การประมาณค่าพารามิเตอร์ของตัวแบบความถดถอยพหุนาม ด้วยการหดและการทำโทษ

นางสาวรัชเนตร ปานดำ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2549

PARAMETER ESTIMATION OF POLYNOMIAL REGRESSION MODEL WITH SHRINKAGE AND PENALIZATION

Miss Ratchanate Pandam

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 2006

หัวข้อวิทยานิพนธ์	การประมาณค่าพารามิเตอร์ของตัวแบบความถดถอยพหุนามด้วยการหดและการทำโทษ
โดย	นางสาวรัชเนตร ปานคำ
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.ธีระพร วีระถาวร

คณะพาณิชศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้วิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

อนุชา คุณพนิชกิจ .... คณบดีคณะพาณิชยศาสตร์และการบัญชี

(รองศาสตราจารย์ ดร. ดนุชา คุณพนิชกิจ)

คณะกรรมการสอบวิทยานิพนธ์

.... ประธานกรรมการ

(อาจารย์ ดร. อรุณี กำลัง)

.... อาจารย์ที่ปรึกษาวิทยานิพนธ์

(รองศาสตราจารย์ ดร.ธีระพร วีระถาวร)

.... กรรมการ

(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

รัชเนตร ปานคำ : การประมาณค่าพารามิเตอร์ของตัวแบบความถดถอยพหุนามด้วยการหดและการทำโทษ (PARAMETER ESTIMATION OF POLYNOMIAL REGRESSION MODEL WITH SHRINKAGE AND PENALIZATION) อ.ที่ปรึกษา : รศ.ดร.ธีระพร วีระถาวร, 74 หน้า.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความถูกต้องของการประมาณค่าสัมประสิทธิ์การถดถอยพหุนาม โดยจะเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอย 4 วิธี ได้แก่ วิธีกำลังสองน้อยสุด (Ordinary Least Squares method (OLS)) วิธีการ์ร็อตที่ไม่เป็นลบ (Nonnegative Garrote method (NG)) วิธีแจ็คไนฟ์ริดจ์ (Jackknifed Ridge method (JR)) และวิธีลิวไทป์ (Liu-Type method (LT)) เกณฑ์ที่ใช้ในการตัดสินใจคือ ค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (average root mean squares error (ARMSE)) และส่วนที่ใช้ประกอบในการพิจารณาเปรียบเทียบได้แก่ ค่าอัตราส่วนผลต่างของค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ratio of different average root mean squares error (DIFF)) สถานการณ์ที่ศึกษาคือ กำหนด $\beta = (1, 1, \ldots, 1)'$ ตัวแปรอิสระที่ศึกษามีการแจกแจงแบบปกติที่มีค่าเฉลี่ย 5 และความแปรปรวน 4 กำลังสูงสุดของตัวแปรอิสระที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบถดถอยพหุนาม (MB) คือ 2 3 4 5 และ 6 ขนาดตัวอย่างที่ใช้คือ 5p 10p 15p และ 20p ส่วนค่าคลาดเคลื่อนสุ่มในตัวแปรตามมีการแจกแจงแบบปกติที่มีค่าเฉลี่ย 0 และความแปรปรวนเท่ากับ 2 4 6 8 และ 10 ข้อมูลที่ใช้ในการวิจัยใช้โปรแกรมคอมพิวเตอร์จำลองด้วยเทคนิคมอนติคาร์โลกระทำซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์ ผลการวิจัยสามารถสรุปได้ดังนี้

ในทุกกรณีวิธี LT ให้ค่า ARMSE น้อยที่สุด รองลงมาคือวิธี JR วิธี NG และวิธี OLS ตามลำดับ โดยที่ค่า ARMSE มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น และมีแนวโน้มเพิ่มขึ้นเมื่อเลขชี้กำลังสูงสุดของตัวแปรอิสระที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบการถดถอยพหุนาม (MB) และความแปรปรวนของความคลาดเคลื่อน (σ_{ε}^{2}) เพิ่มขึ้น

ค่า ARMSE แปรผันตามปัจจัยต่อไปนี้จากมากไปน้อยคือ เลขชี้กำลังสูงสุดของตัวแปรอิสระที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบการถดถอยพหุนาม (MB) และความแปรปรวนของความคลาดเคลื่อน (σ_{ε}^{2}) แต่แปรผกผันกับขนาดตัวอย่าง

ภาควิชา.......สถิติ....... ลายมือชื่อนิสิต.......รัชเนตร ปานคำ.......
สาขาวิชา.......สถิติ....... ลายมือชื่ออาจารย์ที่ปรึกษา..............
ปีการศึกษา.......2549.......

4782343526 : MAJOR STATISTICS

KEY WORD : POLYNOMIAL REGRESSION / SHRINKAGE / PENALIZATION / ORDINARY LEAST SQUARES METHOD / NONNEGATIVE GARROTE METHOD / JACKKNIFED RIDGE METHOD / LIU-TYPE METHOD

RATCHANATE PANDAM : PARAMETER ESTIMATION OF POLYNOMIAL REGRESSION MODEL WITH SHRINKAGE AND PENALIZATION THESIS ADVISOR : ASSOC. PROF. THEERAPORN VERATHAWORN, Ph.D., 74 pp.

The objective of this research is to compare on the accuracy of polynomial-regression-coefficient estimator. This research compares four approaches consisting of polynomial-regression-coefficient estimation methods, Ordinary Least Squares method (OLS), Nonnegative Garrote method (NG), Jackknifed Ridge method (JR) and Liu-Type method (LT). The criterion for making decision is Average Root Mean Squares Error (ARMSE) and use Ratio of Different Average Root Mean Squares Error (DIFF) to support decision. As for the case study, we specify $\underset{\sim}{\beta} = (1, 1, \ldots, 1)'$, the distribution of independent variables are assumed to be normal distribution with mean equal to 5 and variance equal to 4, highest degree of independent variables for dependent variable building in polynomial regression model (MB) are 2, 3, 4, 5 and 6 respectively and the sample sizes are 5p, 10p, 15p and 20p. The distribution of error in the dependent variable is the normal distribution with mean equal to 0 and variance equal to 2, 3, 4, 5 and 6, respectively. The data for this research is simulated by using the Monte Carlo simulation technique with 1,000 repetitions for each case. The results of this research are as follows:

In all cases, LT method has the smallest ARMSE and JR method has a smaller ARMSE than NG method and OLS method, respectively. The ARMSE decreases when sample size increase but it increases when highest degree of independent variables for dependent variable building in polynomial regression model (MB) or variance of error in the dependent variable (σ_{ε}^2) increases.

The ARMSE varies with, most to least, respectively, highest degree of independent variables for dependent variable building in polynomial regression model (MB) and variance of error in the dependent variables (σ_{ε}^2) but converses to sample size.

Department .... Statistics ....

Field of study .... Statistics ....

Academic year .... 2006 ....

Student's signature ...Ratchanate Pandam...

Advisor's signature ...Theeraporn Verath...

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความช่วยเหลือ และเอาใจใส่เป็นอย่างดียิ่งของรองศาสตราจารย์ ดร.ธีระพร วีระถาวร อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้วิจัยขอกราบขอบพระคุณต่อท่านอาจารย์เป็นอย่างสูง ที่กรุณาให้คำปรึกษาและแนะนำเกี่ยวกับวิทยานิพนธ์ด้วยดีเสมอมา

ผู้วิจัยขอกราบขอบพระคุณอาจารย์ ดร. อรุณี กำลัง ประธานกรรมการสอบวิทยานิพนธ์ และรองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา กรรมการสอบวิทยานิพนธ์ ที่กรุณาให้คำแนะนำ ตรวจสอบ และแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น

สุดท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณ คุณแม่ คุณยาย และคุณย่า ที่ช่วยส่งเสริม สนับสนุน และให้กำลังใจเสมอมาจนสำเร็จการศึกษา รวมทั้งขอบคุณเพื่อนๆทุกคนที่คอยห่วงใยและให้กำลังใจผู้วิจัยมาโดยตลอด

สารบัญ

หน้า

สารบัญตาราง

สารบัญภาพ

<u>รูปที่</u> หน้า

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การวิเคราะห์ความถดถอย (regression analysis) คือการศึกษาความสัมพันธ์ระหว่างตัวแปรตัวหนึ่งที่เรียกว่า ตัวแปรตาม (dependent variable) กับตัวแปรอื่นอีกหนึ่งตัวหรือมากกว่าหนึ่งตัวที่เรียกว่า ตัวแปรอิสระ (independent variables) เพื่อที่จะประมาณหรือพยากรณ์ค่าจริงของตัวแปรตาม ลักษณะของความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระอาจมีความสัมพันธ์กันเป็นเชิงเส้นหรือเส้นโค้ง ในกรณีที่ความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระมีความสัมพันธ์กันเป็นเส้นโค้ง วิธีทางสถิติที่นำมาหารูปแบบความสัมพันธ์คือการวิเคราะห์ความถดถอยพหุนาม (polynomial regression analysis) ซึ่งจากการศึกษาข้อมูลจริงพบว่าข้อมูลที่มีรูปแบบความสัมพันธ์กันเป็นเส้นโค้งนั้นมีมากมาย เช่น ดัชนีราคาหุ้น ราคาน้ำมัน ราคาทองคำ ฯลฯ เนื่องจากวิธีการวิเคราะห์ความถดถอยพหุนามอาจใช้แนวความคิดของวิธีการวิเคราะห์ความถดถอยเชิงเส้น (linear regression analysis method) โดยให้แต่ละพจน์ของพหุนามเป็นตัวแปรอิสระตัวหนึ่ง ดังนั้นการวิเคราะห์ความถดถอยพหุนามจึงเป็นกรณีพิเศษของการวิเคราะห์ความถดถอยเชิงเส้น

การวิเคราะห์ความถดถอยพหุนามโดยปกติจะเกิดปัญหาตัวแปรอิสระมีความสัมพันธ์กันสูง (และมักไม่นำมาพิจารณา) จึงส่งผลให้ค่าคลาดเคลื่อนของตัวแบบที่ได้มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูงกว่าปกติ ในการแก้ปัญหาดังกล่าวจึงควรปรับตัวประมาณของพารามิเตอร์โดยการเลือกใช้ตัวประมาณที่แก้ปัญหาความสัมพันธ์ของตัวแปรอิสระหรือพหุสัมพันธ์ (multicollinearity) ที่เกิดขึ้น วิธีการพื้นฐานที่นิยมนำมาใช้ในการแก้ปัญหาข้างต้นคือวิธีการหด (shrinkage method) และวิธีการทำโทษ (penalization method) ซึ่งเป็นวิธีการที่ช่วยเพิ่มความถูกต้องในการพยากรณ์ค่าจริงของตัวแปรตามให้มีความถูกต้องมากยิ่งขึ้น นอกจากนี้ทั้ง 2 วิธีการยังเป็นวิธีการพื้นฐานในการสร้างตัวประมาณค่าพารามิเตอร์แบบต่างๆ เช่น ตัวประมาณสไตน์ (stein estimator) ตัวประมาณริดจ์ (ridge estimator) เป็นต้น

ตัวแบบการถดถอยพหุนามโดยทั่วไปมีรูปแบบดังนี้

(1.1) $$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p + \varepsilon_i \qquad ; \quad i = 1, 2, \ldots, n$$

จาก (1.1) มีข้อมูล n ชุด ดังนั้นสามารถเขียนสมการรูปทั่วไป (General Linear Model) ได้ดังนี้

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

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

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

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

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

จะได้ว่า $\underset{\sim}{y}$ เป็นเวกเตอร์ของตัวแปรตามขนาด $n \times 1$ ที่ขึ้นอยู่กับตัวแปรอิสระ X โดยที่ $\underset{\sim}{y}$ มีการแจกแจงแบบปกติหลายตัวแปร ด้วยค่าเฉลี่ย $X\underset{\sim}{\beta}$ และความแปรปรวน $\sigma_\varepsilon^2 I_n$ ซึ่งเขียนได้เป็น $\underset{\sim}{y} \sim N_n(X\underset{\sim}{\beta}, \sigma_\varepsilon^2 I_n)$

โดยที่ $\underset{\sim}{\beta}$ เป็นเวกเตอร์ของพารามิเตอร์ที่ไม่ทราบค่าขนาด $(p+1)$ ซึ่งเขียนได้เป็น $\underset{\sim}{\beta} = (\beta_0, \beta_1, ..., \beta_p)'$

$\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของความคลาดเคลื่อนสุ่มในตัวแปรตามขนาด $n \times 1$ โดยที่ $\varepsilon_i \; i.i.d. \; N(0, \sigma_\varepsilon^2)$

σ_ε^2 เป็นพารามิเตอร์ที่ไม่ทราบค่า

I_n เป็นเมทริกซ์เอกลักษณ์จากข้อมูลจำนวน n ชุด

และ p เป็นจำนวนพจน์พหุนามในตัวแบบโดยไม่นับพจน์ค่าคงที่

ในการวิเคราะห์ความถดถอยที่มีตัวแปรอิสระหลายตัวสิ่งสำคัญเป็นอย่างมากคือการสร้างตัวแบบให้มีความเหมาะสมเพื่อให้ได้ค่าพยากรณ์ที่ใกล้เคียงกับค่าจริงมากที่สุด ซึ่งตัวแบบที่เหมาะสมนั้นควรสร้างมาจากวิธีการประมาณค่าพารามิเตอร์ที่เหมาะสม ผู้วิจัยจึงได้ทำการศึกษาวิธีการประมาณค่าพารามิเตอร์ที่เหมาะสมจากงานวิจัยต่าง ๆ ดังนี้

นพมาศ อัครจันทโชติ (พ.ศ. 2539) ได้ศึกษาเกี่ยวกับการเปรียบเทียบวิธีการสร้างตัวแบบในการวิเคราะห์ความถดถอยพหุนามกรณีที่มี 2 ตัวแปรอิสระ ซึ่งเกิดอันตรกิริยา (interaction) โดยทำการเปรียบเทียบวิธีการสร้างตัวแบบ 4 วิธี ได้แก่ วิธีการสร้างตัวแบบด้วยวิธีกำลังสองน้อยสุดสามัญ (Ordinary Least Squares method (OLS)) การสร้างตัวแบบด้วยวิธีกำจัดตัวแปรอิสระย้อนหลัง (Backward Elimination method (BE)) การสร้างตัวแบบด้วยวิธีการถดถอยขั้นบันได (Stepwise Regression method (SW)) และการสร้างตัวแบบด้วยวิธีตัวแบบหลักเกณฑ์ดี (Well-Formulated Method (WF))

สมลักษณ์ ศิริชื่นวิจิตร (พ.ศ.2548) ได้ศึกษาเกี่ยวกับการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของตัวแบบถดถอยพหุนาม (A COMPARISION ON PARAMETERS-ESTIMATION OF POLYNOMIAL REGRESSION MODEL) โดยทำการเปรียบเทียบวิธีการ

ประมาณค่าพารามิเตอร์ 3 วิธี ได้แก่ วิธีกำลังสองน้อยสุดสามัญ (Ordinary Least Squares method (OLS)) วิธีริดจ์สามัญ (Ridge Ordinary Least Squares method (ROLS)) และวิธีริดจ์ที่มีค่าสัมบูรณ์น้อยสุด (Ridge Least Absolute Value method (RLAV))

ในการวิจัยครั้งนี้ผู้วิจัยได้สนใจทำการศึกษาและเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยของตัวแบบพหุนาม 4 วิธี คือ

1. วิธีกำลังสองน้อยสุด (Ordinary Least Squares method (OLS))
2. วิธีการ์ร็อตที่ไม่เป็นลบ (Nonnegative Garrote method (NG))
3. วิธีแจ็คไนฟริดจ์ (Jackknifed Ridge method (JR))
4. วิธีลิวไทป์ (Liu-Type method (LT))

โดยวิธีแรกเป็นแนวคิดแบบวิธีแบบฉบับ (Classical method) ส่วนสามวิธีหลังเป็นแนวคิดที่อาศัยวิธีการหด (shrinkage method) และวิธีการทำโทษ (penalization method) เป็นวิธีการพื้นฐานในการสร้างตัวประมาณค่าพารามิเตอร์

1.2 วัตถุประสงค์การวิจัย

เพื่อเปรียบเทียบความถูกต้องของค่าพยากรณ์ของตัวแบบที่ได้จากวิธีการประมาณค่าพารามิเตอร์ที่เหมาะสม 4 วิธีดังนี้

1. วิธีกำลังสองน้อยสุด (Ordinary Least Squares method (OLS))
2. วิธีการ์ร็อตที่ไม่เป็นลบ (Nonnegative Garrote method (NG))
3. วิธีแจ็คไนฟริดจ์ (Jackknifed Ridge method (JR))
4. วิธีลิวไทป์ (Liu-Type method (LT))

1.3 สมมติฐานการวิจัย

การประมาณค่าพารามิเตอร์ของตัวแบบความถดถอยพหุนามจะมีความถูกต้องในการพยากรณ์ตัวแปรตาม เมื่อใช้วิธีการประมาณค่าพารามิเตอร์โดยวิธีลิวไทป์ (LT) มากกว่าวิธีกำลังสองน้อยสุด (OLS) วิธีการ์ร็อตที่ไม่เป็นลบ (NG) และวิธีแจ็คไนฟริดจ์ (JR) (เนื่องจากวิธีการประมาณค่าพารามิเตอร์โดยวิธีลิวไทป์เป็นวิธีการที่สามารถแก้ปัญหาสภาพความเป็นเมทริกซ์ที่ไม่ดี (ill-conditioned) ได้อย่างสมบูรณ์ จึงมีประสิทธิภาพในการช่วยแก้ปัญหาพหุสัมพันธ์ในตัวแปรอิสระโดยตรงทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าน้อยลงกว่าวิธีอื่นๆ)

1.4 ข้อตกลงเบื้องต้น

1. รูปแบบทั่วไปของสมการถดถอยพหุนามมีรูปแบบดังสมการ (1.1)

2. ความคลาดเคลื่อนสุ่มในตัวแปรตามเป็นตัวแปรสุ่มที่มีการแจกแจงเหมือนกันและเป็นอิสระซึ่งกันและกันแบบ $N_n(\underset{\sim}{0}, \sigma_\varepsilon^2 I_n)$

1.5 ขอบเขตของการวิจัย

1. ตัวแบบของความถดถอยพหุนามที่สนใจศึกษาเป็นดังนี้

ตัวแบบที่มีกำลังสูงสุดของตัวแปรอิสระเป็น 6 มีรูปแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \beta_5 x_i^5 + \beta_6 x_i^6 + \varepsilon_i \quad , \; i = 1,2,\ldots,n$$

ตัวแบบที่มีกำลังสูงสุดของตัวแปรอิสระเป็น 5 มีรูปแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \beta_5 x_i^5 + \varepsilon_i \quad , \; i = 1,2,\ldots,n$$

ตัวแบบที่มีกำลังสูงสุดของตัวแปรอิสระเป็น 4 มีรูปแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \varepsilon_i \quad , \; i = 1,2,\ldots,n$$

ตัวแบบที่มีกำลังสูงสุดของตัวแปรอิสระเป็น 3 มีรูปแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \varepsilon_i \quad , \; i = 1,2,\ldots,n$$

ตัวแบบที่มีกำลังสูงสุดของตัวแปรอิสระเป็น 2 มีรูปแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i \quad , \; i = 1,2,\ldots,n$$

2. การวิจัยครั้งนี้กำหนดให้ $\underset{\sim}{\beta}' = (1,\ 1,\ \ldots,1)_{1\times(p+1)}$ [1]ในประชากรทุกรูปแบบที่ศึกษาเมื่อ p เป็นจำนวนพจน์พหุนามในตัวแบบโดยไม่นับพจน์ค่าคงที่

3. ขนาดตัวอย่าง (n) ที่ศึกษาคือ 5p 10p 15p และ 20p [2]

4. กำหนดตัวแปรอิสระ $x_i \sim N\left(\mu, \sigma^2\right)$ [3], $i = 1,2,\ldots,n$ เมื่อ μ และ σ^2 เป็นค่าพารามิเตอร์ ซึ่งในที่นี้กำหนดให้ $\mu = 5$ และ $\sigma^2 = 4$

---

[1]การกำหนดค่าพารามิเตอร์จะกำหนดเป็นเท่าใดก็ได้จากงานวิจัยที่ผ่านมาพบว่าโดยส่วนใหญ่แล้วจะกำหนดเป็น $\underset{\sim}{\beta}' = (1, 1, \ldots, 1)_{1\times(p+1)}$ ซึ่งได้มีผู้ทำการลองเปลี่ยนค่าพารามิเตอร์เป็นค่าอื่นและพบว่าไม่ได้ทำให้การสรุปผลเปลี่ยนแปลงไป

[2] Raykov,T. and Widaman, K.F. "Structural Equation Modeling", Issues in applied structural equation modeling research 2 (1995): 289-318 (ขนาดตัวอย่างที่น้อยที่สุดในการประมาณค่าพารามิเตอร์สำหรับการวิเคราะห์ความถดถอยพหุคูณคือ n = 5p เมื่อ p คือจำนวนตัวแปรอิสระ โดยในงานวิจัยครั้งนี้จะทำการศึกษาเมื่อเพิ่มขนาดตัวอย่างทีละ 5p และ p คือจำนวนพจน์พหุนามในตัวแบบโดยไม่นับพจน์ค่าคงที่)

[3]Seber, G.A.F Linear Regression Analysis (New York, 1997c), pp. 214-216. (เมื่อกำลังของตัวแปรอิสระมีค่าสูง(>6) และค่า x ที่ยกกำลังต่างๆในเมทริกซ์ X มีค่าเพิ่มสูงขึ้น ดังนั้นการหาค่าผกผันของเมทริกซ์ $X'X$ จะเกิดความคลาดเคลื่อนสูง ทำให้ตัวประมาณของพารามิเตอร์ที่ได้มีค่าความคลาดเคลื่อนสูงด้วย นอกจากนั้นกรณีที่ $x \in (0,1)$ จะมีผลทำให้ค่าของ x ที่ยกกำลังต่างๆในเมทริกซ์ X มีค่าลดต่ำลงซึ่งทำให้เกิดปัญหาเช่นเดียวกับกรณีข้างต้น)

5. ในการวิจัยครั้งนี้จะศึกษาเมื่อความคลาดเคลื่อนสุ่มในตัวแปรตามมีการแจกแจง $N(\mu,\sigma^2)$ โดยกำหนด $\mu = 0$ และ $\sigma^2 = 2, 4, 6, 8$ และ 10 เพื่อศึกษาว่าเมื่อความแปรปรวนเพิ่มขึ้นจะส่งผลต่อตัวประมาณที่ศึกษาหรือไม่

6. ตัวแปรตาม $(\underset{\sim}{y})$ ที่ใช้สำหรับคำนวณค่าตัวประมาณสัมประสิทธิ์การถดถอยพหุนามด้วยวิธีต่างๆ สร้างจากตัวแปรอิสระ (X) และความคลาดเคลื่อน $(\underset{\sim}{\varepsilon})$ โดยที่ตัวแปรตาม $(\underset{\sim}{y})$ มีการแจกแจงแบบ $N_n(X\underset{\sim}{\beta},\sigma^2 I_n)$

7. ข้อมูลที่ใช้ในการศึกษาวิจัยครั้งนี้ใช้โปรแกรมคอมพิวเตอร์จำลองด้วยเทคนิคมอนติคาร์โลกระทำซ้ำ 1,000 รอบ ในแต่ละสถานการณ์

1.6 เกณฑ์การตัดสินใจ

เกณฑ์ที่ใช้ในการตัดสินใจว่า วิธีประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามวิธีการใดให้ค่าพารามิเตอร์ใกล้เคียงกับค่าจริงมากที่สุดนั้นจะพิจารณาจากค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (average root mean squares error (ARMSE)) และส่วนประกอบที่ใช้ในการพิจารณาเปรียบเทียบจะใช้ค่าอัตราส่วนผลต่างของค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ratio of different average root mean squares error (DIFF)) ของการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธีต่างๆ

$$RMSE_i = \sqrt{\frac{1}{p+1}\sum_{j=0}^{p}\left(\hat{\beta}_{ij}-\beta_j\right)^2}$$

$$ARMSE = \frac{1}{1{,}000}\sum_{i=1}^{1{,}000} RMSE_i$$

$$DIFF = \frac{ARMSE_k - ARMSE_{\min}}{ARMSE_{\min}} \times 100 \quad ,k = 1, 2, 3, 4$$

เมื่อ β_j แทนสัมประสิทธิ์ความถดถอยพหุนามตัวที่ j

$\hat{\beta}_{ij}$ แทนตัวประมาณสัมประสิทธิ์ความถดถอยพหุนามตัวที่ j ในการประมาณครั้งที่ i

$RMSE_i$ แทนค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์ความถดถอยพหุนามในการประมาณครั้งที่ i

ARMSE แทนค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์ความถดถอยพหุนาม

DIFF แทนอัตราส่วนของผลต่างของค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธีต่างๆ

ARMSE $_k$ แทนค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์ความถดถอยพหุนามโดยวิธีที่ k , k = 1, 2, 3, 4

และ ARMSE $_{min}$ แทนค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์ความถดถอยพหุนามจากวิธีที่มีค่าน้อยที่สุด

ผู้วิจัยพิจารณาว่าตัวประมาณค่าพารามิเตอร์ใดให้ค่า ARMSE และค่า DIFF ต่ำสุดจะเป็นตัวประมาณค่าพารามิเตอร์ที่ดีที่สุด และมีประสิทธิภาพที่สุด

1.7 วิธีดำเนินการวิจัย

1. ศึกษาโปรแกรม MATLAB (Matrix Laboratory Version 7) ซึ่งเป็นโปรแกรมที่จะใช้ในการวิเคราะห์ในครั้งนี้
2. สร้างโปรแกรมที่ใช้ในการสร้างข้อมูลซึ่งใช้ในการทดสอบ โดยวิธีที่เลือกใช้สร้างข้อมูลคือ วิธีมอนติคาร์โล (Simulation by Monte Carlo Method)
3. ศึกษาวิธีการสร้างตัวประมาณที่ต้องการเปรียบเทียบ และเขียนโปรแกรมสร้างตัวประมาณ ดังนี้
   - 3.1 วิธีกำลังสองน้อยสุด (OLS)
   - 3.2 วิธีการ์ร็อตที่ไม่เป็นลบ (NG)
   - 3.3 วิธีแจ็คไนฟริดจ์ (JR)
   - 3.4 วิธีลิวไทป์ (LT)
4. เขียนโปรแกรมสำหรับใช้เป็นเกณฑ์ในการตัดสินใจผลที่ได้จากข้อมูลที่ใช้ในการวิเคราะห์ในแต่ละตัวประมาณ
5. วิเคราะห์ผลจากข้อมูลที่ได้จากการเขียนโปรแกรม

1.8 ประโยชน์ของการวิจัย

เพื่อเป็นแนวทางในการประมาณค่าสัมประสิทธิ์ของตัวแบบถดถอยพหุนามที่ใช้ตัวประมาณค่าสัมประสิทธิ์ที่สามารถแก้ปัญหาพหุสัมพันธ์ (multicollinearity) และทำให้ได้สมการที่ใช้พยากรณ์ค่าของตัวแปรตามมีความคลาดเคลื่อนลดลง

บทที่ 2
ทฤษฎีและสถิติที่ใช้ในการวิจัย

การวิจัยในครั้งนี้ผู้วิจัยได้เสนอวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนาม 4 วิธี คือ วิธีกำลังสองน้อยสุด (Ordinary Least Squares method (OLS)) วิธีการ์ร็อตที่ไม่เป็นลบ (Nonnegative Garrote method (NG)) วิธีแจ็คไนฟริดจ์ (Jackknifed Ridge method (JR)) และวิธีลิวไทป์ (Liu-Type method (LT))

ตัวแบบความถดถอยพหุนามมีรูปแบบทั่วไป ดังนี้

(2.1) $$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p + \varepsilon_i \qquad ; \quad i = 1, 2, \ldots, n$$

เมื่อ β_j เป็นพารามิเตอร์ไม่ทราบค่าของพจน์พหุนามที่ j ; j = 0, 1,…, p

ตัวแบบความถดถอยพหุนามในกรณีนี้เป็นตัวแบบเชิงเส้นในพารามิเตอร์ หมายความว่าในแต่ละเทอมของตัวแบบมีพารามิเตอร์เพียง 1 ตัว คูณด้วยค่าคงที่ของตัวแปรอิสระ ดังนั้นในการประมาณค่าพารามิเตอร์เราจึงสามารถใช้วิธีกำลังสองน้อยสุดในการประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ ดังจะกล่าวดังต่อไปนี้

2.1 การประมาณค่าสัมประสิทธิ์ด้วยวิธีกำลังสองน้อยสุด (OLS)

ในปี ค.ศ.1777-1855 คาร์ล เฟรดริก เกาส์ (Carl Friedrich Gauss) และในปี ค.ศ.1855-1922 อังเดร แอนดรีวิช มาร์คอฟ (Andrei Andreevich Markov) ได้เสนอวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณที่มีรากฐานมาจากทฤษฎีการประมาณเชิงเส้นโดยมีหลักการคือทำให้ผลรวมกำลังสองของความคลาดเคลื่อน (Sum of Square Error (SSE)) มีค่าน้อยที่สุด ซึ่งมีรายละเอียดดังนี้ จาก (2.1) สามารถเขียนสมการรูปเมทริกซ์ได้ดังนี้

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

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

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

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

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

n เป็นขนาดตัวอย่างของตัวแปรอิสระ

และ p เป็นจำนวนพจน์พหุนามในตัวแบบโดยไม่นับพจน์ค่าคงที่

โดยมีข้อกำหนดว่าค่าลำดับชั้น(rank) ของเมทริกซ์ X เท่ากับ $p+1$ เมื่อ $p+1<n$ ความคลาดเคลื่อนเป็นตัวแปรสุ่มที่เป็นอิสระกันและมีการแจกแจงเดียวกันที่มีค่าเฉลี่ย $E(\underset{\sim}{\varepsilon})=\underset{\sim}{0}$ และเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วม $E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}')=\sigma^2 I_n$

ให้ $\hat{\underset{\sim}{\beta}}$ เป็นเวกเตอร์ของตัวประมาณกำลังสองน้อยสุดของพารามิเตอร์ $\underset{\sim}{\beta}$

และ $\underset{\sim}{e}$ เป็นเวกเตอร์ของเศษเหลือ (Residuals) ที่เป็นตัวประมาณของความคลาดเคลื่อน $\underset{\sim}{\varepsilon}$ เมื่อแทนที่ $\underset{\sim}{\beta}$ ด้วย $\hat{\underset{\sim}{\beta}}$ และแทนที่ $\underset{\sim}{\varepsilon}$ ด้วย $\underset{\sim}{e}$ ในตัวแบบ (2.2) จะได้ว่า

(2.3) $$\underset{\sim}{y}=X\hat{\underset{\sim}{\beta}}+\underset{\sim}{e}$$

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

จากตัวแบบในสมการที่ (2.3) จะได้ผลรวมกำลังสองของความคลาดเคลื่อนอยู่ในรูปแบบดังนี้

$$\begin{aligned} SSE &= (\underset{\sim}{e}'\underset{\sim}{e}) \\ &= (\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}})'(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}) \\ &= \underset{\sim}{y}'\underset{\sim}{y}-\hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y}-\underset{\sim}{y}'X\hat{\underset{\sim}{\beta}}+\hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}} \\ &= \underset{\sim}{y}'\underset{\sim}{y}-2\hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y}+\hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}} \end{aligned}$$

หาค่า $\hat{\underset{\sim}{\beta}}$ ที่ทำให้ผลรวมกำลังสองของความคลาดเคลื่อนมีค่าต่ำที่สุด โดยการหาค่าอนุพันธ์ (Differentiate) ของ SSE เทียบกับค่า $\hat{\underset{\sim}{\beta}}$ แล้วกำหนดให้เท่ากับ $\underset{\sim}{0}$ จะได้ว่า

$$\frac{\partial}{\partial\hat{\underset{\sim}{\beta}}}\left(\underset{\sim}{y}'\underset{\sim}{y}-2\hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y}+\hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}}\right)=\underset{\sim}{0}$$

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

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

(2.4) $$\hat{\underset{\sim}{\beta}}=\left(X'X\right)^{-1}X'\underset{\sim}{y}$$

จากตัวประมาณค่าสัมประสิทธิ์ความถดถอยจากสมการ (2.4) มีคุณสมบัติที่ไม่เอนเอียงของ $\underset{\sim}{\beta}$ กล่าวคือ

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

ตัวประมาณ $\underset{\sim}{\beta}$ ที่ได้เป็นตัวประมาณไม่เอนเอียงและมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยสุด แต่ในการประมาณ $\underset{\sim}{\beta}$ ด้วยวิธีกำลังสองน้อยสุด มีข้อจำกัดข้อหนึ่งคือ ตัวแปรอิสระจะต้องไม่มีความสัมพันธ์กัน แต่ในทางปฏิบัติเป็นไปได้น้อยมากจึงทำให้สมาชิกของเมทริกซ์สหสัมพันธ์ของตัวแปรอิสระเกิดเงื่อนไขที่ไม่ดี (ill-conditioned) ทำให้การประมาณค่า $\underset{\sim}{\beta}$ ด้วยวิธีกำลังสองน้อยสุดไม่สามารถให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าต่ำสุดได้ ดังนั้นจึงควรพิจารณาคุณสมบัติของตัวประมาณค่า $\underset{\sim}{\beta}$ สองประการ คือ เมทริกซ์ความแปรปรวนร่วมของตัวประมาณ $\hat{\underset{\sim}{\beta}}$ และค่าเฉลี่ยยกกำลังสองของความแตกต่างระหว่าง $\hat{\underset{\sim}{\beta}}$ และ $\underset{\sim}{\beta}$ ซึ่งสามารถเขียนให้อยู่ในรูปเมทริกซ์ $X'X$ และ σ^2 ดังต่อไปนี้

กำหนดให้ความแปรปรวนร่วมของตัวประมาณ $\hat{\underset{\sim}{\beta}}$ มีค่าเป็น $\text{cov}(\hat{\underset{\sim}{\beta}})$ ดังนี้

$$\text{cov}(\hat{\underset{\sim}{\beta}}) = \sigma^2\left(X'X\right)^{-1} \tag{2.5}$$

ให้ $d(\hat{\underset{\sim}{\beta}}, \underset{\sim}{\beta})$ เป็นระยะทางระหว่าง $\hat{\underset{\sim}{\beta}}$ และ $\underset{\sim}{\beta}$ ดังนั้น

$$\left\{d(\hat{\underset{\sim}{\beta}}, \underset{\sim}{\beta})\right\}^2 = (\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})'(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})$$

และเราจะได้ค่าเฉลี่ยยกกำลังสองของระยะทางระหว่าง $\hat{\underset{\sim}{\beta}}$ และ $\underset{\sim}{\beta}$ จะอยู่ในรูปของ

$$E\left[d(\hat{\underset{\sim}{\beta}}, \underset{\sim}{\beta})\right]^2 = \sigma^2 trace\left(X'X\right)^{-1} \tag{2.6}$$
$$= E\left[(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})'(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})\right]$$
$$= E\left[\hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}} - 2\hat{\underset{\sim}{\beta}}'\underset{\sim}{\beta} + \underset{\sim}{\beta}'\underset{\sim}{\beta}\right]$$
$$= E\left[\hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}}\right] - \underset{\sim}{\beta}'\underset{\sim}{\beta}$$

ค่าของ $E\left[d(\underset{\sim}{\hat{\beta}},\underset{\sim}{\beta})\right]^2$ จะมีค่าสมมูล (equivalent) กับค่าเฉลี่ยของ $\underset{\sim}{\hat{\beta}}$ และ $\underset{\sim}{\beta}$ ซึ่งสามารถเขียนให้อยู่ในรูปของ

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

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

(2.7) $$Var\left[d(\underset{\sim}{\hat{\beta}},\underset{\sim}{\beta})\right]^2=2\sigma^4 trace\left(X'X\right)^{-2}$$

จะเห็นได้ว่าจากสมการที่ (2.5) (2.6) และ (2.7) ต่างก็อยู่ในรูปของฟังก์ชันเมทริกซ์ $X'X$ ดังนั้นเพื่อความสะดวกในการเปรียบเทียบการประมาณค่า $\underset{\sim}{\beta}$ จึงควรแปลงค่าเมทริกซ์ $\left(X'X\right)^{-1}$ ให้อยู่ในรูปฟังก์ชันของค่าเฉพาะ (eigenvalue) ของเมทริกซ์ $X'X$ โดยใช้คุณสมบัติข้อหนึ่งของค่าเฉพาะในเมทริกซ์ $X'X$ กล่าวคือ ค่า λ_j เป็นค่าเฉพาะของเมทริกซ์ $X'X$ เมื่อ $j=0,\ 1,\ldots,\ p$ และ $\sum_{j=0}^{p}\lambda_j=trace\left(X'X\right)$ เมื่อ p เป็นพจน์พหุนามในตัวแบบโดยไม่นับพจน์ค่าคงที่

สมมติให้ค่าเฉพาะของเมทริกซ์ $X'X$ มีค่าเป็น

$$\left(\lambda_{\max}=\lambda_0\right)\geq\lambda_1\geq\lambda_2\geq\ldots\geq\left(\lambda_p=\lambda_{\min}\right)\geq 0$$

จากสมการ (2.6) ค่าเฉลี่ยยกกำลังสองของระยะทางระหว่าง $\underset{\sim}{\hat{\beta}}$ และ $\underset{\sim}{\beta}$ อาจเขียนอยู่ในรูปฟังก์ชันของค่าเฉพาะของเมทริกซ์ $X'X$ ดังนี้

(2.8) $$E\left[d(\underset{\sim}{\hat{\beta}},\underset{\sim}{\beta})\right]^2=\sigma^2\sum_{j=0}^{p}\left(\frac{1}{\lambda_j}\right)$$

จากสมการ (2.7) ค่าความแปรปรวนยกกำลังสองของระยะทางระหว่าง $\underset{\sim}{\hat{\beta}}$ และ $\underset{\sim}{\beta}$ อาจเขียนอยู่ในรูปฟังก์ชันของค่าเฉพาะของเมทริกซ์ $X'X$ ดังนี้

(2.9) $$Var\left[d(\underset{\sim}{\hat{\beta}},\underset{\sim}{\beta})\right]^2=2\sigma^4\sum_{j=0}^{p}\left(\frac{1}{\lambda_j}\right)^2$$

แต่ในกรณีที่ตัวแปรอิสระมีสภาพไม่เหมาะสม ค่าเฉพาะของเมทริกซ์ $X'X$ บางตัวจะมีค่าน้อยมากๆทำให้ระยะทางระหว่าง $\underset{\sim}{\hat{\beta}}$ ไปยัง $\underset{\sim}{\beta}$ มีค่ามาก ดังนั้นค่าจากสมการ (2.8) และ (2.9) ซึ่งก็คือค่า $E\left[d(\underset{\sim}{\hat{\beta}},\underset{\sim}{\beta})\right]^2$ และ $Var\left[d(\underset{\sim}{\hat{\beta}},\underset{\sim}{\beta})\right]^2$ จะมีค่าสูงตามไปด้วย ตัวประมาณ $\underset{\sim}{\hat{\beta}}$ จึงเป็นตัวประมาณที่ไม่เหมาะสม

2.2 การประมาณค่าสัมประสิทธิ์ด้วยวิธีการ์ร็อตที่ไม่เป็นลบ (NG)

ใน ค.ศ. 1995 ลีโอ บลีแมน (Leo Breiman) ได้เสนอวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณในการแก้ปัญหาการเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ โดยนำข้อดีของวิธีกำลังสองน้อยสุดและได้อาศัยวิธีการหด (shrinkage method) เข้ามาช่วยประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีการ์ร็อตที่ไม่เป็นลบ โดยตัวประมาณสัมประสิทธิ์ความถดถอยแต่ละตัวจะได้จากค่าคงที่ c แต่ละตัวคูณกับสัมประสิทธิ์จากวิธีกำลังสองน้อยสุดแต่ละตัว ซึ่งสามารถเขียนได้อยู่ในรูป

$$\hat{\beta}_{NG_j} = c_j \hat{\beta}_j \quad ; \; j = 0,1,\ldots,p$$

เมื่อ $\hat{\beta}_j$ เป็นตัวประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุด
และ c_j เป็นพารามิเตอร์ที่หดตัว (shrinkage parameter)

ค่าของ c_j แต่ละค่าหาได้จากรูปแบบของ

(2.10) $$y_i = \sum_{j=0}^{p} c_j \beta_j X_{ji} + \varepsilon_i \quad ; \; i = 1,2,\ldots,n$$

จากสมการ (2.10) จะได้ผลรวมค่าคลาดเคลื่อนกำลังสองอยู่ในรูปแบบดังนี้

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \sum_{j=0}^{p} c_j \beta_j X_{ij})^2$$

นำหลักการของการหาค่าสัมประสิทธิ์ของวิธีกำลังสองน้อยที่สุดมาใช้ กล่าวคือจะทำให้ผลรวมของค่าคลาดเคลื่อนกำลังสองมีค่าต่ำที่สุด ภายใต้เงื่อนไข $c_j \geq 0$ สำหรับทุกค่า j และ $\sum_{j=0}^{p} c_j = s$

การประมาณค่าพารามิเตอร์ที่หดตัว c

จากตัวแบบในสมการ (2.2) สามารถเขียนให้อยู่ในรูปแบบใหม่โดยใช้เมทริกซ์เชิงตั้งฉาก (Orthogonal matrix) ได้ดังนี้

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

เมื่อ $Z = XQ$, $\underset{\sim}{\alpha} = Q' \underset{\sim}{\beta}$ และ $Q'X'XQ = Z'Z = \Lambda$

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

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

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

Q เป็นเมทริกซ์เชิงตั้งฉากขนาด $(p+1) \times (p+1)$ โดยที่แต่ละแนวตั้งของเมทริกซ์คือ เวกเตอร์เจาะจง (eigenvector) ที่สอดคล้องกับค่าเจาะจง (eigenvalue) $\lambda_0, \lambda_1, \ldots, \lambda_p$ ตามลำดับ

Z เป็นเมทริกซ์ซึ่งแต่ละแนวตั้งของ Z คือตัวแปรอิสระชุดใหม่ซึ่งเป็นผลบวกเชิงเส้น (linear combination) ของตัวแปรอิสระ

และ Λ เป็นเมทริกซ์ทแยงมุมขนาด $(p+1) \times (p+1)$ โดยที่สมาชิกในแนวเส้นทแยงมุมเป็นค่าเจาะจงของเมทริกซ์ $X'X$

จากสมการ (2.11) เราสามารถใช้วิธีกำลังสองน้อยสุดเพื่อประมาณค่า $\underset{\sim}{\alpha}$ จะได้ว่า

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

และค่าของ c_j แต่ละค่าหาได้จากรูปแบบของ

$$y_i = \sum_{j=0}^{p} c_j \alpha_j Z_{ij} + \varepsilon_i \quad ; \; i = 1, 2, \ldots, n$$

กำหนดให้ $$P_{ij} = \alpha_j Z_{ij}$$

จะได้ว่า

$$y_i = \sum_{j=0}^{p} c_j P_{ij} + \varepsilon_i$$

หรือสามารถเขียนให้อยู่ในรูปเมทริกซ์

(2.12) $$\underset{\sim}{y} = P \underset{\sim}{c} + \underset{\sim}{\varepsilon}$$

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

P เป็นเมทริกซ์ของเมทริกซ์ซึ่งแต่ละแนวตั้งของ P คือตัวแปรอิสระชุดใหม่ซึ่งเป็นผลบวกเชิงเส้นของ Z คูณกับตัวประมาณกำลังสองน้อยสุดในตัวแบบเชิงตั้งฉากขนาด $n \times (p+1)$

$\underset{\sim}{c}$ เป็นเวกเตอร์ของค่าพารามิเตอร์ที่หดตัวขนาด $(p+1) \times 1$

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

n เป็นขนาดตัวอย่างของตัวแปรอิสระ

และ p เป็นจำนวนพจน์พหุนามในตัวแบบโดยไม่นับพจน์ค่าคงที่ โดยที่ $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ และ $\operatorname{cov}(\underset{\sim}{\varepsilon}) = \sigma^2 I_n$

จากสมการ (2.12) จะได้ผลรวมค่าคลาดเคลื่อนกำลังสองอยู่ในรูปแบบดังนี้

$$
\begin{aligned}
\underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon} &= (\underset{\sim}{y} - P\underset{\sim}{c})'(\underset{\sim}{y} - P\underset{\sim}{c}) \\
&= \underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{c}'P'\underset{\sim}{y} - \underset{\sim}{y}'P\underset{\sim}{c} + \underset{\sim}{c}'P'P\underset{\sim}{c} \qquad (2.13) \\
&= \underset{\sim}{y}'\underset{\sim}{y} - 2\underset{\sim}{c}'P'\underset{\sim}{y} + \underset{\sim}{c}'P'P\underset{\sim}{c}
\end{aligned}
$$

ผู้วิจัยได้นำหลักการของการหาค่าสัมประสิทธิ์ของวิธีกำลังสองน้อยสุดมาใช้ กล่าวคือจะทำให้ผลรวมของค่าคลาดเคลื่อนกำลังสองมีค่าต่ำที่สุดภายใต้เงื่อนไข $c_j \geq 0$ สำหรับทุกค่า j และ $\sum_{j=0}^{p} c_j = s$ โดยจะได้ค่าประมาณสัมประสิทธิ์ความถดถอยพหุนามจากการหาอนุพันธ์อันดับที่ 1 ของสมการ (2.13) เทียบกับค่า $\underset{\sim}{c}$ และให้เท่ากับศูนย์ ซึ่งผลที่ได้จะอยู่ในรูปแบบดังนี้

$$\frac{\partial}{\partial \underset{\sim}{c}}[-2\underset{\sim}{c}'P'\underset{\sim}{y} + \underset{\sim}{c}'P'P\underset{\sim}{c} + 2\lambda^*\underset{\sim}{c}] = \underset{\sim}{0}$$

$$P'P\underset{\sim}{c} = P'\underset{\sim}{y} - \lambda^*$$

กล่าวคือ

$$
\begin{aligned}
\alpha_j^2\lambda_j c_j &= \alpha_j \underset{\sim}{Z}_j' \underset{\sim}{y} - \lambda^* \\
&= \alpha_j^2\lambda_j - \lambda^* \\
c_j &= 1 - \frac{\lambda^*}{\alpha_j^2\lambda_j}
\end{aligned}
$$

จากเงื่อนไข $c_j \geq 0$ สำหรับทุกค่า j จะได้ว่า

$$c_j = \left(1 - \frac{\lambda^*}{\alpha_j^2\lambda_j}\right)^+$$

เมื่อ $\left(1 - \frac{\lambda^*}{\alpha_j^2\lambda_j}\right)^+$ แสดงเฉพาะส่วนที่เป็นค่าบวกของ $\left(1 - \frac{\lambda^*}{\alpha_j^2\lambda_j}\right)$

จะได้ว่าตัวประมาณ $\hat{c}_j$ อยู่ในรูปของ

$$\hat{c}_j = \left(1 - \frac{\lambda^*}{\hat{\alpha}_j^2\lambda_j}\right)^+$$

ในปี ค.ศ.2006 เลียลา โมฮัมมาดี (Leila Mohammadi) และซาร่า เอ เวน เด เกียร์ (Sara A. van de Geer) ได้ศึกษาตัวประมาณการ์ร็อตที่ไม่เป็นลบในตัวแบบถดถอยเชิงเส้น โดยพบว่าเมื่อกำหนดให้ค่า $\lambda^* = c^*\sqrt{\frac{\log n}{n}}$ และ $c^* > \sqrt{8}$ จะทำให้ตัวประมาณการ์ร็อตที่ไม่เป็นลบจะเป็นตัว

ประมาณที่คงเส้นคงวา (consistency) และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจะมีค่าลู่เข้าสู่ศูนย์ ในงานวิจัยครั้งนี้ผู้วิจัยจึงเลือกใช้ค่า $c^* = 3$ เพื่อป้องกันปัญหาการคำนวณเลขทศนิยมและเป็นค่าที่ใกล้เคียงกับ $\sqrt{8}$ มากที่สุด

ดังนั้นตัวประมาณสัมประสิทธิ์ความถดถอยของวิธีการ์ร็อตที่ไม่เป็นลบจะมีลักษณะดังนี้

(2.14)
$$\hat{\alpha}_{NG_j} = \hat{c}_j \hat{\alpha}_j \qquad ; \quad j = 0, 1, \ldots, p$$
$$= \left(1 - \frac{\lambda^*}{\hat{\alpha}_j^2 \lambda_j}\right)^+ \hat{\alpha}_j$$

พิจารณาจากคุณสมบัติ $\underset{\sim}{\alpha} = Q' \underset{\sim}{\beta}$ เราสามารถเขียนสมการได้เป็น $\underset{\sim}{\beta} = Q \underset{\sim}{\alpha}$ เนื่องจาก Q เป็นเมทริกซ์ตั้งฉาก ดังนั้น $\hat{\beta}_{\sim NG}$ จะอยู่ในรูปของสมการข้างล่างนี้

(2.15)
$$\hat{\beta}_{\sim NG} = Q \hat{\alpha}_{\sim NG}$$

2.3 การประมาณค่าสัมประสิทธิ์ด้วยวิธีแจ็คไนฟริดจ์ (JR)

ในปี ค.ศ.1956 คิวนูลี (Quenouille) ได้เสนอวิธีแจ็คไนฟ (jackknife method) ซึ่งเป็นวิธีการที่ช่วยลดความเอนเอียงของตัวประมาณ โดยมีหลักการคือนำตัวประมาณเริ่มต้น (original estimator) ลบด้วยความเอนเอียงของตัวประมาณนั้นและเรียกว่าตัวประมาณนี้ว่าตัวประมาณแจ็คไนฟซึ่งเขียนได้ในรูปแบบ ดังต่อไปนี้

$$\hat{\beta}_{\sim JAC} = \underset{\sim}{\hat{\beta}} - \underset{\sim}{\hat{b}}$$

เมื่อ $\underset{\sim}{\hat{\beta}}$ เป็นตัวประมาณเริ่มต้น

และ $\underset{\sim}{\hat{b}}$ เป็นความเอนเอียงของตัวประมาณ $\underset{\sim}{\hat{\beta}}$

ในปี ค.ศ. 1970 โฮเอิล (Hoerl) และเคนนาร์ด (Kennard) ได้ศึกษาหาตัวประมาณค่าพารามิเตอร์ในตัวแบบพหุคูณเชิงเส้นที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าตัวประมาณกำลังสองน้อยสุดเมื่อข้อมูลเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ โดยตัวประมาณนี้สร้างจากหลักการนำค่าคงที่ค่าหนึ่งที่มากกว่าศูนย์ (k) มาบวกกับสมาชิกทุกตัวในแนวเส้นทแยงมุมของเมทริกซ์ $X'X$ ซึ่งทำให้เมทริกซ์ $X'X$ ลดสภาพความเป็นเมทริกซ์ที่ไม่ดี (ill-conditioned) ได้แล้วจะหาอินเวอร์สของเมทริกซ์ดังกล่าวต่อไป

สมการปกติ (Normal Equation) ของตัวประมาณริดจ์สามารถเขียนได้ในรูปแบบของ

$$\left(X'X + kI_{p+1} \right) \hat{\underset{\sim}{\beta}}_R = X' \underset{\sim}{y}$$

ดังนั้น ตัวประมาณริดจ์ (Ridge Estimator) เขียนได้ในรูปของ

$$\hat{\underset{\sim}{\beta}}_R = \left(X'X + kI_{p+1} \right)^{-1} X' \underset{\sim}{y} \quad ; \; k > 0$$

เมื่อ k เป็นค่าคงที่หรือเรียกว่าเป็นพารามิเตอร์เอนเอียง (biased parameter)

และ I_{p+1} เป็นเมทริกซ์เอกลักษณ์ (identity matrix) ขนาด $(p+1)\times(p+1)$

ในปี ค.ศ.2003 เจอเกน กรอฟ (Jurgen Grob) ได้เสนอวิธีแจ็คไนฟในตัวประมาณริดจ์ ดังนี้ จากตัวแบบในสมการ(2.2) จะได้ว่า

(2.16) $$E_{(i)} \underset{\sim}{y} = E_{(i)} X \underset{\sim}{\beta} + E_{(i)} \underset{\sim}{\varepsilon}$$

เมื่อ $E_{(i)}$ เป็นเมทริกซ์ขนาด $(n-1)\times n$ ที่เกิดจากการตัดแถวที่ i จากเมทริกซ์เอกลักษณ์ขนาด $n \times n$

โดยมีข้อกำหนดว่าค่าลำดับชั้น(rank) ของเมทริกซ์ $E_{(i)}X$ เท่ากับ p ความคลาดเคลื่อนเป็นตัวแปรสุ่มที่เป็นอิสระกันและมีการแจกแจงเดียวกันที่มีค่าเฉลี่ย $E(E_{(i)} \underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ และเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วม $Cov(E_{(i)} \underset{\sim}{\varepsilon}) = \sigma^2 I_{n-1}$

ดังนั้นตัวประมาณริดจ์จากสมการ (2.16) จะมีลักษณะ ดังนี้

$$\hat{\underset{\sim}{\beta}}_{R,(i)} = (X'E'_{(i)}E_{(i)}X + kI_p)^{-1} X'E'_{(i)}E_{(i)} \underset{\sim}{y} \quad ; \quad k > 0$$

ค่าเฉลี่ยของ $\hat{\underset{\sim}{\beta}}_{R,(i)}$ คือ

$$\hat{\underset{\sim}{\beta}}_{R,(\bullet)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\underset{\sim}{\beta}}_{R,(i)}$$

และค่าประมาณของความเอนเอียงของ $\hat{\underset{\sim}{\beta}}_{R,(i)}$ มีค่าเป็น

$$\hat{\underset{\sim}{b}} = (n-1)(\hat{\underset{\sim}{\beta}}_{R,(\bullet)} - \hat{\underset{\sim}{\beta}}_R)$$

ดังนั้นตัวประมาณแจ็คไนฟริดจ์ (Jackknifed Ridge Estimator) เขียนได้ในรูปของ

$$
\begin{aligned}
\underset{\sim}{\hat{\beta}}_{JR} &= \underset{\sim}{\hat{\beta}}_{R} - \underset{\sim}{\hat{b}} \\
&= \underset{\sim}{\hat{\beta}}_{R} - [(n-1)(\underset{\sim}{\hat{\beta}}_{R,(\bullet)} - \underset{\sim}{\hat{\beta}}_{R})] \\
&= \underset{\sim}{\hat{\beta}}_{R} - n\,\underset{\sim}{\hat{\beta}}_{R,(\bullet)} + \underset{\sim}{\hat{\beta}}_{R,(\bullet)} + n\,\underset{\sim}{\hat{\beta}}_{R} - \underset{\sim}{\hat{\beta}}_{R} \\
&= (n-1)\,\underset{\sim}{\hat{\beta}}_{R,(\bullet)} + n\underset{\sim}{\hat{\beta}}_{R}
\end{aligned}
\tag{2.17}
$$

ในการวิเคราะห์ความถดถอยพหุคูณเมื่อข้อมูลมีพหุสัมพันธ์นั้นการเลือกค่า k ที่เหมาะสมจะสามารถทำให้ตัวประมาณริดจ์มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าตัวประมาณกำลังสองน้อยสุดได้ แต่ในหลายกรณีเราไม่ทราบค่า k ที่เหมาะสมว่าควรมีค่าเป็นเท่าไรซึ่งมีนักวิจัยได้เสนอวิธีการหาค่า k ที่เหมาะสมหลายวิธีด้วยกัน โดยเริ่มต้นในปี ค.ศ. 1970 โฮเอิล (Hoerl) และเคนนาร์ด (Kennard) ใช้เทคนิคการหาค่า k โดยการพิจารณาจากกราฟ (graphical technique) หรือเรียกว่า Ridge Trace แต่เป็นวิธีที่อาจได้คำตอบที่ไม่แน่นอนเพราะขึ้นอยู่กับการเลือกกำหนดของผู้ใช้ ดังนั้นจึงมีผู้คิดค้นวิธีการหาค่า k ที่เหมาะสมโดยไม่ใช้กราฟ (nongraphical technique) ไว้หลายวิธี ซึ่งในปี ค.ศ. 1981 กิบบอนส์ (Gibbons) ได้ทำการวิจัยเพื่อเปรียบเทียบวิธีการต่างๆในการหาค่า k ที่เหมาะสม ผลการวิจัยพบว่าโดยภาพรวมแล้ววิธีการประมาณค่า k ที่คิดค้นโดยโฮเอิล (Hoerl) เคนนาร์ด (Kennard) และ บลาดวิน (Bladwin) ในปี ค.ศ. 1975 ซึ่งแทนด้วย k_{HKB} เป็นวิธีการหนึ่งที่มีประสิทธิภาพที่ดีในการประมาณค่า และเป็นวิธีที่ผู้วิจัยได้นำมาเปรียบเทียบในการวิเคราะห์ครั้งนี้ ซึ่งเขียนได้ในรูปแบบ ดังต่อไปนี้

$$k_{HKB} = \frac{(p+1)s^2}{\underset{\sim}{\hat{\beta}}'\,\underset{\sim}{\hat{\beta}}}$$

เมื่อ $s^2 = \dfrac{(\underset{\sim}{y} - \underset{\sim}{X}\,\underset{\sim}{\hat{\beta}})'(\underset{\sim}{y} - \underset{\sim}{X}\,\underset{\sim}{\hat{\beta}})}{n-(p+1)}$ เป็นค่าประมาณพารามิเตอร์ σ^2

2.4 การประมาณค่าสัมประสิทธิ์ด้วยวิธีลิวไทป์ (LT)

ในปี ค.ศ.1993 คีเจียน ลิว (Kejian Liu) ได้เสนอวิธีประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณในกรณีเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระโดยนำหลักการของวิธีริดจ์รีเกรสชันและวิธีของสไตน์มาผสมผสานกัน กล่าวคือการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธีริดจ์มีประสิทธิภาพในทางปฏิบัติแต่ประสบความยุ่งยากในการหาค่าพารามิเตอร์ k ที่เหมาะสม ส่วนค่าประมาณสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธีของสไตน์ที่อยู่ในรูปของ $\underset{\sim}{\hat{\beta}}_S = c\,\underset{\sim}{\hat{\beta}}$; $0 < c < 1$ เป็นฟังก์ชันเชิงเส้นของค่าคงที่ c ดังนั้นการคำนวณหาค่า c ที่เหมาะสมจึงไม่

ยุ่งยากมากนัก แต่อัตราส่วนการลดลงของค่าสัมประสิทธิ์ความถดถอยแต่ละตัวเท่ากันซึ่งไม่เหมาะสมในทางปฏิบัติ ดังนั้นข้อดีของการประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณด้วยวิธีนี้คือ ตัวประมาณสัมประสิทธิ์ความถดถอยพหุคูณเป็นฟังก์ชันเชิงเส้นของ d ทำให้การคำนวณหาค่า d ที่เหมาะสมสะดวกกว่าการหาค่า k ที่เหมาะสมจากวิธีริดจ์ นอกจากนี้ตัวประมาณสัมประสิทธิ์ความถดถอยพหุคูณวิธีนี้มีประสิทธิภาพในทางปฏิบัติเมื่อตัวแปรอิสระมีพหุสัมพันธ์กันซึ่งมีรูปแบบของตัวประมาณ คือ

(2.18) $$\hat{\underset{\sim}{\beta}}_L = (X'X + I)^{-1}(X'\underset{\sim}{y} + d\,\hat{\underset{\sim}{\beta}}) \quad ; \quad 0 < d < 1$$

ต่อมาในปี ค.ศ.1995-2001 ฟิคริ อัคเดนิส (Fikri Akdeniz) และเซลาฮัททิน คาซิรานลา (Selahattin Kaciranlar) ได้เรียกตัวประมาณ $\hat{\underset{\sim}{\beta}}_L$ ว่าตัวประมาณลิว (Liu estimator) พร้อมทั้งตรวจสอบคุณสมบัติทางสถิติ

ในปี ค.ศ.2003-2004 คีเจียน ลิว (Kejian Liu) ได้ศึกษาตัวแบบความถดถอยเชิงเส้นเมื่อเกิดพหุสัมพันธ์ขั้นรุนแรงพบว่าการใช้พารามิเตอร์ที่หดตัว (shrinkage parameter) จากวิธีริดจ์เพียงตัวเดียวอาจจะไม่สามารถแก้ปัญหาสภาพความเป็นเมทริกซ์ที่ไม่ดี (ill-conditioned) ได้ทั้งหมดจึงเสนอตัวประมาณใหม่โดยใช้พารามิเตอร์ที่หดตัว 2 ตัวขึ้นมาและเรียกตัวประมาณนี้ว่า ตัวประมาณลิวไทป์ (Liu-Type estimator) ซึ่งเป็นตัวประมาณที่สามารถแก้ปัญหาสภาพความเป็นเมทริกซ์ที่ไม่ดีได้อย่างสมบูรณ์แบบ โดยมีรูปแบบดังนี้

(2.19) $$\hat{\underset{\sim}{\beta}}_{LT} = (X'X + kI)^{-1}(X'\underset{\sim}{y} - d\,\hat{\underset{\sim}{\beta}}) \quad ; \quad k > 0,\ -\infty < d < \infty$$

การประมาณค่าพารามิเตอร์ที่หดตัว (shrinkage parameter) d และ k

จากสมการ (2.11) เราสามารถใช้วิธีกำลังสองน้อยสุดเพื่อประมาณค่า $\underset{\sim}{\alpha}$ จะได้ว่า

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

เมทริกซ์ความแปรปรวนร่วมของ $\underset{\sim}{\alpha}$ มีค่าเป็น

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

ในทำนองเดียวกันจะได้ตัวประมาณริดจ์ คือ

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

และจากสมการ (2.19) ตัวประมาณลิว ไทป์ คือ

$$\hat{\underset{\sim}{\alpha}}_{LT} = (Z'Z + kI)^{-1}(Z'\underset{\sim}{y} - d\,\hat{\underset{\sim}{\alpha}})$$

(2.20)

$$= (\Lambda + kI)^{-1}(Z'\underset{\sim}{y} - d\,\hat{\underset{\sim}{\alpha}})$$

พิจารณาคุณสมบัติ $\underset{\sim}{\alpha} = Q'\underset{\sim}{\beta}$ เราสามารถเขียนสมการได้เป็น $\underset{\sim}{\beta} = Q\underset{\sim}{\alpha}$ เนื่องจาก Q เป็นเมทริกซ์ตั้งฉาก ดังนั้น $\hat{\underset{\sim}{\beta}}$ จะอยู่ในรูปของสมการข้างล่างนี้

(2.21) $$\hat{\underset{\sim}{\beta}} = Q\hat{\underset{\sim}{\alpha}}, \quad \hat{\underset{\sim}{\beta}}_R = Q\hat{\underset{\sim}{\alpha}}_R \text{ และ } \hat{\underset{\sim}{\beta}}_{LT} = Q\hat{\underset{\sim}{\alpha}}_{LT}$$

จากสมการ (2.21) แสดงว่าทุกๆค่าของ $\hat{\underset{\sim}{\alpha}}$ เราสามารถหาค่า $\hat{\underset{\sim}{\beta}}$ ได้ นั่นคือคุณสมบัติของ $\hat{\underset{\sim}{\beta}}$ จะสอดคล้องกับ $\hat{\underset{\sim}{\alpha}}$ ดังนั้นเมื่อเราศึกษาคุณสมบัติของ $\hat{\underset{\sim}{\alpha}}$ ก็จะสามารถทราบถึงคุณสมบัติของ $\hat{\underset{\sim}{\beta}}$ ได้เช่นเดียวกัน ในกรณีนี้จึงศึกษาคุณสมบัติของ $\hat{\underset{\sim}{\beta}}_{LT}$ โดยอาศัยคุณสมบัติของ $\hat{\underset{\sim}{\alpha}}_{LT}$

ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ $\hat{\underset{\sim}{\beta}}_{LT}$ จะอยู่ในรูปของ

$$MSE(\hat{\underset{\sim}{\alpha}}_{LT}) = trace[\text{cov}(\hat{\underset{\sim}{\alpha}}_{LT})] + \hat{\underset{\sim}{b}}^2(\hat{\underset{\sim}{\alpha}}_{LT})$$

เมื่อ $\hat{\underset{\sim}{b}}(\hat{\underset{\sim}{\alpha}}_{LT})$ เป็นความเอนเอียงของตัวประมาณ $\hat{\underset{\sim}{\alpha}}_{LT}$

และ $\text{cov}(\hat{\underset{\sim}{\alpha}}_{LT})$ เป็นความแปรปรวนร่วมของตัวประมาณ $\hat{\underset{\sim}{\alpha}}_{LT}$

โดยที่ $\hat{\underset{\sim}{b}}(\hat{\underset{\sim}{\alpha}}_{LT}) = E(\hat{\underset{\sim}{\alpha}}_{LT}) - \underset{\sim}{\alpha}$

$E(\hat{\underset{\sim}{\alpha}}_{LT}) = (\Lambda + kI)^{-1}(\Lambda - dI)\underset{\sim}{\alpha}$

และ $\text{cov}(\hat{\underset{\sim}{\alpha}}_{LT}) = \sigma^2(\Lambda + kI)^{-1}(\Lambda - dI)\Lambda^{-1}(\Lambda - dI)(\Lambda + kI)^{-1}$

จะได้ว่า

(2.22) $$MSE(\hat{\underset{\sim}{\alpha}}_{LT}) = \sigma^2 \sum_{j=0}^{p} \frac{(d+k)^2 \alpha_j^{\,2}}{(\lambda_j + k)^2} + \sigma^2 \sum_{j=0}^{p} \frac{(d - \lambda_j)^2}{\lambda_j(\lambda_j + k)^2}$$

เราสามารถหาค่าพารามิเตอร์ที่หดตัว d โดยการหาอนุพันธ์อันดับที่ 1 ของ $MSE(\hat{\underset{\sim}{\alpha}}_{LT})$ เทียบกับค่า d และกำหนดให้ k เป็นค่าคงที่ดังนั้นค่า d มีค่าเท่ากับ

$$d_{opt} = \sum_{j=0}^{p} \frac{(\sigma^2 - k\alpha_j^{\,2})}{(\lambda_j + k)^2} \Big/ \sum_{j=0}^{p} \frac{(\lambda_j \alpha_j^{\,2} + \sigma^2)}{\lambda_j(\lambda_j + k)^2}$$

เนื่องจากในทางปฏิบัติ α_j และ σ^2 ไม่ทราบค่า เราจึงแทนค่าพารามิเตอร์ทั้งสองด้วยตัวประมาณ $\hat{\alpha}_{R_j}$ และ $\hat{\sigma}_R^2$ ตามลำดับ ดังนั้นค่าประมาณของ d ที่เหมาะสมซึ่งทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่มีค่าต่ำที่สุด คือ

$$\hat{d} = \sum_{j=0}^{p} \frac{(\hat{\sigma}_R^2 - \hat{k}\alpha_{Rj}^2)}{(\lambda_j + \hat{k})^2} \Bigg/ \sum_{j=0}^{p} \frac{(\lambda_j \alpha_{Rj}^2 + \hat{\sigma}_R^2)}{\lambda_j(\lambda_j + \hat{k})^2}$$

เมื่อ $\hat{\sigma}_R^2 = \dfrac{(\underset{\sim}{y} - X\underset{\sim}{\hat{\beta}}_R)'(\underset{\sim}{y} - X\underset{\sim}{\hat{\beta}}_R)}{n-(p+1)}$

การหาค่า k นั้นจะพิจารณาจาก condition number (CN) โดยในปี ค.ศ.1980 เบลสเล (Belsley) ได้ศึกษาเรื่อง CN ไว้และพบว่าค่า CN ประมาณ 10 จะไม่เกิดปัญหาการมีสหสัมพันธ์ในตัวแปรอิสระ ดังนั้นค่า k ที่เหมาะสม มีค่าเท่ากับ

$$\hat{k} = \frac{\lambda_0 - 100\lambda_p}{99}$$

จากสมการ (2.20) จะได้ว่า

(2.23) $$\underset{\sim}{\hat{\alpha}}_{LT} = (\Lambda + \hat{k}I)^{-1}(Z'\underset{\sim}{y} - \hat{d}\,\underset{\sim}{\hat{\alpha}}_R)$$

และจากคุณสมบัติในสมการที่ (2.21) จะได้ว่า

(2.24) $$\underset{\sim}{\hat{\beta}}_{LT} = Q\,\underset{\sim}{\hat{\alpha}}_{LT}$$

คุณสมบัติทางสถิติของตัวประมาณลิวไทป์

<u>ทฤษฎีที่ 2.1</u> $MSE(\underset{\sim}{\hat{\alpha}}_{LT}) \le MSE(\underset{\sim}{\hat{\alpha}}_R)$

สำหรับ $k > 0$ และ $d = \sum_{j=0}^{p} \frac{(\sigma^2 - k\alpha_j{}^2)}{(\lambda_j + k)^2} \Big/ \sum_{j=0}^{p} \frac{(\lambda_j\alpha_j{}^2 + \sigma^2)}{\lambda_j(\lambda_j + k)^2}$

<u>พิสูจน์</u> จากคุณสมบัติในสมการ (2.21) ถ้าเราสามารถพิสูจน์ได้ว่า $MSE(\underset{\sim}{\hat{\alpha}}_{LT}) \le MSE(\underset{\sim}{\hat{\alpha}}_R)$ ก็จะสามารถพิสูจน์ได้ว่า $MSE(\underset{\sim}{\hat{\beta}}_{LT}) \le MSE(\underset{\sim}{\hat{\beta}}_R)$

จากสมการ (2.22)

$$MSE(\underset{\sim}{\hat{\alpha}}_{LT}) = \sigma^2 \sum_{j=0}^{p} \frac{(d+k)^2 \alpha_j^{\,2}}{(\lambda_j + k)^2} + \sigma^2 \sum_{j=0}^{p} \frac{(d-\lambda_j)^2}{\lambda_j(\lambda_j + k)^2}$$
$$= g(d)$$

ค่า d ที่ทำให้ $MSE(\underset{\sim}{\hat{\alpha}}_{LT})$ มีค่าน้อยที่สุดคือ

$$d = \sum_{j=0}^{p} \frac{(\sigma^2 - k\alpha_j^{\,2})}{(\lambda_j + k)^2} \Big/ \sum_{j=0}^{p} \frac{(\lambda_j \alpha_j^{\,2} + \sigma^2)}{\lambda_j(\lambda_j + k)^2}$$

เมื่อ d = 0 จะได้ว่า

$$g(0) = MSE(\underset{\sim}{\hat{\alpha}}_{R})$$

จาก $g(d) \le g(0)$ นั่นก็คือ $MSE(\underset{\sim}{\hat{\alpha}}_{LT}) \le MSE(\underset{\sim}{\hat{\alpha}}_{R})$ จึงสรุปได้ว่า $MSE(\underset{\sim}{\hat{\beta}}_{LT}) \le MSE(\underset{\sim}{\hat{\beta}}_{R})$

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง เพื่อต้องการศึกษาและเปรียบเทียบวิธีประมาณค่าสัมประสิทธิ์ความถดถอยในตัวแบบพหุนามกรณีที่มีตัวแปรอิสระ 1 ตัวและกำลังสูงสุดของตัวแปรอิสระไม่เกิน 6 โดยใช้เกณฑ์ค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเป็นเกณฑ์ในการตัดสินใจ และใช้อัตราส่วนผลต่างของค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเป็นส่วนที่ใช้ประกอบในการพิจารณาการตัดสินใจ ซึ่งวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามที่นำมาใช้มี 4 วิธี ดังนี้

1. วิธีกำลังสองน้อยสุด (OLS)
2. วิธีการ์ร็อตที่ไม่เป็นลบ (NG)
3. วิธีแจ็คไนฟริดจ์ (JR)
4. วิธีลิวไทป์ (LT)

และข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากการจำลองด้วยเทคนิคมอนติคาร์โล(Monte Carlo Simulation Technique) ในการหาข้อสรุปของปัญหาที่ศึกษาตลอดจนทำการเขียนโปรแกรม MATLAB (Matrix Laboratory Version 7) ในการประมวลผลและวิเคราะห์ข้อมูล ซึ่งขั้นตอนในการวิจัยมีรายละเอียดดังนี้

3.1 แผนการทดลอง

ผู้วิจัยได้กำหนดสถานการณ์ต่างๆ สำหรับการวิจัยครั้งนี้ไว้ดังนี้

1. เลือกตัวอย่างอย่างสุ่มของ ε_i จากประชากรที่มีการแจกแจงเดียวกัน โดยการวิจัยในครั้งนี้สนใจศึกษาเฉพาะการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 2 4 6 8 และ 10

2. กำหนดขนาดตัวอย่างที่ใช้ในการศึกษาคือ 5p 10p 15p และ 20p
เมื่อ p คือจำนวนพจน์พหุนามในตัวแบบโดยไม่นับพจน์ค่าคงที่

3. ตัวแปรอิสระที่ใช้ในการศึกษาครั้งนี้เท่ากับ 1 ตัวแปร ที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 5 และความแปรปรวนเท่ากับ 4

4. กำลังสูงสุดของตัวแปรอิสระที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบถดถอยพหุนาม คือ 2 3 4 5 และ 6

3.2 ขั้นตอนการวิจัย

<u>ขั้นตอนในการวิจัยมีดังนี้</u>

1. สร้างข้อมูลที่ใช้ในงานวิจัย

1.1 สร้างความคลาดเคลื่อนสุ่มในตัวแปรตาม ($\underset{\sim}{\varepsilon}$) ที่มีการแจกแจงที่กำหนด และให้มีขนาดตัวอย่าง (n) ตามที่กำหนดไว้

1.2 สร้างข้อมูลตัวแปรอิสระ (X) ที่มีการแจกแจงตามที่กำหนด และให้มีกำลังสูงสุดของตัวแปรอิสระที่ใช้เป็นตัวแบบเต็มรูปเพื่อใช้ในการสร้างตัวแปรตาม

1.3 สร้างข้อมูลตัวแปรตาม (y) ที่ใช้สำหรับหาตัวประมาณค่าพารามิเตอร์ จากตัวแปรอิสระ (พจน์พหุนามของ X) และความคลาดเคลื่อนที่มีลักษณะการแจกแจงตามที่ต้องการศึกษา ($\underset{\sim}{\varepsilon}$) โดยให้ตัวแปรตามมีความสัมพันธ์เชิงเส้นในพารามิเตอร์กับตัวแปรอิสระ

2. การประมาณค่าพารามิเตอร์โดยวิธีทั้ง 4 วิธี ได้แก่

2.1 วิธีกำลังสองน้อยสุด (OLS)

2.2 วิธีการ์ร็อตที่ไม่เป็นลบ (NG)

2.3 วิธีแจ็คไนฟริดจ์ (JR)

2.4 วิธีลิวไทป์ (LT)

3. คำนวณหาค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณพารามิเตอร์ในแต่ละวิธี

4. สรุปผลในรูปตารางและรูปภาพ

<u>รายละเอียดของแต่ละขั้นตอนมีดังนี้</u>

1. การสร้างข้อมูลในงานวิจัย

1.1 <u>การสร้างข้อมูลของความคลาดเคลื่อน</u> ($\underset{\sim}{\varepsilon}$)

ในขั้นตอนนี้สร้างค่าความคลาดเคลื่อนให้มีการแจกแจงปกติด้วยค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 2 4 6 8 และ 10

1.2 <u>การสร้างข้อมูลของตัวแปรอิสระ</u> (X)

สร้างตัวแปรอิสระ (x_i) ให้มีการแจกแจงปกติด้วยค่าเฉลี่ย (μ) เท่ากับ 5 และความแปรปรวน (σ^2) เท่ากับ 4 และการยกกำลังตัวแปรอิสระให้กับกำลังสูงสุดที่กำหนดในแต่ละขั้นตอน

1.3 <u>การสร้างข้อมูลตัวแปรตาม</u> $(\underset{\sim}{y})$

สร้างตัวแปรตาม $(\underset{\sim}{y})$ จากรูปแบบความสัมพันธ์ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ โดยกำหนดให้ $\underset{\sim}{\beta} = \underset{\sim}{1}$ จากตัวแบบความถดถอยพหุนาม ซึ่งอยู่ในรูปแบบดังนี้

1) ตัวแบบที่มีกำลังสูงสุดของตัวแปรอิสระเป็น 6 มีรูปแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \beta_5 x_i^5 + \beta_6 x_i^6 + \varepsilon_i \quad , \; i = 1,2,\ldots,n$$

2) ตัวแบบที่มีกำลังสูงสุดของตัวแปรอิสระเป็น 5 มีรูปแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \beta_5 x_i^5 + \varepsilon_i \quad , \; i = 1,2,\ldots,n$$

3) ตัวแบบที่มีกำลังสูงสุดของตัวแปรอิสระเป็น 4 มีรูปแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \varepsilon_i \quad , \; i = 1,2,\ldots,n$$

4) ตัวแบบที่มีกำลังสูงสุดของตัวแปรอิสระเป็น 3 มีรูปแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \varepsilon_i \quad , \; i = 1,2,\ldots,n$$

5) ตัวแบบที่มีกำลังสูงสุดของตัวแปรอิสระเป็น 2 มีรูปแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i \quad , \; i = 1,2,\ldots,n$$

2. การประมาณค่าพารามิเตอร์ด้วยวิธีการประมาณค่าพารามิเตอร์แบบต่างๆ

เมื่อสร้างข้อมูลของตัวแปรอิสระ (X) และตัวแปรตาม $(\underset{\sim}{y})$ ได้แล้ว ขั้นตอนต่อไปคือการนำข้อมูลของตัวแปรอิสระและตัวแปรตามที่ได้ไปประมาณค่าพารามิเตอร์ $(\underset{\sim}{\beta})$ ซึ่งในงานวิจัยครั้งนี้ได้ทำการประมาณค่าพารามิเตอร์ในสมการถดถอยพหุนาม 4 วิธี ดังนี้

2.1 <u>วิธีกำลังสองน้อยสุด (OLS)</u>

ผู้วิจัยประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ ด้วยตัวประมาณกำลังสองน้อยสุด $\underset{\sim}{\hat{\beta}}$ จากสมการ (2.4) ในบทที่ 2 ซึ่งมีรูปแบบของตัวประมาณดังนี้

$$\underset{\sim}{\hat{\beta}} = \left(X'X\right)^{-1} X'\underset{\sim}{y}$$

2.2 <u>วิธีการ์ร็อตที่ไม่เป็นลบ (NG)</u>

ผู้วิจัยประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ ด้วยตัวประมาณการ์ร็อตที่ไม่เป็นลบ $\underset{\sim}{\hat{\beta}}_{NG}$ จากสมการ (2.14) ในบทที่ 2 ซึ่งมีรูปแบบของตัวประมาณดังนี้

$$\hat{\alpha}_{NG_j} = \hat{c}_j \hat{\alpha}_j \qquad ; \quad j = 0, 1, \ldots, p$$
$$= \left(1 - \frac{\lambda^*}{\hat{\alpha}_j^2 \lambda_j}\right)^+ \hat{\alpha}_j$$

เมื่อ $\hat{\underset{\sim}{\alpha}}$ แทน ตัวประมาณสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธีกำลังสองน้อยสุดจากตัวแบบเชิงตั้งฉาก

และ $\lambda^* = c^* \sqrt{\dfrac{\log n}{n}}$ ซึ่ง $c^* = 3$

โดยที่ $\left(1 - \dfrac{\lambda^*}{\hat{\alpha}_j^2 \lambda_j}\right)^+$ แสดงเฉพาะส่วนที่เป็นค่าบวกของ $\left(1 - \dfrac{\lambda^*}{\hat{\alpha}_j^2 \lambda_j}\right)$

และจากสมการ (2.15) จะได้ว่า

$$\hat{\underset{\sim}{\beta}}_{NG} = Q \hat{\underset{\sim}{\alpha}}_{NG}$$

2.3 <u>วิธีแจ็คไนฟริดจ์ (JR)</u>

ผู้วิจัยประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ ด้วยตัวประมาณแจ็คไนฟริดจ์ $\hat{\underset{\sim}{\beta}}_{JR}$ จากสมการ (2.16) ในบทที่ 2 ซึ่งมีรูปแบบของตัวประมาณดังนี้

$$\hat{\underset{\sim}{\beta}}_{JR} = \hat{\underset{\sim}{\beta}}_R - \hat{\underset{\sim}{b}}$$
$$= (n-1)\hat{\underset{\sim}{\beta}}_{R,(\bullet)} + n\hat{\underset{\sim}{\beta}}_R$$

เมื่อ $\hat{\underset{\sim}{\beta}}_R = \left(X'X + kI_{p+1}\right)^{-1} X'\underset{\sim}{y} \quad ; \; k > 0$

และ $\hat{\underset{\sim}{\beta}}_{R,(\bullet)} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \hat{\underset{\sim}{\beta}}_{R,(i)}$

โดยที่ $\hat{\underset{\sim}{\beta}}_{R,(i)} = (X'E'_{(i)}E_{(i)}X + kI_p)^{-1} X'E'_{(i)}E_{(i)}\underset{\sim}{y} \quad ; \quad k > 0$

และ $E_{(i)}$ เป็นเมทริกซ์ขนาด $(n-1) \times n$ ที่เกิดจากการตัดแถวที่ i จากเมทริกซ์เอกลักษณ์ขนาด $n \times n$

โดยใช้ค่า k_{HKB} ที่ได้จากวิธีการประมาณค่า k ที่คิดค้นโดย โฮเอิล (Hoerl) เคนนาร์ด (Kennard) และ บลาดวิน (Bladwin) อิงพื้นฐานจากตัวประมาณกำลังสองน้อยสุด (OLS) ซึ่งมีรูปแบบค่า k_{HKB} เป็นดังนี้

$$k_{HKB} = \frac{(p+1)s^2}{\hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}}}$$

เมื่อ $s^2 = \dfrac{(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}})'(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}})}{n-(p+1)}$ เป็นค่าประมาณพารามิเตอร์ σ^2

2.4 <u>วิธีลิวไทป์ (LT)</u>

ผู้วิจัยประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ ด้วยตัวประมาณลิวไทป์ $\hat{\beta}_{\sim LT}$ จากสมการ (2.23) ในบทที่ 2 ซึ่งมีรูปแบบของตัวประมาณดังนี้

$$\hat{\alpha}_{\sim LT} = (\Lambda + \hat{k}I)^{-1}(Z'\underset{\sim}{y} - \hat{d}\,\hat{\alpha}_{\sim R})$$

เมื่อ $\hat{\alpha}_{\sim R}$ แทนตัวประมาณสัมประสิทธิ์ความถดถอยพหุนามวิธีริดจ์จากตัวแบบเชิงตั้งฉาก

$$\hat{k} = \frac{\lambda_0 - 100\lambda_p}{99}$$

และ $$\hat{d} = \sum_{j=0}^{p} \frac{(\hat{\sigma}_R^2 - \hat{k}\alpha_{Rj}^2)}{(\lambda_j + \hat{k})^2} \Bigg/ \sum_{j=0}^{p} \frac{(\lambda_j\alpha_{Rj}^2 + \hat{\sigma}_R^2)}{\lambda_j(\lambda_j + \hat{k})^2}$$

โดยที่ $\hat{\sigma}_R^2 = \dfrac{(\underset{\sim}{y} - X\hat{\beta}_{\sim R})'(\underset{\sim}{y} - X\hat{\beta}_{\sim R})}{n-(p+1)}$

และจากสมการ (2.24) จะได้ว่า

$$\hat{\beta}_{\sim LT} = Q\hat{\alpha}_{\sim LT}$$

3. การคำนวณหาค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองค่าสัมประสิทธิ์ความถดถอยพหุนามในแต่ละวิธี

เมื่อได้ค่าประมาณของพารามิเตอร์จากวิธีประมาณทั้ง 4 วิธี แล้วนำมาหาค่าผลรวมกำลังสองของความคลาดเคลื่อนจากการประมาณค่าพารามิเตอร์แต่ละวิธี โดยนำค่าประมาณที่ได้จากตัวประมาณแต่ละวิธีลบออกด้วยค่าจริงของพารามิเตอร์ ซึ่งกำหนดให้ $\underset{\sim}{\beta} = \underset{\sim}{1}$ แล้วยกกำลังสองสะสมไว้ในแต่ละครั้ง และทำการทดลองเช่นเดิมจนครบ 1,000 ครั้ง แล้วจึงคำนวณหาค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของตัวประมาณความถดถอยพหุนาม ($ARMSE$) ซึ่งมีสูตรในการคำนวณดังนี้

$$RMSE_i = \sqrt{\frac{1}{p+1}\sum_{j=0}^{p}\left(\hat{\beta}_{ij} - \beta_j\right)^2}$$

$$ARMSE = \frac{1}{1{,}000}\sum_{i=1}^{1{,}000} RMSE_i$$

$$DIFF\ = \frac{ARMSE_k - ARMSE_{\min}}{ARMSE_{\min}} \times 100 \quad ,k = 1,2,3,4$$

เมื่อ β_j แทน สัมประสิทธิ์ความถดถอยพหุนามตัวที่ j

$\hat{\beta}_{ij}$ แทน ตัวประมาณสัมประสิทธิ์ความถดถอยพหุนามตัวที่ j ในการประมาณครั้งที่ i

$RMSE_i$ แทน ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์ความถดถอยพหุนามในการประมาณครั้งที่ i

ARMSE แทน ค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์ความถดถอยพหุนาม

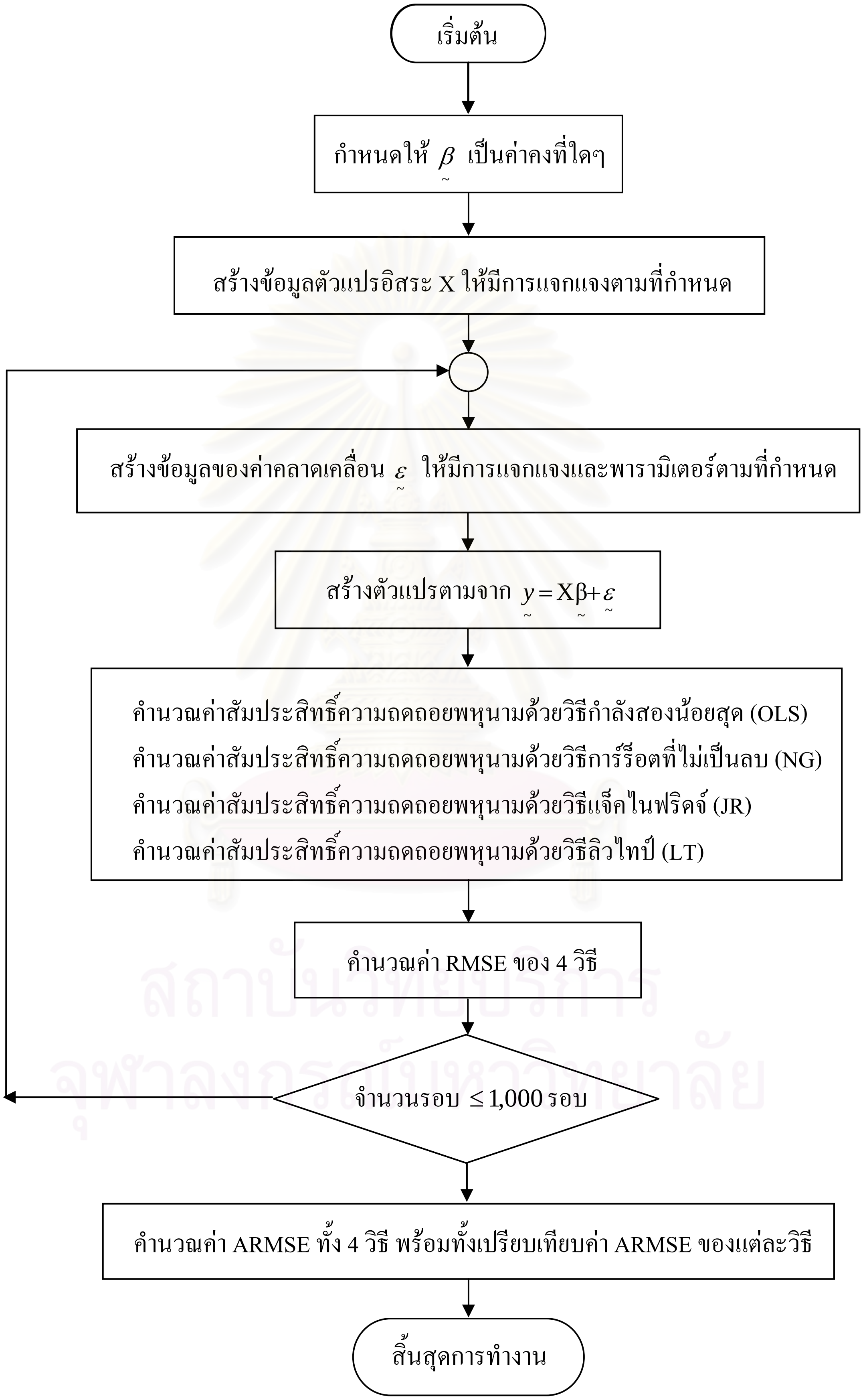
DIFF แทน อัตราส่วนของผลต่างของค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธีต่างๆ

$ARMSE_k$ แทน ค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์การถดถอยพหุนามโดยวิธีที่ k , k = 1, 2, 3, 4

และ $ARMSE_{min}$ แทน ค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์ความถดถอยพหุนามจากวิธีที่มีค่าน้อยที่สุด

จากนั้นจึงทำการเปรียบเทียบค่า *ARMSE* ที่ได้จากวิธีประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามทั้ง 4 วิธี เพื่อพิจารณาว่าวิธีประมาณใดมีประสิทธิภาพสูงสุดในแต่ละสถานการณ์ และทำการทดลองเช่นนี้โดยเปลี่ยนระดับความคลาดเคลื่อน กำลังสูงสุดของตัวแปรอิสระและขนาดตัวอย่าง จนกระทั่งครบทุกรูปแบบของสถานการณ์ที่ต้องการศึกษา เพื่อให้ง่ายแก่การเข้าใจจึงได้แสดงผังงานซึ่งแสดงขั้นตอนการวิจัยทั้งหมดในรูปที่ 3.1 ดังต่อไปนี้

<u>รูปที่ 3.1</u> **ผังงานแสดงขั้นตอนในการดำเนินการวิจัย**



บทที่ 4
ผลการวิจัย

การวิจัยครั้งนี้มีจุดประสงค์เพื่อที่จะเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธีกำลังสองน้อยสุด (OLS) วิธีการ์ร็อตที่ไม่เป็นลบ (NG) วิธีแจ็คไนฟริดจ์ (JR) และวิธีลิวไทป์ (LT) โดยเกณฑ์ที่ใช้ในการตัดสินว่า วิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามวิธีการใดทำให้ได้ค่าประมาณที่ใกล้เคียงกับค่าจริงมากที่สุดภายใต้สถานการณ์ที่กำหนดจะพิจารณาจากค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ซึ่งสถานการณ์ต่างๆ เหล่านี้จะอาศัยวิธีการจำลองด้วยเทคนิคการจำลองมอนติคาร์โลโดยทำการทดลองซ้ำ 1,000 ครั้งในแต่ละสถานการณ์

การนำเสนอผลการวิจัยจะนำเสนอในรูปตารางและกราฟ โดยใช้สัญลักษณ์แทนความหมายต่างๆ ดังนี้

σ_{ε}^{2} แทน ความแปรปรวนของความคลาดเคลื่อนในตัวแบบ

n แทน ขนาดตัวอย่าง

p แทน จำนวนพจน์พหุนามในตัวแบบโดยไม่นับพจน์ค่าคงที่

MB แทน กำลังสูงสุดของตัวแปรอิสระที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบ

ARMSE แทน ค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

SD แทน ค่าส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนกำลังสอง

DIFF แทน ค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยคลาดเคลื่อนกำลังสองของการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธีต่างๆ

OLS แทน การประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธีกำลังสองน้อยสุด

NG แทน การประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธีการ์ร็อตที่ไม่เป็นลบ

JR แทน การประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธีแจ็คไนฟริดจ์

LT แทน การประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธีลิวไทป์

การวิจัยจะใช้สถานการณ์ที่การแจกแจงของความคลาดเคลื่อนสุ่มในตัวแปรตาม (ε) เป็นแบบปกติซึ่งมีค่าเฉลี่ย 0 และความแปรปรวนเป็น 2 4 6 8 และ 10 ขนาดตัวอย่างที่ใช้คือ 5p 10p 15p และ 20p กำลังสูดสุดของตัวแปรอิสระที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบเริ่มต้น(MB) คือ 2 3 4 5 และ 6

ค่าจากตารางในแต่ละกรณีและแต่ละวิธีประมาณจะแสดงค่าตัวเลข 3 ค่า ได้แก่ ARMSE ค่าส่วนเบี่ยงเบนมาตรฐานของ ARMSE จากการกระทำซ้ำ 1,000 รอบ ซึ่งแสดงอยู่ในวงเล็บ และค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยคลาดเคลื่อนกำลังสอง โดยมีสูตรดังนี้

ก) $$RMSE_i = \sqrt{\frac{1}{p+1}\sum_{j=0}^{p}\left(\hat{\beta}_{ij} - \beta_j\right)^2}$$

ข) $$ARMSE = \frac{1}{1{,}000}\sum_{i=1}^{1{,}000} RMSE_i$$

ค) $$SD = \sqrt{\frac{1}{1{,}000}\sum_{i=1}^{1{,}000}\left(RMSE_i - ARMSE\right)^2}$$

และ ง) $$DIFF = \frac{ARMSE_k - ARMSE_{\min}}{ARMSE_{\min}} \times 100 \quad , k = 1,2,3,4$$

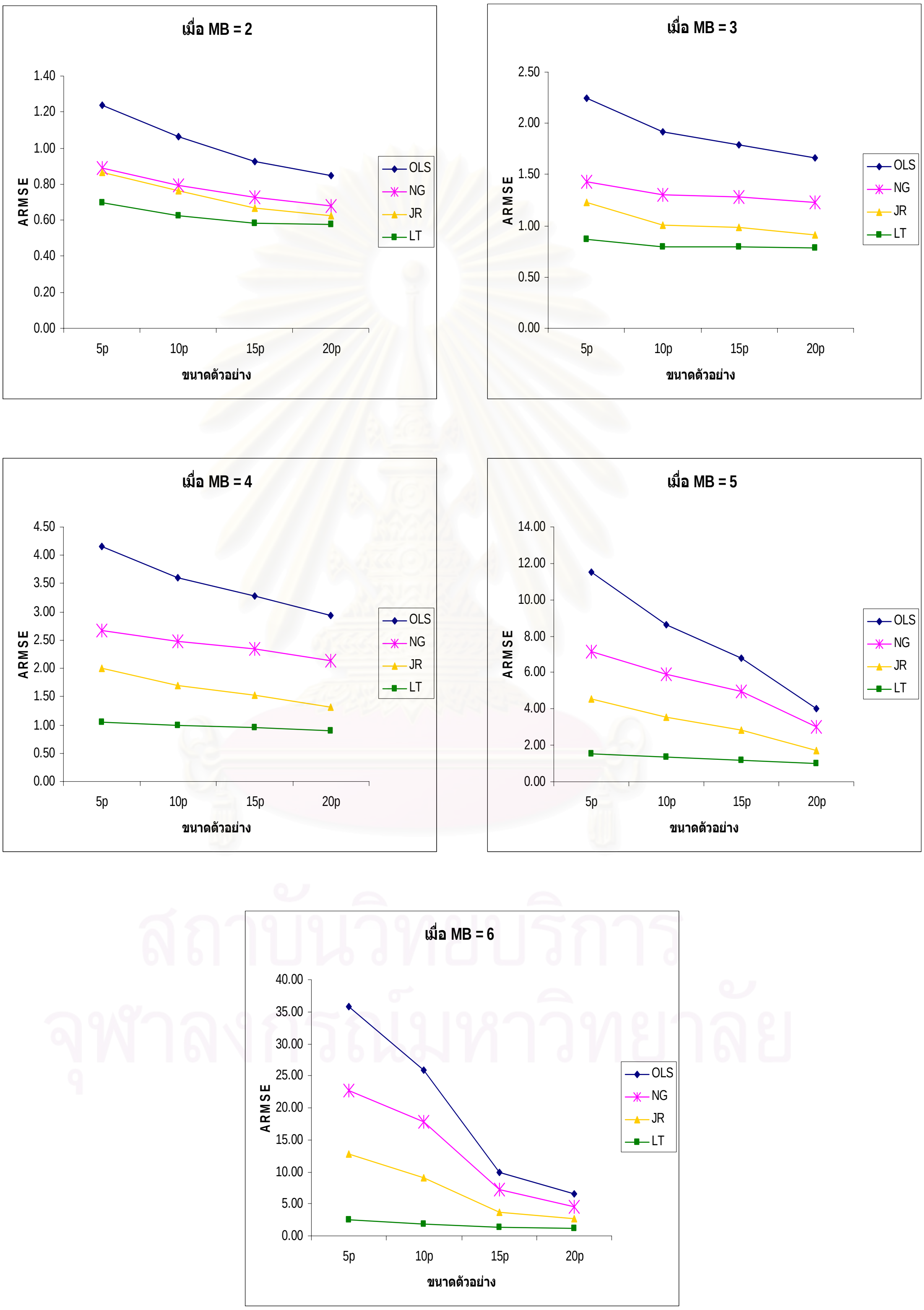
เมื่อ $ARMSE_k$ หมายถึงค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจากวิธีประมาณที่ k

และ $ARMSE_{\min}$ หมายถึงค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจากวิธีประมาณที่มีค่าต่ำสุดจากทั้ง 4 วิธี

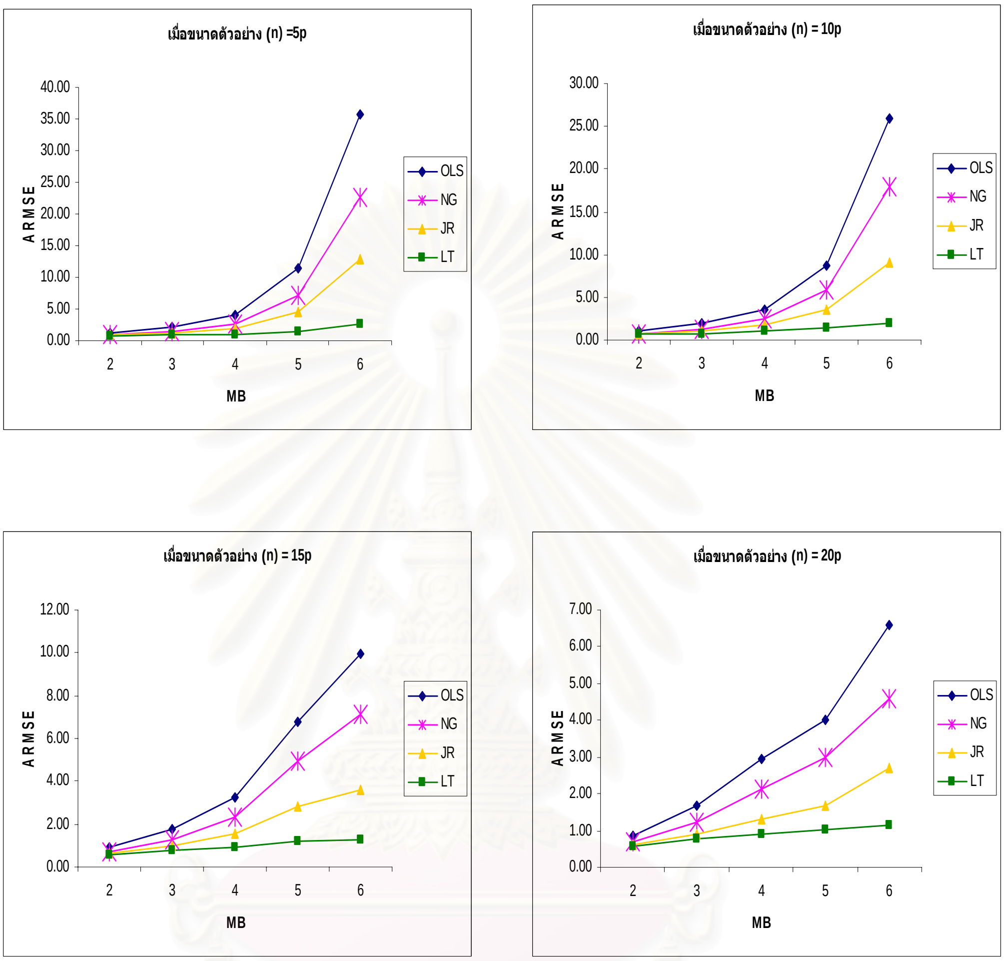
ตารางที่ 4.1 แสดงค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อ $\sigma_{\varepsilon}^{2} = 2$

MB	วิธี	ขนาดตัวอย่าง (n) 5p				10p				15p				20p			
		OLS	NG	JR	LT	OLS	NG	JR	LT	OLS	NG	JR	LT	OLS	NG	JR	LT
2	ARMSE	1.236	0.890	0.865	0.697	1.061	0.793	0.766	0.624	0.928	0.727	0.669	0.585	0.849	0.680	0.627	0.579
	(SD)	(0.299)	(0.212)	(0.193)	(0.151)	(0.228)	(0.169)	(0.154)	(0.121)	(0.175)	(0.133)	(0.114)	(0.096)	(0.132)	(0.099)	(0.089)	(0.080)
	DIFF	77.161	27.667	24.055	0.000	70.214	27.130	22.813	0.000	58.643	24.223	14.334	0.000	46.425	17.288	8.273	0.000
3	ARMSE	2.247	1.432	1.224	0.871	1.912	1.299	1.012	0.793	1.793	1.284	0.981	0.796	1.663	1.231	0.913	0.780
	(SD)	(0.546)	(0.342)	(0.291)	(0.204)	(0.421)	(0.269)	(0.199)	(0.148)	(0.320)	(0.220)	(0.154)	(0.120)	(0.247)	(0.180)	(0.111)	(0.091)
	DIFF	158.062	64.485	40.539	0.000	141.123	63.737	27.547	0.000	125.067	61.270	23.190	0.000	113.096	57.806	17.047	0.000
4	ARMSE	4.158	2.677	2.011	1.049	3.595	2.480	1.689	0.989	3.274	2.344	1.519	0.951	2.933	2.133	1.313	0.889
	(SD)	(1.035)	(0.641)	(0.480)	(0.236)	(0.777)	(0.532)	(0.353)	(0.203)	(0.647)	(0.455)	(0.291)	(0.171)	(0.523)	(0.355)	(0.172)	(0.109)
	DIFF	296.500	155.307	91.799	0.000	263.495	150.794	70.836	0.000	244.413	146.513	59.731	0.000	229.761	139.894	47.665	0.000
5	ARMSE	11.520	7.160	4.575	1.523	8.653	5.932	3.536	1.343	6.793	4.936	2.846	1.202	4.026	3.007	1.693	1.009
	(SD)	(2.874)	(1.782)	(1.083)	(0.335)	(1.804)	(1.183)	(0.672)	(0.250)	(1.244)	(0.885)	(0.477)	(0.198)	(0.494)	(0.365)	(0.196)	(0.104)
	DIFF	656.402	370.098	200.420	0.000	544.334	341.668	163.254	0.000	464.912	310.471	136.677	0.000	299.068	198.097	67.833	0.000
6	ARMSE	35.759	22.625	12.821	2.604	25.913	17.852	9.105	1.915	9.924	7.149	3.629	1.266	6.603	4.566	2.708	1.127
	(SD)	(8.831)	(5.291)	(2.943)	(0.590)	(5.565)	(3.596)	(1.814)	(0.369)	(1.812)	(1.004)	(0.457)	(0.154)	(0.801)	(0.545)	(0.296)	(0.117)
	DIFF	1273.023	768.722	392.282	0.000	1253.230	832.268	375.456	0.000	683.739	464.603	186.574	0.000	486.012	305.247	140.328	0.000

<u>รูปที่ 4.1.1</u> **แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี OLS วิธี NG วิธี JR และวิธี LT ระหว่างค่า ARMSE และขนาดตัวอย่าง (n) เมื่อ** $\sigma_{\varepsilon}^{2} = 2$



<u>รูปที่ 4.1.2</u> แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี OLS วิธี NG วิธี JR และวิธี LT ระหว่างค่า ARMSE และเลขชี้กำลังสูงสุด (MB) เมื่อ $\sigma_{\varepsilon}^{2} = 2$



จากตารางที่ 4.1 และกราฟรูปที่ 4.1.1 – 4.1.2 ผลสรุปคือ

ค่า ARMSE ของ 4 วิธี เรียงลำดับจากน้อยไปมากได้แก่ วิธี LT วิธี JR วิธี NG และวิธี OLS ตามลำดับ สำหรับทุกกรณี

<u>เมื่อเลขชี้กำลังสูงสุด (MB) ของตัวแบบคงที่</u>

การเพิ่มขนาดตัวอย่าง (n) ส่งผลให้ค่า ARMSE ของทุกวิธีมีแนวโน้มลดลงเพราะขนาดตัวอย่างที่เพิ่มขึ้นจะส่งผลทำให้ค่าคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามลดลงซึ่งวิธีการประมาณค่าพารามิเตอร์ที่ดีที่สุดคือ วิธี LT สำหรับทุกขนาดตัวอย่าง รองลงมาคือวิธี JR วิธี NG และวิธี OLS ตามลำดับ เพราะวิธี LT เป็นวิธีการที่สามารถแก้ปัญหาตัวแปรอิสระมีพหุสัมพันธ์กัน (multicollinearity) ได้อย่างเต็มรูปแบบ ดังนั้นเมื่อขนาดตัวอย่างน้อยโอกาสที่จะเกิดค่าผิดปกติในตัวแปรอิสระมากทำให้วิธี LT จะมีประสิทธิภาพมาก และเมื่อขนาดตัวอย่างมากขึ้นโอกาสที่จะเกิดค่าผิดปกติในตัวแปรอิสระลดลงประสิทธิภาพของวิธี LT นี้จึงลดลงด้วย กล่าวคือวิธี LT เป็นวิธีการที่ดีเมื่อขนาดตัวอย่างไม่มากนัก

<u>เมื่อขนาดตัวอย่าง (n) ของตัวแบบคงที่</u>

เมื่อเลขชี้กำลังสูงสุด (MB) เพิ่มขึ้นส่งผลให้ ARMSE ของทุกวิธีมีแนวโน้มเพิ่มขึ้นเพราะเลขชี้กำลังสูงสุด (MB) ที่เพิ่มขึ้นจะทำให้ตัวแปรอิสระมีความสัมพันธ์กัน (multicollinearity) มากขึ้น ทำให้ค่าคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามเพิ่มขึ้น ซึ่งวิธีการประมาณค่าพารามิเตอร์ที่ดีที่สุด คือวิธี LT สำหรับทุกค่าเลขชี้กำลังสูงสุด (MB) รองลงมาคือวิธี JR วิธี NG และวิธี OLS ตามลำดับ เพราะวิธี LT เป็นวิธีการที่สามารถแก้ปัญหาตัวแปรอิสระมีพหุสัมพันธ์กันได้อย่างเต็มรูปแบบ

จากข้างต้นสรุปได้ว่าปัจจัยที่ส่งผลต่อค่า ARMSE ได้แก่ ขนาดตัวอย่าง (n) และเลขชี้กำลังสูงสุด (MB) ซึ่งค่า ARMSE จะแปรผกผันกับขนาดตัวอย่าง (n) (รูปที่ 4.1.1) แต่จะแปรผันตามเลขชี้กำลังสูงสุด (MB) (รูปที่ 4.1.2) ซึ่งวิธี LT จะมีความคงเส้นคงวามากกว่าวิธี OLS วิธี NG และวิธี JR เพราะค่า ARMSE ของวิธี LT ลู่เข้าสู่ศูนย์มากกว่าอีก 3 วิธีในทุกกรณี ดังนั้นวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามวิธี LT จะประมาณค่าพารามิเตอร์ได้ใกล้เคียงกับพารามิเตอร์ที่แท้จริงมากกว่าอีก 3 วิธี

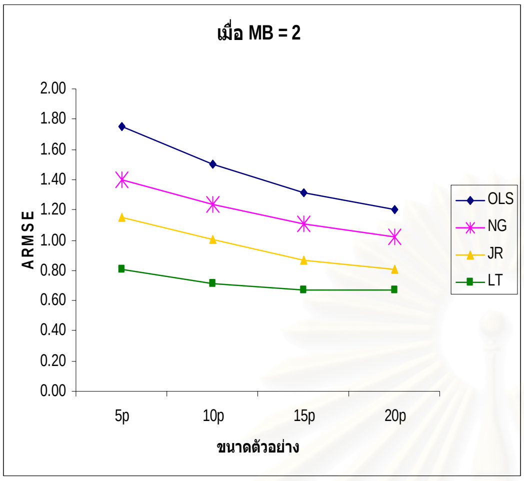
พิจารณาค่า DIFF ของแต่ละวิธีการที่ใช้ประมาณค่าพารามิเตอร์ในทุกขนาดตัวอย่าง จะได้ว่า ประสิทธิภาพของวิธี LT โดดเด่นกว่าอีก 3 วิธีอย่างชัดเจน โดยประสิทธิภาพของวิธี LT จะลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นกล่าวคือ วิธี LT มีประสิทธิภาพดีกว่าวิธี OLS อย่างน้อยประมาณ 77.16% 70.21% 58.64% และ 46.43% ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น วิธี LT มีประสิทธิภาพดีกว่าวิธี NG อย่างน้อยประมาณ 27.67% 27.13% 24.22% และ 17.29% ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น และวิธี LT มีประสิทธิภาพดีกว่าวิธี JR อย่างน้อยประมาณ 24.06% 22.81% 14.33% และ

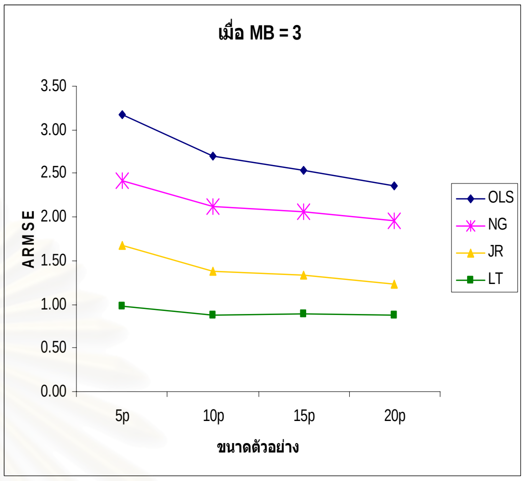
8.27% ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น แต่วิธี LT มีประสิทธิภาพสูงขึ้นเมื่อเลขชี้กำลังสูงสุด (MB) เพิ่มขึ้น

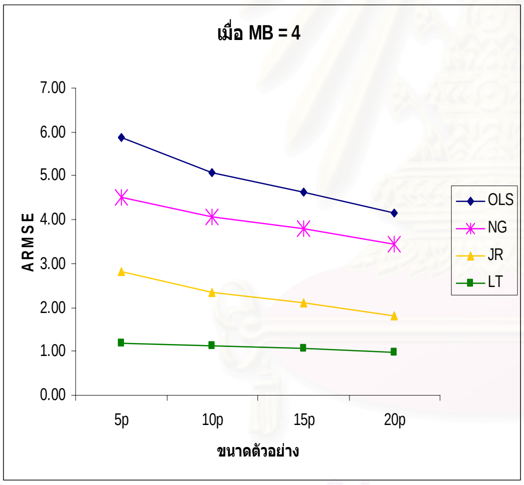
ตารางที่ 4.2 แสดงค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อ $\sigma_{\varepsilon}^{2} = 4$

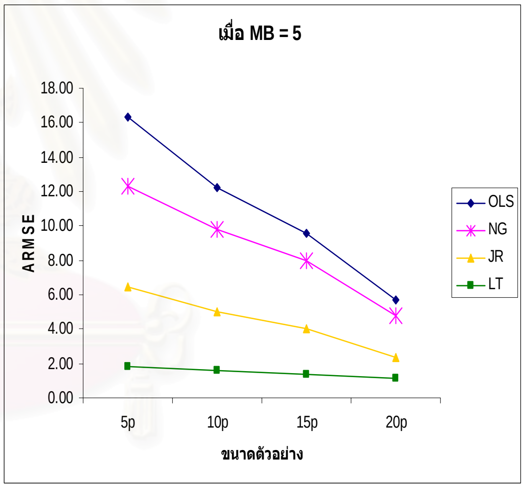
MB	วิธี	ขนาดตัวอย่าง (n)															
		5p				10p				15p				20p			
		OLS	NG	JR	LT	OLS	NG	JR	LT	OLS	NG	JR	LT	OLS	NG	JR	LT
2	ARMSE	1.747	1.398	1.151	0.804	1.501	1.235	1.001	0.717	1.312	1.107	0.868	0.668	1.200	1.022	0.807	0.667
	(SD)	(0.420)	(0.335)	(0.261)	(0.180)	(0.274)	(0.222)	(0.171)	(0.121)	(0.213)	(0.165)	(0.123)	(0.093)	(0.164)	(0.136)	(0.097)	(0.069)
	DIFF	117.407	73.970	43.150	0.000	109.424	72.308	39.672	0.000	96.476	65.708	30.025	0.000	79.888	53.235	21.018	0.000
3	ARMSE	3.178	2.413	1.682	0.985	2.704	2.127	1.377	0.879	2.535	2.062	1.331	0.894	2.351	1.964	1.232	0.878
	(SD)	(0.775)	(0.568)	(0.380)	(0.221)	(0.568)	(0.439)	(0.278)	(0.172)	(0.479)	(0.346)	(0.222)	(0.140)	(0.350)	(0.267)	(0.133)	(0.093)
	DIFF	222.491	144.855	70.658	0.000	207.575	141.902	56.608	0.000	183.627	130.650	48.929	0.000	167.923	123.725	40.411	0.000
4	ARMSE	5.880	4.512	2.813	1.201	5.084	4.071	2.355	1.112	4.630	3.792	2.111	1.058	4.147	3.438	1.820	0.972
	(SD)	(1.386)	(1.039)	(0.609)	(0.256)	(1.034)	(0.823)	(0.456)	(0.196)	(0.809)	(0.661)	(0.318)	(0.154)	(0.603)	(0.494)	(0.252)	(0.119)
	DIFF	389.788	275.812	134.266	0.000	357.026	266.025	111.687	0.000	337.780	258.542	99.575	0.000	326.757	253.793	87.284	0.000
5	ARMSE	16.292	12.339	6.457	1.843	12.238	9.827	4.982	1.589	9.606	7.964	4.002	1.391	5.693	4.810	2.368	1.122
	(SD)	(4.006)	(2.535)	(1.291)	(0.356)	(2.357)	(1.767)	(0.792)	(0.251)	(1.300)	(1.024)	(0.510)	(0.176)	(0.718)	(0.598)	(0.270)	(0.115)
	DIFF	783.802	569.361	250.282	0.000	670.412	518.628	213.623	0.000	590.788	472.731	187.761	0.000	407.478	328.701	111.044	0.000
6	ARMSE	50.571	38.847	18.128	3.381	36.647	29.600	12.870	2.398	14.034	11.612	5.119	1.466	9.338	7.383	3.810	1.282
	(SD)	(12.534)	(9.490)	(4.388)	(0.797)	(8.631)	(6.894)	(2.950)	(0.461)	(2.610)	(1.846)	(0.769)	(0.217)	(1.361)	(0.988)	(0.479)	(0.147)
	DIFF	1395.741	1048.980	436.173	0.000	1428.487	1134.568	436.787	0.000	857.038	691.871	249.059	0.000	628.424	475.931	197.159	0.000

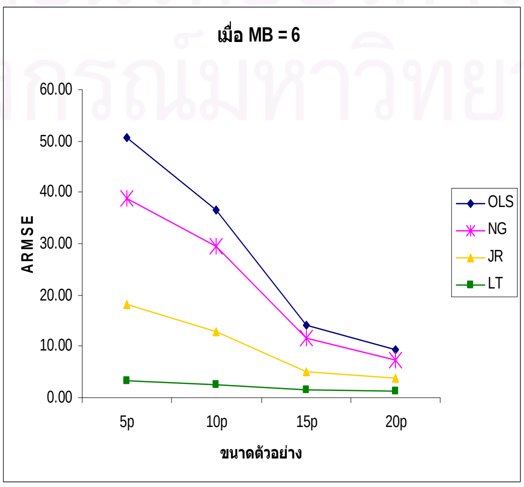
<u>รูปที่ 4.2.1</u> **แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี OLS วิธี NG วิธี JR และวิธี LT ระหว่างค่า ARMSE และขนาดตัวอย่าง (n) เมื่อ** $\sigma_{\varepsilon}^{2} = 4$



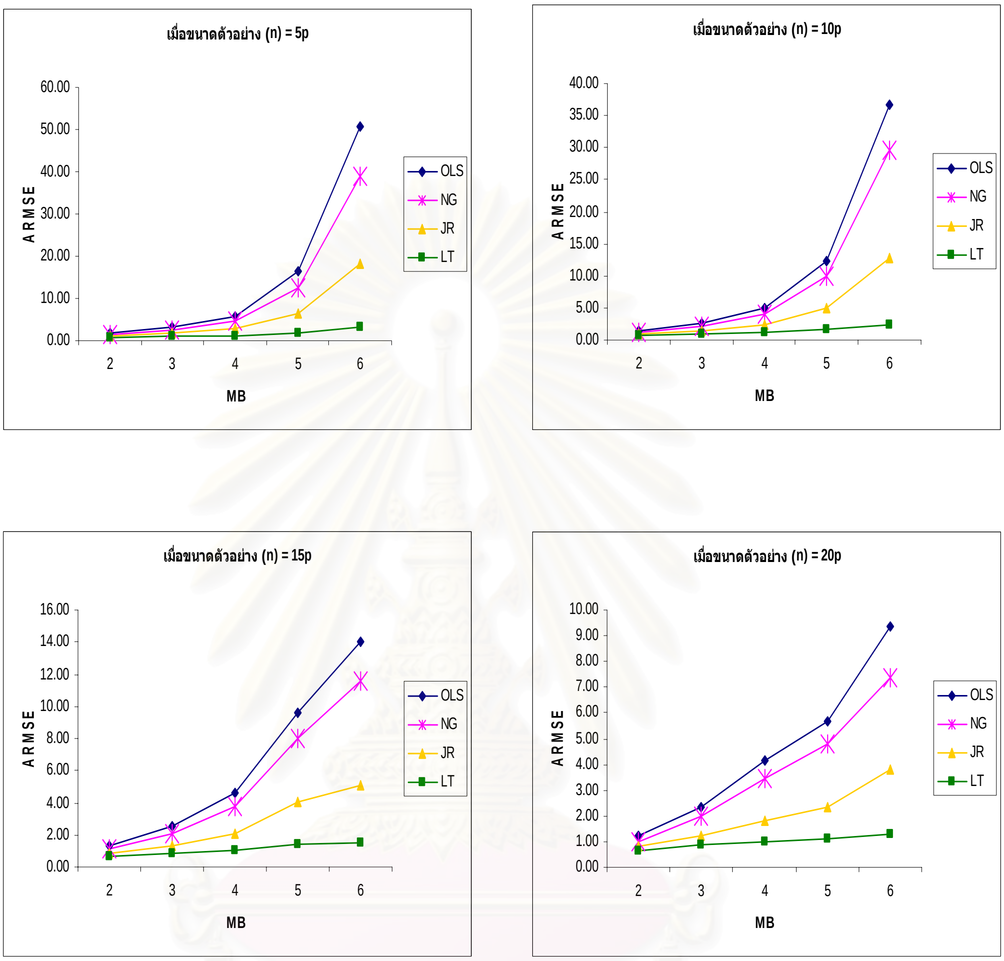








<u>รูปที่ 4.2.2</u> **แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วย วิธี OLS วิธี NG วิธี JR และวิธี LT ระหว่างค่า ARMSE และเลขชี้กำลังสูงสุด (MB) เมื่อ** $\sigma_{\varepsilon}^{2} = 4$



จากตารางที่ 4.2 และกราฟรูปที่ 4.2.1 – 4.2.2 ผลสรุปคือ

ค่า ARMSE ของ 4 วิธี เรียงลำดับจากน้อยไปมากได้แก่ วิธี LT วิธี JR วิธี NG และวิธี OLS ตามลำดับ สำหรับทุกกรณี

<u>เมื่อเลขชี้กำลังสูงสุด (MB) ของตัวแบบคงที่</u>

การเพิ่มขนาดตัวอย่าง (n) ส่งผลให้ค่า ARMSE ของทุกวิธีมีแนวโน้มลดลงเพราะขนาดตัวอย่างที่เพิ่มขึ้นจะส่งผลทำให้ค่าคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามลดลงซึ่งวิธีการประมาณค่าพารามิเตอร์ที่ดีที่สุดคือ วิธี LT สำหรับทุกขนาดตัวอย่าง รองลงมาคือวิธี JR วิธี NG และวิธี OLS ตามลำดับ เพราะวิธี LT เป็นวิธีการที่สามารถแก้ปัญหาตัวแปรอิสระมีพหุสัมพันธ์กัน (multicollinearity) ได้อย่างเต็มรูปแบบ ดังนั้นเมื่อขนาดตัวอย่างน้อยโอกาสที่จะเกิดค่าผิดปกติในตัวแปรอิสระมากทำให้วิธี LT จะมีประสิทธิภาพมาก และเมื่อขนาดตัวอย่างมากขึ้นโอกาสที่จะเกิดค่าผิดปกติในตัวแปรอิสระลดลงประสิทธิภาพของวิธี LT นี้จึงลดลงด้วย กล่าวคือวิธี LT เป็นวิธีการที่ดีเมื่อขนาดตัวอย่างไม่มากนัก

<u>เมื่อขนาดตัวอย่าง (n) ของตัวแบบคงที่</u>

เมื่อเลขชี้กำลังสูงสุด (MB) เพิ่มขึ้นส่งผลให้ ARMSE ของทุกวิธีมีแนวโน้มเพิ่มขึ้นเพราะเลขชี้กำลังสูงสุด (MB) ที่เพิ่มขึ้นจะทำให้ตัวแปรอิสระมีความสัมพันธ์กัน (multicollinearity) มากขึ้น ทำให้ค่าคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามเพิ่มขึ้น ซึ่งวิธีการประมาณค่าพารามิเตอร์ที่ดีที่สุด คือวิธี LT สำหรับทุกค่าเลขชี้กำลังสูงสุด (MB) รองลงมาคือวิธี JR วิธี NG และวิธี OLS ตามลำดับ เพราะวิธี LT เป็นวิธีการที่สามารถแก้ปัญหาตัวแปรอิสระมีพหุสัมพันธ์กันได้อย่างเต็มรูปแบบ

เมื่อพิจารณาตารางที่ 4.1 และตารางที่ 4.2 จะเห็นได้ว่าเมื่อความแปรปรวนของความคลาดเคลื่อน (σ_ε^2) เพิ่มขึ้นทำให้ค่า ARMSE ของทุกกรณีเพิ่มขึ้นด้วย ทั้งนี้เพราะว่าเมื่อ σ_ε^2 เพิ่มขึ้นจะส่งผลให้การกระจายของตัวแปรตามเพิ่มขึ้น ส่งผลให้ค่าประมาณของพารามิเตอร์ที่ได้สูงขึ้น

จากข้างต้นสรุปได้ว่าปัจจัยที่ส่งผลต่อค่า ARMSE ได้แก่ ขนาดตัวอย่าง (n) เลขชี้กำลังสูงสุด (MB) และความแปรปรวนของความคลาดเคลื่อน (σ_ε^2) ซึ่งค่า ARMSE จะแปรผกผันกับขนาดตัวอย่าง (n) (รูปที่ 4.2.1) แต่จะแปรผันตามเลขชี้กำลังสูงสุด (MB) (รูปที่ 4.2.2) และความแปรปรวนของความคลาดเคลื่อน (σ_ε^2) ซึ่งวิธี LT จะมีความคงเส้นคงวามากกว่าวิธี OLS วิธี NG และวิธี JR เพราะค่า ARMSE ของวิธี LT ลู่เข้าสู่ศูนย์มากกว่าอีก 3 วิธีในทุกกรณี ดังนั้นวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามวิธี LT จะประมาณค่าพารามิเตอร์ได้ใกล้เคียงกับพารามิเตอร์ที่แท้จริงมากกว่าอีก 3 วิธี

พิจารณาค่า DIFF ของแต่ละวิธีการที่ใช้ประมาณค่าพารามิเตอร์ในทุกขนาดตัวอย่าง จะได้ว่าประสิทธิภาพของวิธี LT โดดเด่นกว่าอีก 3 วิธีอย่างชัดเจน โดยประสิทธิภาพของวิธี LT จะลดลง

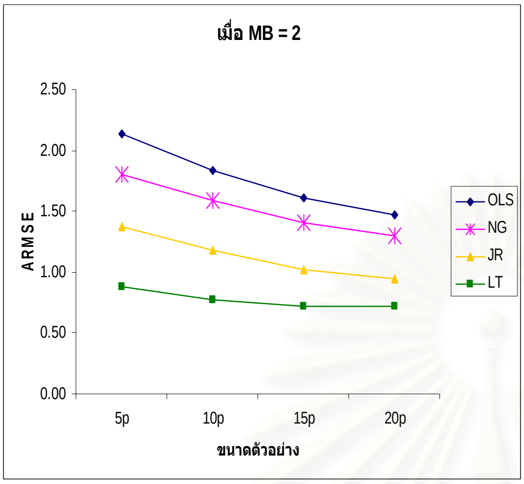
เมื่อขนาดตัวอย่างเพิ่มขึ้นกล่าวคือ วิธี LT มีประสิทธิภาพดีกว่าวิธี OLS อย่างน้อยประมาณ 117.41% 109.42% 96.48% และ 79.89% ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น วิธี LT มีประสิทธิภาพดีกว่าวิธี NG อย่างน้อยประมาณ 73.97% 72.31% 65.71%และ 53.23% ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น และวิธี LT มีประสิทธิภาพดีกว่าวิธี JR อย่างน้อยประมาณ 43.15% 39.67% 30.02% และ21.02% ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น นอกจากนี้วิธี LT มีประสิทธิภาพสูงขึ้นเมื่อเลขชี้กำลังสูงสุด (MB) เพิ่มขึ้น

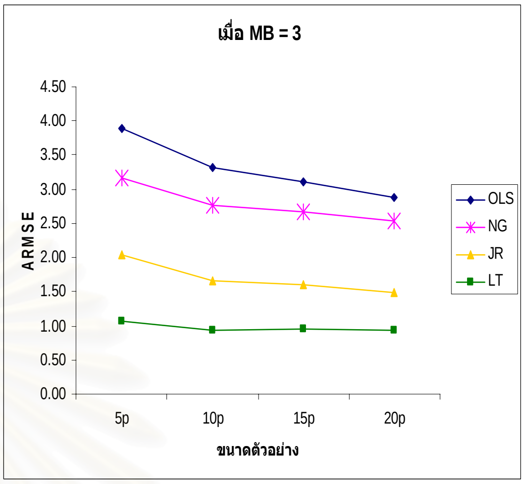
ส่วนค่า DIFF ของทั้งสองตาราง (ตารางที่ 4.1 - 4.2) จะสรุปได้ว่า ประสิทธิภาพของวิธี LT เพิ่มขึ้นเมื่อความแปรปรวนของความคลาดเคลื่อน (σ_{ε}^{2}) สูงขึ้นในทุกกรณีที่ศึกษา เพราะว่าวิธี LT เป็นวิธีที่สามารถแก้ปัญหาตัวแปรอิสระมีพหุสัมพันธ์กันได้อย่างเต็มรูปแบบ ดังนั้นเมื่อตัวแปรตามมีการเปลี่ยนแปลงไปจึงไม่ส่งผลต่อการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามมากนัก ในขณะที่วิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามวิธีการอื่นๆมีค่า ARMSE เพิ่มขึ้นมาก

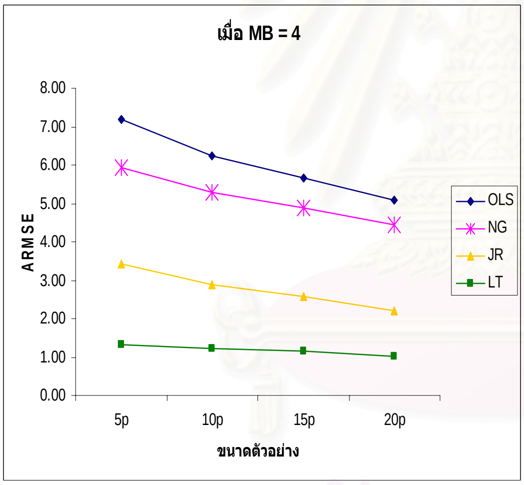
ตารางที่ 4.3 แสดงค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อ $\sigma_{\varepsilon}^{2} = 6$

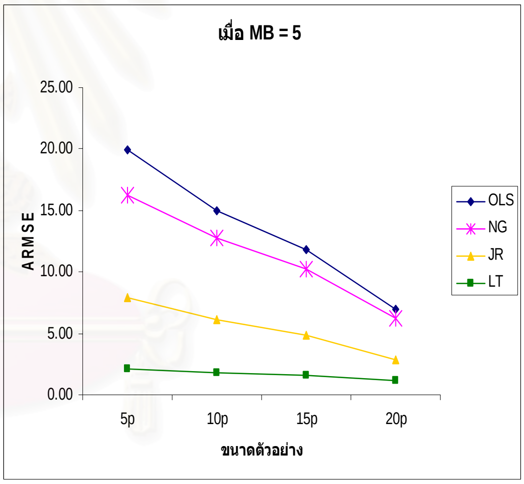
MB	วิธี	ขนาดตัวอย่าง (n)															
		5p				10p				15p				20p			
		OLS	NG	JR	LT	OLS	NG	JR	LT	OLS	NG	JR	LT	OLS	NG	JR	LT
2	ARMSE	2.140	1.797	1.372	0.875	1.838	1.587	1.183	0.775	1.607	1.407	1.022	0.719	1.470	1.298	0.944	0.715
	(SD)	(0.507)	(0.379)	(0.282)	(0.176)	(0.347)	(0.262)	(0.190)	(0.116)	(0.211)	(0.182)	(0.132)	(0.090)	(0.180)	(0.152)	(0.104)	(0.076)
	DIFF	144.435	105.291	56.724	0.000	137.111	104.657	52.622	0.000	123.626	95.826	42.215	0.000	105.544	81.531	32.076	0.000
3	ARMSE	3.892	3.173	2.036	1.066	3.312	2.768	1.661	0.936	3.105	2.662	1.604	0.958	2.880	2.528	1.482	0.940
	(SD)	(0.906)	(0.720)	(0.452)	(0.220)	(0.678)	(0.544)	(0.319)	(0.178)	(0.574)	(0.463)	(0.277)	(0.154)	(0.460)	(0.396)	(0.173)	(0.097)
	DIFF	264.957	197.581	90.951	0.000	253.820	195.662	77.465	0.000	224.169	177.988	67.478	0.000	206.526	169.017	57.707	0.000
4	ARMSE	7.202	5.935	3.429	1.316	6.226	5.293	2.867	1.204	5.671	4.897	2.567	1.136	5.080	4.431	2.211	1.030
	(SD)	(1.573)	(1.277)	(0.701)	(0.260)	(1.105)	(0.890)	(0.480)	(0.200)	(0.913)	(0.678)	(0.304)	(0.131)	(0.580)	(0.496)	(0.239)	(0.110)
	DIFF	447.348	351.064	160.610	0.000	416.938	339.430	138.027	0.000	399.173	331.054	125.957	0.000	393.107	330.114	114.639	0.000
5	ARMSE	19.954	16.294	7.902	2.091	14.988	12.805	6.093	1.778	11.765	10.271	4.890	1.536	6.973	6.172	2.887	1.206
	(SD)	(4.536)	(3.380)	(1.629)	(0.404)	(2.817)	(2.380)	(1.123)	(0.297)	(1.717)	(1.347)	(0.630)	(0.197)	(0.887)	(0.740)	(0.323)	(0.124)
	DIFF	854.189	679.170	277.879	0.000	742.780	620.029	242.623	0.000	666.050	568.772	218.407	0.000	478.336	411.860	139.454	0.000
6	ARMSE	61.937	51.140	22.200	3.980	44.883	38.448	15.760	2.771	17.188	15.009	6.263	1.622	11.437	9.523	4.656	1.400
	(SD)	(14.814)	(11.867)	(5.080)	(0.812)	(9.144)	(7.743)	(3.121)	(0.527)	(3.076)	(2.575)	(1.054)	(0.254)	(1.563)	(1.162)	(0.543)	(0.156)
	DIFF	1456.284	1184.989	457.817	0.000	1519.740	1287.514	468.748	0.000	959.941	825.567	286.236	0.000	716.718	580.039	232.506	0.000

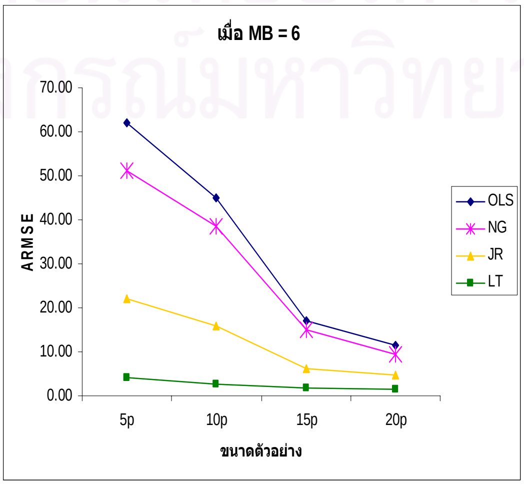
<u>รูปที่ 4.3.1</u> แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี OLS วิธี NG วิธี JR และวิธี LT ระหว่างค่า ARMSE และขนาดตัวอย่าง (n) เมื่อ $\sigma_{\varepsilon}^{2} = 6$



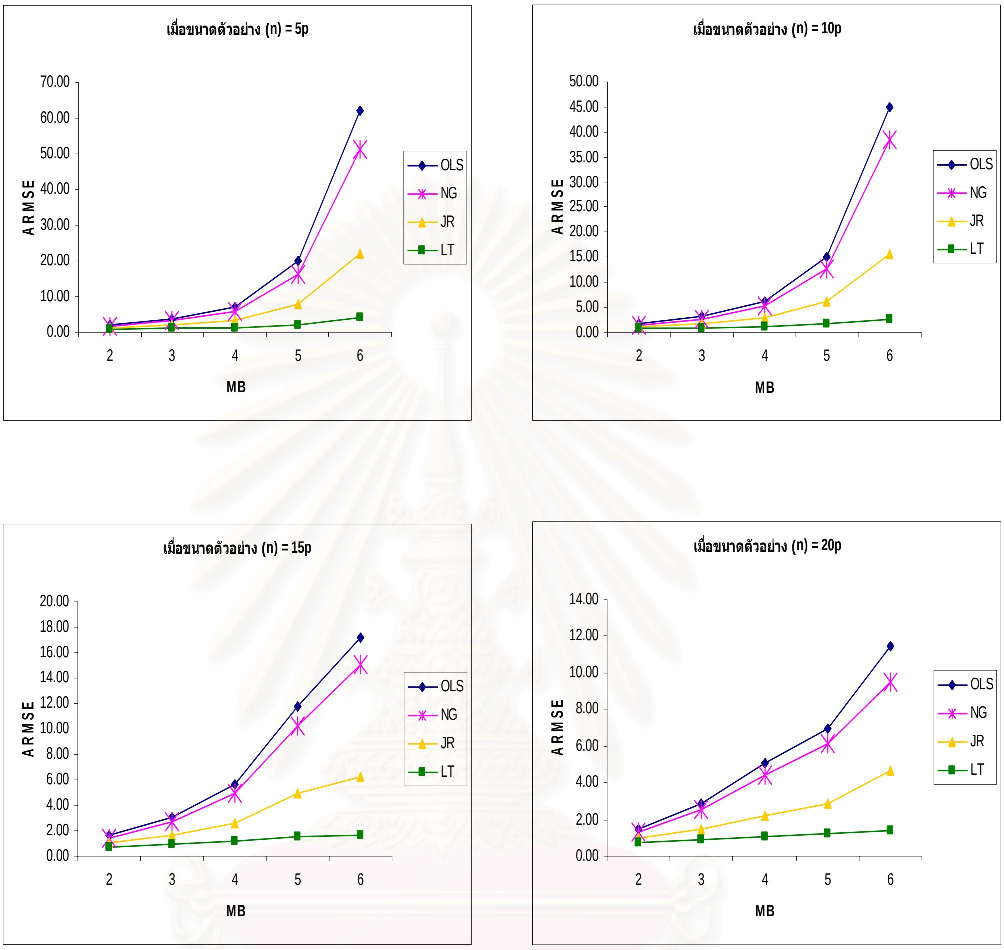








<u>รูปที่ 4.3.2</u> **แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี OLS วิธี NG วิธี JR และวิธี LT ระหว่างค่า ARMSE และเลขชี้กำลังสูงสุด (MB) เมื่อ** $\sigma_{\varepsilon}^{2} = 6$



จากตารางที่ 4.3 และกราฟรูปที่ 4.3.1 – 4.3.2 ผลสรุปคือ

ค่า ARMSE ของ 4 วิธี เรียงลำดับจากน้อยไปมากได้แก่ วิธี LT วิธี JR วิธี NG และวิธี OLS ตามลำดับ สำหรับทุกกรณี

เมื่อเลขชี้กำลังสูงสุด (MB) ของตัวแบบคงที่

การเพิ่มขนาดตัวอย่าง (n) ส่งผลให้ค่า ARMSE ของทุกวิธีมีแนวโน้มลดลงเพราะขนาดตัวอย่างที่เพิ่มขึ้นจะส่งผลทำให้ค่าคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามลดลงซึ่งวิธีการประมาณค่าพารามิเตอร์ที่ดีที่สุดคือ วิธี LT สำหรับทุกขนาดตัวอย่าง รองลงมาคือวิธี JR วิธี NG และวิธี OLS ตามลำดับ เพราะวิธี LT เป็นวิธีการที่สามารถแก้ปัญหาตัวแปรอิสระมีพหุสัมพันธ์กัน (multicollinearity) ได้อย่างเต็มรูปแบบ ดังนั้นเมื่อขนาดตัวอย่างน้อยโอกาสที่จะเกิดค่าผิดปกติในตัวแปรอิสระมากทำให้วิธี LT จะมีประสิทธิภาพมาก และเมื่อขนาดตัวอย่างมากขึ้นโอกาสที่จะเกิดค่าผิดปกติในตัวแปรอิสระลดลงประสิทธิภาพของวิธี LT นี้จึงลดลงด้วย กล่าวคือวิธี LT เป็นวิธีการที่ดีเมื่อขนาดตัวอย่างไม่มากนัก

เมื่อขนาดตัวอย่าง (n) ของตัวแบบคงที่

เมื่อเลขชี้กำลังสูงสุด (MB) เพิ่มขึ้นส่งผลให้ ARMSE ของทุกวิธีมีแนวโน้มเพิ่มขึ้นเพราะเลขชี้กำลังสูงสุด (MB) ที่เพิ่มขึ้นจะทำให้ตัวแปรอิสระมีความสัมพันธ์กัน (multicollinearity) มากขึ้น ทำให้ค่าคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามเพิ่มขึ้น ซึ่งวิธีการประมาณค่าพารามิเตอร์ที่ดีที่สุด คือวิธี LT สำหรับทุกค่าเลขชี้กำลังสูงสุด (MB) รองลงมาคือวิธี JR วิธี NG และวิธี OLS ตามลำดับ เพราะวิธี LT เป็นวิธีการที่สามารถแก้ปัญหาตัวแปรอิสระมีพหุสัมพันธ์กันได้อย่างเต็มรูปแบบ

เมื่อพิจารณาตารางที่ 4.2 และตารางที่ 4.3 จะเห็นได้ว่าเมื่อความแปรปรวนของความคลาดเคลื่อน (σ_ε^2) เพิ่มขึ้นทำให้ค่า ARMSE ของทุกกรณีเพิ่มขึ้นด้วย ทั้งนี้เพราะว่าเมื่อ σ_ε^2 เพิ่มขึ้นจะส่งผลให้การกระจายของตัวแปรตามเพิ่มขึ้น ส่งผลให้ค่าประมาณของพารามิเตอร์ที่ได้สูงขึ้น

จากข้างต้นสรุปได้ว่าปัจจัยที่ส่งผลต่อค่า ARMSE ได้แก่ ขนาดตัวอย่าง (n) เลขชี้กำลังสูงสุด (MB) และความแปรปรวนของความคลาดเคลื่อน (σ_ε^2) ซึ่งค่า ARMSE จะแปรผกผันกับขนาดตัวอย่าง (n) (รูปที่ 4.3.1) แต่จะแปรผันตามเลขชี้กำลังสูงสุด (MB) (รูปที่ 4.3.2) และความแปรปรวนของความคลาดเคลื่อน (σ_ε^2) ซึ่งวิธี LT จะมีความคงเส้นคงวามากกว่าวิธี OLS วิธี NG และวิธี JR เพราะค่า ARMSE ของวิธี LT ลู่เข้าสู่ศูนย์มากกว่าอีก 3 วิธีในทุกกรณี ดังนั้นวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามวิธี LT จะประมาณค่าพารามิเตอร์ได้ใกล้เคียงกับพารามิเตอร์ที่แท้จริงมากกว่าอีก 3 วิธี

พิจารณาค่า DIFF ของแต่ละวิธีการที่ใช้ประมาณค่าพารามิเตอร์ในทุกขนาดตัวอย่าง จะได้ว่าประสิทธิภาพของวิธี LT โดดเด่นกว่าอีก 3 วิธีอย่างชัดเจน โดยประสิทธิภาพของวิธี LT จะลดลง

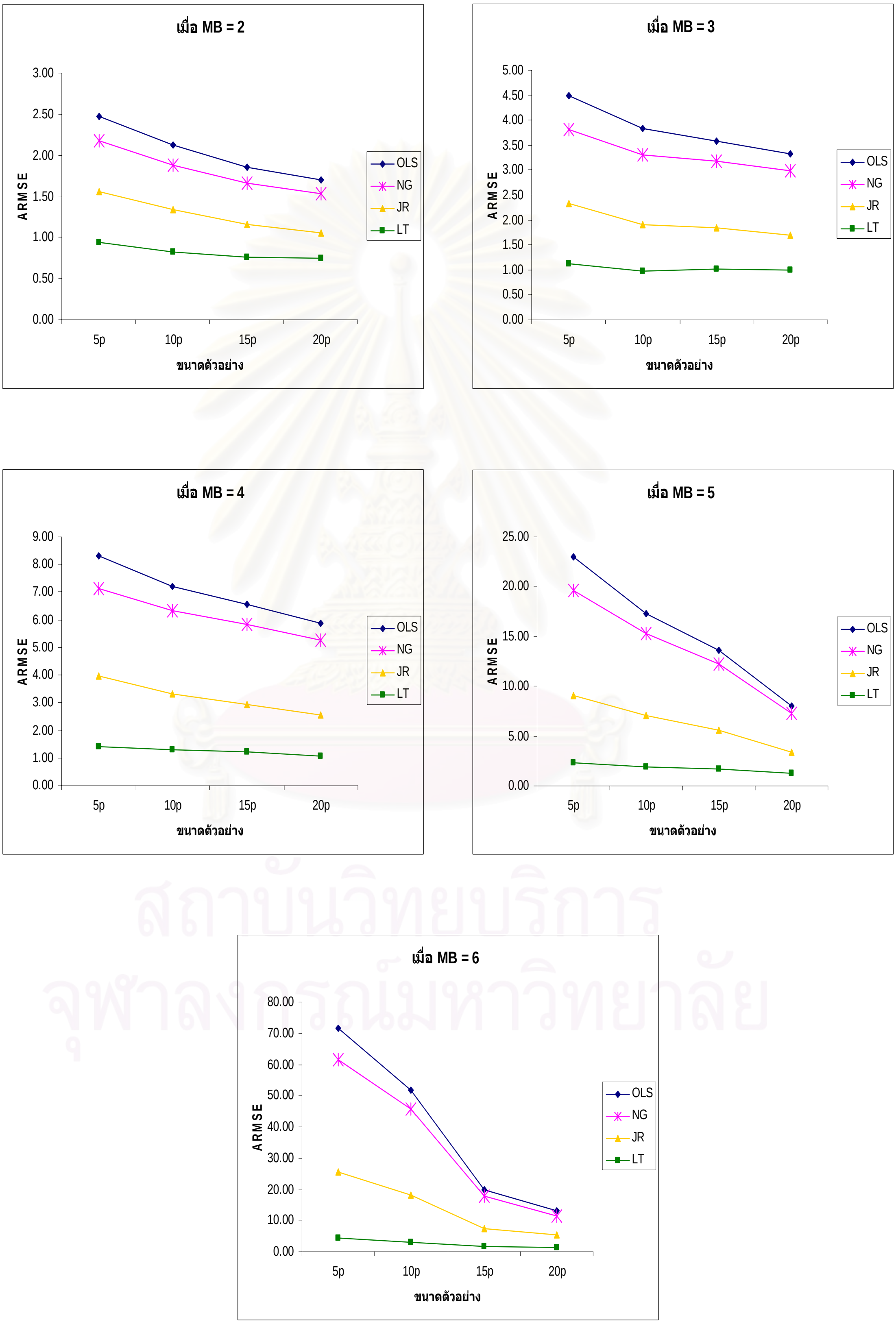
เมื่อขนาดตัวอย่างเพิ่มขึ้นกล่าวคือ วิธี LT มีประสิทธิภาพดีกว่าวิธี OLS อย่างน้อยประมาณ 144.43% 137.11% 123.63% และ 105.54% ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้นวิธี LT มีประสิทธิภาพดีกว่าวิธี NG อย่างน้อยประมาณ 105.29% 104.66% 95.83% และ 81.53% ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น และวิธี LT มีประสิทธิภาพดีกว่าวิธี JR อย่างน้อยประมาณ 56.72% 52.62% 42.22% และ 32.08% ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น นอกจากนี้วิธี LT มีประสิทธิภาพสูงขึ้นเมื่อเลขชี้กำลังสูงสุด (MB) เพิ่มขึ้น

ส่วนค่า DIFF ของทั้งสองตาราง (ตารางที่ 4.2 - 4.3) จะสรุปได้ว่า ประสิทธิภาพของวิธี LT เพิ่มขึ้นเมื่อความแปรปรวนของความคลาดเคลื่อน (σ_{ε}^{2}) สูงขึ้นในทุกกรณีที่ศึกษา เพราะว่าวิธี LT เป็นวิธีที่สามารถแก้ปัญหาตัวแปรอิสระมีพหุสัมพันธ์กันได้อย่างเต็มรูปแบบ ดังนั้นเมื่อตัวแปรตามมีการเปลี่ยนแปลงไปจึงไม่ส่งผลต่อการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามมากนัก ในขณะที่วิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามวิธีการอื่นๆมีค่า ARMSE เพิ่มขึ้นมาก

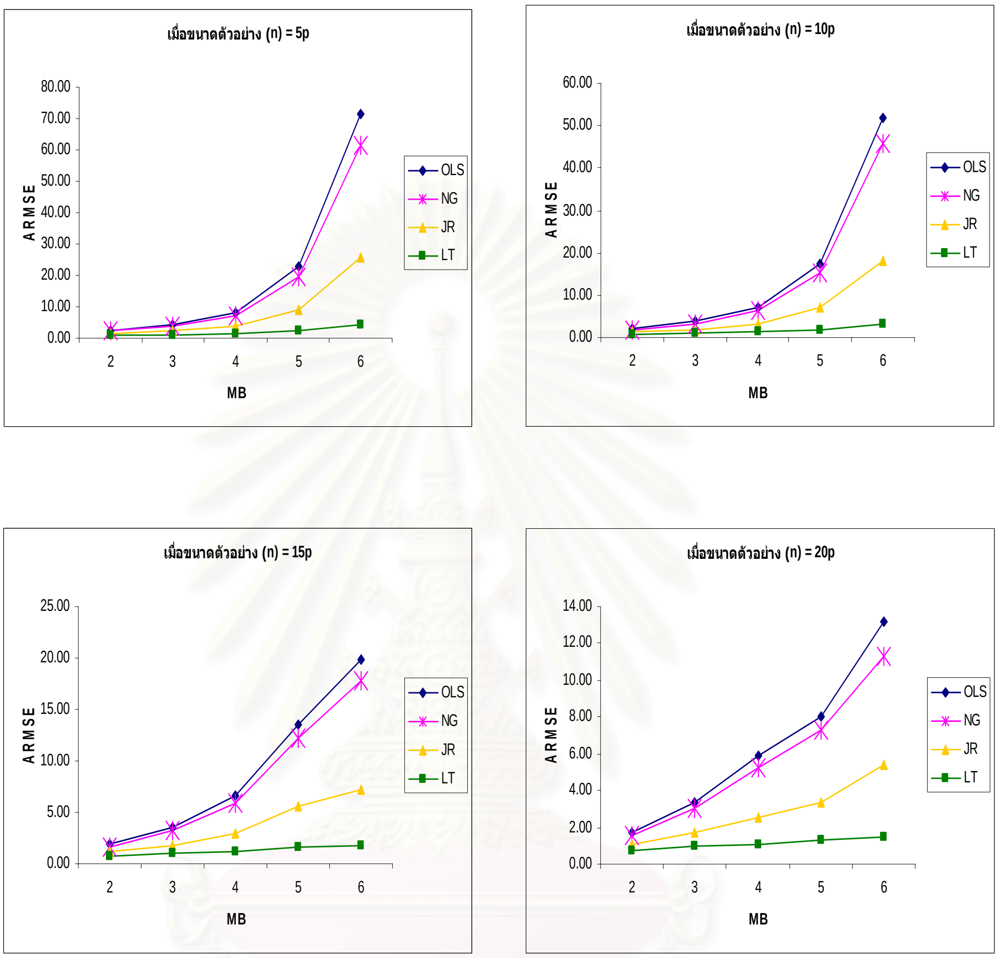
ตารางที่ 4.4 แสดงค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อ $\sigma_{\varepsilon}^{2} = 8$

MB	วิธี	ขนาดตัวอย่าง (n)															
		5p				10p				15p				20p			
		OLS	NG	JR	LT	OLS	NG	JR	LT	OLS	NG	JR	LT	OLS	NG	JR	LT
2	ARMSE	2.471	2.173	1.560	0.934	2.123	1.883	1.338	0.821	1.856	1.662	1.153	0.756	1.697	1.532	1.061	0.751
	(SD)	(0.597)	(0.454)	(0.307)	(0.181)	(0.382)	(0.325)	(0.221)	(0.133)	(0.293)	(0.247)	(0.164)	(0.103)	(0.227)	(0.199)	(0.125)	(0.088)
	DIFF	164.711	132.788	67.154	0.000	158.548	129.351	63.014	0.000	145.379	119.794	52.413	0.000	125.956	103.973	41.233	0.000
3	ARMSE	4.494	3.817	2.336	1.133	3.824	3.308	1.902	0.982	3.585	3.169	1.836	1.009	3.325	2.996	1.694	0.988
	(SD)	(1.087)	(0.876)	(0.524)	(0.250)	(0.842)	(0.691)	(0.350)	(0.176)	(0.630)	(0.538)	(0.308)	(0.166)	(0.448)	(0.373)	(0.192)	(0.110)
	DIFF	296.567	236.798	106.177	0.000	289.515	236.898	93.744	0.000	255.302	214.083	81.923	0.000	236.470	203.161	71.372	0.000
4	ARMSE	8.316	7.132	3.950	1.413	7.189	6.315	3.300	1.282	6.548	5.825	2.953	1.202	5.865	5.259	2.541	1.078
	(SD)	(1.999)	(1.649)	(0.823)	(0.271)	(1.358)	(1.116)	(0.563)	(0.209)	(1.004)	(0.842)	(0.424)	(0.168)	(0.752)	(0.641)	(0.264)	(0.109)
	DIFF	488.508	404.727	179.506	0.000	460.903	392.666	157.432	0.000	444.973	384.729	145.714	0.000	443.940	387.731	135.677	0.000
5	ARMSE	23.041	19.612	9.121	2.301	17.307	15.288	7.031	1.939	13.585	12.201	5.640	1.659	8.052	7.311	3.326	1.276
	(SD)	(5.706)	(4.782)	(2.137)	(0.494)	(3.352)	(2.952)	(1.334)	(0.337)	(2.208)	(1.675)	(0.755)	(0.218)	(1.057)	(0.905)	(0.389)	(0.136)
	DIFF	901.304	752.288	296.376	0.000	792.389	688.285	262.519	0.000	718.966	635.532	239.999	0.000	530.962	472.894	160.607	0.000
6	ARMSE	71.518	61.372	25.633	4.486	51.827	45.800	18.197	3.087	19.847	17.851	7.229	1.753	13.206	11.313	5.371	1.501
	(SD)	(17.356)	(14.141)	(5.508)	(0.922)	(10.473)	(8.425)	(3.347)	(0.507)	(3.017)	(2.681)	(1.083)	(0.249)	(1.852)	(1.533)	(0.670)	(0.165)
	DIFF	1494.249	1268.078	471.400	0.000	1578.825	1383.593	489.453	0.000	1031.915	918.079	312.262	0.000	780.121	653.943	257.924	0.000

<u>รูปที่ 4.4.1</u> **แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี OLS วิธี NG วิธี JR และวิธี LT ระหว่างค่า ARMSE และขนาดตัวอย่าง (n) เมื่อ** $\sigma_{\varepsilon}^{2} = 8$



<u>รูปที่ 4.4.2</u> **แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี OLS วิธี NG วิธี JR และวิธี LT ระหว่างค่า ARMSE และเลขชี้กำลังสูงสุด (MB) เมื่อ** $\sigma_{\varepsilon}^{2} = 8$



จากตารางที่ 4.4 และกราฟรูปที่ 4.4.1 – 4.4.2 ผลสรุปคือ

ค่า ARMSE ของ 4 วิธี เรียงลำดับจากน้อยไปมากได้แก่ วิธี LT วิธี JR วิธี NG และวิธี OLS ตามลำดับ สำหรับทุกกรณี

<u>เมื่อเลขชี้กำลังสูงสุด (MB) ของตัวแบบคงที่</u>

การเพิ่มขนาดตัวอย่าง (n) ส่งผลให้ค่า ARMSE ของทุกวิธีมีแนวโน้มลดลงเพราะขนาดตัวอย่างที่เพิ่มขึ้นจะส่งผลทำให้ค่าคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามลดลงซึ่งวิธีการประมาณค่าพารามิเตอร์ที่ดีที่สุดคือ วิธี LT สำหรับทุกขนาดตัวอย่าง รองลงมาคือวิธี JR วิธี NG และวิธี OLS ตามลำดับ เพราะวิธี LT เป็นวิธีการที่สามารถแก้ปัญหาตัวแปรอิสระมีพหุสัมพันธ์กัน (multicollinearity) ได้อย่างเต็มรูปแบบ ดังนั้นเมื่อขนาดตัวอย่างน้อยโอกาสที่จะเกิดค่าผิดปกติในตัวแปรอิสระมากทำให้วิธี LT จะมีประสิทธิภาพมาก และเมื่อขนาดตัวอย่างมากขึ้นโอกาสที่จะเกิดค่าผิดปกติในตัวแปรอิสระลดลงประสิทธิภาพของวิธี LT นี้จึงลดลงด้วย กล่าวคือวิธี LT เป็นวิธีการที่ดีเมื่อขนาดตัวอย่างไม่มากนัก

<u>เมื่อขนาดตัวอย่าง (n) ของตัวแบบคงที่</u>

เมื่อเลขชี้กำลังสูงสุด (MB) เพิ่มขึ้นส่งผลให้ ARMSE ของทุกวิธีมีแนวโน้มเพิ่มขึ้นเพราะเลขชี้กำลังสูงสุด (MB) ที่เพิ่มขึ้นจะทำให้ตัวแปรอิสระมีความสัมพันธ์กัน (multicollinearity) มากขึ้น ทำให้ค่าคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามเพิ่มขึ้น ซึ่งวิธีการประมาณค่าพารามิเตอร์ที่ดีที่สุด คือวิธี LT สำหรับทุกค่าเลขชี้กำลังสูงสุด (MB) รองลงมาคือวิธี JR วิธี NG และวิธี OLS ตามลำดับ เพราะวิธี LT เป็นวิธีการที่สามารถแก้ปัญหาตัวแปรอิสระมีพหุสัมพันธ์กันได้อย่างเต็มรูปแบบ

เมื่อพิจารณาตารางที่ 4.3 และตารางที่ 4.4 จะเห็นได้ว่าเมื่อความแปรปรวนของความคลาดเคลื่อน (σ_ε^2) เพิ่มขึ้นทำให้ค่า ARMSE ของทุกกรณีเพิ่มขึ้นด้วย ทั้งนี้เพราะว่าเมื่อ σ_ε^2 เพิ่มขึ้นจะส่งผลให้การกระจายของตัวแปรตามเพิ่มขึ้น ส่งผลให้ค่าประมาณของพารามิเตอร์ที่ได้สูงขึ้น

จากข้างต้นสรุปได้ว่าปัจจัยที่ส่งผลต่อค่า ARMSE ได้แก่ ขนาดตัวอย่าง (n) เลขชี้กำลังสูงสุด (MB) และความแปรปรวนของความคลาดเคลื่อน (σ_ε^2) ซึ่งค่า ARMSE จะแปรผกผันกับขนาดตัวอย่าง (n) (รูปที่ 4.4.1) แต่จะแปรผันตามเลขชี้กำลังสูงสุด (MB) (รูปที่ 4.4.2) และความแปรปรวนของความคลาดเคลื่อน (σ_ε^2) ซึ่งวิธี LT จะมีความคงเส้นคงวามากกว่าวิธี OLS วิธี NG และวิธี JR เพราะค่า ARMSE ของวิธี LT ลู่เข้าสู่ศูนย์มากกว่าอีก 3 วิธีในทุกกรณี ดังนั้นวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามวิธี LT จะประมาณค่าพารามิเตอร์ได้ใกล้เคียงกับพารามิเตอร์ที่แท้จริงมากกว่าอีก 3 วิธี

พิจารณาค่า DIFF ของแต่ละวิธีการที่ใช้ประมาณค่าพารามิเตอร์ในทุกขนาดตัวอย่าง จะได้ว่าประสิทธิภาพของวิธี LT โดดเด่นกว่าอีก 3 วิธีอย่างชัดเจน โดยประสิทธิภาพของวิธี LT จะลดลง

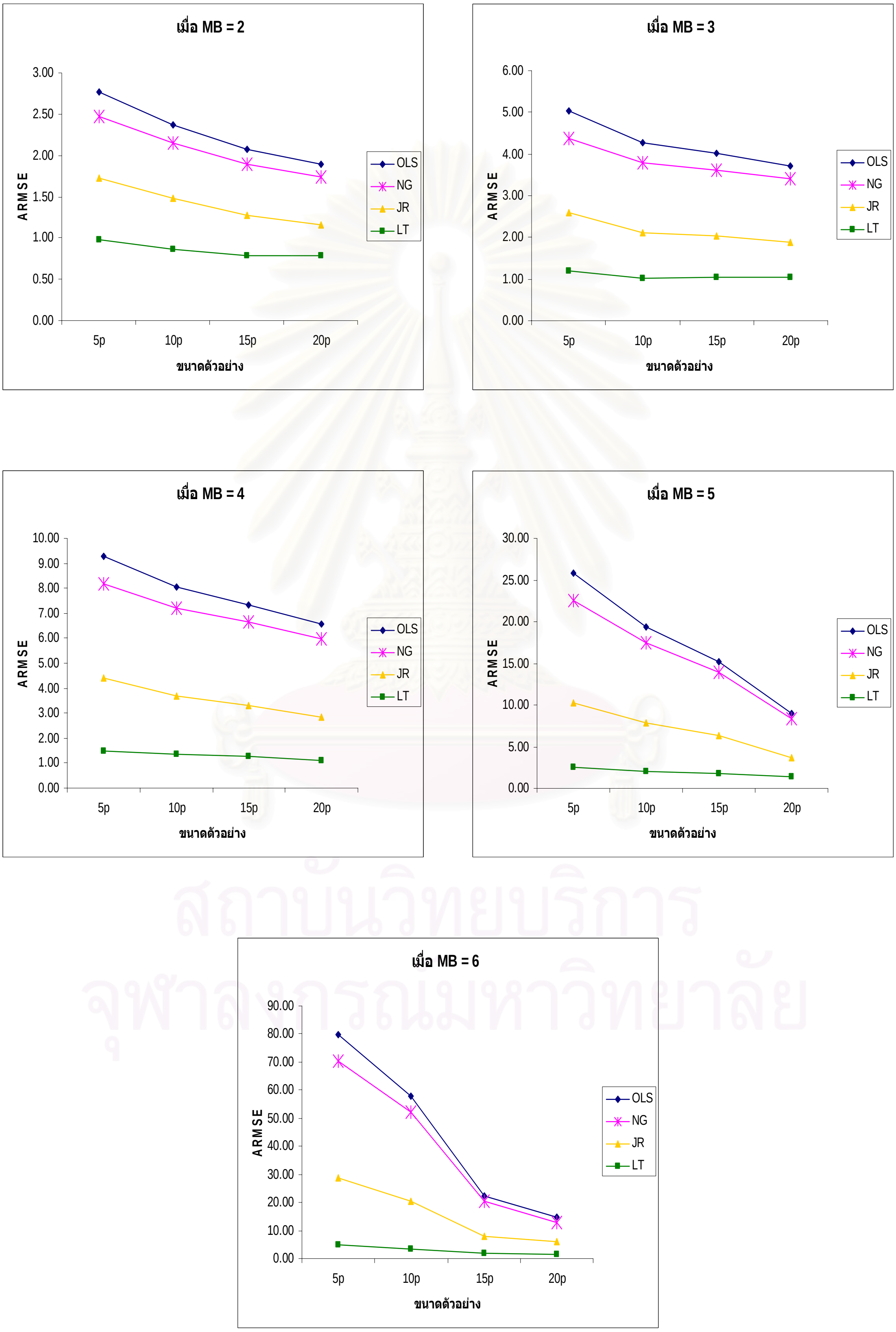
เมื่อขนาดตัวอย่างเพิ่มขึ้นกล่าวคือ วิธี LT มีประสิทธิภาพดีกว่าวิธี OLS อย่างน้อยประมาณ 164.71% 158.55% 145.38% และ 125.96% ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้นวิธี LT มีประสิทธิภาพดีกว่าวิธี NG อย่างน้อยประมาณ 132.79% 129.35% 119.79% และ 103.97% ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น และวิธี LT มีประสิทธิภาพดีกว่าวิธี JR อย่างน้อยประมาณ 67.15% 63.01% 52.41% และ 41.23% ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น นอกจากนี้วิธี LT มีประสิทธิภาพสูงขึ้นเมื่อเลขชี้กำลังสูงสุด (MB) เพิ่มขึ้น

ส่วนค่า DIFF ของทั้งสองตาราง (ตารางที่ 4.3 - 4.4) จะสรุปได้ว่า ประสิทธิภาพของวิธี LT เพิ่มขึ้นเมื่อความแปรปรวนของความคลาดเคลื่อน (σ_{ε}^{2}) สูงขึ้นในทุกกรณีที่ศึกษา เพราะว่าวิธี LT เป็นวิธีที่สามารถแก้ปัญหาตัวแปรอิสระมีพหุสัมพันธ์กันได้อย่างเต็มรูปแบบ ดังนั้นเมื่อตัวแปรตามมีการเปลี่ยนแปลงไปจึงไม่ส่งผลต่อการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามมากนัก ในขณะที่วิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามวิธีการอื่นๆมีค่า ARMSE เพิ่มขึ้นมาก

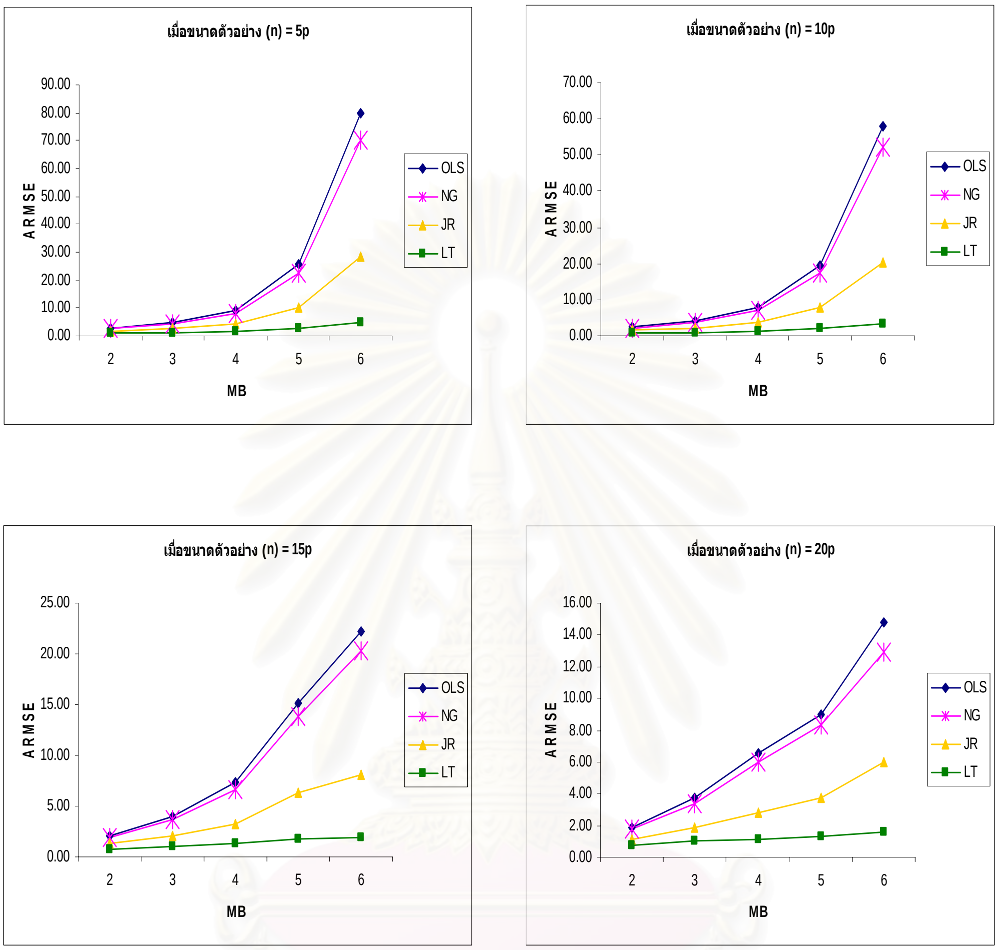
ตารางที่ 4.5 แสดงค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อ $\sigma_{\varepsilon}^{2} = 10$

MB	วิธี	ขนาดตัวอย่าง (n)															
		5p				10p				15p				20p			
		OLS	NG	JR	LT	OLS	NG	JR	LT	OLS	NG	JR	LT	OLS	NG	JR	LT
2	ARMSE	2.763	2.468	1.727	0.984	2.373	2.144	1.476	0.860	2.075	1.888	1.268	0.787	1.897	1.738	1.164	0.781
	(SD)	(0.654)	(0.521)	(0.355)	(0.197)	(0.448)	(0.354)	(0.238)	(0.128)	(0.272)	(0.244)	(0.164)	(0.099)	(0.232)	(0.204)	(0.128)	(0.083)
	DIFF	180.834	150.846	75.563	0.000	176.018	149.384	71.656	0.000	163.580	139.799	61.128	0.000	142.967	122.517	49.027	0.000
3	ARMSE	5.024	4.380	2.601	1.192	4.276	3.783	2.115	1.021	4.008	3.613	2.040	1.053	3.718	3.406	1.881	1.030
	(SD)	(1.169)	(0.994)	(0.577)	(0.246)	(0.875)	(0.743)	(0.406)	(0.194)	(0.741)	(0.629)	(0.352)	(0.169)	(0.594)	(0.534)	(0.219)	(0.106)
	DIFF	321.652	267.590	118.312	0.000	318.653	270.391	107.109	0.000	280.610	243.120	93.752	0.000	260.961	230.641	82.650	0.000
4	ARMSE	9.298	8.181	4.409	1.499	8.038	7.209	3.681	1.350	7.321	6.638	3.293	1.259	6.558	5.985	2.833	1.120
	(SD)	(2.031)	(1.760)	(0.902)	(0.297)	(1.427)	(1.212)	(0.616)	(0.224)	(1.179)	(0.919)	(0.390)	(0.145)	(0.749)	(0.670)	(0.306)	(0.120)
	DIFF	520.225	445.694	194.090	0.000	495.385	434.022	172.674	0.000	481.386	427.086	161.455	0.000	485.291	434.193	152.829	0.000
5	ARMSE	25.761	22.519	10.195	2.487	19.350	17.456	7.857	2.082	15.189	13.889	6.301	1.768	9.002	8.308	3.713	1.338
	(SD)	(5.857)	(4.671)	(2.101)	(0.480)	(3.637)	(3.244)	(1.448)	(0.348)	(2.217)	(1.821)	(0.811)	(0.226)	(1.145)	(0.996)	(0.416)	(0.138)
	DIFF	935.993	805.614	309.998	0.000	829.484	738.505	277.414	0.000	759.252	685.710	256.441	0.000	572.745	520.903	177.446	0.000
6	ARMSE	79.960	70.321	28.658	4.933	57.944	52.240	20.344	3.366	22.190	20.330	8.080	1.870	14.765	12.876	6.001	1.589
	(SD)	(19.125)	(16.318)	(6.558)	(1.006)	(11.805)	(10.520)	(4.029)	(0.640)	(3.971)	(3.488)	(1.360)	(0.293)	(2.018)	(1.571)	(0.699)	(0.177)
	DIFF	1521.019	1325.609	480.980	0.000	1621.245	1451.806	504.325	0.000	1086.568	987.108	332.036	0.000	829.204	710.318	277.626	0.000

<u>รูปที่ 4.5.1</u> **แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี OLS วิธี NG วิธี JR และวิธี LT ระหว่างค่า ARMSE และขนาดตัวอย่าง (n) เมื่อ** $\sigma_{\varepsilon}^{2} = 10$



<u>รูปที่ 4.5.2</u> **แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี OLS วิธี NG วิธี JR และวิธี LT ระหว่างค่า ARMSE และเลขชี้กำลังสูงสุด (MB) เมื่อ** $\sigma_{\varepsilon}^{2} = 10$



จากตารางที่ 4.5 และกราฟรูปที่ 4.5.1 – 4.5.2 ผลสรุปคือ

ค่า ARMSE ของ 4 วิธี เรียงลำดับจากน้อยไปมากได้แก่ วิธี LT วิธี JR วิธี NG และวิธี OLS ตามลำดับ สำหรับทุกกรณี

<u>เมื่อเลขชี้กำลังสูงสุด (MB) ของตัวแบบคงที่</u>

การเพิ่มขนาดตัวอย่าง (n) ส่งผลให้ค่า ARMSE ของทุกวิธีมีแนวโน้มลดลงเพราะขนาดตัวอย่างที่เพิ่มขึ้นจะส่งผลทำให้ค่าคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามลดลงซึ่งวิธีการประมาณค่าพารามิเตอร์ที่ดีที่สุดคือ วิธี LT สำหรับทุกขนาดตัวอย่าง รองลงมาคือวิธี JR วิธี NG และวิธี OLS ตามลำดับ เพราะวิธี LT เป็นวิธีการที่สามารถแก้ปัญหาตัวแปรอิสระมีพหุสัมพันธ์กัน (multicollinearity) ได้อย่างเต็มรูปแบบ ดังนั้นเมื่อขนาดตัวอย่างน้อยโอกาสที่จะเกิดค่าผิดปกติในตัวแปรอิสระมากทำให้วิธี LT จะมีประสิทธิภาพมาก และเมื่อขนาดตัวอย่างมากขึ้นโอกาสที่จะเกิดค่าผิดปกติในตัวแปรอิสระลดลงประสิทธิภาพของวิธี LT นี้จึงลดลงด้วย กล่าวคือวิธี LT เป็นวิธีการที่ดีเมื่อขนาดตัวอย่างไม่มากนัก

<u>เมื่อขนาดตัวอย่าง (n) ของตัวแบบคงที่</u>

เมื่อเลขชี้กำลังสูงสุด (MB) เพิ่มขึ้นส่งผลให้ ARMSE ของทุกวิธีมีแนวโน้มเพิ่มขึ้นเพราะเลขชี้กำลังสูงสุด (MB) ที่เพิ่มขึ้นจะทำให้ตัวแปรอิสระมีความสัมพันธ์กัน (multicollinearity) มากขึ้น ทำให้ค่าคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามเพิ่มขึ้น ซึ่งวิธีการประมาณค่าพารามิเตอร์ที่ดีที่สุด คือวิธี LT สำหรับทุกค่าเลขชี้กำลังสูงสุด (MB) รองลงมาคือวิธี JR วิธี NG และวิธี OLS ตามลำดับ เพราะวิธี LT เป็นวิธีการที่สามารถแก้ปัญหาตัวแปรอิสระมีพหุสัมพันธ์กันได้อย่างเต็มรูปแบบ

เมื่อพิจารณาตารางที่ 4.4 และตารางที่ 4.5 จะเห็นได้ว่าเมื่อความแปรปรวนของความคลาดเคลื่อน (σ_ε^2) เพิ่มขึ้นทำให้ค่า ARMSE ของทุกกรณีเพิ่มขึ้นด้วย ทั้งนี้เพราะว่าเมื่อ σ_ε^2 เพิ่มขึ้นจะส่งผลให้การกระจายของตัวแปรตามเพิ่มขึ้น ส่งผลให้ค่าประมาณของพารามิเตอร์ที่ได้สูงขึ้น

จากข้างต้นสรุปได้ว่าปัจจัยที่ส่งผลต่อค่า ARMSE ได้แก่ ขนาดตัวอย่าง (n) เลขชี้กำลังสูงสุด (MB) และความแปรปรวนของความคลาดเคลื่อน (σ_ε^2) ซึ่งค่า ARMSE จะแปรผกผันกับขนาดตัวอย่าง (n) (รูปที่ 4.5.1) แต่จะแปรผันตามเลขชี้กำลังสูงสุด (MB) (รูปที่ 4.5.2) และความแปรปรวนของความคลาดเคลื่อน (σ_ε^2) ซึ่งวิธี LT จะมีความคงเส้นคงวามากกว่าวิธี OLS วิธี NG และวิธี JR เพราะค่า ARMSE ของวิธี LT ลู่เข้าสู่ศูนย์มากกว่าอีก 3 วิธีในทุกกรณี ดังนั้นวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามวิธี LT จะประมาณค่าพารามิเตอร์ได้ใกล้เคียงกับพารามิเตอร์ที่แท้จริงมากกว่าอีก 3 วิธี

พิจารณาค่า DIFF ของแต่ละวิธีการที่ใช้ประมาณค่าพารามิเตอร์ในทุกขนาดตัวอย่าง จะได้ว่าประสิทธิภาพของวิธี LT โดดเด่นกว่าอีก 3 วิธีอย่างชัดเจน โดยประสิทธิภาพของวิธี LT จะลดลง

เมื่อขนาดตัวอย่างเพิ่มขึ้นกล่าวคือ วิธี LT มีประสิทธิภาพดีกว่าวิธี OLS อย่างน้อยประมาณ 180.83% 176.02% 163.58% และ 142.97% ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้นวิธี LT มีประสิทธิภาพดีกว่าวิธี NG อย่างน้อยประมาณ 150.85% 149.38% 139.80% และ 122.52% ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น และวิธี LT มีประสิทธิภาพดีกว่าวิธี JR อย่างน้อยประมาณ 75.56% 71.66% 61.13% และ49.03% ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น นอกจากนี้วิธี LT มีประสิทธิภาพสูงขึ้นเมื่อเลขชี้กำลังสูงสุด (MB) เพิ่มขึ้น

ส่วนค่า DIFF ของทั้งสองตาราง (ตารางที่ 4.4 - 4.5) จะสรุปได้ว่า ประสิทธิภาพของวิธี LT เพิ่มขึ้นเมื่อความแปรปรวนของความคลาดเคลื่อน (σ_{ε}^{2}) สูงขึ้นในทุกกรณีที่ศึกษา เพราะว่าวิธี LT เป็นวิธีที่สามารถแก้ปัญหาตัวแปรอิสระมีพหุสัมพันธ์กันได้อย่างเต็มรูปแบบ ดังนั้นเมื่อตัวแปรตามมีการเปลี่ยนแปลงไปจึงไม่ส่งผลต่อการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามมากนัก ในขณะที่วิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามวิธีการอื่นๆมีค่า ARMSE เพิ่มขึ้นมาก

ผลสรุปของตารางที่ 4.1 – 4.5

ผลสรุปโดยรวมของข้อมูลในตารางที่ 4.1 – 4.5 เป็นดังนี้

1. ขนาดตัวอย่าง (n)

ค่า ARMSE มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น เพราะเมื่อขนาดตัวอย่างเพิ่มขึ้น หมายความว่าเราได้ข้อมูลมากขึ้นโอกาสที่จะเกิดค่าผิดปกติในตัวแปรอิสระลดลง จึงทำให้ค่า ARMSE ลดลง และส่งผลให้การประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามถูกต้องมากขึ้น

2. เลขชี้กำลังสูงสุด (MB)

ค่า ARMSE มีแนวโน้มเพิ่มขึ้นเมื่อเลขชี้กำลังสูงสุด (MB) เพิ่มขึ้น เพราะเมื่อเลขชี้กำลังสูงสุดเพิ่มขึ้น หมายความว่าตัวแปรอิสระมีความสัมพันธ์กันมากขึ้น จึงทำให้ค่า ARMSE เพิ่มขึ้น และส่งผลให้ความถูกต้องในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามลดลง

3. ความแปรปรวนของความคลาดเคลื่อน (σ_ε^2)

ค่า ARMSE มีแนวโน้มเพิ่มขึ้นเมื่อความแปรปรวนของค่าคลาดเคลื่อนเพิ่มขึ้น เพราะเมื่อความแปรปรวนของความคลาดเคลื่อนเพิ่มขึ้น หมายความว่าความคลาดเคลื่อนมีกระจายมากขึ้น จึงทำให้ค่า ARMSE เพิ่มขึ้น และส่งผลให้ความถูกต้องในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามลดลง

<u>ผู้วิจัยสามารถสรุปผลเกี่ยวกับการเปลี่ยนแปลงของค่า ARMSE ได้ดังนี้</u>

1. แปรผกผันกับขนาดตัวอย่าง
2. แปรผันตามเลขชี้กำลังสูงสุดและความแปรปรวนของความคลาดเคลื่อน

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบความถูกต้องของการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนาม ซึ่งวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามที่นำมาใช้มี 4 วิธี ดังนี้

1. วิธีกำลังสองน้อยสุด (Ordinary Least Squares method (OLS))
2. วิธีการ์ร็อตที่ไม่เป็นลบ (Nonnegative Garrote method (NG))
3. วิธีแจ็คไนฟริดจ์ (Jackknifed Ridge method (JR))
4. วิธีลิวไทป์ (Liu-Type method (LT))

โดยวิธีแรกเป็นแนวคิดแบบวิธีแบบฉบับ (Classical method) ซึ่งวิธีกำลังสองน้อยสุดเป็นวิธีพื้นฐานที่นิยมใช้กันอย่างแพร่หลาย แต่ไม่ได้พิจารณาถึงการเกิดปัญหาพหุสัมพันธ์ (multicollinearity) วิธีการ์ร็อตที่ไม่เป็นลบเป็นวิธีที่พัฒนามาจากตัวประมาณสไตน์ซึ่งอาศัยวิธีการหด (shrinkage method) โดยจะทำการหดตัวประมาณกำลังสองน้อยสุดที่อาจจะมีค่าสูงกว่าความเป็นจริงให้ต่ำลง แต่เป็นวิธีการที่ไม่ได้ช่วยแก้ปัญหาพหุสัมพันธ์ เพียงแต่ช่วยลดค่าคลาดเคลื่อนของตัวประมาณเท่านั้น ส่วนวิธีแจ็คไนฟริดจ์และวิธีลิวไทป์เป็นวิธีที่พัฒนามาจากตัวประมาณริดจ์ซึ่งอาศัยวิธีการทำโทษ (penalization method) ในการประมาณกำลังสองน้อยสุดวิธีการนี้สามารถช่วยแก้ปัญหาพหุสัมพันธ์ที่เกิดขึ้น โดยวิธีแจ็คไนฟริดจ์จะเป็นวิธีที่ช่วยลดความเอนเอียงให้กับตัวประมาณริดจ์แต่ไม่อาจช่วยแก้ปัญหาพหุสัมพันธ์ได้อย่างสมบูรณ์ได้ในขณะที่วิธีลิวไทป์เป็นวิธีการที่อาศัยพารามิเตอร์ในการประมาณค่าถึง 2 ตัว จึงทำให้สามารถแก้ปัญหาพหุสัมพันธ์ในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามได้อย่างสมบูรณ์ ผู้วิจัยได้ทำการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนาม 4 วิธีข้างต้น เพื่อศึกษาเปรียบเทียบว่าวิธีการใดเป็นวิธีการที่เหมาะสมที่สุด ซึ่งมีการกำหนดสถานการณ์ต่างๆในการวิจัยครั้งนี้ไว้ดังนี้

1. การแจกแจงของความคลาดเคลื่อนสุ่มในตัวแปรตาม (ε) เป็นแบบปกติที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 2 4 6 8 และ 10
2. ขนาดตัวอย่างที่ศึกษาเป็น 5p 10p 15p และ 20p
3. กำลังสูงสุดของตัวแปรอิสระที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบถดถอยพหุนามเป็น 2 3 4 5 และ 6

การสรุปผลว่าตัวประมาณใดเป็นตัวประมาณที่ดีที่สุดจะพิจารณาจากค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) โดยที่ตัวประมาณใดให้ค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง และค่าอัตราส่วนผลต่างของค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าต่ำสุดจะเป็นตัวประมาณที่ดีที่สุด ผลการวิจัยมีข้อสรุปดังนี้

5.1 สรุปผลการวิจัย

5.1.1 ผลการเปรียบเทียบค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

จากการเปรียบเทียบค่า ARMSE ของทั้ง 4 วิธีพบว่า วิธีที่ดีที่สุดคือวิธี LT รองลงมาคือวิธี JR วิธี NG และวิธี OLS ตามลำดับ สำหรับทุกขนาดตัวอย่าง เลขชี้กำลังสูงสุด และความแปรปรวนของค่าคลาดเคลื่อน เพราะวิธี LT เป็นวิธีการที่สามารถแก้ปัญหาพหุสัมพันธ์ (multicollinearity) ได้อย่างสมบูรณ์

5.1.2 ปัจจัยที่มีผลต่อค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

1. <u>ขนาดตัวอย่าง</u> (n)

ค่า ARMSE มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น เพราะเมื่อขนาดตัวอย่างเพิ่มขึ้นหมายความว่าเราได้ข้อมูลมากขึ้นโอกาสที่จะเกิดค่าผิดปกติในตัวแปรอิสระลดลง จึงทำให้ค่า ARMSE ลดลง และส่งผลให้การประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามถูกต้องมากขึ้น

2. <u>เลขชี้กำลังสูงสุดของตัวแปรอิสระที่ใช้ในการสร้างตัวแปรตามในตัวแบบถดถอยพหุนาม</u> (MB)

ค่า ARMSE มีแนวโน้มเพิ่มขึ้นเมื่อเลขชี้กำลังสูงสุด (MB) เพิ่มขึ้น เพราะเมื่อเลขชี้กำลังสูงสุดเพิ่มขึ้น หมายความว่าตัวแปรอิสระมีความสัมพันธ์กันมากขึ้น จึงทำให้ค่า ARMSE เพิ่มขึ้น และส่งผลให้ความถูกต้องในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามลดลง

3. <u>ความแปรปรวนของความคลาดเคลื่อน</u> (σ_{ε}^{2})

ค่า ARMSE มีแนวโน้มเพิ่มขึ้นเมื่อความแปรปรวนของค่าคลาดเคลื่อนเพิ่มขึ้น เพราะเมื่อความแปรปรวนของความคลาดเคลื่อนเพิ่มขึ้น หมายความว่าความคลาดเคลื่อนมีกระจายมากขึ้น จึงทำให้ค่า ARMSE เพิ่มขึ้น และส่งผลให้ความถูกต้องในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามลดลง

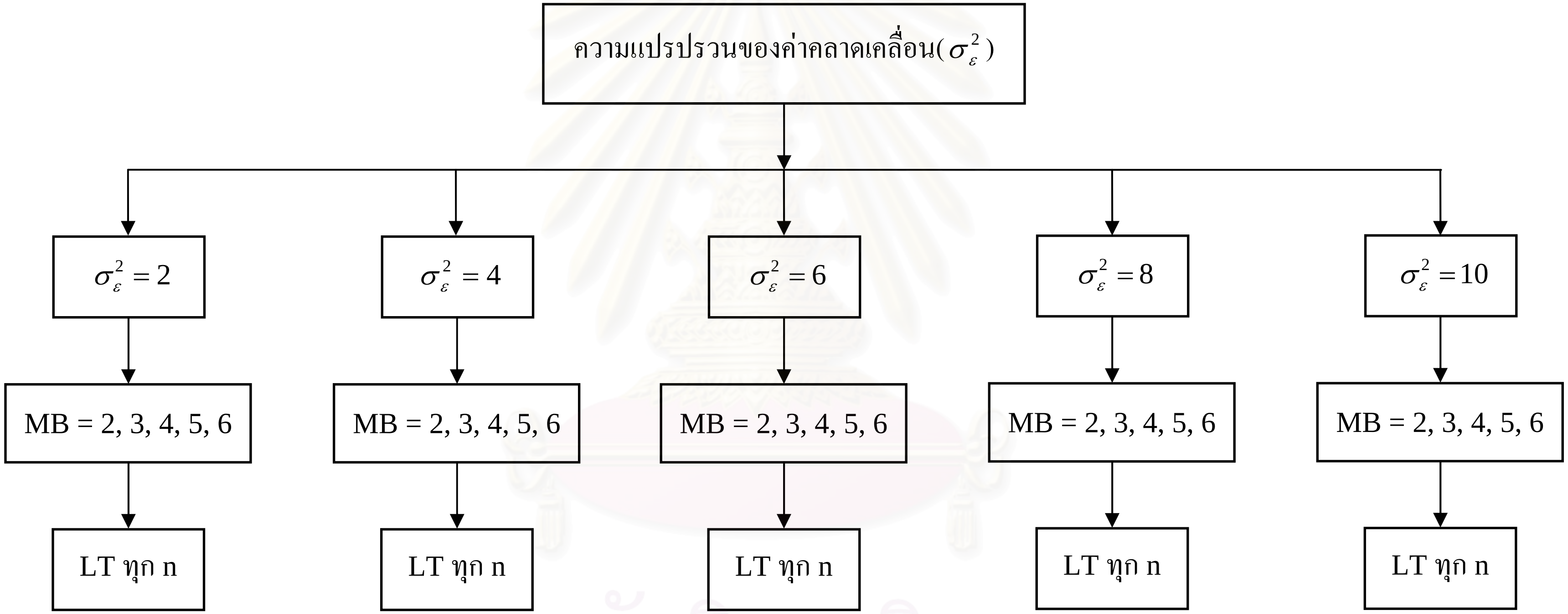
<u>ผู้วิจัยสามารถสรุปผลเกี่ยวกับการเปลี่ยนแปลงของค่า ARMSE ได้ดังนี้</u>

1. แปรผกผันกับขนาดตัวอย่าง
2. แปรผันตามเลขชี้กำลังสูงสุดและความแปรปรวนของความคลาดเคลื่อน

5.1.3 ผลสรุปการเลือกวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนาม

การเลือกวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามในการวิจัยครั้งนี้จะสรุปได้ว่า การประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี LT จะให้ผลดีทุกกรณี เพราะว่าสำหรับการวิเคราะห์ความถดถอยพหุนามนั้นได้อาศัยพื้นฐานมาจากการวิเคราะห์ความถดถอยเชิงเส้น ดังนั้นเทอมของตัวแปรอิสระจึงมีความสัมพันธ์กัน วิธีการที่ควรนำมาประมาณค่าพารามิเตอร์จึงควรเป็นวิธีการที่สามารถแก้ปัญหาพหุสัมพันธ์ได้อย่างสมบูรณ์ซึ่งตรงกับคุณสมบัติของวิธี LT และเพื่อให้ง่ายแก่การเข้าใจผู้วิจัยจึงได้แสดงแผนผังซึ่งแสดงการเลือกวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามในรูปที่ 5.1 ดังต่อไปนี้

รูปที่ 5.1 **แผนผังสรุปผลวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนาม สำหรับใช้ในเชิงทฤษฎี**



เมื่อ n คือขนาดตัวอย่าง เท่ากับ 5p 10p 15p และ 20p

โดยที่ p คือจำนวนพจน์พหุนามในตัวแบบโดยไม่นับพจน์ค่าคงที่

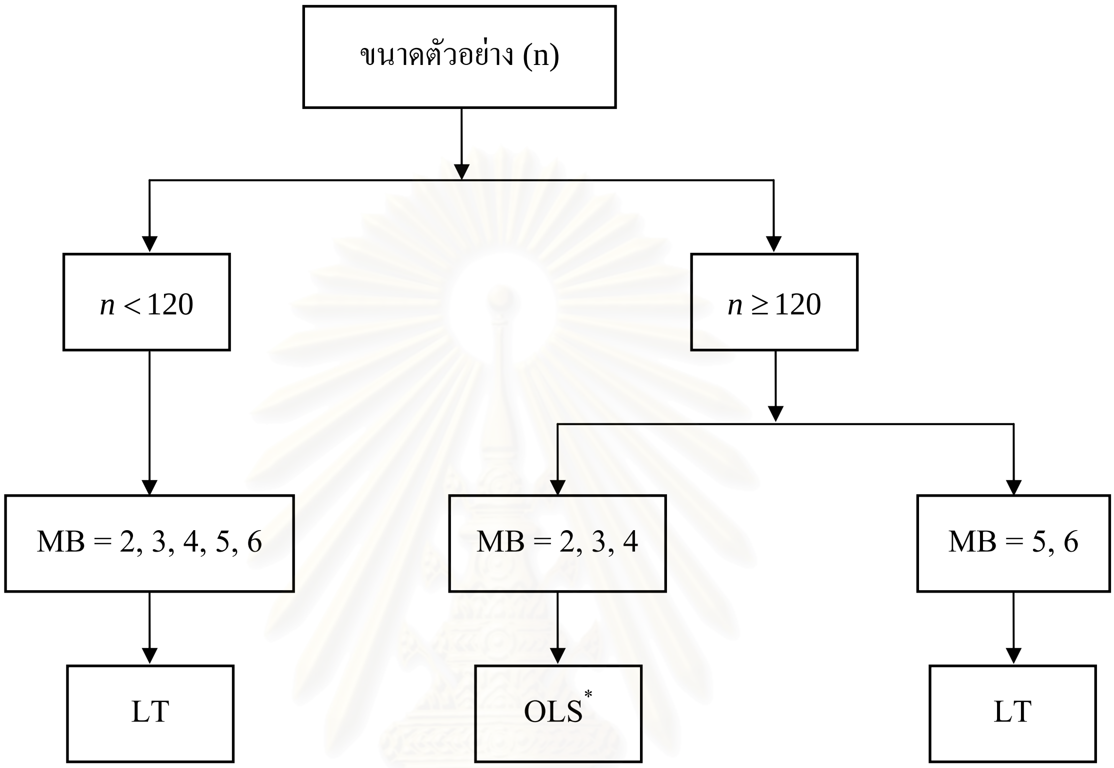
5.2 ข้อเสนอแนะ

5.2.1 การนำไปใช้ประโยชน์

การวิจัยครั้งนี้เป็นการเปรียบเทียบการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธีกำลังสองน้อยสุด (OLS) วิธีการ์ร็อตที่ไม่เป็นลบ (NG) วิธีแจ็คไนฟริดจ์ (JR) และวิธีลิวไทป์ (LT) เพื่อจะได้ใช้เป็นแนวทางในการเลือกวิธีการประมาณค่าสัมประสิทธิ์พหุนามให้เหมาะแต่ละสถานการณ์ ดังนั้นในการนำไปใช้ควรเลือกวิธี LT เพราะเป็นวิธีที่ให้ค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยที่สุดในทุกกรณีที่ทำการศึกษา ซึ่งส่งผลให้ได้ตัวประมาณที่ดีและใกล้เคียงกับค่าจริงมากที่สุด

โดยทั่วไปในทางปฏิบัติจริงผู้ทำการวิจัยส่วนใหญ่มักจะเก็บขนาดตัวอย่างข้อมูลมาในรูปของจำนวนเต็มสิบหรือจำนวนเต็มร้อยและไม่ทราบล่วงหน้าว่ากำลังสูงสุดของตัวแปรอิสระในตัวแบบพหุนามควรมีค่าเท่าไร ผู้วิจัยจึงได้ทำการสร้างขอบเขตของขนาดตัวอย่างใหม่เพื่อให้เหมาะกับการนำไปใช้ประโยชน์กับงานวิจัยอื่นๆ และสรุปสถานการณ์ต่างๆในรูปแบบแผนผัง เพื่อใช้เป็นแนวทางในการเลือกวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามในรูปที่ 5.2 ดังต่อไปนี้

<u>รูปที่ 5.2</u> **แผนผังสรุปผลวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุนาม สำหรับใช้ในเชิงปฏิบัติ**



เมื่อ OLS* คือประสิทธิภาพของวิธี OLS ต่างจากวิธี LT ไม่ถึง 50% แสดงว่าใช้วิธี OLS แทนวิธี LT ในกรณีนั้นได้

5.2.2 การศึกษาการวิจัย

เพื่อเป็นแนวทางให้ผู้ที่สนใจศึกษาเพิ่มเติม โดยทำการศึกษาในกรณีต่างๆ ดังต่อไปนี้

1. การวิจัยครั้งนี้กำหนดให้ $\varepsilon_i \sim N(0,\sigma_\varepsilon^2)$ โดยที่ $\sigma_\varepsilon^2 = 2, 4, 6, 8$ และ10 ซึ่งผลการวิจัยภายใต้ขอบเขตที่กำหนดพบว่าวิธี LT เป็นวิธีที่ดีที่สุด ผู้วิจัยเห็นว่าครั้งต่อไปควรพิจารณากรณีที่ตัวแปรอิสระมีค่าคลาดเคลื่อนหรือตัวแปรแฝงเพื่อศึกษาดูว่าวิธี LT ยังคงเป็นวิธีที่ดีที่สุดหรือไม่

2. การวิจัยครั้งนี้มีขอบเขตงานวิจัยที่ตัวแปรอิสระมีการแจกแจงแบบปกติเมื่อ $\mu = 5$ และ $\sigma^2 = 4$ ซึ่งผลการวิจัยพบว่าวิธี LT เป็นวิธีที่สุดในทุกกรณีที่ศึกษา ผู้วิจัยเห็นว่าครั้งต่อไปควรพิจารณากรณีที่กำหนดให้ค่าตัวแปรอิสระมีค่าคงที่ที่เป็นจำนวนเต็มใดๆ และตัวแปรตามมีการแจกแจงแบบปกติหรือกรณีที่ตัวแปรอิสระมีการแจกแจงแบบต่างๆ เช่น การแจกแจงแบบหางยาว การแจกแจงแบบเบ้ขวาและการแจกแจงแบบเบ้ซ้าย เป็นต้น เพื่อศึกษาดูว่าวิธี LT ยังคงเป็นวิธีที่ดีที่สุดหรือไม่

3. ผู้วิจัยได้ทดลองศึกษากรณีตัวแปรอิสระมีการแจกแจงแบบปกติด้วย $\mu = 0.5$ และ σ^2 ที่ระดับต่างๆ เพื่อศึกษาดูว่าหากตัวแปรอิสระมีค่าต่ำมากจะส่งผลต่อการวิจัยมากน้อยเพียงไร ซึ่งพบว่าค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าสูงกว่าปกติ ทั้งนี้เมื่อได้ทำการตรวจสอบแล้วพบว่าความคลาดเคลื่อนนี้มาจากขั้นตอนในการคำนวณค่า $(X'X)^{-1}$ แต่ทั้งนี้วิธี LT ก็ยังให้ผลดีกว่าวิธีอื่นๆ เมื่อค่าสัมประสิทธิ์ความผันแปร (Coefficient of Variation (CV)) ที่มีค่าไม่เกิน 50% นอกจากนี้ผู้วิจัยได้ผู้วิจัยได้ทำการศึกษากรณีที่ตัวแปรอิสระมีค่าสูงมากๆ ซึ่งพบว่าค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าสูงกว่าปกติ เมื่อได้ทำการตรวจสอบแล้วพบว่าความคลาดเคลื่อนนี้มาจากขั้นตอนในการคำนวณค่า $(X'X)^{-1}$ เช่นเดียวกัน

รายการอ้างอิง

ภาษาไทย

นพมาศ อัครจันทโชติ. การเปรียบเทียบวิธีการสร้างตัวแบบในการวิเคราะห์ความถดถอยพหุนาม กรณีที่มี 2 ตัวแปรอิสระ ซึ่งเกิดอันตรกิริยา. วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2539.

สมลักษณ์ ศิริชื่นวิจิตร. การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของตัวแบบถดถอยพหุนาม. วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2548.

ภาษาต่างประเทศ

Breiman, L. Better Subset Regression Using the Nonnegative Garrote. Technometrics 37 (1995): 373-384.

GroB, J. Linear Regression. Germany: Springer-Verlag Berlin Heidelberge, 2003.

Liu, K. More on Liu-Type Estimator in Linear Regression. Communications in Statistics Theory and Methods 33 (2004): 2723-2733.

Liu, K. Using Liu-Type Estimator to Combat Collinearity. Communications in Statistics Theory and Methods 32 (2003): 1009-1020.

Mohammadi, L. and van de Geer, S.A. On Nonnegative Garrote Estimator in A Linear Regression Model. Technical Report MI 12, University of Leiden, 2002.

Raykov, T. and Widaman, K.F. Issues in applied structural equation modeling research. Structural Equation Modeling: A Multidisciplinary Journal 2, 4 (1995): 289- 318.

Yuan, M. and Lin, Y. On the Nonnegative Garrote Estimator. Georgia Institute of Technology and University of Wisconsin, 2005.

บรรณานุกรม

ภาษาไทย

จิรายุส พู่มนตรี. การเปรียบเทียบตัวประมาณริดจ์สำหรับการวิเคราะห์ถดถอยแบบริดจ์. วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2534.

ธีระพร วีระถาวร. ตัวแบบเชิงเส้น ทฤษฎีและการประยุกต์. พิมพ์ครั้งที่ 1. กรุงเทพมหานคร: วิทยพัฒน์, 2541.

ปัทมวดี นันทนาเนตร์. การเปรียบเทียบตัวประมาณการถดถอยเมื่อเกิดพหุสัมพันธ์และหรือมีค่าผิดปกติ. วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2544.

มนัส สังวรศิลป์. คู่มือการใช้งาน MATLAB ฉบับสมบูรณ์. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร: อินโฟเพรส, 2543.

สมพล จารุธนศักดิ์กูร. การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในการวิเคราะห์ความถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด วิธีริดจ์รีเกรสชันที่ใช้ข้อมูลสนเทศโดยหลักเกณฑ์ และวิธีลิวคีเจียนทั่วไป เมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ. วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2539.

อัชฌา อระวีพร. การเปรียบเทียบการประมาณค่าสัมประสิทธิ์ในการถดถอยเชิงเส้นพหุคูณเมื่อเกิดพหุสัมพันธ์. วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2541.

ภาษาต่างประเทศ

Hocking, R.R. Method and Applications of Linear Models: Regression and the Analysis of Variance. A Wiley – Interscience Publication John Wiley and Sons, 1996.

Liu, K. A New Class of Biased Estimate in Linear Regression. Communications in Statistics Theory and Methods 22 (1993): 393-402.

Luenberger, D.G. Linear and Nonlinear Programming. Kluwer Academic Publishers, 2003.

ภาคผนวก

รายละเอียดของโปรแกรมที่ใช้ในงานวิจัย

ในงานวิจัยครั้งนี้ได้ใช้โปรแกรม MATLAB สำหรับหาค่าประมาณสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธีกำลังสองน้อยสุด (OLS) วิธีการ์ร็อตที่ไม่เป็นลบ (NG) วิธีแจ็คไนฟริดจ์ (JR) และวิธีลิวไทป์ (LT) ซึ่งเป็นโปรแกรมที่ใช้การพัฒนาโปรแกรมบน Windows ส่งผลให้การพัฒนาโปรแกรมสามารถทำได้ง่ายและสะดวก

รายละเอียดทั้งหมดของโปรแกรมที่ใช้ในงานวิจัยมีดังนี้

ตารางแสดงลักษณะการทำงานของโปรแกรมที่ใช้ในงานวิจัย

ลำดับที่	ชื่อโปรแกรม	การทำงานของโปรแกรม	ชื่อโปรแกรมย่อย (ฟังก์ชันที่เรียกใช้)
โปรแกรมหลัก			
1	Main Program	- รวบรวมค่า ARMSE SDและ DIFF เพื่อแสดงค่าในตาราง	- product
ฟังก์ชัน			
1	random	- สร้างตัวแปรสุ่ม	
2	normal	- สร้างข้อมูลที่มีการแจกแจงปกติ	- random
3	data	- สร้างค่าคลาดเคลื่อน $\underset{\sim}{\varepsilon}$ - กำหนดค่าพารามิเตอร์ $\underset{\sim}{\beta}$ - สร้างตัวแปรตาม	- normal
4	product	- กำหนดจำนวนรอบ - สร้างข้อมูลตัวแปรอิสระ - ประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี OLS NG JR และLT - คำนวณค่า ARMSE SD และ DIFF	- normal - data - OLS - NG - JR - LT
5	OLS	- ประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี OLS - คำนวณค่า MSE เมื่อประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี OLS	

ลำดับที่	ชื่อโปรแกรม	การทำงานของโปรแกรม	ชื่อโปรแกรมย่อย (ฟังก์ชันที่เรียกใช้)
ฟังก์ชัน			
6	NG	- ประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี NG - คำนวณค่า MSE เมื่อประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี NG	
7	JR	- ประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี JR - คำนวณค่า MSE เมื่อประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี JR	
8	LT	- ประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี LT - คำนวณค่า MSE เมื่อประมาณค่าสัมประสิทธิ์ความถดถอยพหุนามด้วยวิธี LT	

การแสดงผลในรูปตารางตามความแปรปรวนของค่าคลาดเคลื่อนที่เปลี่ยนไป

```
%   Main program
            sigmae = sqrt(2);
%******************* p = 2 ********************
p1=2;
[p1n1]=product(p1,5*p1,sigmae);
[p1n2]=product(p1,10*p1,sigmae);
[p1n3]=product(p1,15*p1,sigmae);
[p1n4]=product(p1,20*p1,sigmae);
P2=[p1n1 p1n2 p1n3 p1n4];
%******************* p = 3 ********************
p2=3;
[p2n1]=product(p2,5*p2,sigmae);
[p2n2]=product(p2,10*p2,sigmae);
[p2n3]=product(p2,15*p2,sigmae);
[p2n4]=product(p2,20*p2,sigmae);
P3=[p2n1 p2n2 p2n3 p2n4];
%******************* p = 4 ********************
p3=4;
[p3n1]=product(p3,5*p3,sigmae);
[p3n2]=product(p3,10*p3,sigmae);
[p3n3]=product(p3,15*p3,sigmae);
[p3n4]=product(p3,20*p3,sigmae);
P4=[p3n1 p3n2 p3n3 p3n4];
%******************* p = 5 ********************
p4=5;
[p4n1]=product(p4,5*p4,sigmae);
[p4n2]=product(p4,10*p4,sigmae);
[p4n3]=product(p4,15*p4,sigmae);
[p4n4]=product(p4,20*p4,sigmae);
P5=[p4n1 p4n2 p4n3 p4n4];
```

```
%******************* p = 6 ********************
p5=6;
[p5n1]=product(p5,5*p5,sigmae);
[p5n2]=product(p5,10*p5,sigmae);
[p5n3]=product(p5,15*p5,sigmae);
[p5n4]=product(p5,20*p5,sigmae);
P6=[p5n1 p5n2 p5n3 p5n4];
%******************* Conclusion ********************
TABLE=[P2;P3;P4;P5;P6];
```

 # การสร้างเลขสุ่ม #

```
function [r,seed]=random(seed)
seed=seed*16807;
if seed<0
    seed=seed+((2^31)-1);
end
flt=rem(seed,((2^31)-1));
seed=flt;
r=flt*(1/((2^31)-1));
```

 # การสร้างข้อมูลให้มีการแจกแจงปกติ #

```
function[temp,seed,mm,z2]=normal(mean,sigma,seed,mm,z2)
if mm~=1
    [r1,seed] = random(seed);
    [r2,seed] = random(seed);
    z1=sqrt(-2*log(r1))*cos(2*pi*r2);
    z2=sqrt(-2*log(r1))*sin(2*pi*r2);
    temp=(z1*sigma)+mean;
    mm=1;
else
   temp=(z2*sigma)+mean;
    mm=0;
```

```
end
end
```

 # การสร้างข้อมูล #

```
function[b,X,y,seed,mm,z2]=data(X,sigmae,meanx,sigmax,seed,mm,z2,n,p,col)
for i=1:n
  [e(i,1),seed,mm,z2]=normal(0,sigmae,seed,mm,z2);
end
b=ones(col,1);
y=(X*b)+e;
```

 # การกำหนดค่าตัวแปร จำนวนรอบ และผลลัพธ์ที่ต้องการแสดงค่าในตาราง #

```
function [RESULT]=product(p,n,sigmae)
it=1000;
meanx=5;
sigmax=2;
seed=65539;
z2=0;
col=p+1;
mm=0;
for i=1:n
  [XS(i,1),seed,mm,z2]=normal(meanx,sigmax,seed,mm,z2);
end
for i=1:n
   for j=1:p
      XO(i,j)=XS(i,1).^j  ;
       end
end
X=[ones(n,1),XO];
%**************************************************
for i=1:it
[b,X,y,seed,mm,z2]=data(X,sigmae,meanx,sigmax,seed,mm,z2,n,p,col);
```

```
%****************** OLS *************************
[ols, mse_ols]=OLS(X,y,p,n,col,b);
RMSEols(i)=sqrt(mse_ols);

%****************** NG **************************
 [ng,mse_ng,c]=NG(X,y,p,n,col,b,ols);
RMSEng(i)=sqrt(mse_ng);
%****************** JR **************************
 [jr,rid,mse_jr]=JR(X,y,p,n,col,ols,b);
RMSEjr(i)=sqrt(mse_jr);
%****************** LT **************************
 [lt,mse_lt]=LT(X,y,p,n,col,rid,b);
RMSElt(i)=sqrt(mse_lt);
end
%****************** Criterion **************************
ARMSE_ols =mean(RMSEols);
ARMSE_ng =mean(RMSEng);
ARMSE_jr =mean(RMSEjr);
ARMSE_lt =mean(RMSElt);
ARMSE=[ ARMSE_ols ARMSE_ng ARMSE_jr ARMSE_lt];

SD_ols =std(RMSE_ols);
SD_ng =std(RMSE_ng);
SD_jr =std(RMSE_jr);
SD_lt =std(RMSE_lt);
SD=[ SD_ols SD_ng SD_jr SD_lt];

ARMSE_min=min(ARMSE);
DIFF_ols=100*( ARMSE_ols- ARMSE_min)/ ARMSE_min;
DIFF_ng=100* ( ARMSE_ng- ARMSE_min)/ ARMSE_min;
DIFF_jr=100* ( ARMSE_jr- ARMSE_min)/ ARMSE_min;
DIFF_lt=100* ( ARMSE_lt- ARMSE_min)/ ARMSE_min;
```

```
DIFF=[ DIFF_ols DIFF_ng DIFF_jr DIFF_lt];
RESULT=[ARMSE;SD;DIFF];
```

 # การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุด (OLS) #

```
function[ols,mse_ols]=OLS(X,y,p,n,col,b)
ols=(inv(X'*X))*(X'*y);
mse_ols=sum((ols-b).^2)/col;
```

 # การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีการ์ร็อตที่ไม่เป็นลบ (NG) #

```
function[ng,mse_ng,c]=NG(X,y,p,n,col,b,ols)
XtX=X'*X;
[Q,lam]=eig(XtX);
al_ols=Q'*ols;
for j=1:col
  lamda=3*sqrt(log10(n)/n);
  c(j,1)=1-(lamda/(lam(j,j)*(al_ols(j,1))^2));
        if c(j,1)<0
          c(j,1)=0;
        end
   al_ng(j,1)=c(j,1)*al_ols(j,1);
end
ng=Q*al_ng;
mse_ng=sum((ng-b).^2)/col;
```

 # การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีแจ็คไนฟริดจ์ (JR) #

```
function[jr,rid,mse_jr]=JR(X,y,p,n,col,ols,b)
var=(y-X*ols)'*(y-X*ols)/(n-col);
k=col*var/(ols'*ols);
rid=inv(X'*X+k*eye(col))*X'*y;
s1=0;
for i=1:n
  E=eye(n);
```

```
  Et=eye(n);
  E(i,:)=[];
  Et(:,i)=[];
  jac_rid=inv(X'*Et*E*X+k*eye(col))*X'*Et*E*y;
  s1=s1+jac_rid;
end
avg_jac_rid=s1./n;
bias=(n-1)*(avg_jac_rid-rid);
jr=rid-bias;
mse_jr=sum((jr-b).^2)/col;
```

 # การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีลิวไทป์ (LT) #

```
function[lt,mse_lt]=LT(X,y,p,n,col,rid,b)
XtX=X'*X;
[Q,lam]=eig(XtX);
C = max(lam,[],1);
lam1=max(C);
lamp=min(C);
Z=X*Q;
k=(lam1-100*lamp)/99;
al_rid=Q'*rid;
var=(y-X*rid)'*(y-X*rid)/(n-col);
for i=1:col
   D1(i,1)=(var-k*(al_rid(i,1)^2))/(C(1,i)+k)^2;
   D2(i,1)=((C(1,i)*al_rid(i,1)^2)+var)/(C(1,i)*(C(1,i)+k)^2);
end
d1=sum(D1);
d2=sum(D2);
d=d1/d2;
al_lt=inv(lam+(k*eye(col)))*((Z'*y)-(d*al_rid));
lt=Q*al_lt;
mse_lt=sum((lt-b).^2)/col;
```

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวรัชเนตร ปานดำ เกิดเมื่อวันที่ 7 ธันวาคม 2524 สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาคณิตศาสตร์ ภาคคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล ในปีการศึกษา 2546 และเข้าศึกษาต่อในหลักสูตรสถิติศาสตรมหาบัณฑิต (สต.ม.) สาขาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปี พ.ศ.2547