 4 วิธีดังต่อไปนี้

- วิธีกำลังสองน้อยที่สุด (Ordinary Least Square Method-OLS)
- วิธีกำลังสองน้อยที่สุดหลังจากแก้ไขด้วยวิธีทำให้เป็นแนวตั้งเชิงตั้งฉาก (Orthogonal Least Square Method-OLS-ORT)
- วิธีริดจ์รีเกรสชั่น (Ridge Regression Method-RID)
- วิธีความถดถอยองค์ประกอบหลัก (principle Component Regression Method-PCR)

โดยใช้ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองของการประมาณค่าสัมประสิทธิ์การถดถอยเป็นค่าเปรียบเทียบภายใต้สถานการณ์ที่กำหนด ซึ่งสถานการณ์ต่าง ๆ เหล่านี้จะอาศัยวิธีการจำลองด้วยเทคนิคการจำลองมอนติคาร์โลโดยทำการทดลองซ้ำ โดยจำนวนครั้งในแต่ละสถานการณ์จะต่างกัน แต่จะทำการจำลองจนกว่าค่า AMSE ที่ได้ในแต่ละรอบความแตกต่างกันไม่เกิน 5% ของค่า AMSE เดิม

เกณฑ์ที่ใช้ในการตัดสินว่าตัวแบบใดเป็นตัวแบบที่ดีที่สุดจะใช้ค่าเฉลี่ยรากของความคลาดเคลื่อนยกกำลังสองของการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณต่ำสุดจะถือว่าเป็นตัวประมาณที่ดีที่สุด สามารถคำนวณได้ดังนี้

$$MSE_m = \frac{\sum_{p=1}^{p+1} MSE_p}{p+1}$$

$$MSEm = \sqrt{MSEm}$$

เมื่อ m คือ วิธีการประมาณค่าสัมประสิทธิ์ความถดถอย

$$AMSE_m = \frac{\sum_{i=1}^{loops} MSE_i}{loops}$$

โดย $loops$ คือจำนวนรอบที่ทำซ้ำ

$$SD = \sqrt{\frac{\sum_{j=1}^{loops}\left(MSE_m - AMSE\right)^2}{loops}}$$

ซึ่ง SD คือ หมายถึง ค่าส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนกำลังสองของ AMSE

เพื่อความสะดวกและรวดเร็วในการนำเสนอผลการวิจัยจึงขอใช้สัญลักษณ์ต่าง ๆ ในตารางรูปภาพและการสรุปผลซึ่งมีความหมาย ดังนี้

OLS หมายถึง ตัวประมาณกำลังสองน้อยสุดสามัญ

OLS-ORT หมายถึง ตัวประมาณกำลังสองน้อยสุดสามัญหลังจากแก้ไข X ด้วยวิธีทำให้เป็นแนวตั้งเชิงตั้งฉาก

RID หมายถึง ตัวประมาณริดจ์

PCR หมายถึง ตัวประมาณความถดถอยองค์ประกอบหลัก

σ_ε^2 หมายถึง ความแปรปรวนของความคลาดเคลื่อนในตัวแบบ

σ_ε หมายถึง ค่าความเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนในตัวแบบ

n หมายถึง ขนาดตัวอย่าง

p หมายถึง จำนวนตัวแปรอิสระ

correlation หมายถึง ระดับความสัมพันธ์ที่เกิดขึ้นระหว่างตัวแปรอิสระ

AMSE หมายถึง ค่าเฉลี่ยรากของค่าคลาดเคลื่อนกำลังสอง

SD หมายถึง ค่าส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนกำลังสองของ AMSE

Diff หมายถึง สัดส่วนความแตกต่างของวิธีค่าต่ำสุดกับวิธีอื่น

การวิจัยจะใช้สถานการณ์ที่การแจกแจงของความคลาดเคลื่อนสุ่มในตัวแปรตาม เป็นแบบปกติซึ่งมี ค่าเฉลี่ย 0 และความแปรปรวนเป็น 1,5,10 และมีการกำหนดสถานการณ์อื่นดังนี้

- ตัวแปรอิสระ(p) คือ 3,6,9
- ขนาดตัวอย่างที่ใช้คือระดับน้อย (5p,10p) ระดับปานกลาง(15p,20p) ระดับมาก (25p,30p) โดย p คือจำนวนตัวแปรอิสระ
- ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ คือ
    - ระดับน้อย : ช่วง 0.15-0.30
    - ระดับปานกลาง : ช่วง 0.31-0.50 และ ช่วง 0.51-0.65
    - ระดับมาก : ช่วง 0.65-0.85 และ ช่วง 0.86-0.99

ค่าจากตารางในแต่ละกรณีและแต่ละตัวประมาณจะแสดงค่าตัวเลข 3 ค่า ได้แก่ AMSE ค่าส่วนเบี่ยงเบนมาตรฐานของ AMSE จากการกระทำซ้ำ n รอบ ซึ่งแสดงอยู่ในวงเล็บ และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย ซึ่งเป็นเกณฑ์ประกอบการตัดสินใจหาได้จากผลต่างของวิธีซึ่งให้ค่าเฉลี่ยรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดกับตัวประมาณที่เหลืออีก 2 ตัวประมาณ แล้วหารด้วยค่าเฉลี่ยรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดโดยมีสูตรดังนี้

$$DIFF = \frac{(AMSE_m - AMSE_{\min})}{AMSE_{\min}} \times 100\ \% \qquad ; \quad m = 1, 2, 3, 4$$

โดย $DIFF$ แทนเปอร์เซ็นต์ของอัตราส่วนผลต่างค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจากการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณคูณดังวิธีต่าง ๆ

ผู้วิจัยเสนอผลการวิจัยโดยแบ่งเป็น 2 ส่วนคือ

ส่วนที่ 1 ผลการเปรียบเทียบตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณ ในทุกกรณีของ n และ p

ส่วนที่ 2 การวิเคราะห์ผล โดยพิจารณาจากตารางผลการเปรียบเทียบ และ กราฟ

4.1 ส่วนที่ 1 : การเปรียบเทียบตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณ ในทุกกรณีของ n และ p

ตารางที่	จำนวนตัวแปรอิสระ		ขนาดตัวอย่าง	
4.1.1	น้อย	3	น้อย	5p
4.1.2				10p
4.1.3			ปานกลาง	15p
4.1.4				20p
4.1.5			มาก	25p
4.1.6				30p
4.1.7	ปานกลาง	6	น้อย	5p
4.1.8				10p
4.1.9			ปานกลาง	15p
4.1.10				20p
4.1.11			มาก	25p
4.1.12				30p
4.1.13	มาก	9	น้อย	5p
4.1.14				10p
4.1.15			ปานกลาง	15p
4.1.16				20p
4.1.17			มาก	25p
4.1.18				30p

ตารางที่ 4.1.1 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์ เกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 3 และ ขนาดตัวอย่าง เท่ากับ 5p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.2660	0.2959	0.2876	0.3589	1.4671	1.2914	1.4056	1.3236	2.8189	2.3749	2.6337	2.4407
	(SD)	0.0381	0.0332	0.0372	0.0616	0.0304	0.0634	0.0987	0.0491	0.2641	0.1116	0.1234	0.3317
	%DIFF	0.0000	11.2207	8.1195	34.8933	13.6057	0.0000	8.8414	2.4915	18.6959	0.0000	10.8986	2.7718
0.31-0.5	AMSE	0.2896	0.3378	0.2992	0.3764	1.4906	1.3228	1.4214	1.3492	2.9730	2.3846	2.6432	2.4658
	(SD)	0.0436	0.0587	0.0466	0.0814	0.2107	0.1057	0.2823	0.1038	0.3476	0.2455	0.2594	0.4444
	%DIFF	0.0000	16.6336	3.2974	29.9612	12.6793	0.0000	7.4537	1.9938	24.6716	0.0000	10.8424	3.4031
0.51-0.65	AMSE	0.3370	0.4146	0.3435	0.4214	1.9135	1.6247	1.7811	1.7375	2.9923	2.5090	2.6565	2.5917
	(SD)	0.0452	0.0366	0.0506	0.0961	0.3036	0.1709	0.3620	0.2655	0.2977	0.1969	0.2487	0.4631
	%DIFF	0.0000	23.0119	1.9362	25.0371	17.7743	0.0000	9.6296	6.9445	19.2637	0.0000	5.8778	3.2961
0.66-0.85	AMSE	0.3584	0.4304	0.3624	0.5129	1.9523	1.6677	1.7963	1.6361	4.0984	3.4689	3.9595	3.6349
	(SD)	0.0221	0.0382	0.0431	0.0817	0.1097	0.1688	0.1602	0.1332	0.7846	0.3809	0.8391	0.9078
	%DIFF	0.0000	20.0823	1.1230	43.1013	19.3252	1.9330	9.7948	0.0000	18.1494	0.0000	14.1430	4.7854
0.85-0.99	AMSE	0.6227	0.5566	0.6348	0.3367	3.9307	3.1037	3.8640	1.4044	7.4643	5.7962	7.2651	2.8493
	(SD)	0.0508	0.0692	0.0623	0.0808	0.4190	0.1956	0.4425	0.1131	1.3568	1.1049	1.2775	0.3819
	%DIFF	84.9547	65.3349	88.5638	0.0000	179.8914	121.0022	175.1366	0.0000	161.9702	103.4281	154.9798	0.0000

43

จากตารางที่ 4.1.1 : กรณีนี้ทั้งตัวแปรอิสระและขนาดตัวอย่าง อยู่ในระดับน้อย พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.65)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.66-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5 พบว่า วิธี PCR จะดีสุด รองลงมาคือวิธี OLS-ORT ,วิธี RID และวิธี OLS ตามลำดับ ซึ่งการที่ วิธี PCR ดีกว่า OLS-ORT เพราะ ระดับสหสัมพันธ์ในช่วงนี้ นับว่ามากระดับหนึ่ง ซึ่งย่อมมีปัญหาพหุสัมพันธ์ในระดับพอควร ซึ่งเมื่อมาพิจารณา จำนวนตัวแปรอิสระที่น้อย และ การกระจายข้อมูลที่ปานกลาง จึงทำให้มีโอกาสตัดแกนที่ไม่จำเป็นออกได้ ในการกระบวนการของวิธี PCR ดังนั้น เมื่อมีการตัดแกนออก วิธี PCR ย่อมดีกว่า จากความแปรปรวนที่ลดลง

3. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุด อย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ

ตารางที่ 4.1.2 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 3 และ ขนาดตัวอย่าง เท่ากับ 10p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.1994	0.2175	0.2027	0.2241	0.9058	0.7739	0.8918	0.8158	1.8823	1.5911	1.8762	1.6463
	(SD)	0.0256	0.0220	0.0285	0.0250	0.1662	0.1692	0.1839	0.1668	0.0234	0.0173	0.0893	0.1074
	%DIFF	0.0000	9.0647	1.6550	12.3872	17.0479	0.0000	15.2452	5.4145	18.3071	0.0000	17.9221	3.4741
0.31-0.5	AMSE	0.2041	0.2182	0.2051	0.2504	0.9747	0.8940	0.9263	0.9125	2.1692	1.7913	1.9398	1.8315
	(SD)	0.0470	0.0540	0.0491	0.0033	0.1482	0.1229	0.1554	0.0762	0.1809	0.1585	0.2414	0.1970
	%DIFF	0.0000	6.9468	0.4901	22.7150	9.0210	0.0000	3.6129	2.0665	21.0998	0.0000	8.2931	2.2442
0.51-0.65	AMSE	0.2290	0.2768	0.2349	0.2895	1.1553	1.0442	1.1164	1.0918	2.3194	1.9505	2.1764	1.9859
	(SD)	0.0277	0.0226	0.0195	0.0594	0.1183	0.1159	0.1032	0.1038	0.2435	0.1944	0.1660	0.1390
	%DIFF	0.0000	20.8602	2.5652	26.3836	10.6376	0.0000	6.9122	4.5610	18.9159	0.0000	11.5831	1.8175
0.66-0.85	AMSE	0.2917	0.3900	0.2975	0.3969	1.3730	1.1788	1.3289	1.2326	2.6527	2.2047	2.5324	2.2846
	(SD)	0.0255	0.0123	0.0268	0.0329	0.1012	0.0956	0.1210	0.1395	0.4284	0.2588	0.4771	0.4264
	%DIFF	0.0000	33.7019	2.0057	36.0675	16.4677	0.0000	12.7267	4.5575	20.3202	0.0000	14.8648	3.6218
0.85-0.99	AMSE	0.5089	0.4633	0.5045	0.3353	2.2206	1.8503	2.1353	1.6730	4.8590	3.7838	4.7914	3.4084
	(SD)	0.0727	0.0492	0.0742	0.0925	0.4351	0.3569	0.4176	0.0514	1.2907	0.9044	1.2594	0.9786
	%DIFF	51.7632	38.1645	50.4436	0.0000	32.7301	10.5947	27.6345	0.0000	42.5588	11.0125	40.5770	0.0000

จากตารางที่ 4.1.2 : กรณีนี้ทั้งตัวแปรอิสระและขนาดตัวอย่าง อยู่ในระดับน้อย พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS ประมาณ 9-18%

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุด อย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน ซึ่งจะพบว่า ประสิทธิภาพของ PCR ดีกว่า OLS ประมาณ 40%

ตารางที่ 4.1.3 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 3 และ ขนาดตัวอย่าง เท่ากับ 15p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.1456	0.1854	0.1487	0.1894	0.6943	0.6072	0.7013	0.6340	1.3598	1.1460	1.3583	1.2431
	(SD)	0.0249	0.0181	0.0290	0.0211	0.0452	0.0289	0.0449	0.0271	0.1084	0.0788	0.0383	0.1202
	%DIFF	0.0000	27.3446	2.1470	30.1099	14.3486	0.0000	15.4891	4.4096	18.6492	0.0000	18.5227	8.4662
0.31-0.5	AMSE	0.1568	0.1936	0.1573	0.2162	0.8141	0.7394	0.8008	0.7600	1.6533	1.2862	1.4638	1.3836
	(SD)	0.0151	0.0353	0.0171	0.0214	0.0700	0.1164	0.0553	0.1303	0.3194	0.2002	0.3272	0.2720
	%DIFF	0.0000	23.4928	0.3509	37.9107	10.0991	0.0000	8.3037	2.7792	28.5409	0.0000	13.8020	7.5725
0.51-0.65	AMSE	0.1935	0.2679	0.1973	0.2720	0.9000	0.8251	0.8822	0.8554	1.8205	1.6331	1.7244	1.6465
	(SD)	0.0311	0.0456	0.0329	0.0345	0.0651	0.0500	0.0738	0.0883	0.3348	0.2254	0.3496	0.2944
	%DIFF	0.0000	38.4188	1.9507	40.5503	9.0873	0.0000	6.9238	3.6786	11.4734	0.0000	5.5890	0.8205
0.66-0.85	AMSE	0.2004	0.2963	0.2011	0.3138	1.0252	0.9048	1.0046	0.9957	2.0774	1.8390	2.0298	1.9070
	(SD)	0.0066	0.0270	0.0091	0.0274	0.0937	0.0561	0.0948	0.0696	0.4896	0.3277	0.4269	0.2867
	%DIFF	0.0000	47.8662	0.3868	56.6134	13.3009	0.0000	11.0270	10.0406	12.9639	0.0000	10.3782	3.7018
0.85-0.99	AMSE	0.3285	0.3048	0.3198	0.2516	2.0286	1.6281	2.0150	1.9697	3.6195	2.8647	3.4668	2.2561
	(SD)	0.0357	0.0255	0.0345	0.0698	0.2367	0.1412	0.2199	0.2450	0.2816	0.2512	0.2429	0.1992
	%DIFF	30.5545	21.1347	27.1165	0.0000	24.5962	0.0000	23.7654	20.9784	60.4277	26.9743	53.6628	0.0000

47

การสรุปผล : กรณีนี้ตัวแปรอิสระน้อย ขนาดตัวอย่างอยู่ในระดับปานกลาง พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุด อย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน

อย่างไรก็ตาม ถ้าพหุสัมพันธ์สูง แต่ไม่สามารถลดองค์ประกอบออกไปได้ วิธี OLS-ORT จะดีสุดแทน ซึ่งจากตารางจะพบว่า ในกรณีความแปรปรวนของข้อมูล คือ 25 ซึ่งนับว่าข้อมูลกระจายปานกลาง โดยที่ ความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) ไม่สามารถตัดแกนออกไปได้ ทำให้วิธี OLS-ORT มีประสิทธิภาพดีกว่า (สามารถดูข้อมูลเพิ่มเติมจากตารางที่ 4.2.3 ได้)

ตารางที่ 4.1.4 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 3 และ ขนาดตัวอย่าง เท่ากับ 20p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.1313	0.1603	0.1329	0.1668	0.6310	0.5857	0.6222	0.6010	1.3786	1.1526	1.3531	1.2845
	(SD)	0.0094	0.0136	0.0154	0.0332	0.1186	0.1122	0.1112	0.1441	0.1568	0.1484	0.1164	0.1148
	%DIFF	0.0000	22.0952	1.2762	27.0667	7.7258	0.0000	6.2233	2.6037	19.6052	0.0000	17.3951	11.4391
0.31-0.5	AMSE	0.1393	0.1821	0.1489	0.2160	0.7524	0.6433	0.7456	0.7041	1.4222	1.1919	1.3180	1.2107
	(SD)	0.0145	0.0131	0.0151	0.0226	0.1170	0.1066	0.1241	0.0828	0.1164	0.0629	0.1344	0.1774
	%DIFF	0.0000	30.7361	6.9479	55.0808	16.9607	0.0000	15.9153	9.4559	19.3217	0.0000	10.5774	1.5731
0.51-0.65	AMSE	0.1583	0.2249	0.1605	0.2743	0.7621	0.6858	0.7553	0.7169	1.5990	1.4731	1.5683	1.4821
	(SD)	0.0315	0.0292	0.0314	0.0363	0.1090	0.0824	0.0848	0.1479	0.1834	0.1674	0.1509	0.1688
	%DIFF	0.0000	42.0853	1.3902	73.3491	11.1220	0.0000	10.1341	4.5385	8.5449	0.0000	6.4609	0.6076
0.66-0.85	AMSE	0.1864	0.2564	0.1881	0.3093	0.8791	0.7975	0.8562	0.8031	1.4957	1.2958	1.4164	1.3309
	(SD)	0.0260	0.0210	0.0241	0.0622	0.1452	0.0951	0.1421	0.1533	0.2534	0.1491	0.2550	0.1819
	%DIFF	0.0000	37.5352	0.8985	65.8844	10.2257	0.0000	7.3636	0.6991	15.4265	0.0000	9.3049	2.7048
0.85-0.99	AMSE	0.3083	0.2720	0.2904	0.2026	1.8108	1.4606	1.7824	1.6900	2.8864	2.6498	2.8162	1.9498
	(SD)	0.0709	0.0269	0.0715	0.0413	0.3138	0.2413	0.2879	0.2610	0.8623	0.4720	0.8680	0.5414
	%DIFF	52.1718	34.2300	43.3490	0.0000	23.9730	0.0000	22.0303	15.7059	48.0382	35.9020	44.4390	0.0000

49

การสรุปผล : กรณีนี้ตัวแปรอิสระน้อย ขนาดตัวอย่างอยู่ในระดับปานกลางพบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุด อย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน

อย่างไรก็ตาม ถ้าพหุสัมพันธ์สูง แต่ไม่สามารถลดองค์ประกอบออกไปได้ วิธี OLS-ORT จะดีสุดแทน ซึ่งจากตารางจะพบว่า ในกรณีความแปรปรวนของข้อมูล คือ 25 ซึ่งนับว่าข้อมูลกระจายปานกลาง โดยที่ ความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) ไม่สามารถตัดแกนออกไปได้ ทำให้วิธี OLS-ORT มีประสิทธิภาพดีกว่า (สามารถดูข้อมูลเพิ่มเติมจากตารางที่ 4.2.3 ได้)

ตารางที่ 4.1.5 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และ วิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 3 และ ขนาดตัวอย่าง เท่ากับ 25p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.1019	0.1378	0.1027	0.1388	0.5921	0.4922	0.5778	0.5463	1.1995	0.9504	1.1928	1.0621
	(SD)	0.0268	0.0113	0.0246	0.0241	0.0558	0.0440	0.0708	0.0355	0.1865	0.2292	0.1687	0.1583
	%DIFF	0.0000	35.2393	0.8098	36.2699	20.3139	0.0000	17.4032	11.0027	26.2114	0.0000	25.5143	11.7615
0.31-0.5	AMSE	0.1299	0.1585	0.1387	0.2016	0.6128	0.5046	0.6080	0.5784	1.2343	1.1090	1.2040	1.1112
	(SD)	0.0163	0.0088	0.0159	0.0251	0.0696	0.0570	0.0410	0.0605	0.1339	0.1107	0.1050	0.0998
	%DIFF	0.0000	22.0362	6.7937	55.1963	21.4417	0.0000	20.4806	14.6099	11.2990	0.0000	8.5689	0.1984
0.51-0.65	AMSE	0.1328	0.2046	0.1416	0.2457	0.7392	0.6479	0.7238	0.6893	1.2868	1.1607	1.2266	1.1692
	(SD)	0.0110	0.0208	0.0136	0.0064	0.0732	0.0852	0.0753	0.0640	0.2687	0.2049	0.2680	0.2496
	%DIFF	0.0000	54.1243	6.6855	85.0847	14.0873	0.0000	11.7066	6.3781	10.8689	0.0000	5.6800	0.7388
0.66-0.85	AMSE	0.1668	0.2055	0.1682	0.2930	0.7487	0.6529	0.7306	0.6920	1.5609	1.2420	1.5371	1.2965
	(SD)	0.0261	0.0210	0.0235	0.0150	0.0673	0.0636	0.0642	0.0785	0.2263	0.2159	0.2168	0.1750
	%DIFF	0.0000	23.1865	0.8243	75.6595	14.6774	0.0000	11.8974	5.9889	25.6774	0.0000	23.7671	4.3903
0.85-0.99	AMSE	0.2626	0.2121	0.2542	0.1195	1.5503	1.2489	1.5317	1.3739	2.5646	2.2972	2.5282	1.7718
	(SD)	0.0332	0.0529	0.0343	0.0304	0.2594	0.1805	0.2578	0.2085	0.4440	0.3174	0.4112	0.4898
	%DIFF	119.7239	77.4315	112.6543	0.0000	24.1267	0.0000	22.6375	10.0066	44.7481	29.6557	42.6951	0.0000

51

การสรุปผล : กรณีนี้ตัวแปรอิสระน้อย ขนาดตัวอย่างมาก พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุด อย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน

อย่างไรก็ตาม ถ้าพหุสัมพันธ์สูง แต่ไม่สามารถลดองค์ประกอบออกไปได้ วิธี OLS-ORT จะดีสุดแทน ซึ่งจากตารางจะพบว่า ในกรณีความแปรปรวนของข้อมูล คือ 25 ซึ่งนับว่าข้อมูลกระจายปานกลาง โดยที่ ความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) ไม่สามารถตัดแกนออกไปได้ ทำให้วิธี OLS-ORT มีประสิทธิภาพดีกว่า (สามารถดูข้อมูลเพิ่มเติมจากตารางที่ 4.2.3 ได้)

ตารางที่ 4.1.6 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 3 และ ขนาดตัวอย่าง เท่ากับ 30p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.1017	0.1378	0.1133	0.1444	0.4906	0.4094	0.4706	0.4533	1.0147	0.8916	1.0122	0.9177
	(SD)	0.0113	0.0183	0.0151	0.0332	0.0900	0.0868	0.0726	0.0956	0.0999	0.1142	0.1371	0.0841
	%DIFF	0.0000	35.4966	11.4061	41.9862	19.8400	0.0000	14.9548	10.7291	13.8074	0.0000	13.5270	2.9331
0.31-0.5	AMSE	0.1129	0.1443	0.1175	0.1875	0.5868	0.5021	0.5691	0.5293	1.1403	1.1007	1.0707	1.0015
	(SD)	0.0121	0.0120	0.0127	0.0203	0.0748	0.0614	0.0786	0.0551	0.0785	0.0630	0.0868	0.1280
	%DIFF	0.0000	27.7839	4.0071	66.0394	16.8683	0.0000	13.3333	5.4020	13.8539	9.9024	6.9070	0.0000
0.51-0.65	AMSE	0.1138	0.1898	0.1237	0.2280	0.6013	0.5095	0.5760	0.5396	1.3258	1.1211	1.3006	1.1916
	(SD)	0.0091	0.0075	0.0075	0.0159	0.0650	0.0643	0.0700	0.0183	0.1285	0.1414	0.0976	0.0976
	%DIFF	0.0000	66.7472	8.6756	100.3075	18.0079	0.0000	13.0520	5.9028	18.2597	0.0000	16.0118	6.2954
0.66-0.85	AMSE	0.1539	0.2027	0.1543	0.2671	0.6972	0.6406	0.6923	0.6585	1.4750	1.1603	1.4484	1.2357
	(SD)	0.0122	0.0381	0.0129	0.0203	0.1511	0.1234	0.1382	0.1178	0.1970	0.1720	0.1886	0.2309
	%DIFF	0.0000	31.7465	0.2437	73.5500	8.8312	0.0000	8.0585	2.7824	27.1223	0.0000	24.8319	6.4962
0.85-0.99	AMSE	0.2013	0.1154	0.2000	0.0815	1.2969	1.0551	1.2906	1.1943	2.3000	1.9010	2.2505	1.3010
	(SD)	0.0593	0.0330	0.0554	0.0245	0.1100	0.0629	0.1076	0.1234	0.4156	0.3195	0.4116	0.3681
	%DIFF	146.9181	41.5517	145.3542	0.0000	22.9125	0.0000	22.3225	13.1907	76.7852	46.1184	72.9785	0.0000

การสรุปผล : กรณีนี้ตัวแปรอิสระน้อย ขนาดตัวอย่างมาก พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุด อย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน

อย่างไรก็ตาม ถ้าพหุสัมพันธ์สูง แต่ไม่สามารถลดองค์ประกอบออกไปได้ วิธี OLS-ORT จะดีสุดแทน ซึ่งจากตารางจะพบว่า ในกรณีความแปรปรวนของข้อมูล คือ 25 ซึ่งนับว่าข้อมูลกระจายปานกลาง โดยที่ ความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) ไม่สามารถตัดแกนออกไปได้ ทำให้วิธี OLS-ORT มีประสิทธิภาพดีกว่า (สามารถดูข้อมูลเพิ่มเติมจากตารางที่ 4.2.3 ได้)

ตารางที่ 4.1.7 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 6 และ ขนาดตัวอย่าง เท่ากับ 5p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.2042	0.2534	0.2074	0.2587	0.9368	0.8846	0.9202	0.9134	2.1340	1.9401	1.9405	1.9403
	(SD)	0.0039	0.0073	0.0094	0.0155	0.1131	0.0994	0.0831	0.1228	0.0423	0.0734	0.0706	0.1381
	%DIFF	0.0000	24.1092	1.5795	26.7173	5.9010	0.0000	4.0272	3.2529	9.9957	0.0000	0.0232	0.0116
0.31-0.5	AMSE	0.2456	0.2741	0.2463	0.3220	1.1685	1.1064	1.1311	1.1167	2.6744	2.3911	2.5629	2.4953
	(SD)	0.0146	0.0229	0.0192	0.0137	0.0907	0.0924	0.0958	0.0986	0.2329	0.1208	0.2801	0.2924
	%DIFF	0.0000	11.5839	0.2850	31.1177	5.6107	0.0000	2.2370	0.9332	11.8491	0.0000	7.1829	4.3589
0.51-0.65	AMSE	0.2822	0.3146	0.2887	0.3743	1.3788	1.2383	1.3561	1.3113	2.6848	2.3998	2.6145	2.5409
	(SD)	0.0355	0.0266	0.0356	0.0347	0.1233	0.0973	0.1156	0.1022	0.2680	0.1447	0.2869	0.2695
	%DIFF	0.0000	11.4734	2.3213	32.6393	11.3527	0.0000	9.5154	5.8954	11.8794	0.0000	8.9499	5.8798
0.66-0.85	AMSE	0.3349	0.3587	0.3445	0.4423	1.6339	1.3868	1.6182	1.5347	3.3499	2.9908	3.2377	3.1135
	(SD)	0.0275	0.0253	0.0240	0.0386	0.1776	0.1531	0.1700	0.1625	0.4861	0.2998	0.5049	0.5781
	%DIFF	0.0000	7.1221	2.8742	32.0866	17.8204	0.0000	16.6865	10.6670	12.0050	0.0000	8.2536	4.1000
0.85-0.99	AMSE	5.9062	1.7483	5.1016	0.5292	23.7665	7.6186	20.4233	1.8369	56.3364	28.1138	54.3845	4.0044
	(SD)	0.8445	0.5070	1.4795	0.0918	6.8923	2.1332	5.7185	0.0914	11.6894	2.5312	11.6534	0.2494
	%DIFF	1016.0667	230.3647	864.0259	0.0000	1193.8185	314.7488	1011.8200	0.0000	1306.8700	602.0765	1258.1270	0.0000

กรณีนี้ตัวแปรอิสระปานกลาง ขนาดตัวอย่างน้อย พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS ประมาณ 6-18%

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุด อย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน ซึ่งจะพบว่า ประสิทธิภาพของ PCR fuดีกว่า OLS ประมาณ 1000% ขึ้นไป

ตารางที่ 4.1.8 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 6 และ ขนาดตัวอย่าง เท่ากับ 10p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.1455	0.1831	0.1461	0.1951	0.6941	0.6525	0.6799	0.6580	1.3819	1.3045	1.3452	1.3147
	(SD)	0.0118	0.0249	0.0096	0.0040	0.0441	0.0327	0.0370	0.0309	0.1655	0.1346	0.1385	0.1761
	%DIFF	0.0000	25.8247	0.4296	34.0550	6.3793	0.0000	4.1992	0.8352	5.9295	0.0000	3.1219	0.7800
0.31-0.5	AMSE	0.1495	0.1886	0.1506	0.1970	0.7296	0.6828	0.7069	0.7012	1.6321	1.5002	1.5804	1.5152
	(SD)	0.0229	0.0254	0.0230	0.0174	0.0354	0.0188	0.0399	0.0628	0.1638	0.1904	0.1749	0.1751
	%DIFF	0.0000	26.1538	0.7358	31.7893	6.8541	0.0000	3.5296	2.6875	8.7941	0.0000	5.3478	1.0049
0.51-0.65	AMSE	0.1677	0.2229	0.1706	0.2323	0.9086	0.8330	0.8993	0.8974	1.7681	1.6044	1.7315	1.6600
	(SD)	0.0088	0.0084	0.0103	0.0361	0.0993	0.0938	0.0990	0.1159	0.2351	0.1896	0.2198	0.2167
	%DIFF	0.0000	32.9705	1.7447	38.5774	9.0726	0.0000	7.9532	7.7281	10.2082	0.0000	7.9253	3.4687
0.66-0.85	AMSE	0.2060	0.2500	0.2075	0.2814	1.1361	0.9924	1.1136	1.1122	2.3107	1.9780	2.3046	2.2708
	(SD)	0.0319	0.0176	0.0297	0.0421	0.0885	0.0484	0.0883	0.1242	0.1390	0.1202	0.1475	0.0838
	%DIFF	0.0000	21.3445	0.7281	36.5732	14.4746	0.0000	12.2075	12.0639	16.8242	0.0000	16.5120	14.8045
0.85-0.99	AMSE	5.0081	0.7005	4.9880	0.4830	19.6842	7.1882	19.3473	1.8312	47.4133	24.5220	46.3432	3.2767
	(SD)	0.4199	0.1523	0.3718	0.0749	5.7084	2.1388	5.2238	0.1142	3.0421	1.2197	3.0801	0.5046
	%DIFF	936.9222	45.0282	932.7708	0.0000	974.9491	292.5471	956.5511	0.0000	1346.9840	648.3757	1314.3239	0.0000

การสรุปผล : กรณีนี้ตัวแปรอิสระปานกลาง ขนาดตัวอย่างน้อย พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS ประมาณ 6-14%

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุด อย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน ซึ่งจะพบว่า ประสิทธิภาพของ PCR fuดีกว่า OLS ประมาณ 1000%

ตารางที่ 4.1.9 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 6 และ ขนาดตัวอย่าง เท่ากับ 15p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.1166	0.1551	0.1188	0.1592	0.6002	0.5611	0.5907	0.5752	1.1463	1.1018	1.1050	1.1040
	(SD)	0.0107	0.0150	0.0114	0.0102	0.0516	0.0335	0.0424	0.0501	0.0784	0.0635	0.0488	0.0423
	%DIFF	0.0000	32.9904	1.8221	36.5273	6.9732	0.0000	5.2756	2.5219	4.0388	0.0000	0.2927	0.2019
0.31-0.5	AMSE	0.1169	0.1563	0.1192	0.1818	0.6058	0.5720	0.5982	0.5897	1.1823	1.1251	1.1545	1.1325
	(SD)	0.0175	0.0150	0.0172	0.0178	0.0729	0.0611	0.0767	0.0956	0.0340	0.0142	0.0301	0.0257
	%DIFF	0.0000	33.6541	1.9707	55.4843	5.9093	0.0000	4.5894	3.0902	5.0819	0.0000	2.6109	0.6577
0.51-0.65	AMSE	0.1515	0.1839	0.1519	0.2220	0.7409	0.6891	0.7290	0.6952	1.4414	1.3454	1.4047	1.3519
	(SD)	0.0262	0.0219	0.0267	0.0331	0.0836	0.0687	0.0771	0.0551	0.0889	0.1455	0.0944	0.1203
	%DIFF	0.0000	21.3531	0.2310	46.5347	7.5134	0.0000	5.7829	0.8852	7.1336	0.0000	4.4076	0.4850
0.66-0.85	AMSE	0.1701	0.2340	0.1741	0.2516	0.8241	0.7561	0.8172	0.8136	1.7541	1.5666	1.7519	1.7127
	(SD)	0.0224	0.0175	0.0229	0.0074	0.0924	0.0887	0.0860	0.0826	0.1009	0.1060	0.0842	0.1028
	%DIFF	0.0000	37.5717	2.3372	47.9494	8.9968	0.0000	8.0776	7.5982	11.9690	0.0000	11.8333	9.3262
0.85-0.99	AMSE	4.7180	0.6640	4.2299	0.3967	14.9902	6.8416	13.0405	1.5454	41.5859	21.7374	39.4807	3.1446
	(SD)	0.8018	0.1926	0.8395	0.0391	1.9473	1.0407	2.0182	0.1025	11.2282	5.8691	10.2650	0.1852
	%DIFF	1089.4491	67.4020	966.4125	0.0000	870.0040	342.7113	743.8405	0.0000	1222.4649	591.2659	1155.5162	0.0000

การสรุปผล : กรณีนี้ตัวแปรอิสระปานกลาง ขนาดตัวอย่างปานกลาง พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS ประมาณ 6-8%

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุด อย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน ซึ่งจะพบว่า ประสิทธิภาพของ PCR fuดีกว่า OLS ประมาณ 1000%

ตารางที่ 4.1.10 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 6 และ ขนาดตัวอย่าง เท่ากับ 20p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.0901	0.1221	0.0902	0.1381	0.4819	0.4573	0.4760	0.4654	0.8711	0.8371	0.8505	0.8448
	(SD)	0.0071	0.0114	0.0069	0.0104	0.0387	0.0244	0.0441	0.0275	0.1065	0.1000	0.1148	0.1098
	%DIFF	0.0000	35.5358	0.1943	53.3592	5.3742	0.0000	4.0840	1.7768	4.0708	0.0000	1.6098	0.9289
0.31-0.5	AMSE	0.1031	0.1422	0.1036	0.1568	0.5530	0.5345	0.5446	0.5411	0.9975	0.9313	0.9803	0.9377
	(SD)	0.0104	0.0138	0.0101	0.0105	0.0305	0.0376	0.0376	0.0375	0.1481	0.1291	0.1361	0.1428
	%DIFF	0.0000	37.9486	0.4365	52.0369	3.4612	0.0000	1.8849	1.2348	7.1055	0.0000	5.2533	0.6818
0.51-0.65	AMSE	0.1282	0.1546	0.1284	0.1952	0.6698	0.6224	0.6674	0.6336	1.2607	1.1655	1.2291	1.2067
	(SD)	0.0114	0.0073	0.0105	0.0115	0.0845	0.0420	0.0800	0.0970	0.0843	0.0681	0.0849	0.1057
	%DIFF	0.0000	20.5773	0.1365	52.2528	7.6117	0.0000	7.2220	1.7915	8.1750	0.0000	5.4593	3.5351
0.66-0.85	AMSE	0.1526	0.1924	0.1530	0.2235	0.7354	0.6579	0.7286	0.7110	1.4964	1.3132	1.4816	1.4501
	(SD)	0.0074	0.0129	0.0087	0.0298	0.0801	0.0655	0.0770	0.0690	0.1053	0.1002	0.1130	0.1037
	%DIFF	0.0000	26.0691	0.2949	46.4526	11.7723	0.0000	10.7425	8.0635	13.9509	0.0000	12.8277	10.4289
0.85-0.99	AMSE	4.3359	0.6138	3.3693	0.3286	12.9624	6.6191	12.3651	1.4646	33.1772	18.2785	30.1525	2.3844
	(SD)	1.2574	0.1780	0.9771	0.0443	3.4998	1.8533	3.3386	0.1772	9.2896	5.3008	7.8397	0.3719
	%DIFF	1219.5922	86.7991	925.4356	0.0000	785.0622	351.9434	744.2791	0.0000	1291.4266	666.5859	1164.5749	0.0000

61

การสรุปผล : กรณีนี้ตัวแปรอิสระปานกลาง ขนาดตัวอย่างปานกลาง พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS ประมาณ 6-11%

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุด อย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน ซึ่งจะพบว่า ประสิทธิภาพของ PCR fuดีกว่า OLS ประมาณ 800-1000%

ตารางที่ 4.1.11 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 6 และ ขนาดตัวอย่าง เท่ากับ 25p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.0797	0.1087	0.0810	0.1315	0.4199	0.3983	0.4187	0.4093	0.8665	0.8143	0.8488	0.8283
	(SD)	0.0076	0.0059	0.0094	0.0062	0.0244	0.0285	0.0117	0.0228	0.0886	0.0954	0.0773	0.0930
	%DIFF	0.0000	36.3437	1.5679	64.9420	5.4426	0.0000	5.1412	2.7684	6.4073	0.0000	4.2368	1.7131
0.31-0.5	AMSE	0.0947	0.1306	0.0957	0.1564	0.4921	0.4805	0.4917	0.4905	0.9095	0.8348	0.8866	0.8501
	(SD)	0.0072	0.0108	0.0083	0.0054	0.0241	0.0242	0.0230	0.0228	0.0270	0.0288	0.0320	0.0275
	%DIFF	0.0000	37.9292	1.1358	65.1875	2.4036	0.0000	2.3256	2.0811	8.9488	0.0000	6.2114	1.8329
0.51-0.65	AMSE	0.1070	0.1615	0.1076	0.1803	0.5234	0.4927	0.5175	0.5059	1.0319	0.9559	1.0229	0.9733
	(SD)	0.0109	0.0083	0.0114	0.0050	0.0152	0.0220	0.0204	0.0310	0.1123	0.0901	0.1110	0.1183
	%DIFF	0.0000	50.8879	0.5140	68.5047	6.2208	0.0000	5.0335	2.6740	7.9510	0.0000	7.0095	1.8282
0.66-0.85	AMSE	0.1414	0.1932	0.1417	0.2137	0.7143	0.6598	0.7123	0.7080	1.3950	1.2310	1.3921	1.3548
	(SD)	0.0195	0.0184	0.0193	0.0217	0.0286	0.0465	0.0237	0.0155	0.2233	0.1419	0.2414	0.2574
	%DIFF	0.0000	36.5918	0.1768	51.0871	8.2563	0.0000	7.9570	7.2977	13.3205	0.0000	13.0869	10.0589
0.85-0.99	AMSE	2.8995	0.5231	1.9082	0.2726	12.2898	5.3902	12.0360	1.1397	20.5416	12.3738	18.0666	2.1511
	(SD)	0.2330	0.0961	0.2360	0.0593	3.5640	1.4357	3.2497	0.1286	5.7517	3.2172	5.0587	0.1486
	%DIFF	963.5397	91.8845	599.9358	0.0000	978.3557	372.9594	956.0884	0.0000	854.9358	475.2301	739.8784	0.0000

การสรุปผล : กรณีนี้ตัวแปรอิสระปานกลาง ขนาดตัวอย่างมาก พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS ประมาณ 2-7%

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุดอย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน ซึ่งจะพบว่า ประสิทธิภาพของ PCR fuดีกว่า OLS ประมาณ 800-900%

ตารางที่ 4.1.12 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 6 และ ขนาดตัวอย่าง เท่ากับ 30p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.0727	0.0950	0.0729	0.1223	0.3815	0.3402	0.3743	0.3559	0.7266	0.6734	0.7157	0.7115
	(SD)	0.0069	0.0116	0.0068	0.0081	0.0154	0.0139	0.0115	0.0213	0.0533	0.0543	0.0527	0.0513
	%DIFF	0.0000	30.7190	0.2408	68.2835	12.1326	0.0000	10.0088	4.6149	7.8962	0.0000	6.2776	5.6576
0.31-0.5	AMSE	0.0751	0.0998	0.0755	0.1458	0.4173	0.3871	0.4101	0.4028	0.8458	0.7574	0.8322	0.7874
	(SD)	0.0056	0.0030	0.0055	0.0103	0.0273	0.0200	0.0274	0.0167	0.0462	0.0088	0.0445	0.0490
	%DIFF	0.0000	32.8562	0.5326	94.1079	7.7882	0.0000	5.9412	4.0426	11.6752	0.0000	9.8729	3.9677
0.51-0.65	AMSE	0.0950	0.1255	0.0973	0.1774	0.5011	0.4825	0.5008	0.4998	0.9830	0.9249	0.9705	0.9327
	(SD)	0.0078	0.0184	0.0079	0.0211	0.0549	0.0582	0.0445	0.0428	0.0876	0.0773	0.0938	0.1051
	%DIFF	0.0000	32.0789	2.4474	86.7632	3.8605	0.0000	3.8087	3.5859	6.2764	0.0000	4.9276	0.8406
0.66-0.85	AMSE	0.1200	0.1668	0.1205	0.2092	0.5907	0.5535	0.5871	0.5860	1.2110	1.0678	1.1935	1.1464
	(SD)	0.0114	0.0090	0.0108	0.0115	0.0249	0.0196	0.0224	0.0223	0.1332	0.1130	0.1231	0.1205
	%DIFF	0.0000	38.9919	0.3541	74.3179	6.7254	0.0000	6.0705	5.8672	13.4157	0.0000	11.7721	7.3658
0.85-0.99	AMSE	2.7366	0.4817	1.7388	0.2110	9.9749	4.2850	9.1858	1.1090	16.7963	11.2883	15.7841	2.0700
	(SD)	0.3281	0.1262	0.3272	0.0440	2.6932	1.1570	2.3883	0.1932	1.9230	3.0478	1.9254	0.4017
	%DIFF	1197.1205	128.3090	724.1616	0.0000	799.4477	286.3841	728.2913	0.0000	711.4130	445.3261	662.5169	0.0000

การสรุปผล : กรณีนี้ตัวแปรอิสระปานกลาง ขนาดตัวอย่างมาก พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS ประมาณ 7-12%

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุดอย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน ซึ่งจะพบว่า ประสิทธิภาพของ PCR fuดีกว่า OLS ประมาณ 800%

ตารางที่ 4.1.13 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 9 และ ขนาดตัวอย่าง เท่ากับ 5p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.1773	0.2203	0.1805	0.2475	0.8960	0.8400	0.8835	0.8710	1.8810	1.7370	1.8336	1.7802
	(SD)	0.0056	0.0085	0.0048	0.0147	0.0310	0.0273	0.0356	0.0465	0.1920	0.1604	0.1854	0.2092
	%DIFF	0.0000	24.2668	1.7908	39.6080	6.6575	0.0000	5.1725	3.6814	8.2902	0.0000	5.5613	2.4885
0.31-0.5	AMSE	0.2164	0.2447	0.2190	0.2742	0.9699	0.9046	0.9494	0.9468	1.8582	1.7375	1.8371	1.7837
	(SD)	0.0083	0.0197	0.0102	0.0293	0.0813	0.0371	0.0835	0.0846	0.1338	0.1229	0.1437	0.1344
	%DIFF	0.0000	13.0922	1.2249	26.7391	7.2189	0.0000	4.9498	4.6679	6.9423	0.0000	5.7280	2.6590
0.51-0.65	AMSE	0.2292	0.2581	0.2316	0.3180	1.1433	1.0437	1.1199	1.0953	2.3209	2.0772	2.2839	2.1913
	(SD)	0.0311	0.0231	0.0290	0.0362	0.1304	0.1323	0.1403	0.1411	0.1429	0.1597	0.1341	0.1463
	%DIFF	0.0000	12.6227	1.0583	38.7628	9.5506	0.0000	7.3013	4.9514	11.7297	0.0000	9.9521	5.4942
0.66-0.85	AMSE	0.2717	0.3116	0.2783	0.3723	1.5290	1.3475	1.5199	1.5051	2.6992	2.4317	2.6904	2.6723
	(SD)	0.0102	0.0226	0.0162	0.0327	0.1850	0.1617	0.1591	0.1173	0.1503	0.1415	0.1378	0.1414
	%DIFF	0.0000	14.6761	2.4384	37.0261	13.4736	0.0000	12.8001	11.7017	11.0006	0.0000	10.6398	9.8944
0.85-0.99	AMSE	9.3736	6.0632	8.2162	0.4355	60.2590	45.4417	58.3080	1.6132	134.2222	40.6617	130.3150	2.8696
	(SD)	1.9425	7.7240	2.9173	0.0595	36.9302	52.3115	36.6542	0.1345	64.2315	16.3183	64.5619	0.1151
	%DIFF	2052.4887	1292.3245	1786.7157	0.0000	3635.4865	2716.9529	3514.5414	0.0000	4577.4252	1316.9947	4441.2648	0.0000

67

การสรุปผล : กรณีนี้ตัวแปรอิสระมาก ขนาดตัวอย่างน้อย พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS ประมาณ 7-11%

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุด อย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน ซึ่งจะพบว่า ประสิทธิภาพของ PCR fuดีกว่า OLS ประมาณ 2000% สำหรับจำนวนตัวแปรอิสระเท่ากับ 3, ประสิทธิภาพของ PCR fuดีกว่า OLS ประมาณ 3000% สำหรับจำนวนตัวแปรอิสระเท่ากับ 6 และ ประสิทธิภาพของ PCR fuดีกว่า OLS ประมาณ 4000% สำหรับจำนวนตัวแปรอิสระเท่ากับ 9

ตารางที่ 4.1.14 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์เกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 9 และ ขนาดตัวอย่าง เท่ากับ 10p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.1153	0.1527	0.1159	0.1573	0.5922	0.5706	0.5817	0.5744	1.1124	1.0489	1.0884	1.0531
	(SD)	0.0168	0.0101	0.0183	0.0196	0.0588	0.0483	0.0585	0.0610	0.0412	0.0547	0.0617	0.0498
	%DIFF	0.0000	32.3938	0.4770	36.3833	3.7855	0.0000	1.9453	0.6616	6.0590	0.0000	3.7708	0.4052
0.31-0.5	AMSE	0.1308	0.1779	0.1310	0.1841	0.6231	0.5996	0.6220	0.6135	1.1864	1.1157	1.1652	1.1179
	(SD)	0.0084	0.0101	0.0094	0.0061	0.0523	0.0293	0.0549	0.0667	0.0649	0.0718	0.0758	0.0774
	%DIFF	0.0000	36.0161	0.1912	40.7953	3.9149	0.0000	3.7273	2.3056	6.3323	0.0000	4.4344	0.1927
0.51-0.65	AMSE	0.1617	0.2027	0.1635	0.2175	0.7760	0.7018	0.7681	0.7548	1.4963	1.3688	1.4660	1.4454
	(SD)	0.0037	0.0089	0.0015	0.0183	0.1034	0.0884	0.1078	0.1121	0.1400	0.0904	0.1241	0.1244
	%DIFF	0.0000	25.3517	1.0821	34.4876	10.5807	0.0000	9.4478	7.5597	9.3167	0.0000	7.1012	5.5981
0.66-0.85	AMSE	0.1832	0.2227	0.1857	0.2529	0.9186	0.8536	0.9111	0.9108	2.0025	1.7582	1.9862	1.9633
	(SD)	0.0172	0.0083	0.0181	0.0412	0.0780	0.0900	0.0766	0.0862	0.0855	0.0881	0.0937	0.0990
	%DIFF	0.0000	21.5777	1.3648	38.0374	7.6209	0.0000	6.7364	6.6983	13.8981	0.0000	12.9696	11.6685
0.85-0.99	AMSE	9.2881	4.0808	7.1083	0.2938	35.5873	16.3492	34.6025	1.2113	56.6675	35.2986	51.1218	2.6180
	(SD)	6.2098	2.2543	5.5478	0.0278	11.6881	7.9283	11.6962	0.1180	31.4782	29.6605	32.2835	0.1679
	%DIFF	3061.0908	1288.8624	2319.2121	0.0000	2838.0619	1249.7771	2756.7595	0.0000	2064.5133	1248.2921	1852.6857	0.0000

69

การสรุปผล : กรณีนี้ตัวแปรอิสระมาก ขนาดตัวอย่างน้อย พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS ประมาณ 3-8%

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุด อย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน ซึ่งจะพบว่า ประสิทธิภาพของ PCR fuดีกว่า OLS มาก

ตารางที่ 4.1.15 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 9 และ ขนาดตัวอย่าง เท่ากับ 15p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.0940	0.1249	0.0945	0.1441	0.4937	0.4817	0.4910	0.4845	0.9411	0.9006	0.9013	0.9132
	(SD)	0.0044	0.0080	0.0037	0.0052	0.0382	0.0283	0.0373	0.0308	0.0869	0.0919	0.0725	0.0846
	%DIFF	0.0000	32.9431	0.6120	53.3795	2.4964	0.0000	1.9307	0.5813	4.4131	0.0000	1.0318	1.3231
0.31-0.5	AMSE	0.1092	0.1454	0.1108	0.1553	0.5479	0.5350	0.5477	0.5386	1.0211	0.9528	0.9925	0.9708
	(SD)	0.0103	0.0136	0.0107	0.0124	0.0060	0.0076	0.0074	0.0058	0.0948	0.0821	0.0840	0.1040
	%DIFF	0.0000	33.1654	1.4888	42.3042	2.4064	0.0000	2.3597	0.6729	7.1687	0.0000	4.1669	1.8919
0.51-0.65	AMSE	0.1272	0.1683	0.1278	0.1816	0.6248	0.5838	0.6212	0.6158	1.2310	1.1162	1.2212	1.1970
	(SD)	0.0105	0.0079	0.0114	0.0184	0.0322	0.0387	0.0283	0.0303	0.0650	0.0560	0.0626	0.0642
	%DIFF	0.0000	32.2460	0.4323	42.7589	7.0184	0.0000	6.4060	5.4682	10.2804	0.0000	9.4069	7.2366
0.66-0.85	AMSE	0.1504	0.2022	0.1505	0.2156	0.7672	0.7106	0.7608	0.7561	1.5344	1.3817	1.5278	1.5115
	(SD)	0.0183	0.0174	0.0177	0.0156	0.0412	0.0359	0.0406	0.0682	0.0534	0.0714	0.0639	0.0575
	%DIFF	0.0000	34.4581	0.0499	43.3344	7.9578	0.0000	7.0642	6.4028	11.0500	0.0000	10.5741	9.3962
0.85-0.99	AMSE	8.8744	2.5011	6.3241	0.2781	31.3777	10.3748	31.3661	1.1309	51.6394	11.5065	47.6415	2.1086
	(SD)	1.6635	0.4492	1.3887	0.0636	18.5300	5.0670	18.5360	0.2003	16.2763	2.3217	16.2929	0.2344
	%DIFF	3091.3513	799.4156	2174.2516	0.0000	2674.6960	817.4360	2673.6725	0.0000	2348.9608	445.6885	2159.3645	0.0000

71

การสรุปผล : กรณีนี้ตัวแปรอิสระมาก ขนาดตัวอย่างปานกลาง พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS ประมาณ 3-11%

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุด อย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน ซึ่งจะพบว่า ประสิทธิภาพของ PCR fuดีกว่า OLS มาก

ตารางที่ 4.1.16 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 9 และ ขนาดตัวอย่าง เท่ากับ 20p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.0798	0.1055	0.0801	0.1255	0.4020	0.3956	0.3983	0.3971	0.8094	0.7647	0.7917	0.7750
	(SD)	0.0056	0.0089	0.0054	0.0035	0.0180	0.0153	0.0164	0.0123	0.0566	0.0480	0.0541	0.0631
	%DIFF	0.0000	32.2156	0.3447	57.3488	1.6114	0.0000	0.6761	0.3728	5.8387	0.0000	3.5307	1.3469
0.31-0.5	AMSE	0.0881	0.1268	0.0886	0.1393	0.4620	0.4328	0.4606	0.4403	0.8295	0.7904	0.8134	0.7963
	(SD)	0.0038	0.0058	0.0040	0.0050	0.0689	0.0566	0.0604	0.0569	0.0711	0.0734	0.0811	0.0669
	%DIFF	0.0000	43.9149	0.5106	58.0142	6.7525	0.0000	6.4233	1.7329	4.9535	0.0000	2.9164	0.7560
0.51-0.65	AMSE	0.0984	0.1382	0.0986	0.1608	0.5462	0.5195	0.5448	0.5375	1.1321	1.0503	1.1251	1.0992
	(SD)	0.0078	0.0045	0.0069	0.0061	0.0513	0.0534	0.0501	0.0562	0.0794	0.0605	0.0856	0.0874
	%DIFF	0.0000	40.5186	0.2542	63.4469	5.1299	0.0000	4.8701	3.4697	7.7932	0.0000	7.1267	4.6559
0.66-0.85	AMSE	0.1354	0.1872	0.1359	0.1999	0.7298	0.6606	0.7283	0.7226	1.3328	1.2243	1.3195	1.3013
	(SD)	0.0056	0.0077	0.0059	0.0104	0.0698	0.0383	0.0678	0.0546	0.1028	0.0801	0.1022	0.0865
	%DIFF	0.0000	38.1946	0.3138	47.5725	10.4711	0.0000	10.2403	9.3813	8.8646	0.0000	7.7782	6.2896
0.85-0.99	AMSE	7.7403	2.2342	5.2546	0.2658	27.3363	9.9588	26.1181	1.0973	45.9241	10.9839	45.0383	2.1769
	(SD)	2.1264	1.0583	3.2552	0.0402	13.6312	5.7303	13.7857	0.1402	19.9159	4.8770	19.8110	0.3514
	%DIFF	2812.0673	740.5662	1876.9093	0.0000	2391.1763	807.5547	2280.1563	0.0000	2009.6100	404.5638	1968.9191	0.0000

73

การสรุปผล : กรณีนี้ตัวแปรอิสระมาก ขนาดตัวอย่างปานกลาง พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS ประมาณ 3-11%

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุด อย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน ซึ่งจะพบว่า ประสิทธิภาพของ PCR fuดีกว่า OLS มาก

ตารางที่ 4.1.17 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 9 และ ขนาดตัวอย่าง เท่ากับ 25p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.0700	0.0994	0.0704	0.1149	0.3614	0.3461	0.3560	0.3522	0.6780	0.6455	0.6739	0.6534
	(SD)	0.0060	0.0106	0.0066	0.0068	0.0050	0.0089	0.0075	0.0049	0.0301	0.0305	0.0206	0.0324
	%DIFF	0.0000	42.0658	0.6076	64.1887	4.4059	0.0000	2.8386	1.7551	5.0308	0.0000	4.3879	1.2238
0.31-0.5	AMSE	0.0754	0.1113	0.0755	0.1307	0.3884	0.3723	0.3881	0.3795	0.7884	0.7478	0.7827	0.7577
	(SD)	0.0051	0.0054	0.0052	0.0063	0.0306	0.0239	0.0279	0.0194	0.0807	0.0722	0.0856	0.0710
	%DIFF	0.0000	47.5796	0.1326	73.3090	4.3452	0.0000	4.2445	1.9476	5.4261	0.0000	4.6672	1.3306
0.51-0.65	AMSE	0.0892	0.1236	0.0898	0.1588	0.4780	0.4496	0.4742	0.4666	0.8578	0.8081	0.8496	0.8349
	(SD)	0.0095	0.0092	0.0097	0.0144	0.0363	0.0237	0.0329	0.0300	0.0339	0.0341	0.0323	0.0242
	%DIFF	0.0000	38.5930	0.6726	78.0549	6.3108	0.0000	5.4545	3.7754	6.1506	0.0000	5.1420	3.3166
0.66-0.85	AMSE	0.1127	0.1646	0.1131	0.1885	0.5711	0.5327	0.5582	0.5356	1.1109	1.0114	1.1062	1.0903
	(SD)	0.0092	0.0052	0.0093	0.0051	0.0422	0.0453	0.0443	0.0450	0.1500	0.1269	0.1430	0.1451
	%DIFF	0.0000	46.0617	0.3994	67.2953	7.2035	0.0000	4.7820	0.5444	9.8327	0.0000	9.3704	7.7935
0.85-0.99	AMSE	6.8487	1.6398	4.3488	0.2580	24.6278	8.3294	20.4683	1.0667	43.1013	10.4659	43.0982	1.8245
	(SD)	2.4986	0.5227	2.3856	0.0266	16.4684	15.7347	21.8492	0.1993	34.5420	5.3555	34.3785	0.4021
	%DIFF	2554.2874	535.5198	1585.3987	0.0000	2208.8923	680.8958	1818.9331	0.0000	2262.3933	473.6363	2262.2247	0.0000

75

การสรุปผล : กรณีนี้ตัวแปรอิสระมาก ขนาดตัวอย่างมาก พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS ประมาณ 3-11%

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุดอย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน ซึ่งจะพบว่า ประสิทธิภาพของ PCR fuดีกว่า OLS มากถึงประมาณ 2000%

ตารางที่ 4.1.18 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์เกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 9 และ ขนาดตัวอย่าง เท่ากับ 30p

ระดับความสัมพันธ์		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน											
		1				5				10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR
0.15-0.30	AMSE	0.0629	0.0766	0.0632	0.1036	0.3233	0.3082	0.3177	0.3165	0.6602	0.6229	0.6432	0.6280
	(SD)	0.0036	0.0049	0.0039	0.0107	0.0255	0.0282	0.0247	0.0266	0.0546	0.0398	0.0520	0.0370
	%DIFF	0.0000	21.8377	0.5967	64.7574	4.8917	0.0000	3.0989	2.7095	5.9843	0.0000	3.2631	0.8148
0.31-0.5	AMSE	0.0722	0.0950	0.0723	0.1258	0.3676	0.3566	0.3616	0.3598	0.6968	0.6546	0.6839	0.6640
	(SD)	0.0018	0.0079	0.0010	0.0062	0.0287	0.0242	0.0288	0.0369	0.0971	0.0881	0.1026	0.1211
	%DIFF	0.0000	31.5680	0.1385	74.1087	3.0775	0.0000	1.3810	0.8973	6.4548	0.0000	4.4802	1.4361
0.51-0.65	AMSE	0.0842	0.1128	0.0846	0.1537	0.4226	0.4010	0.4134	0.4074	0.8532	0.8009	0.8428	0.8267
	(SD)	0.0078	0.0114	0.0081	0.0105	0.0463	0.0313	0.0494	0.0654	0.0360	0.0495	0.0341	0.0322
	%DIFF	0.0000	34.0362	0.4752	82.6255	5.3931	0.0000	3.0925	1.6023	6.5299	0.0000	5.2283	3.2150
0.66-0.85	AMSE	0.1102	0.1493	0.1102	0.1765	0.4694	0.4311	0.4689	0.4627	1.1103	0.9783	1.1071	1.0821
	(SD)	0.0090	0.0111	0.0088	0.0089	0.0521	0.0376	0.0512	0.0475	0.0524	0.0575	0.0468	0.0454
	%DIFF	0.0000	35.5424	0.0227	60.2360	8.8842	0.0000	8.7741	7.3301	13.4986	0.0000	13.1689	10.6185
0.85-0.99	AMSE	6.2913	1.5577	2.5038	0.2436	12.8044	6.5988	11.7006	0.9096	41.6965	9.7031	41.4424	1.8185
	(SD)	0.9888	1.4048	4.5472	0.0210	3.4185	11.7171	5.7602	0.1809	26.9243	11.0904	27.1561	0.3625
	%DIFF	2482.3602	539.3843	927.7168	0.0000	1307.6597	625.4363	1186.3048	0.0000	2192.8747	433.5698	2178.9032	0.0000

77

การสรุปผล : กรณีนี้ตัวแปรอิสระมาก ขนาดตัวอย่างมาก พบว่าค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวน เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ซึ่งจะเกิดปัญหาพหุสัมพันธ์สูง เพราะได้มีการตัดบางองค์ประกอบออก จะทำให้ได้ค่า AMSE ต่ำทั้งที่ระดับความสัมพันธ์สูง

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.15-0.85)

1. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 1 พบว่า วิธี OLS จะดีสุด รองลงมาคือวิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ โดยที่ OLS จะดีกว่าวิธี RID ไม่มากนัก แต่จะดีกว่า วิธี OLS-ORT และวิธี PCR มาก

2. กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน เท่ากับ 5,10 พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR ,วิธี RID และวิธี OLS ตามลำดับ โดยวิธี OLS-ORT มีประสิทธิภาพมากกว่าวิธี OLS ประมาณ 3-11%

ในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ อยู่ในช่วง (0.86-0.99) วิธี PCR จะดีที่สุดอย่างชัดเจนมาก รองลงมา คือ OLS-ORT , RID ,OLS ตามลำดับ ซึ่ง PCR จะให้ค่า AMSE ที่ดีที่สุด เพราะมีการตัดองค์ประกอบออก ทำให้ความแปรปรวนลดลง ในทุกค่าความเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน ซึ่งจะพบว่า ประสิทธิภาพของ PCR fuดีกว่า OLS มากถึง 2000%

เมื่อพิจารณาผลโดยรวมจากตาราง**การเปรียบเทียบตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณ ในทุกกรณีของ** n **และ** p **ที่ได้ในตารางที่** 4.1.1 **ถึง** 4.1.18

กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนของตัวแบบ เท่ากับ 1

สำหรับกรณีระดับความสัมพันธ์อยู่ในช่วงประมาณ 0.15-0.80

กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนของตัวแบบ เท่ากับ 1 และระดับความสัมพันธ์ในระดับที่ไม่สูงมาก ช่วงประมาณ 0.15-0.80 พบว่า วิธี OLS จะดีสุด รองลงมาคือ วิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ เสมอ

เนื่องจาก การที่ข้อมูลมีลักษณะการกระจายที่ไม่มาก และ ระดับความสัมพันธ์ของตัวแปร X นั้นไม่ได้สูงมาก นั้นทำให้มีการเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระบ้าง แต่ไม่สูงมากจนกระทบต่อ การประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณ

ถ้าไม่เกิดปัญหาใด โดยทั่วไปวิธี OLS จะเป็นวิธีการประมาณค่าที่มีประสิทธิภาพมากที่สุด จากสมการ $\underset{\sim OLS}{\hat{\beta}} = (X'X)^{-1}X'\underset{\sim}{y}$ พบว่า ถ้าปัญหาพหุสัมพันธ์ไม่มาก ย่อมไม่ส่งผลต่อการประมาณค่ามากนัก เนื่องจากยังสามารถหาค่า $X'X$ ได้ วิธี OLS จึงยังคงเป็นตัวประมาณที่เหมาะสมมากสุด ในสถานการณ์นี้

สาเหตุที่ทั้ง วิธี RID, OLS-ORT,PCR ไม่ได้มีประสิทธิภาพดีกว่า OLS ทั้งที่เป็นวิธีที่พัฒนาขึ้นมาเพื่อใช้แก้ไขปัญหาพหุสัมพันธ์ เพราะ ทั้ง 3 วิธีนี้เหมาะสำหรับ ข้อมูลที่มีการกระจายในระดับหนึ่งด้วย ยิ่งการกระจายข้อมูลเพิ่มขึ้น ยิ่งมีประสิทธิภาพเพิ่มมากขึ้น ดังนั้นในกรณีนี้ที่ข้อมูลมีการกระจายน้อย แล้วปัญหาพหุสัมพันธ์ไม่มาก จึงทำให้ OLS ยังเป็นวิธีที่มีประสิทธิภาพมากสุด สำหรับทุกจำนวนของตัวแปรอิสระ และขนาดตัวอย่าง

สำหรับกรณีระดับความสัมพันธ์อยู่ในช่วงประมาณ 0.85-0.99

พบว่า PCR จะเป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมา คือ OLS-ORT, วิธี RID, วิธี OLS เพราะ วิธี PCR นั้นจะมีการตัดองค์ประกอบบางส่วนออกไป ส่งผลให้ลดความแปรปรวนให้ต่ำลง

กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนของตัวแบบ เท่ากับ 5,10

สำหรับกรณีระดับความสัมพันธ์อยู่ในช่วงประมาณ 0.15-0.85

ส่วนใหญ่ถึง 99% ของสถานการณ์ พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR, วิธี RID และวิธี OLS ตามลำดับ เสมอ เนื่องจาก OLS-ORT เป็นวิธีที่ยังคงรักษาประสิทธิภาพที่ดีของวิธี OLS เดิมไว้ และไปแก้ไขปัญหาพหุสัมพันธ์ที่ข้อมูลเลย คือ การไปแปลงค่า X ที่มีปัญหาพหุสัมพันธ์ โดยทำให้ แต่ละ X ไม่ได้ตั้งฉากกัน ให้เป็น X ที่ตั้งฉากกัน ด้วยวิธี orthogonalization ของ Gram-Schmidt จึงทำให้ OLS-ORT เป็นวิธีที่ดีที่สุดในสถานการณ์นี้

เว้นแต่กรณีที่ จำนวนตัวแปรอิสระเท่ากับ 3 ค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนของตัวแบบ เท่ากับ 5 มีขนาดตัวอย่างน้อย คือ 5p และมีระดับความสัมพันธ์อยู่ในช่วง 0.66-0.85 พบว่า วิธี PCR นั้นให้ค่า AMSE ต่ำสุด ซึ่งมีโอกาสเกิดขึ้นได้ เพราะ เป็นสถานการณ์ที่อยู่ในระดับความสัมพันธ์ระหว่าง X มาก ซึ่งย่อมส่งผลให้มีปัญหาค่าพหุสัมพันธ์สูงระดับหนึ่ง และเมื่อจำนวนตัวแปรอิสระน้อย ข้อมูลมีการกระจายปานกลาง ย่อมทำให้การจัดองค์ประกอบหลักใหม่ ด้วยวิธี PCR สามารถตัดองค์ประกอบที่ไม่จำเป็นออกไปได้ ซึ่งเมื่อมีการตัดองค์ประกอบที่ไม่จำเป็นออกไป ย่อมทำให้ลดความแปรปรวนลงอย่างชัดเจน ทำให้วิธี PCR ดีกว่า OLS-ORT ในสถานการณ์นี้

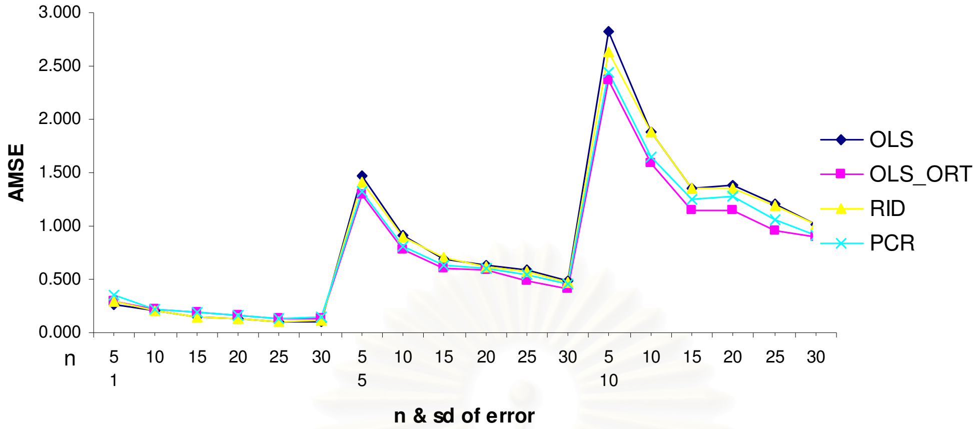
สำหรับกรณีระดับความสัมพันธ์อยู่ในช่วงประมาณ 0.85-0.99

พบว่า PCR จะเป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมา คือ OLS-ORT, วิธี RID, วิธี OLS เพราะ วิธี PCR นั้นจะมีการตัดองค์ประกอบบางส่วนออกไป ส่งผลให้ลดความแปรปรวนให้ต่ำลง

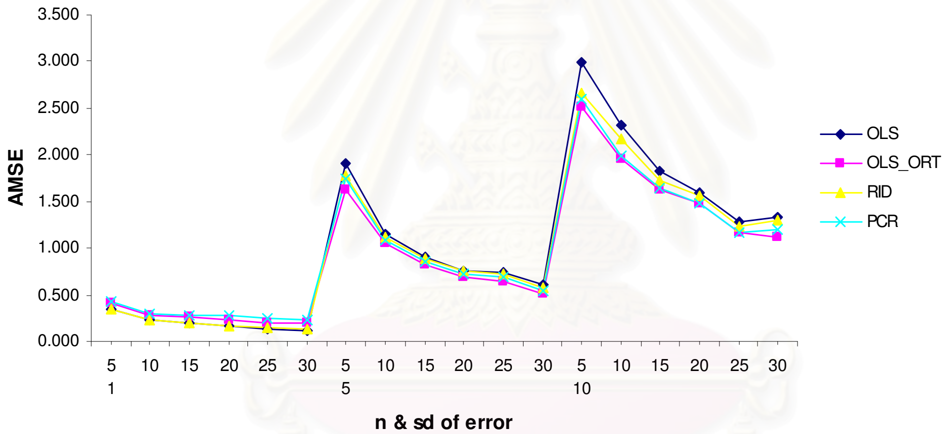
เมื่อพิจารณาค่า Diff พบว่า ถ้าปัญหาพหุสัมพันธ์ไม่สูงมาก จนทำให้วิธี PCR มีการตัดองค์ประกอบบางส่วนออกแล้ว วิธี OLS-ORT กับ PCR จะมีประมาณค่า AMSE ได้ใกล้เคียงกัน โดยวิธี OLS-ORT จะดีกว่าวิธี PCR ไม่มากนัก แต่เมื่อปัญหาพหุสัมพันธ์มีระดับสูง ย่อมทำให้การจัดองค์ประกอบหลักใหม่ ด้วยวิธี PCR สามารถตัดองค์ประกอบที่ไม่จำเป็นออกไปได้ ซึ่งเมื่อมีการตัดองค์ประกอบที่ไม่จำเป็นออกไป ย่อมทำให้ลดความแปรปรวนลงอย่างชัดเจน ทำให้ประสิทธิภาพของ วิธี PCR ดีกว่า OLS-ORT อย่างชัดเจนมาก และถ้าเทียบกับวิธีอื่นด้วย จะพบว่ามีประสิทธิภาพโดดเด่นมาก

สำหรับแนวโน้มของค่า AMSE พบว่า ค่า AMSE ของทุกวิธีแปรผันตามระดับความสัมพันธ์และค่าความแปรปรวนของค่าคลาดเคลื่อนของตัวแบบ เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ด้วยเหตุผลเดิมที่เมื่อปัญหาพหุสัมพันธ์มีระดับสูง ทำให้องค์ประกอบที่ไม่จำเป็นถูกออกไปจากสมการความถดถอย และให้ส่งผลให้ความแปรปรวนลงต่ำลงมาก

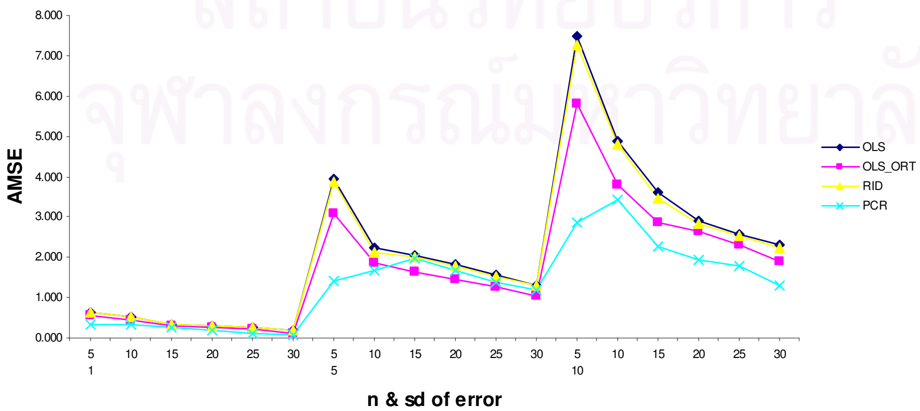
4.2.1 AMSE (Case Correlation in 0.15-0.3,p =3)



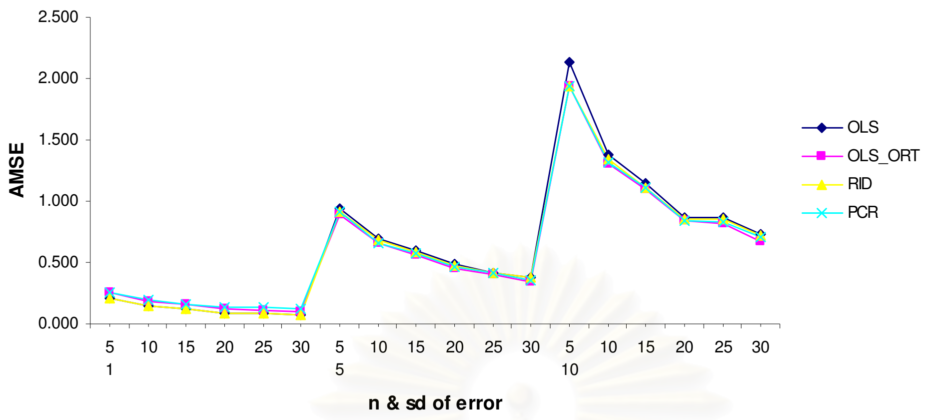
4.2.2 AMSE (Case Correlation in 0.5-0.65,p =3)



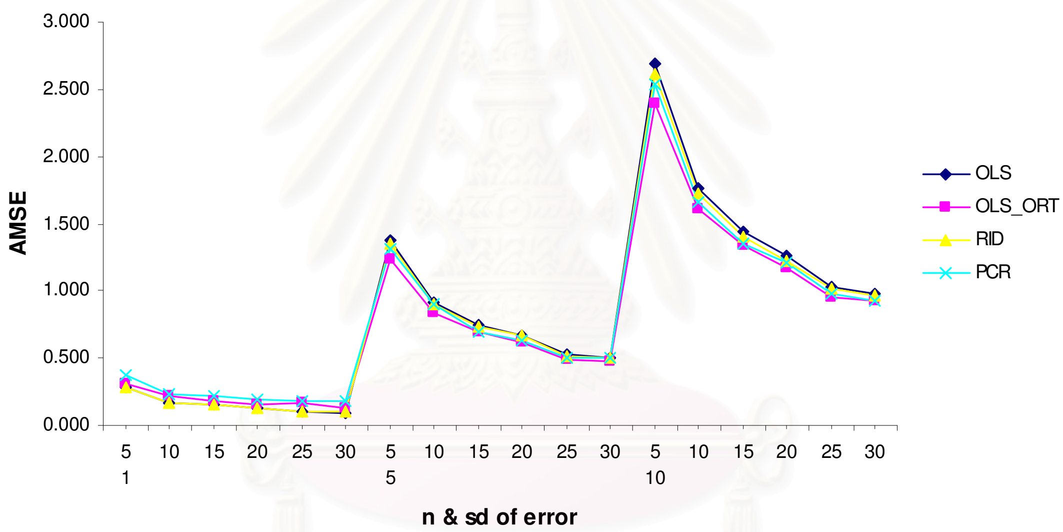
4.2.3 AMSE (Case Correlation in 0.85-0.99,p =3)



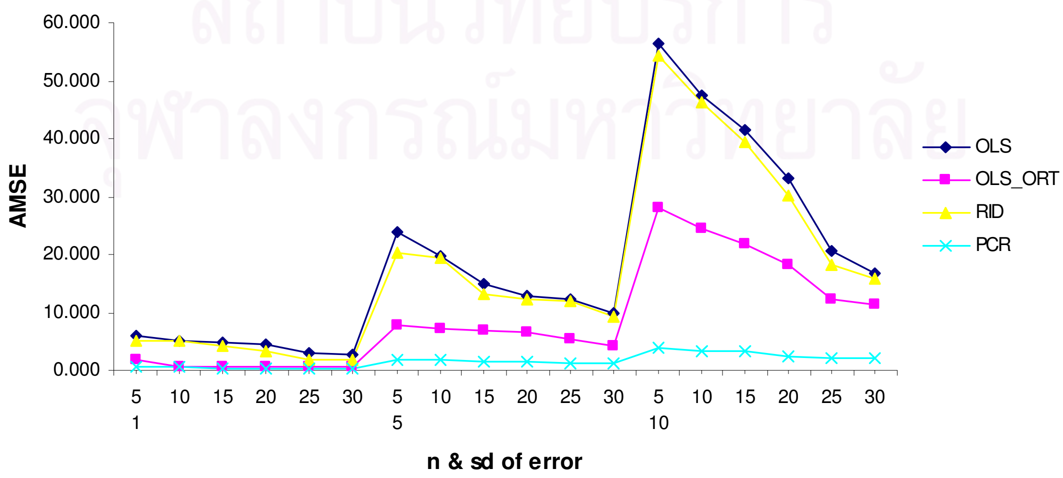
4.2.4 AMSE (Case Correlation in 0.15-0.3,p =6)



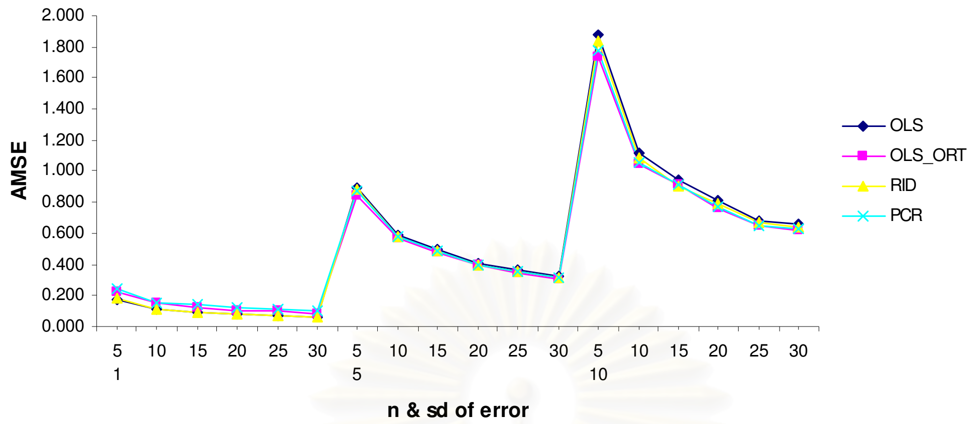
4.2.5 AMSE (Case Correlation in 0.5-0.65,p =6)



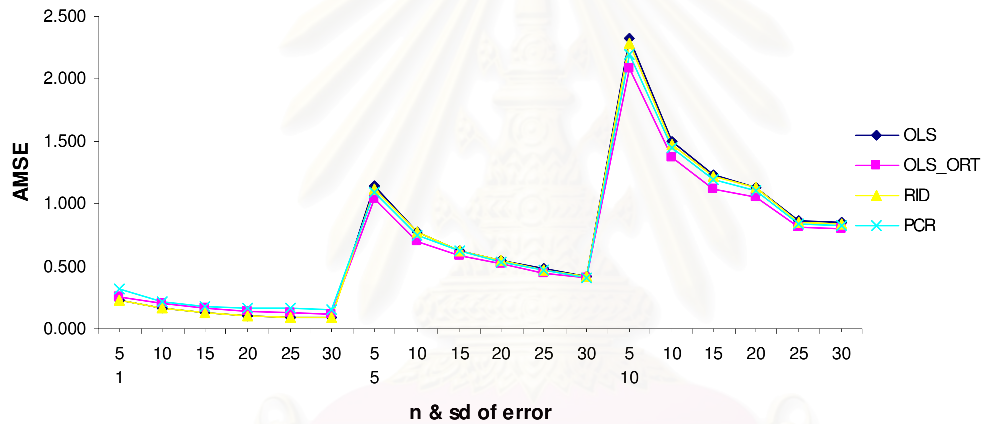
4.2.6 AMSE (Case Correlation in 0.85-0.99,p =6)



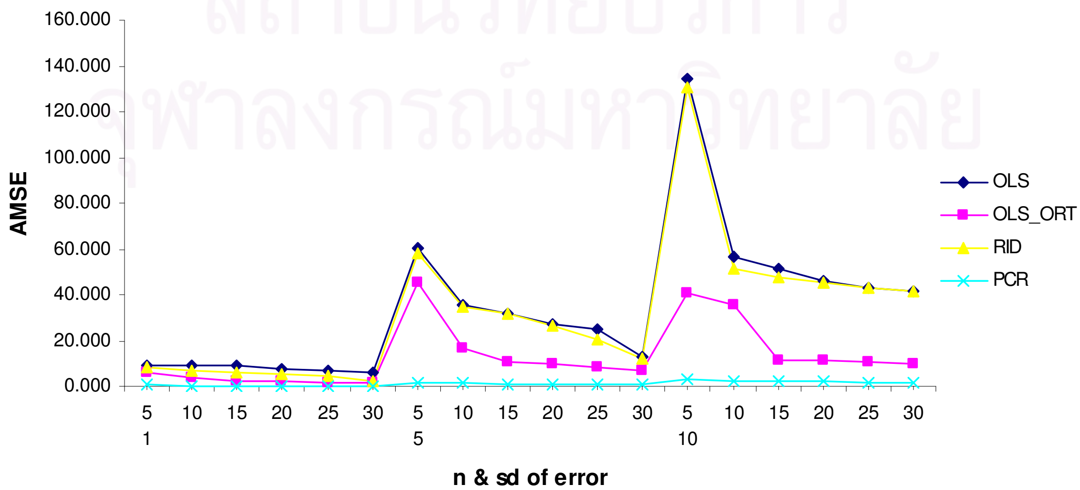
4.2.7 AMSE (Case Correlation in 0.15-0.3,p =9)



4.2.8 AMSE (Case Correlation in 0.5-0.65,p =9)



4.2.9 AMSE (Case Correlation in 0.85-0.99,p =9)



จากรูปที่ 4.2.1-4.2.9 จะเห็นว่าใน p และระดับความสัมพันธ์เดียวกัน พบว่าเมื่อความแปรปรวนของข้อมูลเพิ่มขึ้น ค่า AMSE ในแต่ละวิธีจะสูงขึ้น เนื่องจากความแปรปรวนของความคลาดเคลื่อนส่งผลให้การกระจายของตัวแปรตามเพิ่มขึ้น ทำให้ค่าประมาณของพารามิเตอร์ที่ได้สูงขึ้น และเมื่อพิจารณาในความแปรปรวนเดียวกันพบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้น ค่า AMSE จะลดลง ในสถานการณ์ของความแปรปรวนมาก ก็จะมีสัดส่วนการลดลงของค่า AMSE ในขนาดตัวอย่างที่เพิ่มขึ้น มากกว่า ในสถานการณ์ของความแปรปรวนน้อย

จากกราฟ พบว่ายิ่งตัวแปรอิสระมีความสัมพันธ์กันมากขึ้น ค่า AMSE ก็จะมากขึ้น และในระดับความสัมพันธ์ของตัวแปรอิสระ ในช่วง 0.85-0.99 นั้น ความแตกต่างของค่า AMSE ที่ได้จากแต่ละวิธี จะแตกต่างกันมากยิ่งขึ้น โดยเฉพาะวิธี PCR ในกรณี p เท่ากับ 6 และ 9 จะเห็นได้ชัดว่า ประสิทธิภาพของ PCR จะดีกว่าตัวประมาณตัวอื่นอย่างมาก

ในกรณี กราฟที่ 4.2.3 จะเห็นว่า สำหรับความแปรปรวนของข้อมูลที่ เท่ากับ 25 และ 100 พบว่าเมื่อระดับความสัมพันธ์ของตัวแปรอิสระ ในช่วง 0.85-0.99 ส่งผลให้เกิดปัญหาพหุสัมพันธ์สูง ซึ่ง PCR จะเป็นวิธีที่เหมาะสมในสถานการณ์ลักษณะนี้ แต่แนวโน้มของ ค่า AMSE ของ PCR ไม่ได้ต่ำสุดในบางสถานการณ์ที่เป็นส่วนน้อยมาก เพราะ เนื่องจากวิธี PCR นั้นมีหลักการอยู่ที่ การจัดองค์ประกอบใหม่ของข้อมูล X และเลือกเอาเฉพาะองค์ประกอบที่สามารถอธิบายตัวสมการได้มากไว้ ซึ่งถ้ามีการตัดองค์ประกอบออกไปแล้ว ก็จะลดความแปรปรวนลงได้มาก ดังนั้นจึงมีความเป็นไปได้ที่ในบางสถานการณ์ ไม่สามารถตัดองค์ประกอบออกไปได้เลย จึงทำให้วิธี OLS-ORT มีประสิทธิภาพดีกว่า วิธี PCR ซึ่งสามารถพิจารณาได้จากข้อมูลในตารางที่ 4.2.3

นอกจากนี้มีความเป็นไปได้ที่ วิธี PCR ของ ขนาดตัวอย่างที่มีค่าน้อย จะมีค่า AMSE ต่ำกว่า ขนาดตัวอย่างที่มีค่ามาก ในปัจจัยอื่นๆ เดียวกัน เพราะ ในกรณีนั้น ขนาดตัวอย่างน้อย และตัวแปรอิสระน้อย ข้อมูลอาจจะไม่หลากหลายมากนัก และ สามารถรวมเป็นองค์ประกอบหลักอันเดียวก็ได้ ซึ่งสามารถพิจารณาได้จากข้อมูลในตารางที่ 4.2.1 และ 4.2.2

การเลือกองค์ประกอบมีได้หลายเงื่อนไข โดยในงานวิจัยนี้ได้เลือก Kaiser's Criterion เป็นเกณฑ์นั้นคือ หากพบว่า λ ตัวใดมีค่าต่ำกว่า โดยจำนวนองค์ประกอบหลักที่เราจะตัดออกจะมีค่าเท่ากับจำนวนของค่าเฉพาะที่มีค่าน้อยกว่าก็ให้ตัดทิ้งไปซึ่งมีผลเหมือนกำหนดให้ λ ตัวนั้นมีค่าเป็น 0 จำนวนองค์ประกอบที่เหลือจะนำมาสร้างตัวแบบถดถอยด้วยวิธีกำลังสองน้อยสุด

จากข้อมูล พบว่า ถ้าปัญหาพหุสัมพันธ์เล็กน้อย จะไม่ค่อยส่งผลกระทบต่อการประมาณค่า มากเท่าใด และจากกราฟรูปที่ 4.2.10-4.2.12 จะเห็นได้ชัดว่า ในระดับความสัมพันธ์ของตัวแปรอิสระ ช่วง 0.86-0.99 ที่ทำให้ปัญหาพหุสัมพันธ์สูงมาก จะเป็นกรณีชัดเจนกรณีเดียวที่ส่งผลให้ ค่าประมาณที่ได้จากวิธี OLS นั้นไม่มีประสิทธิภาพ

ตารางที่ 4.2.1 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram−Schmidt orthogonalization วิธีริดจ์รีเกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 3 และ X มีระดับความสัมพันธ์กันในช่วง 0.66-0.85

ขนาดตัวอย่าง		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน												
		1				5					10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	จำนวนองค์ประกอบใหม่	OLS	OLS_ORT	RID	PCR
5p	AMSE	0.3584	0.4304	0.3624	0.5129	1.9523	1.6677	1.7963	1.6361	2.98	4.0984	3.4689	3.9595	3.6349
	(SD)	0.0221	0.0382	0.0431	0.0817	0.1097	0.1688	0.1602	0.1332		0.7846	0.3809	0.8391	0.9078
	%DIFF	0.0000	20.0823	1.1230	43.1013	19.3252	1.9330	9.7948	0.0000		18.1494	0.0000	14.1430	4.7854
10p	AMSE	0.2917	0.3900	0.2975	0.3969	1.3730	1.1788	1.3289	1.2326	3.00	2.6527	2.2047	2.5324	2.2846
	(SD)	0.0255	0.0123	0.0268	0.0329	0.1012	0.0956	0.1210	0.1395		0.4284	0.2588	0.4771	0.4264
	%DIFF	0.0000	33.7019	2.0057	36.0675	16.4677	0.0000	12.7267	4.5575		20.3202	0.0000	14.8648	3.6218
15p	AMSE	0.2004	0.2963	0.2011	0.3138	1.0252	0.9048	1.0046	0.9957	3.00	2.0774	1.8390	2.0298	1.9070
	(SD)	0.0066	0.0270	0.0091	0.0274	0.0937	0.0561	0.0948	0.0696		0.4896	0.3277	0.4269	0.2867
	%DIFF	0.0000	47.8662	0.3868	56.6134	13.3009	0.0000	11.0270	10.0406		12.9639	0.0000	10.3782	3.7018
20p	AMSE	0.1864	0.2564	0.1881	0.3093	0.8791	0.7975	0.8562	0.8031	3.00	1.4957	1.2958	1.4164	1.3309
	(SD)	0.0260	0.0210	0.0241	0.0622	0.1452	0.0951	0.1421	0.1533		0.2534	0.1491	0.2550	0.1819
	%DIFF	0.0000	37.5352	0.8985	65.8844	10.2257	0.0000	7.3636	0.6991		15.4265	0.0000	9.3049	2.7048
25p	AMSE	0.1668	0.2055	0.1682	0.2930	0.7487	0.6529	0.7306	0.6920	3.00	1.5609	1.2420	1.5371	1.2965
	(SD)	0.0261	0.0210	0.0235	0.0150	0.0673	0.0636	0.0642	0.0785		0.2263	0.2159	0.2168	0.1750
	%DIFF	0.0000	23.1865	0.8243	75.6595	14.6774	0.0000	11.8974	5.9889		25.6774	0.0000	23.7671	4.3903
30p	AMSE	0.1539	0.2027	0.1543	0.2671	0.6972	0.6406	0.6923	0.6585	3.00	1.4750	1.1603	1.4484	1.2357
	(SD)	0.0122	0.0381	0.0129	0.0203	0.1511	0.1234	0.1382	0.1178		0.1970	0.1720	0.1886	0.2309
	%DIFF	0.0000	31.7465	0.2437	73.5500	8.8312	0.0000	8.0585	2.7824		27.1223	0.0000	24.8319	6.4962

87

ตารางที่ 4.2.2 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์เกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 3 และ X มีระดับความสัมพันธ์กันในช่วง 0.86-0.99

ขนาดตัวอย่าง		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน												
		1				5					10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	จำนวนองค์ประกอบใหม่	OLS	OLS_ORT	RID	PCR
5p	AMSE	0.6227	0.5566	0.6348	0.3367	3.9307	3.1037	3.8640	1.4044	2.00	7.4643	5.7962	7.2651	2.8493
	(SD)	0.0508	0.0692	0.0623	0.0808	0.4190	0.1956	0.4425	0.1131		1.3568	1.1049	1.2775	0.3819
	%DIFF	84.9547	65.3349	88.5638	0.0000	179.8914	121.0022	175.1366	0.0000		161.9702	103.4281	154.9798	0.0000
10p	AMSE	0.5089	0.4633	0.5045	0.3353	2.2206	1.8503	2.1353	1.6730	2.85	4.8590	3.7838	4.7914	3.4084
	(SD)	0.0727	0.0492	0.0742	0.0925	0.4351	0.3559	0.4176	0.0514		1.2907	0.9044	1.2594	0.9786
	%DIFF	51.7632	38.1645	50.4436	0.0000	32.7301	10.5947	27.6345	0.0000		42.5588	11.0125	40.5770	0.0000
15p	AMSE	0.3285	0.3048	0.3198	0.2516	2.0286	1.6281	2.0150	1.9697	3.00	3.6195	2.8647	3.4668	2.2561
	(SD)	0.0357	0.0255	0.0345	0.0698	0.2367	0.1412	0.2199	0.2450		0.2816	0.2512	0.2429	0.1992
	%DIFF	30.5545	21.1347	27.1165	0.0000	24.5962	0.0000	23.7654	20.9784		60.4277	26.9743	53.6628	0.0000
20p	AMSE	0.3083	0.2720	0.2904	0.2026	1.8108	1.4606	1.7824	1.6900	3.00	2.8864	2.6498	2.8162	1.9498
	(SD)	0.0709	0.0269	0.0715	0.0413	0.3138	0.2413	0.2879	0.2610		0.8623	0.4720	0.8680	0.5414
	%DIFF	52.1718	34.2300	43.3490	0.0000	23.9730	0.0000	22.0303	15.7059		48.0382	35.9020	44.4390	0.0000
25p	AMSE	0.2626	0.2121	0.2542	0.1195	1.5503	1.2489	1.5317	1.3739	3.00	2.5646	2.2972	2.5282	1.7718
	(SD)	0.0332	0.0529	0.0343	0.0304	0.2594	0.1805	0.2578	0.2085		0.4440	0.3174	0.4112	0.4898
	%DIFF	119.7239	77.4315	112.6543	0.0000	24.1267	0.0000	22.6375	10.0066		44.7481	29.6557	42.6951	0.0000
30p	AMSE	0.2013	0.1154	0.2000	0.0815	1.2969	1.0551	1.2906	1.1943	3.00	2.3000	1.9010	2.2505	1.3010
	(SD)	0.0593	0.0330	0.0554	0.0245	0.1100	0.0629	0.1076	0.1234		0.4156	0.3195	0.4116	0.3681
	%DIFF	146.9181	41.5517	145.3542	0.0000	22.9125	0.0000	22.3225	13.1907		76.7852	46.1184	72.9785	0.0000

88

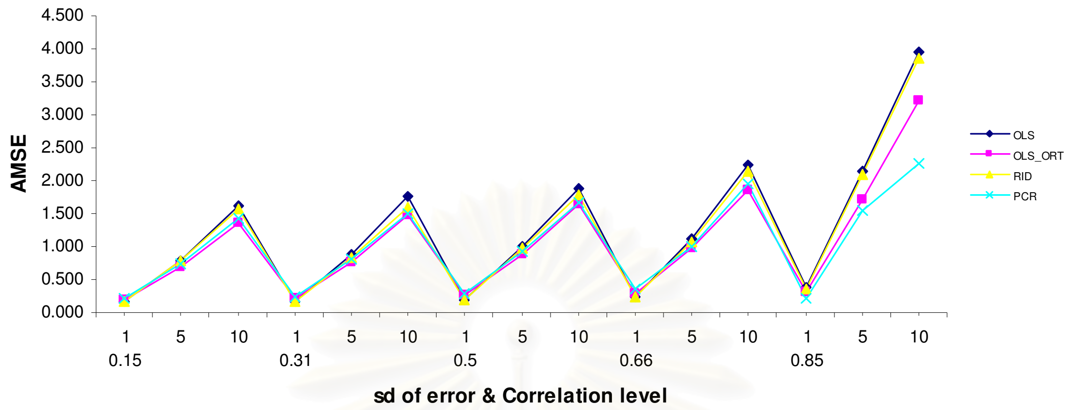
ตารางที่ 4.2.2 การเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุด หลังจากแก้ไข X ด้วย Gram–Schmidt orthogonalization วิธีริดจ์เกรสชั่นสามัญ และวิธีการถดถอยองค์ประกอบหลัก ในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 3 และ X มีระดับความสัมพันธ์กันในช่วง 0.86-0.99

ขนาดตัวอย่าง		ค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน												
		1				5					10			
		OLS	OLS_ORT	RID	PCR	OLS	OLS_ORT	RID	PCR	จำนวนองค์ประกอบใหม่	OLS	OLS_ORT	RID	PCR
5p	AMSE	0.6227	0.5566	0.6348	0.3367	3.9307	3.1037	3.8640	1.4044	2.00	7.4643	5.7962	7.2651	2.8493
	(SD)	0.0508	0.0692	0.0623	0.0808	0.4190	0.1956	0.4425	0.1131		1.3568	1.1049	1.2775	0.3819
	%DIFF	84.9547	65.3349	88.5638	0.0000	179.8914	121.0022	175.1366	0.0000		161.9702	103.4281	154.9798	0.0000
10p	AMSE	0.5089	0.4633	0.5045	0.3353	2.2206	1.8503	2.1353	1.6730	2.85	4.8590	3.7838	4.7914	3.4084
	(SD)	0.0727	0.0492	0.0742	0.0925	0.4351	0.3559	0.4176	0.0514		1.2907	0.9044	1.2594	0.9786
	%DIFF	51.7632	38.1645	50.4436	0.0000	32.7301	10.5947	27.6345	0.0000		42.5588	11.0125	40.5770	0.0000
15p	AMSE	0.3285	0.3048	0.3198	0.2516	2.0286	1.6281	2.0150	1.9697	3.00	3.6195	2.8647	3.4668	2.2561
	(SD)	0.0357	0.0255	0.0345	0.0698	0.2367	0.1412	0.2199	0.2450		0.2816	0.2512	0.2429	0.1992
	%DIFF	30.5545	21.1347	27.1165	0.0000	24.5962	0.0000	23.7654	20.9784		60.4277	26.9743	53.6628	0.0000
20p	AMSE	0.3083	0.2720	0.2904	0.2026	1.8108	1.4606	1.7824	1.6900	3.00	2.8864	2.6498	2.8162	1.9498
	(SD)	0.0709	0.0269	0.0715	0.0413	0.3138	0.2413	0.2879	0.2610		0.8623	0.4720	0.8680	0.5414
	%DIFF	52.1718	34.2300	43.3490	0.0000	23.9730	0.0000	22.0303	15.7059		48.0382	35.9020	44.4390	0.0000
25p	AMSE	0.2626	0.2121	0.2542	0.1195	1.5503	1.2489	1.5317	1.3739	3.00	2.5646	2.2972	2.5282	1.7718
	(SD)	0.0332	0.0529	0.0343	0.0304	0.2594	0.1805	0.2578	0.2085		0.4440	0.3174	0.4112	0.4898
	%DIFF	119.7239	77.4315	112.6543	0.0000	24.1267	0.0000	22.6375	10.0066		44.7481	29.6557	42.6951	0.0000
30p	AMSE	0.2013	0.1154	0.2000	0.0815	1.2969	1.0551	1.2906	1.1943	3.00	2.3000	1.9010	2.2505	1.3010
	(SD)	0.0593	0.0330	0.0554	0.0245	0.1100	0.0629	0.1076	0.1234		0.4156	0.3195	0.4116	0.3681
	%DIFF	146.9181	41.5517	145.3542	0.0000	22.9125	0.0000	22.3225	13.1907		76.7852	46.1184	72.9785	0.0000

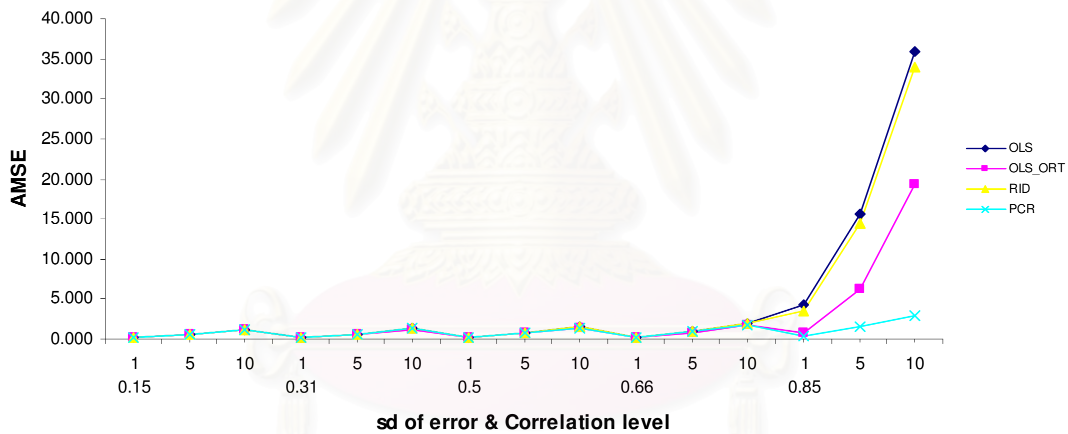
89

รูปที่ 4.2.10-4.2.12 แสดงแนวโน้มค่า AMSE ของแต่ละวิธีโดยแยกตามค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนของตัวแบบและระดับสัมพันธ์ของตัวแปรอิสระ เมื่อ p เท่ากับ 3,6,9 ตามลำดับ

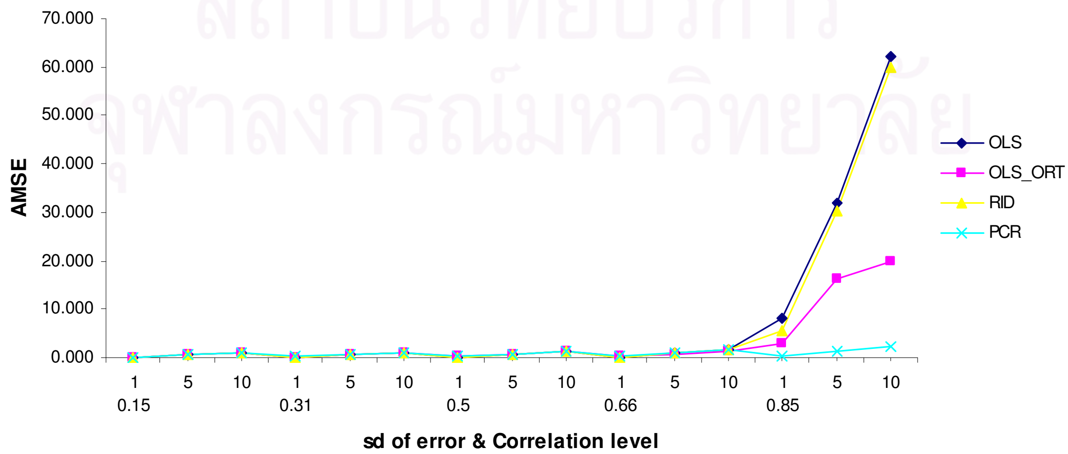
4.2.10 AMSE (p =3)



4.2.11 AMSE (p =6)

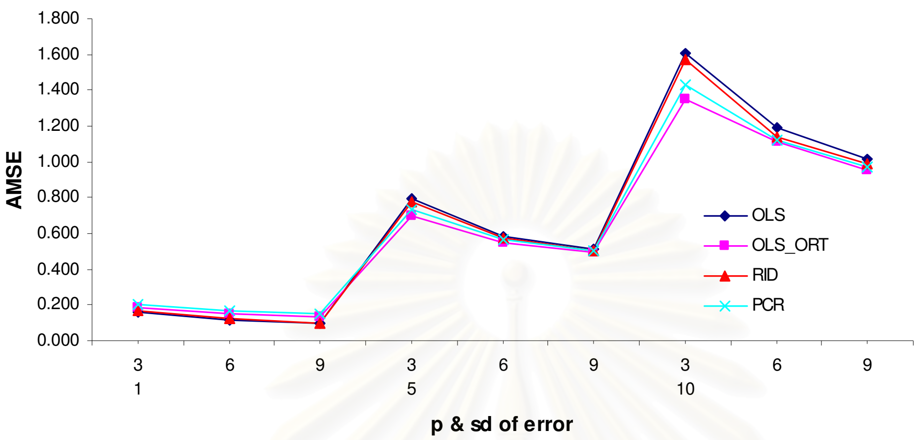


4.2.12 AMSE (p =9)

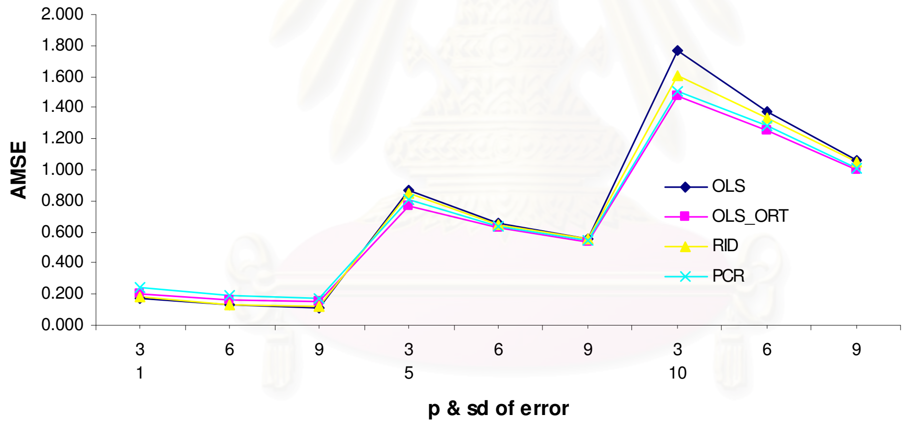


รูปที่ 4.2.13-4.2.17 แสดงแนวโน้มค่า AMSE ของแต่ละวิธีโดยแยกตามค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนของตัวแบบและจำนวนตัวแปรอิสระ ตามระดับความสัมพันธ์ระหว่างตัวแปรอิสระตามลำดับ

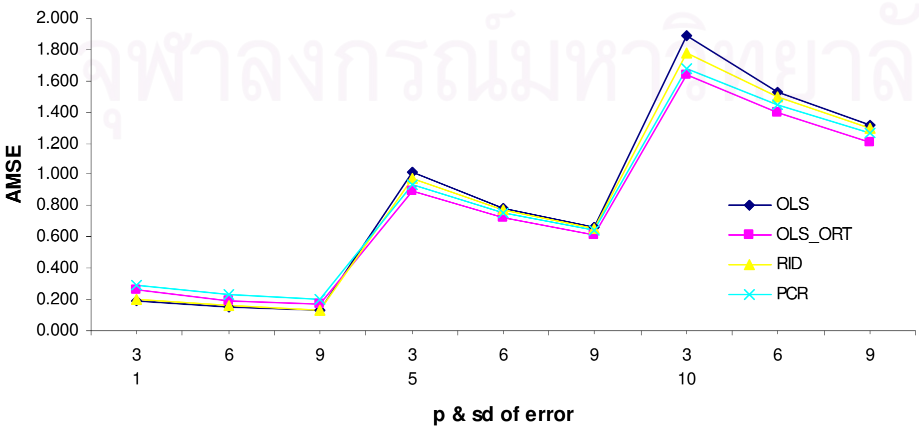
4.2.13 AMSE (Case in correlation : 0.15-0.3)

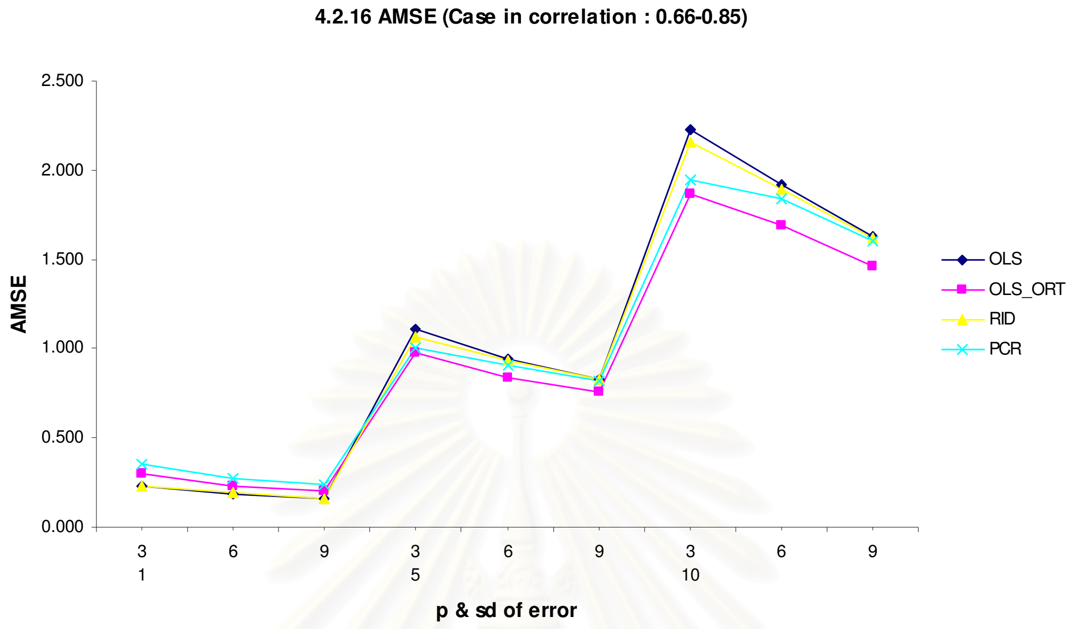


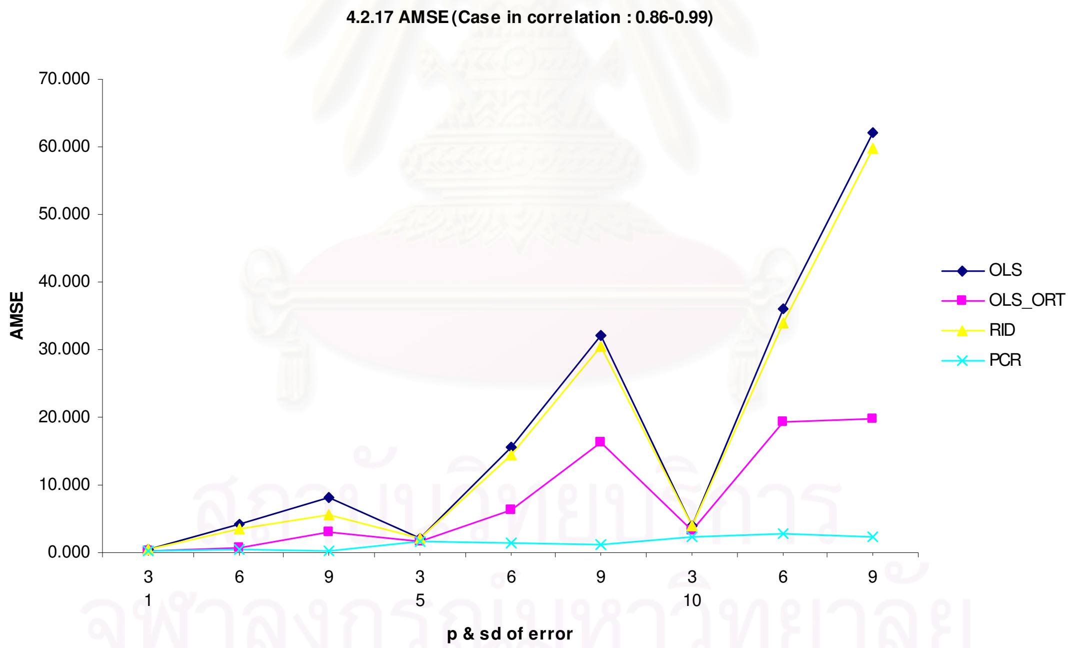
4.2.14 AMSE (Case in correlation : 0.31-0.5)



4.2.15 AMSE (Case in correlation : 0.51-0.65)







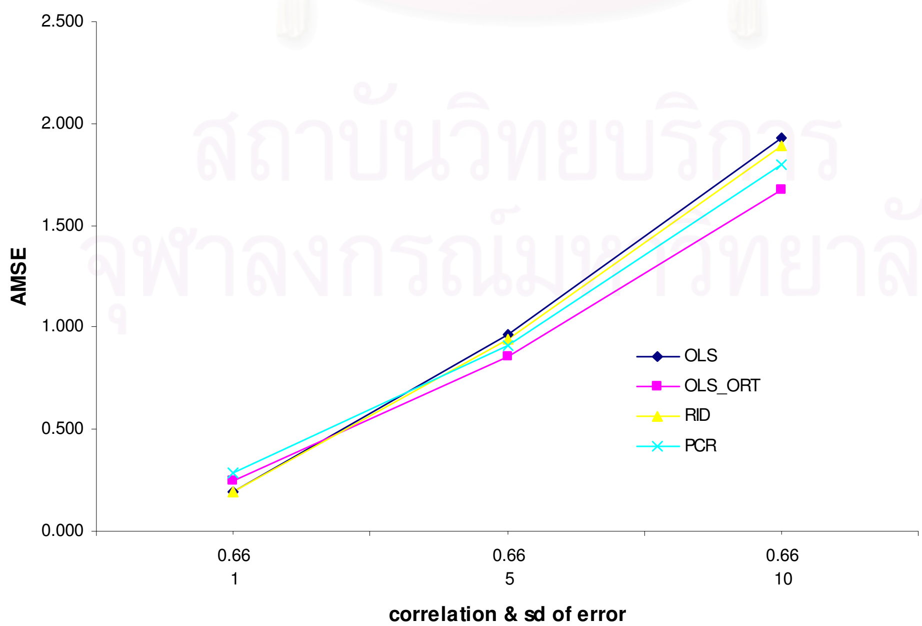
จากกราฟที่ 4.2.1-4.2.9 พบว่า เมื่อมีปัจจัยที่กำหนดในสถานการณ์ที่ส่งผลต่อค่า AMSE เดียวกัน ถ้าขนาดตัวอย่างเพิ่มขึ้น จะมีค่า AMSE ลดลง เช่นเดียวกับ กรณีของ p ซึ่งสามารถพิจารณาได้ จากกราฟที่ 4.2.13-4.2.17 พบว่า ถ้า ถ้าระดับความสัมพันธ์ของตัวแปรอิสระต่ำ เมื่อ p เพิ่มขึ้น ค่า AMSE จะลดลง เพราะ มีปัจจัยที่ไม่มีปัญหาพหุสัมพันธ์กัน มาช่วยอธิบายสมการ

ถดถอยได้ดีขึ้น แต่ในกรณีที่ ระดับความสัมพันธ์ของตัวแปรอิสระสูง จนทำให้เกิดปัญหาพหุสัมพันธ์มาก นั้น ยิ่ง p เพิ่มขึ้น ยิ่งทำให้ ค่า AMSE เพิ่มขึ้น เพราะไม่สามารถบอกได้ว่าการที่มีตัวแปรอิสระหลายตัวจะสามารถอธิบายได้ดีขึ้น ด้วยสาเหตุที่มีปัญหาพหุสัมพันธ์กันสูง

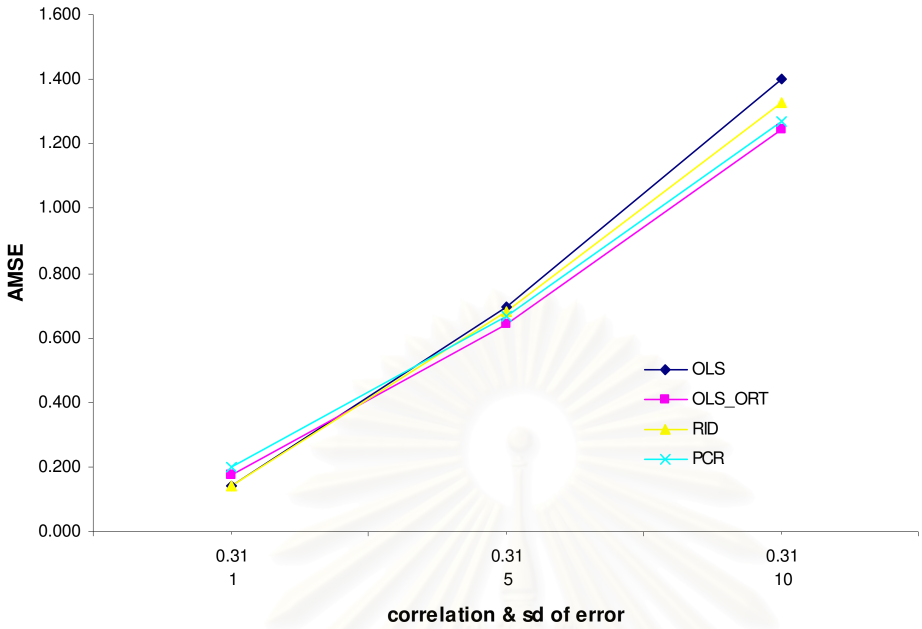
รูปที่ 4.2.18-4.2.22 แสดงแนวโน้มค่าเฉลี่ย AMSEในทุก n และ p ของแต่ละวิธีโดยแยกตามค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนของตัวแบบ แยกตามระดับความสัมพันธ์ของตัวแปรอิสระ

4.2.18 AMSE (Case in correlation : 0.86-0.99)

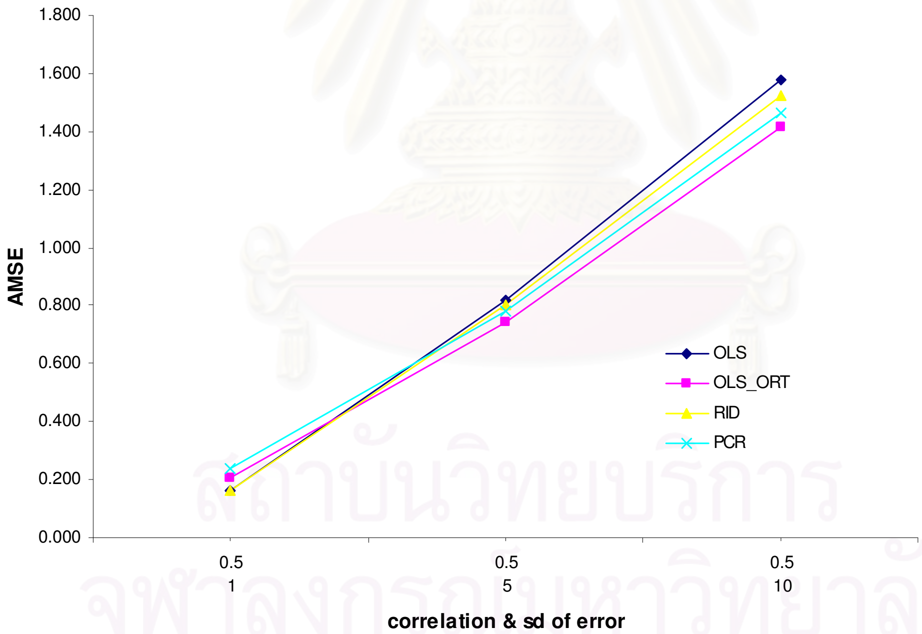
4.2.19 AMSE (Case in correlation : 0.66-0.80)

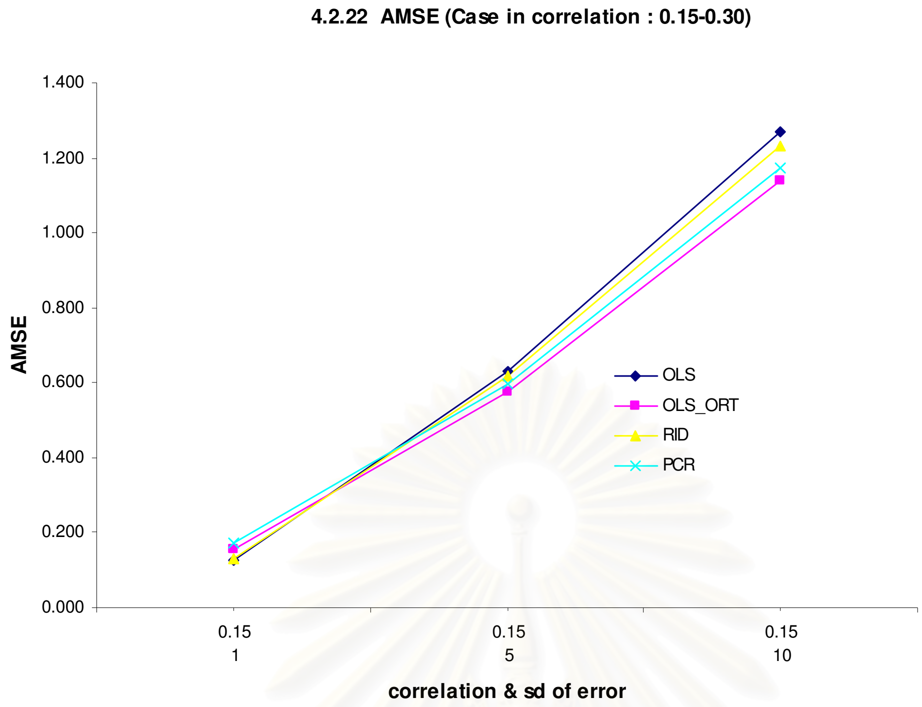


4.2.20 AMSE (Case in correlation : 0.31-0.50)



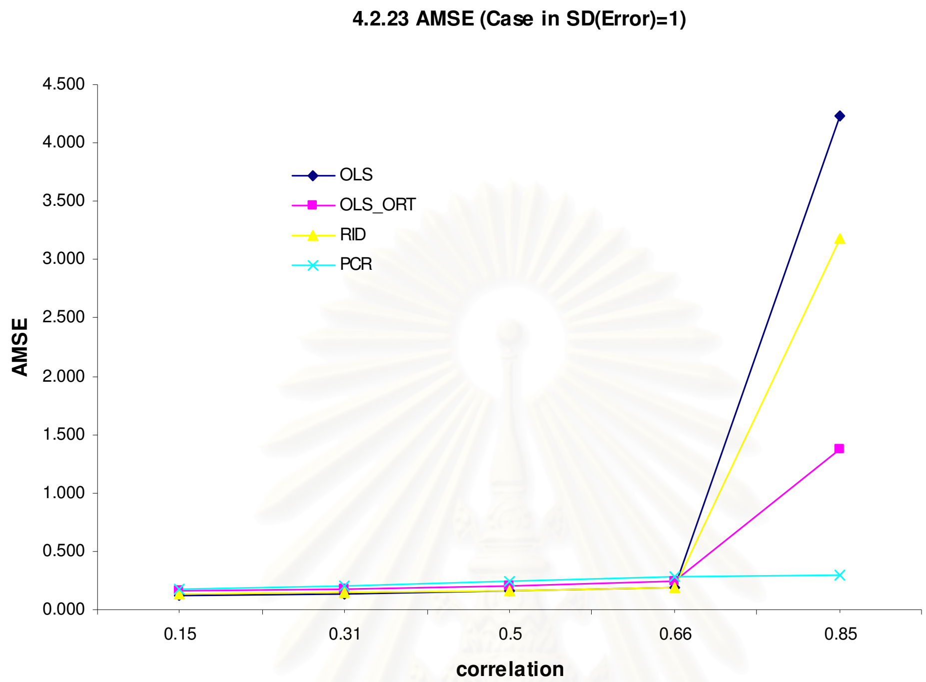
4.2.21 AMSE (Case in correlation : 0.51-0.65)

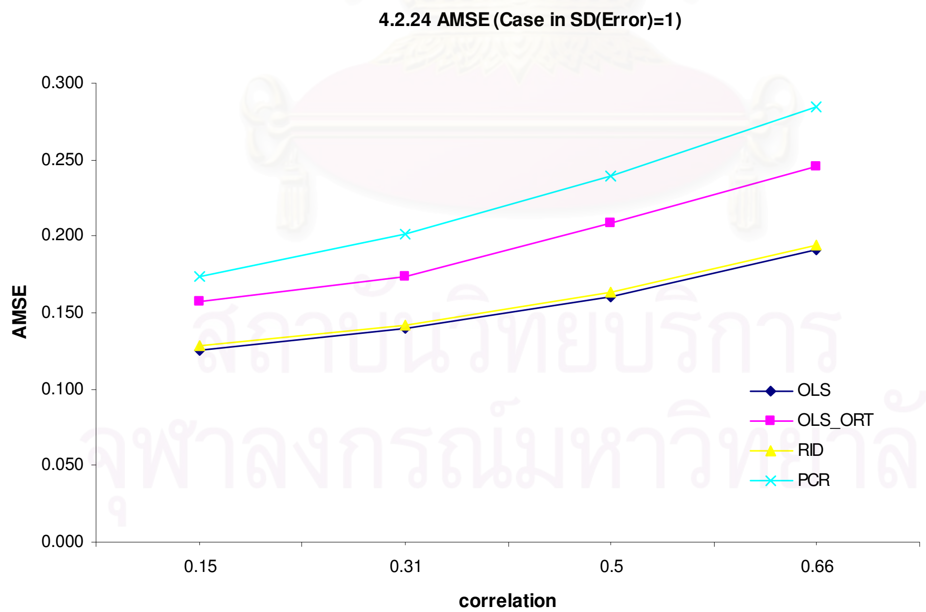




จากกราฟที่ 4.2.18-4.2.22 พบว่ากรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนของตัวแบบ เท่ากับ 5,10 สำหรับความสัมพันธ์อยู่ในช่วงประมาณ 0.15-0.85 นั้น วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR, วิธี RID และวิธี OLS ตามลำดับ เนื่องจาก OLS-ORT เป็นวิธีที่ยังคงรักษาประสิทธิภาพที่ดีของวิธี OLS เดิมไว้ และไปแก้ไขปัญหาพหุสัมพันธ์ที่ข้อมูลเลย คือ การไปแปลงค่า X ที่มีปัญหาพหุสัมพันธ์ โดยทำให้ แต่ละ X ไม่ได้ตั้งฉากกัน ให้เป็น X ที่ตั้งฉากกัน ด้วยวิธี orthogonalization ของ Gram-Schmidt จึงทำให้ OLS-ORT เป็นวิธีที่ดีที่สุดในสถานการณ์นี้ สำหรับกรณีระดับความสัมพันธ์อยู่ในช่วงประมาณ 0.85-0.99 พบว่า PCR จะเป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมา คือ OLS-ORT, วิธี RID, วิธี OLS เพราะ วิธี PCR นั้นจะมีการตัดองค์ประกอบบางส่วนออกไป ส่งผลให้ลดความแปรปรวนให้ต่ำลง นอกจากนี้จะเห็นว่ากรณี ความแปรปรวนของข้อมูลเท่ากับ 1 จะให้ผลที่แตกต่างไปดังกราฟที่ 4.2.23-4.2.24

รูปที่ 4.2.23-4.2.24 แสดงแนวโน้มค่าเฉลี่ย AMSEในทุก n และ p ของแต่ละวิธี ค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนของตัวแบบเท่ากับ 1





กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนของตัวแบบ เท่ากับ 1 และระดับความสัมพันธ์ในระดับที่ไม่สูงมาก ช่วงประมาณ 0.15-0.80 พบว่า วิธี OLS จะดีสุด รองลงมาคือ วิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ สำหรับกรณีระดับความสัมพันธ์อยู่ในช่วงประมาณ 0.85-0.99 พบว่า PCR จะเป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมา คือ OLS-ORT, วิธี RID, วิธี OLS เพราะ วิธี PCR นั้นจะมีการตัดองค์ประกอบบางส่วนออกไป ส่งผลให้ลดความแปรปรวนให้ต่ำลง

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

การวิจัยครั้งนี้วัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระซึ่งในที่นี้จะมี 4 วิธีคือ

- วิธีกำลังสองน้อยที่สุด (Ordinary Least Square Method)
- วิธีกำลังสองน้อยที่สุด หลังจากแก้ไขด้วยวิธีทำให้เป็นแนวตั้งเชิงตั้งฉาก (Orthogonal Least Square Method)
- วิธีริดจ์รีเกรสชั่น (Ridge Regression Method)
- วิธีความถดถอยองค์ประกอบหลัก (principle Component Regression Method)

โดยมีสมมติฐานการวิจัยว่า เมื่อตัวแปรอิสระมีพหุสัมพันธ์กันมากวิธีการประมาณค่าพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุด (Ordinary Least Square Method (OLS)) หลังจากที่ได้ทำการแก้ไขพหุสัมพันธ์ระหว่างตัวแปรอิสระด้วยวิธีแนวตั้งเชิงตั้งฉาก แล้วน่าจะให้ค่าประมาณที่มีความถูกต้องเชื่อถือได้มากกว่าวิธีอื่น ๆ ไม่ว่าขนาดตัวอย่างหรือความแปรปรวนมีค่ามากหรือน้อยก็ตาม โดย สถานการณ์ที่ทำการศึกษามีดังนี้

- ตัวแปรอิสระ (p) ที่ใช้ในการวิจัยมี 3 จำนวน คือ 3, 6 และ 9 ตัวแปร
- ขนาดตัวอย่าง (n) ที่ศึกษาคือ 5p , 10p , 15p , 20p , 25p และ 30p
- ความคลาดเคลื่อนมีการแจกแจงแบบปกติ (Normal Distribution)

  เมื่อ $\mu = 0$, $\sigma = 1, 5$ และ 10
- ระดับพหุสัมพันธ์ของตัวแปรอิสระแบ่งเป็น 3 ระดับ คือ

ระดับต่ำ	ค่า ρ มีค่าอยู่ในช่วง 0.15 ถึง 0.30
ระดับปานกลาง	ค่า ρ มีค่าอยู่ในช่วง 0.31 ถึง 0.50
	ค่า ρ มีค่าอยู่ในช่วง 0.50 ถึง 0.65
ระดับสูง	ค่า ρ มีค่าอยู่ในช่วง 0.61 ถึง 0.85
	ค่า ρ มีค่าอยู่ในช่วง 0.85 ถึง 0.99

5.1 สรุปผลการวิจัย

5.1.1 ปัจจัยที่มีผลต่อค่าเฉลี่ยรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของแต่ละวิธี

1. ขนาดตัวอย่าง

ถ้าขนาดตัวอย่างเพิ่มขึ้น จะมีค่า AMSE ลดลงเนื่องจาก ข้อมูลมีเพิ่มมากขึ้น ก็จะสามารถนำมาอธิบายสมการความถดถอยได้ดีขึ้น

2. จำนวนตัวแปรอิสระ

ถ้าระดับความสัมพันธ์ของตัวแปรอิสระต่ำ เมื่อ p เพิ่มขึ้น ค่า AMSE จะลดลง เพราะ มีปัจจัยที่ไม่มีปัญหาพหุสัมพันธ์กัน มาช่วยอธิบายสมการถดถอยได้ดีขึ้น แต่ในกรณีที่ ระดับความสัมพันธ์ของตัวแปรอิสระสูง จนทำให้เกิดปัญหาพหุสัมพันธ์มาก นั้น ยิ่ง p เพิ่มขึ้น ยิ่งทำให้ค่า AMSE เพิ่มขึ้น เพราะไม่สามารถบอกได้ว่าการที่มีตัวแปรอิสระหลายตัวจะสามารถอธิบายได้ดีขึ้น ด้วยสาเหตุที่มีปัญหาพหุสัมพันธ์กันสูง

3. ค่าความแปรปรวนของค่าคลาดเคลื่อนของตัวแบบ

เมื่อข้อมูลมีการกระจายสูง ย่อมทำให้ค่าจริงที่กระจายมาก เมื่อไปหาระยะแตกต่าง กับค่าประมาณ ก็จะได้มาก ทำให้ AMSE มีค่ามากตามไปด้วย

4. ระดับความสัมพันธ์ของตัวแปรอิสระ

เมื่อระดับความสัมพันธ์ของตัวแปรอิสระ สูงขึ้น โอกาสที่ปัญหาพหุสัมพันธ์ที่จะกระทบต่อการประมาณค่า ก็ย่อมมีมากขึ้น ทำให้ค่า AMSE ที่ได้ในแต่ละวิธีมีมากขึ้น เว้นแต่วิธี PCR ในกรณีที่ระดับความสัมพันธ์ของ X อยู่ในช่วงประมาณ 0.85-0.99 ด้วยเหตุผลเดิมที่เมื่อปัญหาพหุสัมพันธ์มีระดับสูง ทำให้องค์ประกอบที่ไม่จำเป็นถูกออกไปจากสมการความถดถอย และให้ส่งผลให้ความแปรปรวนลงต่ำลงมาก

5.1.2 สรุปผลการวิจัยการเปรียบเทียบตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณ

5.1.2.1 กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนของตัวแบบ เท่ากับ 1

สำหรับกรณีระดับความสัมพันธ์อยู่ในช่วงประมาณ 0.15-0.85

กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนของตัวแบบ เท่ากับ 1 และระดับความสัมพันธ์ในระดับที่ไม่สูงมาก ช่วงประมาณ 0.15-0.80 พบว่า วิธี OLS จะดีสุด รองลงมาคือ วิธี RID, วิธี OLS-ORT และวิธี PCR ตามลำดับ เสมอ

เนื่องจาก การที่ข้อมูลมีลักษณะการกระจายที่ไม่มาก และ ระดับความสัมพันธ์ของตัวแปร X นั้นไม่ได้สูงมาก นั้นทำให้มีการเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระบ้าง แต่ไม่สูงมากจนกระทบต่อ การประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณ

ถ้าไม่เกิดปัญหาใด โดยทั่วไปวิธี OLS จะเป็นวิธีการประมาณค่าที่มีประสิทธิภาพมากที่สุด จากสมการ $\underset{\sim OLS}{\hat{\beta}} = (X'X)^{-1}X'\underset{\sim}{y}$ พบว่า ถ้าปัญหาพหุสัมพันธ์ไม่มาก ย่อมไม่ส่งผลต่อการประมาณค่ามากนัก เนื่องจากยังสามารถหาค่า $X'X$ ได้ วิธี OLS จึงยังคงเป็นตัวประมาณที่เหมาะสมมากสุด ในสถานการณ์นี้

สาเหตุที่ทั้ง วิธี RID, OLS-ORT,PCR ไม่ได้มีประสิทธิภาพดีกว่า OLS ทั้งที่เป็นวิธีที่พัฒนาขึ้นมาเพื่อใช้แก้ไขปัญหาพหุสัมพันธ์ เพราะ ทั้ง 3 วิธีนี้เหมาะสำหรับ ข้อมูลที่มีการกระจายในระดับหนึ่งด้วย ยิ่งการกระจายข้อมูลเพิ่มขึ้น ยิ่งมีประสิทธิภาพเพิ่มมากขึ้น ดังนั้นในกรณีนี้ที่ข้อมูลมีการกระจายน้อย แล้วปัญหาพหุสัมพันธ์ไม่มาก จึงทำให้ OLS ยังเป็นวิธีที่มีประสิทธิภาพมากสุด สำหรับทุกจำนวนของตัวแปรอิสระ และขนาดตัวอย่าง

สำหรับกรณีระดับความสัมพันธ์อยู่ในช่วงประมาณ 0.86-0.99

พบว่า PCR จะเป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมา คือ OLS-ORT, วิธี RID, วิธี OLS เพราะ วิธี PCR นั้นจะมีการตัดองค์ประกอบบางส่วนออกไป ส่งผลให้ลดความแปรปรวนให้ต่ำลง

5.1.2.2 กรณีค่าเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนของตัวแบบ เท่ากับ 5,10

สำหรับกรณีระดับความสัมพันธ์อยู่ในช่วงประมาณ 0.15-0.85

ส่วนใหญ่ถึง 99% ของสถานการณ์ พบว่า วิธี OLS-ORT จะดีสุด รองลงมาคือวิธี PCR, วิธี RID และวิธี OLS ตามลำดับ เสมอ เนื่องจาก OLS-ORT เป็นวิธีที่ยังคงรักษาประสิทธิภาพที่ดีของวิธี OLS เดิมไว้ และไปแก้ไขปัญหาพหุสัมพันธ์ที่ข้อมูลเลย คือ การไปแปลงค่า X ที่มีปัญหาพหุสัมพันธ์ โดยทำให้ แต่ละ X ไม่ได้ตั้งฉากกัน ให้เป็น X ที่ตั้งฉากกัน ด้วยวิธี orthogonalization ของ Gram-Schmidt จึงทำให้ OLS-ORT เป็นวิธีที่ดีที่สุดในสถานการณ์นี้

จะมีบางโอกาสซึ่งน้อยมากที่ พบว่า วิธี PCR นั้นให้ค่า AMSE ต่ำสุด ซึ่งมีโอกาสเกิดขึ้นได้ เพราะ เมื่อมีปัญหาค่าพหุสัมพันธ์สูงระดับหนึ่ง ย่อมทำให้การจัดองค์ประกอบหลักใหม่ ด้วยวิธี PCR สามารถตัดองค์ประกอบที่ไม่จำเป็นออกไปได้ ซึ่งเมื่อมีการตัดองค์ประกอบที่ไม่จำเป็นออกไป ย่อมทำให้ลดความแปรปรวนลงอย่างชัดเจน ทำให้วิธี PCR ดีกว่า OLS-ORT ในสถานการณ์นี้

สำหรับกรณีระดับความสัมพันธ์อยู่ในช่วงประมาณ 0.85-0.99

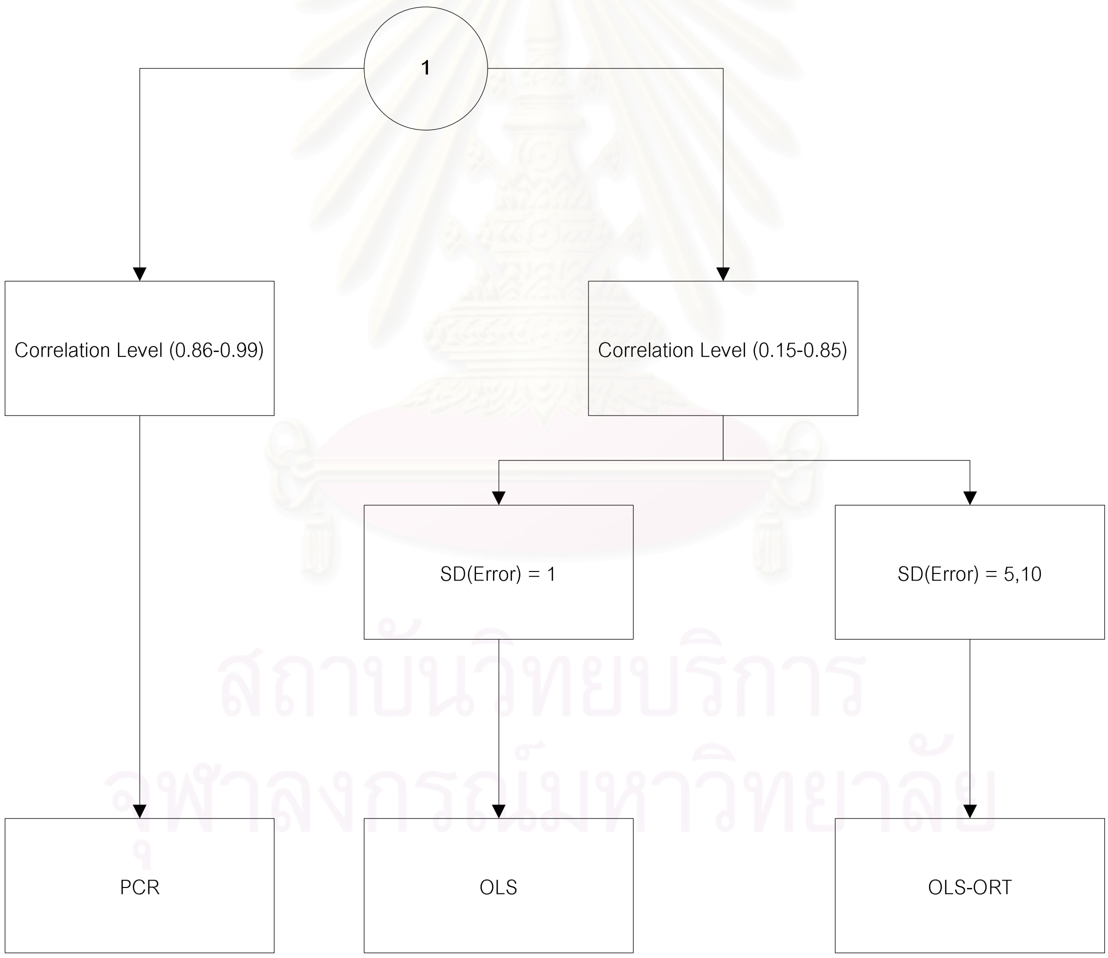
พบว่า PCR จะเป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมา คือ OLS-ORT, วิธี RID, วิธี OLS เพราะ วิธี PCR นั้นจะมีการตัดองค์ประกอบบางส่วนออกไป ส่งผลให้ลดความแปรปรวนให้ต่ำลง

เมื่อพิจารณาค่า Diff พบว่า ถ้าปัญหาพหุสัมพันธ์ไม่สูงมาก จนทำให้วิธี PCR มีการตัดองค์ประกอบบางส่วนออกแล้ว วิธี OLS-ORT กับ PCR จะมีประมาณค่า AMSE ได้ใกล้เคียงกัน โดยวิธี OLS-ORT จะดีกว่าวิธี PCR ไม่มากนัก แต่เมื่อปัญหาพหุสัมพันธ์มีระดับสูง ย่อมทำให้การจัดองค์ประกอบหลักใหม่ ด้วยวิธี PCR สามารถตัดองค์ประกอบที่ไม่จำเป็นออกไปได้ ซึ่งเมื่อมีการตัดองค์ประกอบที่ไม่จำเป็นออกไป ย่อมทำให้ลดความแปรปรวนลงอย่างชัดเจน ทำให้ประสิทธิภาพของ วิธี PCR ดีกว่า OLS-ORT อย่างชัดเจนมาก และถ้าเทียบกับวิธีอื่นด้วย จะพบว่ามีประสิทธิภาพโดดเด่นมาก

อย่างไรก็ตาม ในบางกรณีที่พบได้น้อยที่เมื่อระดับความสัมพันธ์ของตัวแปรอิสระ ในช่วง 0.85-0.99 ส่งผลให้เกิดปัญหาพหุสัมพันธ์สูง ซึ่ง PCR จะเป็นวิธีที่เหมาะสมในสถานการณ์

ลักษณะนี้ แต่แนวโน้มของ ค่า AMSE ของ PCR ไม่ได้ต่ำสุด เพราะ เนื่องจากวิธี PCR นั้นมีหลักการอยู่ที่ การจัดองค์ประกอบใหม่ของข้อมูล X และเลือกเอาเฉพาะองค์ประกอบที่สามารถอธิบายตัวสมการได้มากไว้ ซึ่งถ้ามีการตัดองค์ประกอบออกไปแล้ว ก็จะลดความแปรปรวนลงได้มาก ดังนั้นจึงมีความเป็นไปได้ที่ในบางสถานการณ์ ไม่สามารถตัดองค์ประกอบออกไปได้เลย จึงทำให้วิธี OLS-ORT มีประสิทธิภาพดีกว่า วิธี PCR หรือ เป็นไปได้ที่ วิธี PCR ของ ขนาดตัวอย่างที่มีค่าน้อย จะมีค่า AMSE ต่ำกว่า ขนาดตัวอย่างที่มีค่ามาก ในปัจจัยอื่นๆ เดียวกัน เพราะ ในกรณีนั้น ขนาดตัวอย่างน้อย และตัวแปรอิสระน้อย ข้อมูลอาจจะไม่หลากหลายมากนัก และ สามารถรวมเป็นองค์ประกอบหลักอันเดียวก็ได้

ผังภาพการตัดสินใจเลือกวิธีการประมาณค่า



5.2 ข้อเสนอแนะ

1. การวิจัยครั้งนี้มีขอบเขตการวิจัยที่ $x_i \sim N(\mu_x, \sigma_x^2)$ แต่ถ้า x_i มีฟังก์ชันการแจกแจงเป็นแบบอื่น ๆ เช่น การแจกแจงแบบที โคชี ที่มีลักษณะหางยาว การแจกแจงเบ้ขวาหรือเบ้ซ้าย เป็นต้น และผู้วิจัยได้ทดลองทำกรณีที่ x_i มีค่าต่ำมาก หรือสูงมาก พบว่าให้ค่าค่าเฉลี่ยรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ที่สูงมากกว่าปกติ ทั้งนี้อาจมีผลเนื่องมาจากความคลาดเคลื่อนในขั้นตอนการคำนวณค่า $(X'X)^{-1}$
2. จากผลการวิจัย พบว่า ปัญหาพหุสัมพันธ์ที่กระทบต่อการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ นั้นจะเป็นกรณีที่มีปัญหาพหุสัมพันธ์อย่างมาก โดยจากงานวิจัยพบว่าจะอยู่ในช่วงระดับความสัมพันธ์ที่ 0.86-0.99 จึงพอจะสามารถทำการสนใจเฉพาะลงไปในผลในแต่จุดของช่วงนี้ ด้วยวิธีการประมาณค่าต่างๆ
3. ในแต่ละวิธีของการประมาณค่า จะมีหลายแนวคิด และ หลักการรายละเอียดบางส่วนที่ต่างกัน จึงน่าสนใจที่หาข้อสรุปว่า แนวคิดใดในแต่ละวิธีการประมาณค่า เหมาะกับสถานการณ์ใด

รายการอ้างอิง

ภาษาไทย

กัลยา วานิชย์บัญชา. การใช้ SPSS for Windows ในการวิเคราะห์ข้อมูล เวอร์ชั่น 7 – 10. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร : ซี เค แอนด์ เอส โฟโต้สตูดิโอ, 2543.

จินดา ยาปนเวช. คู่มือเรียนภาษา S-plus. พิมพ์ครั้งที่ 1. กรุงเทพมหานคร : โปรวิชั่น, 2549.

จิตรวี วีระประดิษฐ์. การประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นพหุ เมื่อข้อมูลมีค่าผิดปกติ. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2538.

ประชุม สุวัตถี. ทฤษฎีการอนุมานเชิงสถิติ. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2545.

มนตรี พิริยะกุล. ทฤษฎีสถิติ 2 : Theory of Statistics 2 (ST412). พิมพ์ครั้งที่ 6. กรุงเทพมหานคร : สำนักพิมพ์มหาวิทยาลัยรามคำแหง, 2536.

มนตรี พิริยะกุล. เทคนิคการวิเคราะห์สมการถดถอย : Regression Analysis (ST331). พิมพ์ครั้งที่ 1. กรุงเทพมหานคร : สำนักพิมพ์มหาวิทยาลัยรามคำแหง, 2536.

มานพ วราภักดิ์. การจำลองเบื้องต้น : Introduction to Simulation. กรุงเทพมหานคร : ศูนย์ผลิตตำราเรียน สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2547.

สุชาดา กีระนันทน์. การอนุมานเชิงสถิติ : ทฤษฎีขั้นต้น. พิมพ์ครั้งที่ 3. กรุงเทพมหานคร : โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2534.

สุพล ดุรงค์วัฒนา. การวิเคราะห์เชิงสถิติ : การวิเคราะห์ความถดถอย. กรุงเทพมหานคร.โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2536.

ภาษาต่างประเทศ

Askin, R. G., and Montgomery, D. C. Augmented robust estimaters. Technometrics 22 (1980) : 333 - 341.

Casella, G. and Berger, R. L. Statistical Inference. 2nd ed. Pacific Grove : Duxbury. 2002.

Danodar N.Gujarati. Basic Econometrics. : McGraw Hill. 1995.

G.S.Maddala. Econometrics. Florida : McGraw Hill. 1997

Hamilton, L. C. Modern data analysis : A first course in applied statistics. California : Wadsworth (1990) :116 – 147.

Hoerl, A. E., and Kennard, R. W. Ridge Regression : Applications to nonorthogonal problems. Technometrics 12 (1970) : 69 – 82.

Hoerl, A. E., and Kennard, R. W., and Bladwin, K. F. Ridge Regression : Some simulations. Commun, Stat. 4 (1975) : 105 – 123.

Pfaffenberger, R. C., and Dielman, T. E. Robust regression : analysis and applications. New York : Marcel Dekker (1990) : 243 – 270.

Pfaffenberger, R. C., and Dielman, T. E. A comparison of robust ridge estimators. Proceedings of the American Statistical Association Business and Economic Statistics Section. Las Vegas , Nev., (1985) : 631 – 635.

Pfaffenberger, R. C., and Dielman, T. E. A modified ridge regression estimator using the least absolute value criterion in the multiple linear regression model. Proceedings of the American Institue for Decision Sciences. Toronto, (1984) : 791 – 793.

ภาคผนวก

รายละเอียดของโปรแกรม

กำหนดค่าต่างๆในโปรแกรม

```
ini.n <-
ini.rho <
ini.sd <-

#define variable
#X.x  =  x-generate before add Beta
#X = x after add beta
#x0 = x of b0 (have value = 1)
#B1 = Start B
#Y1 = Start Y

# Keep Result of Run Program

k.MSE.B.Ols <- array(,dim=c(loops))
k.MSE.B.Ortho <- array(,dim=c(loops))
k.MSE.B.Rid <- array(,dim=c(loops))
k.MSE.B.Pcafull <- array(,dim=c(loops))
k.MSE.B.Pcafull2 <- array(,dim=c(loops))
k.MSE.B.Prin <- array(,dim=c(loops))
k.MSE.B.Pcasel <- array(,dim=c(loops))
k.numpoint <- array(,dim=c(loops))
```

สร้างค่า X

```
for(loop in 1:loops)
{ # loops
    # Generate ~ N(0,1)
    Z <- array(rnorm(p*n,0,1),dim=c(p,n))
# Define Variance-Covarinace Matrix
    cov.x <- array(,dim=c(p,p))
    di.cov.x <- array(,dim=c(p))
    for(i in 1:p)
    {
        di.cov.x[i] <- 1
    }
    cov.x <- diag(di.cov.x)
    set.a <- c(1:15)
    prob.a <- c(1/15)
    relation <- rho+(1/100)*sample(set.a,size=1,prob.a)

    for(i in 1:p-1)
    {
        for(j in (i+1):p)
        {
            cov.x[i,j] <- rho+(1/100)*sample(set.a,size=1,prob.a)
            cov.x[j,i] <- cov.x[i,j]
        }
    }
    C.chok <- t(chol(cov.x,pivot=F))
```

```
# Define X matrix from Choleski
    X.xx <- array(,dim=c(p,n))
    for(i in 1:p)
    {
        for(j in 1:n)
        {
            X.xx[i,j] <- C.chok[i,] %*% Z[,j]
        }
    }
    X.x <- round(t(X.xx),dig=4)
```

สร้างค่าความคลาดเคลื่อน และสัมประสิทธิ์การถดถอยพหุคูณตั้งต้น

```
# Generate Error Normal
er <- array(rnorm(n,0,sd),dim=c(n,1))

# Define Design Matrix
x0 <- matrix(c(1),n,1)
X <- round(matrix(c(x0,X.x),n,p+1),dig=4)


 ###########################################################

# Define Beta Parameter By FINDING EIGENVALUE from X'X
XForFindB <- t(X.x) %*% X.x
Ei2 <- eigen(XForFindB)
ei.val2 <- c(Ei2$values)
ei.vec2 <- array(c(Ei2$vectors),dim=c(p,p))
ei.max2 <- max(ei.val2)
ei.min2 <- min(ei.val2)
condition <- ei.max2/ei.min2
for(i in 1:p)
{
	if(ei.val2[i]==ei.max2) B112 <- ei.vec2[,i]
	if(ei.val2[i]==ei.min2)  B222 <- ei.vec2[,i]
}
Bmaxof2 <- matrix(c(1,B112),p+1,1)
Bminof2 <- matrix(c(1,B222),p+1,1)

 Bpoint <- matrix(c(1),pr,1)
Bpoint0 <- matrix(c(1),p,1)
```

กำหนด ค่า Y ตั้งต้น

```
# Data Observations
Ypoint <- X %*% Bpoint + er
Ymaxof2 <- X %*% Bmaxof2 + er
Yminof2 <- X %*% Bminof2 + er

#Case Bstart = Bpoint#

######################Set value in loop########################

B1 <- Bpoint
Y1 <- Ypoint
```

การประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธี OLS

```
################ OLS Estimator################################

Xols <- matrix(c(0),n,p+1)
Xols <- X

B.Ols <- (ginverse(t(Xols) %*% Xols)) %*% (t(Xols)%*%Y1)
Y1.hat.Ols <- Xols %*% B.Ols

        MSE.Y1.hat.Ols <- sum((Y1.hat.Ols-Y1)^2)

MSE.B.Ols <- sqrt((sum((B.Ols-B1)^2))/pr)

k.MSE.B.Ols[loop] <- MSE.B.Ols
```

การประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธี OLS-ORT

```
#################Transform X ########################
        Xortho <- array(0,dim=c(n,p))
         Xortho <- X.x
        Z1 <- matrix(c(Xortho[,1]),n,1)
        ZZ <- matrix(c(0),n,p)
        ZZ <- matrix(c(Z1),n,p)
        Mat.k <- matrix(c(0),n,p)
        Mat.sum <- matrix(c(0),n,p)
        udot <- 0
        ldot <- 0
        dotdot <- 0

        for(i in 2:p)
        { #i
                Mat.sum[,i] <- 0
                for(a in 1:(i-1))
                { # a
                        udot <- Xortho[,i] *ZZ[,a]
                        udot2 <- sum(udot)
                        ldot <- ZZ[,a] *ZZ[,a]
                        ldot2 <- sum(ldot)
                           dotdot <-  udot2/ldot2
                                Mat.k[,a] <- dotdot*ZZ[,a]
                        Mat.sum[,i] <- Mat.sum[,i]+Mat.k[,a]
                } #a

                ZZ[,i] <- 0
                ZZ[,i] <- Xortho[,i]-Mat.sum[,i]
        } #i
```

```
#Add B0 for x
        xor0 <- matrix(c(1),n,1)
        Xor <- round(matrix(c(xor0,ZZ),n,p+1),dig=3)

        #################Estimator Orthogonal########################
        B.Ortho <- (ginverse(t(Xor) %*% Xor)) %*% (t(Xor)%*%Y1)
        Y.hat.Ortho <- Xor %*% B.Ortho
        MSE.B.Ortho <- sqrt((sum((B.Ortho-B1)^2))/pr)

        k.MSE.B.Ortho[loop] <- MSE.B.Ortho
```

การประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธี RID

```
 ############## Ridge Estimator##################
# Compute Trace of (X'X) -1
  tk <- ginverse(t(X)%*%X)
  sum.tk <- 0
  for(i in 1:pr)
  {
        for(j in 1:pr)
        {
              if(i==j)
              {
                    temp <- tk[i,j]
                    sum.tk <- sum.tk+temp
              }
        }
  }
  TT <- sum.tk/pr
  bb <- t(B.Ols)%*%B.Ols

  # Initial Value of Ridge Estimator
  k0 <- (1/bb)*(pr*(1/(n-1))*MSE.Y1.hat.Ols )
  k0.V <- array(,dim=c(pr))
  for(i in 1:pr)
  {
        k0.V[i] <- k0
  }
  k0.M <- diag(k0.V)
  ka.M <- diag(k0.M)

  B.Rid <- ginverse((t(X)%*%X) + ka.M) %*% (t(X)%*%Y1)
  Y.hat.Rid  <- X %*% B.Rid
  MSE.Y.hat.Rid <- (sum((Y.hat.Rid-Y1)^2))/n
  MSE.B.Rid  <- sqrt((sum((B.Rid-B1)^2))/pr)
```

```
k.MSE.B.Rid[loop] <- MSE.B.Rid
```

การประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธี PCR

```
#################Principal Component Regression#################
XTX <- t(X.x) %*% X.x
PPP <-  eigen(XTX)
PP <- matrix(PPP$vectors,p,p)
ZZ <- X.x %*% PP
XP.Z <- matrix(c(ZZ),n,p)

B.new <- ginverse(t(XP.Z)%*%XP.Z)  %*% (t(XP.Z)%*%Y1)
BB.PCA <- PP %*% B.new
B.Pcafull <- matrix(c(1,BB.PCA),pr,1)
MSE.B.Pcafull  <- sqrt((sum((B.Pcafull-B1)^2))/pr)

k.MSE.B.Pcafull[loop] <- MSE.B.Pcafull

#################Principal Component Regression for selecting eigen >=1
##################
what.num<- array(0,dim=c(p))

XTX <- t(X.x) %*% X.x
PPP2 <-  eigen(XTX)
PP2 <- matrix(PPP2$vectors,p,p)


ZZZ2 <- X.x %*% PP
PV <- array(PPP$values,dim=c(p))
num <- 0
for(i in 1:p)
{
    if(PV[i] > 1)
    {
        num <- num +1
        what.num[i] <- num
```

```
        }
}

num_test <- num
k.numpoint <- num_test

ZZ2 <- matrix(ZZZ2,n,num)
XP.Z2 <- matrix(c(ZZ2),n,num)
B.new2 <- ginverse(t(XP.Z2)%*%XP.Z2)  %*% (t(XP.Z2)%*%Y1)
P.num <- matrix(PPP2$vectors,p,num)
BB.PCA2 <- P.num %*% B.new2
B.Pcasel <- matrix(c(1,BB.PCA2),num+1,1)

B.O <- matrix(c(1),num+1,1)
for(i in 1:p)
{
        if(what.num[i] > 0)
        {
                B.O[i+1] <- B1[num+1]
        }
}

MSE.B.Pcasel  <- sqrt((sum((B.Pcasel-B.O)^2))/pr)

k.MSE.B.Pcasel[loop] <- MSE.B.Pcasel
```

การคำนวณค่า AMSE

```
AMSE.B.Ols <- sum(k.MSE.B.Ols)/loops
AMSE.Bmax.Ols <- sum(k.MSE.Bmax.Ols)/loops
AMSE.Bmin.Ols <- sum(k.MSE.Bmin.Ols)/loops

AMSE.B.Ortho <- sum(k.MSE.B.Ortho)/loops
AMSE.Bmax.Ortho <- sum(k.MSE.Bmax.Ortho)/loops
AMSE.Bmin.Ortho <- sum(k.MSE.Bmin.Ortho)/loops


AMSE.B.Rid <- sum(k.MSE.B.Rid)/loops
AMSE.Bmax.Rid <- sum(k.MSE.Bmax.Rid)/loops
AMSE.Bmin.Rid <- sum(k.MSE.Bmin.Rid)/loops

AMSE.Bmin.Pcafull <- sum(k.MSE.Bmin.Pcafull)/loops
AMSE.Bmaxof2.Pcafull <- sum(k.MSE.Bmaxof2.Pcafull)/loops
AMSE.Bminof2.Pcafull <- sum(k.MSE.Bminof2.Pcafull)/loops


AMSE.B.Pcasel <- sum(k.MSE.B.Pcasel)/loops
AMSE.Bmax.Pcasel <- sum(k.MSE.Bmax.Pcasel)/loops

AMSE.Bmin.Pcasel <- sum(k.MSE.Bmin.Pcasel)/loops
AMSE.Bmaxof2.Pcasel <- sum(k.MSE.Bmaxof2.Pcasel)/loops
AMSE.Bminof2.Pcasel <- sum(k.MSE.Bminof2.Pcasel)/loops

AVG.k.numpoint <- sum(k.numpoint )/loops

WW <- round(c(countdo,loops,p,n,rho,sd,relation,condition,AMSE.B.Ols ,AMSE.B.Ortho
      ,AMSE.B.Rid,AMSE.B.Pcafull,AMSE.B.Pcasel,AVG.k.numpoint ,AMSE.Bmaxof2.Ols
      ,AMSE.Bmaxof2.Ortho
      ,AMSE.Bmaxof2.Rid,AMSE.Bmaxof2.Pcafull,AMSE.Bmaxof2.Pcasel,AVG.k.nummaxof2,AMSE.
      Bminof2.Ols ,AMSE.Bminof2.Ortho
      ,AMSE.Bminof2.Rid,AMSE.Bminof2.Pcafull,AMSE.Bminof2.Pcasel,AVG.k.numminof2 ),dig=4)
```

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวเปรมวดี ชูไสว เกิดเมื่อวันที่ 24 ธันวาคม พ.ศ.2523 สำเร็จการศึกษาปริญญาสถิติศาสตรบัณฑิต สาขาเทคโนโลยีสารสนเทศเพื่อธุรกิจ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2544 สำเร็จการศึกษาการจัดการมหาบัณฑิตสาขานวัตกรรมการจัดการ วิทยาลัยการจัดการ(นานาชาติ) มหาวิทยาลัยมหิดล ในปี 2547 และเข้าศึกษาต่อในหลักสูตรสถิติศาสตรมหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2546