การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในการวิเคราะห์ความถดถอยพหุคูณ เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ

นางสาวสิริกัญญา สุขวิเสส

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2549
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A COMPARISON OF PARAMETER ESTIMATION METHOD IN MULTIPLE REGRESSION ANALYSIS WITH MISSING DATA IN INDEPENDENT VARIABLES

Miss Sirikanya Sukwises

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

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

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในการวิเคราะห์ความถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ
โดย	นางสาวสิริกัญญา สุขวิเสส
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร. ธีระพร วีระถาวร

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะพาณิชยศาสตร์และการบัญชี
(รองศาสตราจารย์ ดร. ดนุชา คุณพนิชกิจ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ผกาวดี ศิริรังษี)

..... อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร. ธีระพร วีระถาวร)

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

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สิริกัญญา สุขวิเสส : การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในการวิเคราะห์ความถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ (A COMPARISON OF PARAMETER ESTIMATION METHOD IN MULTIPLE REGRESSION ANALYSIS WITH MISSING DATA IN INDEPENDENT VARIABLES). อ. ที่ปรึกษา : รศ.ดร. ธีระพร วีระถาวร, 99 หน้า.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ โดยจะเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ 5 วิธี ได้แก่ วิธีข้อมูลสมบูรณ์(Complete Case method (CC)) วิธีการถดถอยอันดับที่หนึ่ง-วิธีผสมแบบถ่วงน้ำหนัก (First Order Regression method - Weight Mixed Estimator method(FOR-W)) วิธีการถดถอยอันดับที่หนึ่ง-วิธีผสมแบบสไตน์ (First Order Regression method - Stein Rule Mixed Estimator method (FOR-SR)) วิธีประยุกต์การถดถอยอันดับที่หนึ่ง-วิธีผสมแบบถ่วงน้ำหนัก (Modified First Order Regression method - Weight Mixed Estimator method (MFOR-W)) และวิธีประยุกต์การถดถอยอันดับที่หนึ่ง-วิธีผสมแบบสไตน์ (Modified First Order Regression method - Stein Rule Mixed Estimator method (MFOR-SR)) เกณฑ์ที่ใช้ในการตัดสินใจคือค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Average Root Mean Squares Error (ARMSE)) และส่วนที่ใช้ประกอบในการเปรียบเทียบประสิทธิภาพของวิธีการทั้ง 3 วิธีคือ ค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Ratio of Different Average Root Mean Squares Error (DIFF)) สถานการณ์ที่ศึกษาจะใช้จำนวนตัวแปรอิสระ(p) เท่ากับ 3 ขนาดตัวอย่างเท่ากับ 5p 10p 15p และ 20p ระดับความสัมพันธ์ระหว่างตัวแปรอิสระใช้ 3 ระดับคือ ระดับต่ำ (ρ = 0.1 0.2 0.3) ระดับปานกลาง (ρ =0.4 0.5 0.6) และระดับสูง (ρ =0.7 0.7 0.8) สัดส่วนข้อมูลสูญหายในตัวแปรอิสระใช้ 3 ระดับคือ 5 % 10 % และ 15 % และความคลาดเคลื่อนมีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับศูนย์ และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.5 1 และ 3 ตามลำดับ ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากการจำลองแบบโดยวิธีมอนติคาร์โลซึ่งกระทำซ้ำ 1,000 รอบ ในแต่ละสถานการณ์ ผลของการวิจัยสามารถสรุปได้ดังนี้

การเปรียบเทียบค่า ARMSE ของทั้ง 5 วิธี พบว่า ในทุกกรณี วิธี MFOR-SR จะให้ค่า ARMSE น้อยที่สุด และมีค่าใกล้เคียงกับวิธี MFOR-W โดยที่ค่า ARMSE มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น และมีแนวโน้มเพิ่มขึ้นเมื่อสัดส่วนข้อมูลสูญหายในตัวแปรอิสระ ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ และส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเพิ่มขึ้น

ค่า ARMSE **แปรผันตาม**ปัจจัยต่อไปนี้จากมากไปน้อยคือ ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระ **แต่แปรผกผัน**กับขนาดตัวอย่าง

ภาควิชา .....สถิติ..... ลายมือชื่อนิสิต สิริกัญญา สุขวิเสส

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

ปีการศึกษา .....2549.....

4782416726: MAJOR STATISTICS

KEY WORD : MULTIPLE REGRESSION / MISSING DATA / COMPLETE CASE / FIRST ORDER REGRESSION / MODIFIED FIRST ORDER REGRESSION / MIXED ESTIMATOR / WEIGHT MIXED ESTIMATOR / STEIN RULE MIXED ESTIMATOR

SIRIKANYA SUKWISES : A COMPARISION OF PARAMETER ESTIMATION METHOD IN MULTIPLE REGRESSION ANALYSIS WITH MISSING DATA IN INDEPENDENT VARIABLES. THESIS ADVISOR : ASSOC. PROF. THEERAPORN VERATHAWORN, Ph.D., 99 pp.

The objective of this research is to compare multiple regression coefficients estimating methods with missing data in independent variables. This research compares five multiple regression coefficient estimation methods, which comprise Complete Case (CC), First Order Regression method - Weight Mixed Estimator method (FOR-W), First Order Regression method - Stein Rule Mixed Estimator method (FOR-SR), Modified First Order Regression method - Weight Mixed Estimator method (MFOR-W) and Modified First Order Regression method - Stein Rule Mixed Estimator method (MFOR-SR). The criterion for making decision is Average Root Mean Squares Error (ARMSE) and Ratio of Different Average Root Mean Squares Error (DIFF). As for the case study, we specify the number of independent variables (p) equal to 3. The sample sizes are 5p, 10p, 15p and 20p. The level of correlations among independent variables could be classified into 3 levels for which low levels equal to (0.1, 0.2, 0.3), middle levels equal to (0.4, 0.5, 0.6) and high levels equal to (0.7, 0.7, 0.8). The proportions of missing data in independent variables are 5%, 15% and 25%. The distribution of error is normal distribution with mean equal to 0 and standard deviations equal to 0.5, 1 and 3, 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:

According to the comparison of ARMSE from five referred method, in all cases, MFOR-SR method has a smallest ARMSE and MFOR-W method is nearly to the same. The ARMSE decreases when sample size increases but it increases when the proportions of missing data in independent variables or the level of correlation among independent variables or the standard deviation of the error increases.

The ARMSE varies with, most to least, respectively, the standard deviation of the error, the level of correlation among independent variables and the proportions of missing data in independent variables but converses to sample size.

Department .....Statistics..... Student's Signature Sirikanya Sukwises

Field of study .....Statistics..... Advisor's Signature Theeraporn Verathaworn

Academic year.....2006.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความช่วยเหลือ และเอาใจใส่อย่างดียิ่งของรองศาสตราจารย์ ดร. ธีระพร วีระถาวร อาจารย์ที่ปรึกษาวิทยานิพธ์ ผู้วิจัยขอกราบขอบพระคุณต่อท่านอาจารย์เป็นอย่างสูง ที่กรุณาให้คำปรึกษา และแนะนำเกี่ยวกับวิทยานิพนธ์ด้วยดีเสมอมา

ผู้วิจัยขอกราบขอบพระคุณรองศาสตราจารย์ผกาวดี ศิริรังษี ประธานกรรมการสอบวิทยานิพนธ์ และรองศาสตราจารย์ ดร. กัลยา วานิชย์บัญชา กรรมการสอบวิทยานิพนธ์ที่กรุณาให้คำแนะนำ ตรวจสอบ และแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น

สุดท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณครอบครัว ที่ช่วยส่งเสริม สนับสนุนและให้กำลังใจเสมอมาจนสำเร็จการศึกษา รวมทั้งขอขอบคุณเพื่อนๆ ทุกคนที่คอยห่วงใยและให้กำลังใจผู้วิจัยมาโดยตลอด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง

บทคัดย่อภาษาอังกฤษ..... จ

กิตติกรรมประกาศ..... ฉ

สารบัญ..... ช

สารบัญตาราง..... ฌ

สารบัญรูปภาพ..... ฎ

บทที่ 1 บทนำ..... 1

1.1 ความเป็นมาและความสำคัญของปัญหา..... 1

1.2 วัตถุประสงค์ของการวิจัย..... 4

1.3 สมมติฐานของการวิจัย..... 4

1.4 ขอบเขตของการวิจัย..... 5

1.5 เกณฑ์ที่ใช้ในการตัดสินใจ..... 5

1.6 วิธีดำเนินการวิจัย..... 6

1.7 ประโยชน์ที่คาดว่าจะได้รับ..... 7

บทที่ 2 ทฤษฎีและตัวสถิติที่เกี่ยวข้อง..... 8

2.1 ตัวแบบสมการถดถอยของตัวประมาณแบบผสม..... 8

2.2 วิธีประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ..... 11

2.3 การประมาณค่าข้อมูลสูญหาย..... 13

บทที่ 3 วิธีดำเนินการวิจัย..... 16

3.1 แผนการทดลอง..... 16

3.2 ขั้นตอนในการวิจัย..... 17

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

หน้า

บทที่ 4 ผลการวิจัย..... 22

4.1 ผลการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ
ในกรณีส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 0.5..... 24

4.2 ผลการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ
ในกรณีส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 1..... 41

4.3 ผลการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ
ในกรณีส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 3..... 59

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ..... 78

5.1 ผลการเปรียบเทียบค่าเฉลี่ยรากที่สองของ
ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง..... 78

5.2 ปัจจัยที่มีผลต่อค่าเฉลี่ยรากที่สองของ
ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของแต่ละวิธี..... 79

5.3 ผลสรุปการเลือกวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ..... 80

5.4 ข้อเสนอแนะ..... 82

รายการอ้างอิง..... 85

บรรณานุกรม..... 86

ภาคผนวก..... 87

ประวัติผู้เขียนวิทยานิพนธ์..... 99

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตารางที่ หน้า

4.1.1 แสดงค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %..... 25

4.1.2 แสดงค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %..... 30

4.1.3 แสดงค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %..... 35

4.2.1 แสดงค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %..... 42

4.2.2 แสดงค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %..... 47

4.2.3 แสดงค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %..... 52

ตารางที่ หน้า

4.3.1 แสดงค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %..... 60

4.3.2 แสดงค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %..... 65

4.3.3 แสดงค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %..... 70

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญรูปภาพ

รูปที่ หน้า

3.1 แผนผังการเขียนโปรแกรม 21

4.1.1 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSEและขนาดตัวอย่าง
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 0.5
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %..... 26

4.1.2 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE
และระดับความสัมพันธ์ของตัวแปรอิสระ
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 0.5
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %..... 27

4.1.3 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSEและขนาดตัวอย่าง
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 0.5
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %..... 31

4.1.4 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE
และระดับความสัมพันธ์ของตัวแปรอิสระ
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 0.5
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %..... 32

4.1.5 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSEและขนาดตัวอย่าง
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 0.5
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %..... 36

รูปที่ หน้า

4.1.6 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE
และระดับความสัมพันธ์ของตัวแปรอิสระ
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 0.5
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %..... 37

4.2.1 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSEและขนาดตัวอย่าง
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 1
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %..... 43

4.2.2 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE
และระดับความสัมพันธ์ของตัวแปรอิสระ
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 1
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %..... 44

4.2.3 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSEและขนาดตัวอย่าง
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 1
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %..... 48

4.2.4 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE
และระดับความสัมพันธ์ของตัวแปรอิสระ
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 1
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %..... 49

4.2.5 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSEและขนาดตัวอย่าง
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 1
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %..... 53

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ หน้า

4.2.6 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE
และระดับความสัมพันธ์ของตัวแปรอิสระ
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 1
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %..... 54

4.3.1 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSEและขนาดตัวอย่าง
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 3
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %..... 61

4.3.2 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE
และระดับความสัมพันธ์ของตัวแปรอิสระ
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 3
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %..... 62

4.3.3 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSEและขนาดตัวอย่าง
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่ม
เท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %..... 66

4.3.4 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE
และระดับความสัมพันธ์ของตัวแปรอิสระ
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 3
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %..... 67

4.3.5 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ
เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSEและขนาดตัวอย่าง
เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 3
และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %..... 71

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ หน้า

4.3.6 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และระดับความสัมพันธ์ของตัวแปรอิสระ เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %..... 72

5.1 แผนผังสรุปผลวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระเชิงทฤษฎี 81

5.2 แผนผังสรุปผลวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระเชิงปฏิบัติ 83

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการวิจัยในด้านต่างๆ อาทิเช่น วิศวกรรม อุตสาหกรรม วิทยาศาสตร์ สังคมศาสตร์และในด้านธุรกิจต่างๆ จำเป็นต้องอาศัยความรู้ทางสถิติเข้ามาช่วยในการวิจัยค้นคว้าโดยเฉพาะ เพื่อคาดคะเนสิ่งที่เกิดขึ้นในอนาคตหรือที่เรียกว่าการพยากรณ์ ซึ่งเป็นประโยชน์ในการวางแผนล่วงหน้าได้อย่างใกล้เคียงความเป็นจริง และวิธีการทางสถิติวิธีหนึ่งที่จะใช้กันมากในการอธิบายความสัมพันธ์ระหว่างตัวแปรในการคาดคะเนหรือการพยากรณ์ค่าของตัวแปรคือการวิเคราะห์ความถดถอย ในการวิเคราะห์ความถดถอยจะประกอบไปด้วยตัวแปรอิสระ (independent variable) และตัวแปรตาม (dependent variable) แต่ในความเป็นจริงการใช้ตัวแปรอิสระเพียงตัวเดียวอาจจะไม่สามารถอธิบายตัวแปรตามได้ดีจึงควรใช้ตัวแปรอิสระมากกว่าหนึ่งตัวขึ้นไปมาช่วยในการอธิบายตัวแปรตามให้มีประสิทธิภาพในการพยากรณ์มากขึ้น ซึ่งถ้าใช้ตัวแปรอิสระมากกว่าหนึ่งตัวขึ้นไปในสมการถดถอยนี้เรียกว่า การถดถอยพหุคูณ (multiple regression) แต่ปัญหาที่อาจเกิดขึ้นในการวิเคราะห์การถดถอยพหุคูณคือการที่ค่าสังเกตของตัวแปรบางค่าที่ต้องการศึกษาขาดหายไป ซึ่งการขาดหายไปนี้อาจเกิดขึ้นโดยไม่ได้ตั้งใจหรืออาจเนื่องมาจากค่าใช้จ่ายในการเก็บค่าสังเกตมีจำกัด หรืออาจเกิดจากเวลาหรือสภาวะแวดล้อมที่ต้องทำให้ตัดหรือหยุดการสังเกตค่าสังเกตบางค่า เป็นต้น

การวิเคราะห์ความถดถอยเชิงเส้นทั่วไปมีรูปแบบดังนี้

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

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

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

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

$\underset{\sim}{\varepsilon}$ แทนเวกเตอร์ของความคลาดเคลื่อนขนาด $n \times 1$ ซึ่ง ε_i i.i.d. $N(0, \sigma^2)$

σ^2 แทนพารามิเตอร์ที่ไม่ทราบค่า

I_n แทนเมทริกซ์เอกลักษณ์ขนาด n × n

และ p แทนจำนวนตัวแปรอิสระในสมการ

การวิเคราะห์ความถดถอยสำหรับข้อมูลครบสมบูรณ์ทุกค่าก็ไม่เกิดปัญหาในการวิเคราะห์ แต่ถ้าหากข้อมูลที่รวบรวมมาได้นั้นมีบางค่าสูญหายไปและไม่สามารถเก็บเพิ่มเติมได้จึงทำให้เกิดปัญหาในการวิเคราะห์ข้อมูลได้ ซึ่งทำให้ผู้วิจัยจำเป็นต้องตัดตัวอย่างที่ไม่สมบูรณ์ทิ้งเพราะจะนำไปใช้ร่วมเป็นข้อมูลสำหรับการวิเคราะห์ไม่ได้ ผลกระทบที่ได้รับจากวิธีดังกล่าวก็คือจำนวนตัวอย่างมีน้อยลงกว่าเดิม และอาจมีผลทำให้จำนวนตัวอย่างมีจำนวนใกล้เคียงกับจำนวนตัวแปรอิสระ ถ้าจำนวนตัวอย่างที่ไม่สมบูรณ์ที่ต้องตัดทิ้งมีเป็นจำนวนมากผลที่ตามมาคือค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าสูงมาก ซึ่งทำให้การประมาณค่าขาดความน่าเชื่อถือ สิ่งสำคัญคือทำให้สูญเสียรายละเอียดบางอย่างของตัวแปรอิสระนั้นๆไป ซึ่งอาจจะมีผลกระทบต่อผลสรุปของการวิเคราะห์นั้นๆได้ จุดประสงค์ประการหนึ่งของการวิเคราะห์ความถดถอยคือการพยากรณ์หรือการศึกษาเกี่ยวกับแนวโน้มของตัวแปรตาม ซึ่งการที่จะสามารถพยากรณ์ค่าของตัวแปรตามได้นั้นจำเป็นต้องทราบถึงอิทธิพลของตัวแปรอิสระที่มีต่อตัวแปรตาม กล่าวคือต้องการประมาณค่าพารามิเตอร์ที่เกี่ยวข้องทั้งหมดหรือเวกเตอร์สัมประสิทธิ์การถดถอย วิธีการประมาณเวกเตอร์สัมประสิทธิ์การถดถอยวิธีหนึ่งที่นิยมใช้กันมากคือ วิธีกำลังสองน้อยสุด (Least Squares method) ซึ่งวิธีนี้เป็นวิธีที่ให้ตัวประมาณที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงและมีความแปรปรวนต่ำสุด แต่เมื่อค่าสังเกตขาดหายไปจะทำให้ไม่สามารถประมาณได้ด้วยวิธีดังกล่าว

ในการวิจัยครั้งนี้ผู้วิจัยได้ศึกษาเอกสารประกอบการทำการวิจัยเกี่ยวกับการประมาณค่าสัมประสิทธิ์การถดถอยกรณีที่ข้อมูลของตัวแปรอิสระมีการสูญหาย โดยมีงานวิจัยที่เกี่ยวข้องดังนี้

ในปีการศึกษา พ.ศ. 2532 วิวัฒน์ สกลสนธิเศรษฐ์ ได้ทำการศึกษาเกี่ยวกับการประมาณค่าแบบริดจ์เมื่อมีค่าสังเกตสูญหาย โดยทำการเปรียบเทียบวิธีประมาณค่าสูญหาย 5 วิธีได้แก่ วิธีค่าเฉลี่ย-โฮเอร์ล เคนนาร์ดและบอล์ดวิน(Mean-Hoerl, Kennard and Baldwin (Mean-HKB)) วิธีค่าเฉลี่ย-ลอว์เลสและแวง(Mean-Lawless and Wang (Mean-LW)) วิธีสมการถดถอย-โฮเอร์ล เคนนาร์ดและบอล์ดวิน (Regression-Hoerl, Kennard and Baldwin (REG/HKB)) วิธีสมการถดถอย-ลอว์เลสและแวง (Regression- Lawless and Wang (REG/LW)) และวิธีกำลังสองน้อยสุด (Ordinary Least Squares (OLS))

ในปีการศึกษา พ.ศ. 2533 ชุติมา ชัยมุสิก ได้ทำการศึกษาเกี่ยวกับการวิเคราะห์การถดถอยพหุคูณเมื่อข้อมูลของตัวแปรอิสระสูญหายโดยทำการเปรียบเทียบวิธีการประมาณข้อมูลสูญหายในการวิเคราะห์การถดถอยพหุคูณ 4 วิธีได้แก่ วิธีวิเคราะห์ความถดถอย วิธีความควรจะเป็นสูงสุด(maximum likelihood method) วิธีค่าเฉลี่ย และวิธีค่ามัธยฐาน รวมทั้งศึกษากรณีที่ตัดชุดข้อมูลที่มีค่าสูญหายนั้นทิ้งไป จากการศึกษาพบว่า ถ้าตัวแปรอิสระและ

อัตราส่วนของข้อมูลสูญหายมีจำนวนมากในขณะที่ขนาดตัวอย่างเล็กวิธีประมาณค่าสูญหายโดยวิธีการวิเคราะห์ความถดถอยจะเหมาะสมที่สุด แต่ถ้าตัวแปรอิสระมีจำนวนน้อยและขนาดตัวอย่างใหญ่วิธีประมาณค่าสูญหายโดยวิธีการวิเคราะห์ความถดถอยจะให้ผลใกล้เคียงกับวิธีที่ตัดชุดข้อมูลสูญหายทิ้ง

ในปี ค.ศ. 1996 เทาเท็นเบิร์ก (Toutenburg) ศรีวัสตาวา (Srivastava) และฟีเกอร์ (Fieger) ได้ทำการศึกษาการประมาณค่าพารามิเตอร์ในการถดถอยพหุคูณเมื่อค่าสังเกตของตัวแปรอิสระมีการสูญหาย โดยได้เสนอวิธีการประยุกต์การถดถอยอันดับที่หนึ่ง (modified first order regression method (MFOR)) ในการประมาณค่าข้อมูลสูญหาย และประมาณค่าพารามิเตอร์ด้วยตัวประมาณถดถอยถ่วงน้ำหนักแบบผสม (weight mixed regression estimator (WM))

ในปี ค.ศ. 1998 เทาเท็นเบิร์ก (Toutenburg) และศรีวัสตาวา (Srivastava) ได้ทำการศึกษาปรับปรุงการประมาณค่าสัมประสิทธิ์การถดถอยในตัวแบบการถดถอยเชิงเส้นเมื่อค่าสังเกตบางค่าของบางตัวแปรอิสระเกิดการสูญหาย โดยการพัฒนาวิธีการถดถอยอันดับที่หนึ่ง (first order regression method) ด้วยวิธีการประมาณของสไตน์ (stein - rule mixed estimator (SRM))

ในปี ค.ศ. 2001 เทาเท็นเบิร์ก (Toutenburg) และศรีวัสตาวา (Srivastava) ได้ทำการศึกษาเกี่ยวกับการประมาณค่าสัมประสิทธิ์การถดถอยเมื่อค่าสังเกตมีการสูญหายโดยตัวประมาณข้อมูลสมบูรณ์ (complete case method(CC)) ตัวประมาณถดถอยแบบผสม (mixed regression estimator) และตัวประมาณถดถอยถ่วงน้ำหนักแบบผสม (weight mixed regression estimator (WM))

จากที่กล่าวมาข้างต้น ผู้วิจัยมีความสนใจว่าในกรณีที่สัดส่วนข้อมูลสูญหายในตัวแปรอิสระเพิ่มขึ้นและตัวแปรอิสระมีความสัมพันธ์กัน การประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระทั้ง 5 วิธี คือ

1. วิธีข้อมูลสมบูรณ์ (CC)
2. วิธีที่ประมาณค่าสูญหายด้วยวิธีการถดถอยอันดับที่หนึ่ง (FOR) และประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีผสมแบบถ่วงน้ำหนัก (WM): (FOR-W)
3. วิธีที่ประมาณค่าสูญหายด้วยวิธีการถดถอยอันดับที่หนึ่ง (FOR) และประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีผสมแบบสไตน์ (SRM): (FOR- SR)

4. วิธีที่ประมาณค่าสูญหายด้วยวิธีประยุกต์การถดถอยอันดับที่หนึ่ง (MFOR) และประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีผสมแบบถ่วงน้ำหนัก (WM): (MFOR-W)

5. วิธีที่ประมาณค่าสูญหายด้วยวิธีประยุกต์การถดถอยอันดับที่หนึ่ง (MFOR) และประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีผสมแบบสไตน์ (SRM): (MFOR- SR)

วิธีการใดที่จะทำให้ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณสำหรับสัมประสิทธิ์การถดถอยพหุคูณมีค่าต่ำสุด ซึ่งทำให้ได้ตัวประมาณที่ดีและใกล้เคียงกับค่าจริงมากที่สุด

1.2 วัตถุประสงค์ของการวิจัย

การวิจัยนี้มีวัตถุประสงค์ที่จะเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ โดยมีวิธีการประมาณ 5 วิธี

1. วิธีข้อมูลสมบูรณ์ (CC)
2. วิธีที่ประมาณค่าสูญหายด้วยวิธีการถดถอยอันดับที่หนึ่ง (FOR) และประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีผสมแบบถ่วงน้ำหนัก (WM): (FOR-W)
3. วิธีที่ประมาณค่าสูญหายด้วยวิธีการถดถอยอันดับที่หนึ่ง (FOR) และประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีผสมแบบสไตน์ (SRM): (FOR- SR)
4. วิธีที่ประมาณค่าสูญหายด้วยวิธีประยุกต์การถดถอยอันดับที่หนึ่ง (MFOR) และประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีผสมแบบถ่วงน้ำหนัก (WM): (MFOR-W)
5. วิธีที่ประมาณค่าสูญหายด้วยวิธีประยุกต์การถดถอยอันดับที่หนึ่ง (MFOR) และประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีผสมแบบสไตน์ (SRM): (MFOR- SR)

1.3 สมมติฐานของการวิจัย

เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติ วิธี MFOR-SR และวิธี MFOR-W จะให้ค่าประมาณสัมประสิทธิ์การถดถอยพหุคูณที่มีความถูกต้องมากกว่าวิธี CC วิธี FOR-SR และวิธี FOR-W (เพราะวิธี WM ได้มีการปรับค่าของชุดข้อมูลสูญหายที่ได้จากการประมาณค่าสูญหายนั้นด้วยตัวถ่วงน้ำหนักซึ่งเป็นการลดความแปรปรวนของความคลาดเคลื่อนที่เพิ่มขึ้น ส่วนวิธี SRM นั้นใช้รูปแบบของตัวประมาณของสไตน์ เพื่อลดค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง และการประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR เป็นวิธีที่นำข้อมูลที่มีอยู่ทั้งหมดมาใช้ในการประมาณค่า)

1.4 ขอบเขตของการวิจัย

1. ค่าความคลาดเคลื่อนมีการแจกแจงปกติที่มีค่าเฉลี่ย(μ) มีค่าเท่ากับ 0 ส่วนเบี่ยงเบนมาตรฐาน (σ) เท่ากับ 0.5 1 และ 3 ตามลำดับ

2. จำนวนตัวแปรอิสระ(p) ที่ศึกษา เท่ากับ 3

3. ขนาดตัวอย่างที่ศึกษามี 4 ขนาด คือ 5p 10p 15p และ 20p ตามลำดับ

4. การสูญหายของข้อมูลกำหนดให้ปรากฏขึ้นโดยสุ่ม และกำหนดให้อัตราการสูญหายของข้อมูลซึ่ง มีค่าเท่ากับ 5% 15% และ 25 %

5. ระดับพหุสัมพันธ์ของตัวแปรอิสระ แบ่งเป็น 3 ระดับ คือ

ระดับต่ำ ค่า ρ = (0.1 0.2 0.3)

ระดับปานกลาง ค่า ρ = (0.4 0.5 0.6)

ระดับสูง ค่า ρ = (0.7 0.7 0.8)

โดยค่า ρ ในวงเล็บ คือ ความสัมพันธ์ระหว่าง X_1 กับ X_2, X_1 กับ X_3 และ X_2 กับ X_3 ตามลำดับ

6. ค่าสัมประสิทธิ์การถดถอยพหุคูณที่ใช้ในการศึกษาจะกำหนดให้เป็นค่าคงที่ใดๆชุดหนึ่งคือ $\underset{\sim}{\beta}' = (1,1, \ldots ,1)_{1\times(p+1)}$ เพื่อสร้างค่า $\underset{\sim}{y}$ ขึ้นจากตัวแบบ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

7. การวิจัยครั้งนี้สร้างแบบจำลองข้อมูลให้มีสถานการณ์ตามที่ต้องการศึกษาโดยอาศัยเทคนิคมอนติคาร์โล โดยจะกระทำซ้ำ 1,000 รอบ ในแต่ละสถานการณ์

1.5 เกณฑ์ที่ใช้ในการตัดสินใจ

เกณฑ์ที่ใช้ในการตัดสินใจว่า วิธีประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณวิธีการใดให้ค่าพารามิเตอร์ใกล้เคียงกับค่าจริงมากที่สุดนั้นจะพิจารณาจากค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง(ARMSE) และส่วนที่ใช้ประกอบในการพิจารณาเปรียบเทียบจะใช้ค่าเปอร์เซ็นต์ของอัตราส่วนผลต่างค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) ของการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีต่างๆ

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

$$ARMSE = \frac{1}{1000}\sum_{i=1}^{1000} RMSE_i$$

$$DIFF = \left[\frac{ARMSE_{(k)} - ARMSE_{(\min)}}{ARMSE_{(\min)}} \right] \times 100$$

เมื่อ β_j แทนสัมประสิทธิ์การถดถอยพหุคูณตัวที่ j

$\hat{\beta}_{ij}$ แทนตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณตัวที่ j จากการประมาณครั้งที่ i

$RMSE_i$ แทนค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์การถดถอยพหุคูณจากการประมาณครั้งที่ i

ARMSEแทนค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุคูณ

DIFF แทนเปอร์เซ็นต์ของอัตราส่วนผลต่างค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีต่างๆ

$ARMSE_{min}$แทนค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุคูณจากวิธีที่มีค่าน้อยที่สุด

และ $ARMSE_k$แทนค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุคูณแต่ละวิธี

จากเกณฑ์ที่ได้กล่าวมา ถ้าวิธีใดให้ค่า ARMSE ต่ำสุดจะเป็นวิธีที่ดีที่สุด และค่า DIFF จะใช้วัดว่าวิธีที่ให้ค่า ARMSE ต่ำที่สุดจะดีกว่าวิธีอื่นกี่เปอร์เซ็นต์

1.6 วิธีดำเนินการวิจัย

1. สร้างข้อมูลของความคลาดเคลื่อนที่มีการแจกแจงปกติ โดยมีพารามิเตอร์ตามที่กำหนด

2. สร้างข้อมูลของตัวแปรอิสระ (X) ให้มีระดับความสัมพันธ์ตามที่กำหนด และสร้างข้อมูลของตัวแปรตาม (y) จากรูปแบบความสัมพันธ์ $y = X\beta + \varepsilon$ โดยกำหนดให้ β เป็นค่าคงที่ใดๆ คือ β_j = 1 โดยที่ j = 0, 1, …, p

3. สร้างข้อมูลของตัวแปรอิสระให้เกิดการสูญหายโดยเกิดขึ้นอย่างสุ่ม ตามที่กำหนด

4. ประมาณค่าข้อมูลเพื่อแทนที่ข้อมูลที่สูญหายในตัวแปรอิสระ ด้วยวิธี FOR และวิธี MFOR

5. ประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี CC วิธี FOR-W วิธี FOR-SR วิธี MFOR-W และวิธี MFOR-SR

6. คำนวณหาค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) พร้อมทั้งค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เพื่อเปรียบเทียบประสิทธิภาพของวิธีการทั้ง 5 วิธี

7. สรุปผลที่ได้จากการทดลอง

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1. ทำให้สามารถเลือกใช้วิธีการประมาณค่าพารามิเตอร์ในสมการถดถอยพหุคูณเมื่อตัวแปรอิสระมีข้อมูลสูญหายไปบางส่วนและตัวแปรอิสระมีสหสัมพันธ์กันได้อย่างเหมาะสม

2. ผลการศึกษาจะเป็นแนวทางในการแก้ไขและปรับปรุงวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณในกรณีที่ตัวแปรอิสระมีข้อมูลสูญหายได้ต่อไปในอนาคต

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

ในกรณีที่เกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ การประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุดจะต้องตัดชุดข้อมูลที่สูญหายออก หรือประมาณค่าข้อมูลสูญหายด้วยวิธีต่างๆ ก่อน ซึ่งในการตัดชุดข้อมูลสูญหายออกนั้นจะทำให้ค่าประมาณที่ได้อาจไม่ใกล้เคียงความเป็นจริง เนื่องจากขนาดตัวอย่างที่ลดลง ดังนั้นในการวิจัยครั้งนี้ได้เสนอวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณโดยพิจารณาแบ่งรูปแบบสมการถดถอยเป็น 2 ส่วน คือส่วนที่เป็นข้อมูลสมบูรณ์และส่วนที่มีบางข้อมูลสูญหาย ซึ่งเรียกว่า ตัวแบบสมการถดถอยของตัวประมาณแบบผสม (mixed estimator) ในบทนี้จะกล่าวถึงการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระและวิธีการประมาณค่าข้อมูลสูญหาย ซึ่งมีรายละเอียดดังต่อไปนี้

2.1 ตัวแบบสมการถดถอยของตัวประมาณแบบผสม (mixed estimator)

การวิเคราะห์ความถดถอยเชิงเส้นทั่วไปมีรูปแบบดังนี้

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

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

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

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

$\underset{\sim}{\varepsilon}$ แทนเวกเตอร์ของความคลาดเคลื่อนขนาด $n \times 1$ ซึ่ง ε_i i.i.d. $N(0, \sigma^2)$

σ^2 แทนพารามิเตอร์ที่ไม่ทราบค่า

I_n แทนเมทริกซ์เอกลักษณ์ขนาด n × n

และ p แทนจำนวนตัวแปรอิสระในสมการ

ในการวิเคราะห์ความถดถอยพหุคูณกรณีที่มีข้อมูลสูญหาย มีการแบ่งรูปแบบเป็น 2 ส่วน ดังนี้

1. **ตัวแบบถดถอยเชิงเส้นสำหรับชุดค่าสังเกตที่ไม่มีข้อมูลสูญหายในตัวแปรอิสระ (ชุดข้อมูลที่สมบูรณ์)** ตัวแบบดังกล่าวมีรูปแบบดังนี้

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

เมื่อ $X_c = [Z_c \ , \underset{\sim c}{x}]$ และ $\underset{\sim}{\beta} = \begin{pmatrix} \underset{\sim}{\gamma} \\ \alpha \end{pmatrix}$

ซึ่งสามารถเขียนสมการใหม่ได้ดังนี้

(2.3) $$\underset{\sim c}{y} = Z_c \underset{\sim}{\gamma} + \alpha \underset{\sim c}{x} + \underset{\sim c}{\varepsilon}$$

เมื่อ $\underset{\sim c}{y}$ แทนเวกเตอร์ของตัวแปรตามขนาด $m_c \times 1$

X_c แทนเมทริกซ์ของตัวแปรอิสระขนาด $m_c \times (p+1)$

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

$\underset{\sim c}{\varepsilon}$ แทนเวกเตอร์ของค่าความคลาดเคลื่อนที่เกิดขึ้นขนาด $m_c \times 1$

m_c แทนขนาดตัวอย่างของตัวแปรอิสระที่มีข้อมูลสมบูรณ์(ไม่มีข้อมูลสูญหาย)

p แทนจำนวนตัวแปรอิสระ

Z_c แทนเมทริกซ์ของตัวแปรอิสระที่ไม่มีข้อมูลสูญหาย ขนาด $m_c \times p$

$\underset{\sim c}{x}$ แทนเวกเตอร์ของตัวแปรอิสระที่มีข้อมูลสูญหาย ขนาด $m_c \times 1$

$\underset{\sim}{\gamma}$ แทนเวกเตอร์ของสัมประสิทธิ์การถดถอยพหุคูณของชุดข้อมูลที่ไม่มีข้อมูลสูญหายขนาด $p \times 1$

และ α แทนสัมประสิทธิ์การถดถอยพหุคูณของชุดข้อมูลที่ไม่มีข้อมูลสูญหาย

โดยที่ $E(\underset{\sim c}{\varepsilon}) = \underset{\sim}{0}$, $\text{cov}(\underset{\sim *}{\varepsilon}) = \sigma^2 I_{m_c}$

2. **ตัวแบบถดถอยเชิงเส้นสำหรับชุดค่าสังเกตที่มีข้อมูลสูญหายในตัวแปรอิสระ (ชุดข้อมูลเพิ่มเติมเมื่อทำการประมาณค่าสูญหายแล้ว)** ตัวแบบดังกล่าวมีรูปแบบดังนี้

(2.4) $$\underset{\sim *}{y} = X_* \underset{\sim}{\beta} + \underset{\sim *}{\varepsilon}$$

เมื่อ $X_* = [Z_* \ , \underset{\sim *}{x}]$ และ $\underset{\sim}{\beta} = \begin{pmatrix} \underset{\sim}{\gamma} \\ \alpha \end{pmatrix}$

ซึ่งสามารถเขียนสมการใหม่ได้ดังนี้

(2.5) $$\underset{\sim *}{y} = Z_* \underset{\sim}{\gamma} + \alpha \underset{\sim *}{x} + \underset{\sim *}{\varepsilon}$$

เมื่อ $\underset{\sim *}{y}$ แทนเวกเตอร์ของตัวแปรตามขนาด $m_* \times 1$

X_* แทนเมทริกซ์ของตัวแปรอิสระขนาด $m_* \times (p+1)$

$\underset{\sim *}{\varepsilon}$ แทนเวกเตอร์ของค่าความคลาดเคลื่อนที่เกิดขึ้นขนาด $m_* \times 1$

m_* แทนขนาดตัวอย่างของตัวแปรอิสระที่มีข้อมูลไม่สมบูรณ์

n แทนขนาดตัวอย่างของตัวแปรอิสระ ซึ่ง $n = m_c + m_*$

Z_* แทนเมทริกซ์ของตัวแปรอิสระที่ไม่มีข้อมูลสูญหายขนาด $m_* \times p$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

และ $\underset{\sim}{x}_*$ แทนเวกเตอร์ของตัวแปรอิสระตัวที่ p ซึ่งเป็นข้อมูลสูญหายทั้งหมดขนาด $m_* \times 1$

โดยที่ $E(\underset{\sim}{\varepsilon}_*) = \underset{\sim}{0}$, $\text{cov}(\underset{\sim}{\varepsilon}_*) = \sigma^2 I_{m_*}$

จากสมการ (2.2) และ (2.4) สามารถเขียนเป็นตัวแบบผสม (mixed model) ได้ดังนี้

(2.6) $$\begin{pmatrix} \underset{\sim}{y}_c \\ \underset{\sim}{y}_* \end{pmatrix} = \begin{pmatrix} X_c \\ X_* \end{pmatrix} \underset{\sim}{\beta} + \begin{pmatrix} \underset{\sim}{\varepsilon}_c \\ \underset{\sim}{\varepsilon}_* \end{pmatrix}$$

จากสมการ (2.6) จะได้ตัวประมาณของ $\underset{\sim}{\beta}$ คือ

(2.7) $$\underset{\sim M}{\hat{\beta}} = \left(X_c' X_c + \hat{X}_*' \hat{X}_*\right)^{-1} \left(X_c' \underset{\sim c}{y} + \hat{X}_*' \underset{\sim *}{y}\right)$$

ตัวประมาณ $\underset{\sim M}{\hat{\beta}}$ นี้เรียกว่า ตัวประมาณแบบผสม (mixed estimator) ซึ่งตัวประมาณ $\underset{\sim M}{\hat{\beta}}$ เป็นตัวประมาณเอนเอียง (biased) ซึ่งมีเวกเตอร์ความเอนเอียงและเมทริกซ์ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองดังนี้

$$\begin{aligned} bias\left(\underset{\sim M}{\hat{\beta}}\right) &= E\left(\underset{\sim M}{\hat{\beta}} - \underset{\sim}{\beta}\right) \\ &= \alpha (X_c' X_c + \hat{X}_*' \hat{X}_*)^{-1} \hat{X}_*' \theta \end{aligned}$$

$$\begin{aligned} MSE\left(\underset{\sim M}{\hat{\beta}}\right) &= E\left[\left(\underset{\sim M}{\hat{\beta}} - \underset{\sim}{\beta}\right)\left(\underset{\sim M}{\hat{\beta}} - \underset{\sim}{\beta}\right)'\right] \\ &= \sigma^2 (X_c' X_c + \hat{X}_*' \hat{X}_*)^{-1} + \alpha^2 (X_c' X_c + \hat{X}_*' \hat{X}_*)^{-1} \\ &\quad \cdot \hat{X}_*' \underset{\sim}{\theta} \underset{\sim}{\theta}' \hat{X}_* (X_c' X_c + \hat{X}_*' \hat{X}_*)^{-1} \end{aligned}$$

เมื่อ $\underset{\sim}{\theta} = \frac{1}{\alpha}(X_* - \hat{X}_*)\underset{\sim}{\beta}$

ตัวประมาณ $\underset{\sim M}{\hat{\beta}}$ ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (mean squares error) ต่ำกว่าตัวประมาณกำลังสองน้อยสุดนอกจากตัวประมาณ $\underset{\sim M}{\hat{\beta}}$ เป็นตัวประมาณที่เอนเอียงแล้วยังเป็นตัวประมาณที่ไม่คงเส้นคงวา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2.2 วิธีประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ

2.2.1 วิธีประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีข้อมูลสมบูรณ์ (Complete Case Estimator (CC))

วิธี CC เป็นการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด (OLS) โดยได้ทำการตัดชุดข้อมูลที่มีค่าสูญหายทิ้งไป ชุดข้อมูลที่เหลืออยู่เป็นข้อมูลที่สมบูรณ์จึงสามารถที่จะใช้การประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยสุด ดังนี้

(2.8) $$\underset{\sim CC}{\hat{\beta}} = (X_c' X_c)^{-1} X_c' \underset{\sim c}{y}$$

จากตัวประมาณ $\underset{\sim OLS}{\hat{\beta}}$ เป็นตัวประมาณไม่เอนเอียง (unbiased) และเป็นตัวประมาณที่คงเส้นคงวา ในทำนองเดียวกัน จะได้ว่า $\underset{\sim CC}{\hat{\beta}}$ เป็นตัวประมาณไม่เอนเอียง (unbiased)และเป็นตัวประมาณที่คงเส้นคงวาสำหรับ $\underset{\sim}{\beta}$ ซึ่งมีเมทริกซ์ความแปรปรวนร่วม

$$\operatorname{cov}\left(\underset{\sim CC}{\hat{\beta}}\right) = E\left[\left(\underset{\sim CC}{\hat{\beta}} - \underset{\sim}{\beta}\right)\left(\underset{\sim CC}{\hat{\beta}} - \underset{\sim}{\beta}\right)'\right]$$
$$= \sigma^2 (X_c' X_c)^{-1}$$

แต่ในการที่ตัดชุดข้อมูลที่มีการสูญหายของตัวแปรอิสระบางตัวทิ้งจะทำให้สูญเสียรายละเอียดบางอย่างไป ซึ่งอาจจะมีผลกระทบต่อผลสรุปของการวิเคราะห์นั้นๆ

2.2.2 วิธีประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีผสมแบบถ่วงน้ำหนัก (Weight Mixed Estimator (WM))

ในปี ค.ศ. 1998 เทาเท็นบูรก์ (Toutenburg) ฟีเกอร์ (Fieger) และศริวัสทวา (Srivastava) ได้มีการนำเสนอพัฒนาตัวประมาณ $\underset{\sim M}{\hat{\beta}}$ ด้วยการถ่วงน้ำหนักสำหรับชุดข้อมูลที่ไม่สมบูรณ์ด้วยค่า w ซึ่งจะมีค่าอยู่ระหว่าง 0 ถึง 1 เพื่อเป็นการลดความแปรปรวนของความคลาดเคลื่อนของการประมาณค่าสูญหายที่เพิ่มขึ้นโดยใช้หลักการของตัวประมาณกำลังสองน้อยสุดถ่วงน้ำหนัก ซึ่ง

$$w = \frac{\sigma_{yy \cdot ps}}{\sigma_{yy \cdot s}} = 1 - \rho^2_{yp \cdot s}$$

เมื่อ $\sigma_{yy \cdot ps}$ แทนความแปรปรวนของความคลาดเคลื่อน (residual variances) ของ y บน X_1, X_2,..., X_p

$\sigma_{yy\cdot s}$ แทนความแปรปรวนของความคลาดเคลื่อน (residual variances) ของ y บน X_1, X_2,..., X_{p-1}

และ $\rho_{yp\cdot s}$ แทนสหสัมพันธ์เชิงส่วน (partial correlation) ของ y กับ X_p โดยกำหนดให้ X_1, X_2,..., X_{p-1} มีค่าคงที่

ตัวประมาณในเรื่องนี้อยู่ในรูปของ

(2.9) $$\underset{\sim WM}{\hat{\beta}} = \left(X_c' X_c + w^2 \hat{X}_*' \hat{X}_*\right)^{-1}\left(X_c' \underset{\sim c}{y} + w^2 \hat{X}_*' \underset{\sim *}{y}\right)$$

ซึ่งเวกเตอร์ความเอนเอียง มีค่าเป็น

$$bias\left(\underset{\sim WM}{\hat{\beta}}\right) = E\left(\underset{\sim WM}{\hat{\beta}} - \underset{\sim}{\beta}\right)$$
$$= \alpha w^2 \left(X_c' X_c + w^2 \hat{X}_*' \hat{X}_*\right)^{-1} X_*' \underset{\sim}{\theta}$$

เมื่อ $\underset{\sim}{\theta} = \frac{1}{\alpha}(X_* - \hat{X}_*)\underset{\sim}{\beta}$

และเมทริกซ์ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (mean squares error matrix) คือ

$$MSE\left(\underset{\sim WM}{\hat{\beta}}\right) = E\left[\left(\underset{\sim WM}{\hat{\beta}} - \underset{\sim}{\beta}\right)\left(\underset{\sim WM}{\hat{\beta}} - \underset{\sim}{\beta}\right)'\right]$$
$$= \sigma^2 (X_c' X_c + w^2 \hat{X}_*' \hat{X}_*)^{-1} - \sigma^2 w^2 (X_c' X_c + w^2 \hat{X}_*' \hat{X}_*)^{-1} \hat{X}_*'$$
$$\cdot [(1 - w^2) I + w^2 \theta\theta'] \hat{X}_* (X_c' X_c + w^2 \hat{X}_*' \hat{X}_*)^{-1}$$

ตัวประมาณ $\underset{\sim WM}{\hat{\beta}}$ จะมีความเอนเอียงน้อยกว่าตัวประมาณ $\underset{\sim M}{\hat{\beta}}$ เพราะมีการถ่วงน้ำหนักด้วย w

หมายเหตุ ตัวประมาณ $\underset{\sim WM}{\hat{\beta}}$ จะลดรูปเป็นตัวประมาณ $\underset{\sim CC}{\hat{\beta}}$ เมื่อ $w^2 = 0$ และลดรูปเป็นตัวประมาณแบบผสมเมื่อ $w^2 = 1$

2.2.3 วิธีประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีผสมแบบสไตน์ (Stein - Rule Mixed Estimator (SRM))

ในปี ค.ศ. 1998 เทาเท็นบูรก์ (Toutenburg) และศริวัสทวา (Srivastava) ได้ทำการศึกษาเพื่อปรับปรุงการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีที่ค่าสังเกตบางค่าของตัวแปรอิสระมีการสูญหาย โดยได้เสนอการพิจารณาตัวประมาณลดรูป (shrunken estimator)

ด้วยตัวประมาณ $\underset{\sim M}{\hat{\beta}}$ ซึ่งเราจะประยุกต์โดยวิธีการประมาณของสไตน์ (stein - rule estimation) เพื่อจะลดปัญหาการเกิดพหุสัมพันธ์ในตัวแปรอิสระ จึงได้ตัวประมาณดังนี้

$$(2.10)\quad \underset{\sim SRM}{\hat{\beta}} = \left[1 - h^{*}\frac{\left(\underset{\sim c}{y} - X_c \underset{\sim M}{\hat{\beta}}\right)'\left(\underset{\sim c}{y} - X_c \underset{\sim M}{\hat{\beta}}\right) + \left(\underset{\sim *}{y} - \hat{X}_* \underset{\sim M}{\hat{\beta}}\right)'\left(\underset{\sim *}{y} - \hat{X}_* \underset{\sim M}{\hat{\beta}}\right)}{\underset{\sim M}{\hat{\beta}}' (X_c' X_c + \hat{X}_*' \hat{X}_*) \underset{\sim M}{\hat{\beta}}}\right] \underset{\sim M}{\hat{\beta}}$$

เมื่อ $h^{*} = \left(\frac{h}{m_c + m_* - p + 2}\right)$

และ h เป็นสเกลาร์ที่มีค่ามากกว่า 0

ในปี ค.ศ. 2000 ซาลา (Shalabh) และ เอลาน ที เค วาน (Alan T.K. Wan) ได้ศึกษาเกี่ยวกับตัวประมาณ $\underset{\sim M}{\hat{\beta}}$ และตัวประมาณ $\underset{\sim SRM}{\hat{\beta}}$ พบว่า

$$MSE\left(\underset{\sim SRM}{\hat{\beta}}\right) \leq MSE\left(\underset{\sim M}{\hat{\beta}}\right)$$

เมื่อ $0 < h < 2(p-2)\left[1 + \frac{m_*}{m_c - p + 2}\left(1 + \frac{2}{m_c + m_* - p + 2}\right)\right]^{-1}$ และที่ $p > 2$

จะได้ว่าการลดลงของฟังก์ชันการเสี่ยงมากที่สุด เมื่อ

$$h = (p-2)\left[1 + \frac{m_*}{m_c - p + 2}\left(1 + \frac{2}{m_c + m_* - p + 2}\right)\right]^{-1}$$

ตัวประมาณ $\underset{\sim SRM}{\hat{\beta}}$ นั้นมีความเอนเอียงน้อยกว่าตัวประมาณ $\underset{\sim M}{\hat{\beta}}$ แต่มีประสิทธิภาพมากกว่าตัวประมาณ $\underset{\sim M}{\hat{\beta}}$

2.3 การประมาณค่าข้อมูลที่สูญหาย

ในการวิเคราะห์ความถดถอยพหุคูณ ปัญหาที่พบอยู่เสมอก็คือ ปัญหาอันเนื่องมาจากข้อมูลสูญหายไป วิธีแก้ปัญหาในทางปฏิบัติคือการประมาณค่าสัมประสิทธิ์การถดถอยโดยการตัดชุดข้อมูลที่มีการสูญหายทิ้งไป ซึ่งผลที่ตามมาก็คือขนาดตัวอย่างลดลงทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจากการประมาณค่ามีค่าสูงกว่าความเป็นจริง ด้วยเหตุผลดังกล่าวจึงได้มีการสร้างข้อมูลเพื่อทดแทนข้อมูลที่สูญหายไปด้วยวิธีดังนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2.3.1 วิธีการถดถอยอันดับที่หนึ่ง (First Order Regression method (FOR))

เป็นวิธีที่ใช้การวิเคราะห์ความถดถอยในการประมาณค่าข้อมูลสูญหาย โดยมีขั้นตอนดังนี้

1. จากตัวแบบที่ (2.3) ซึ่งเป็นตัวแบบการถดถอยเชิงเส้นของชุดข้อมูลที่สมบูรณ์ ซึ่งจะสามารถประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยสุด โดยสมการความถดถอยเชิงเส้นแสดงความสัมพันธ์ระหว่าง $\underset{\sim c}{x}$ และตัวแปรอิสระ $X_1, X_2, ..., X_{k-1}$ ในเมทริกซ์ Z_c ดังนี้

(2.11) $$\underset{\sim OLS}{\hat{\beta}} = (Z_c' Z_c)^{-1} Z_c' \underset{\sim c}{x}$$

2. นำสมการ (2.11) มาประมาณข้อมูลสูญหายใน $\underset{\sim *}{x}$ ในสมการ (2.5) ของสมการถดถอยเชิงเส้น จะได้ว่า

(2.12) $$\underset{\sim *}{\hat{x}} = Z_* \underset{\sim OLS}{\hat{\beta}} = Z_* (Z_c' Z_c)^{-1} Z_c' \underset{\sim c}{x}$$

วิธีประมาณค่าข้อมูลสูญหายด้วยวิธี FOR มีข้อดีตรงที่ใช้ความสัมพันธ์ระหว่างตัวแปรอิสระในการประมาณค่าข้อมูลสูญหาย โดยคำนวณจากสมการถดถอยของข้อมูลในชุดข้อมูลที่สมบูรณ์ เป็นการประมาณโดยใช้ความน่าเชื่อถือที่ได้จากข้อมูล ซึ่งวิธีนี้จะเหมาะสมกับข้อมูลทางเศรษฐศาสตร์เนื่องจากในทางเศรษฐศาสตร์ข้อมูลจะมีความสัมพันธ์เกี่ยวโยงกัน

2.3.2 วิธีการประยุกต์การถดถอยอันดับที่หนึ่ง (Modified First Order Regression method (MFOR))

ในปี ค.ศ.1996 เทาเท็นเบิร์ก (Toutenburg) ศรีวัสตาวา (Srivastava) และฟีเกอร์ (Fieger) ได้เสนอการประมาณค่าแทนที่ค่าสังเกตที่สูญหายด้วยวิธีการประยุกต์การถดถอยอันดับที่หนึ่ง (modified first order regression method) จากการวิเคราะห์การถดถอยด้วยตัวแปรอิสระแล้ว ยังนำข้อมูลของตัวแปรที่ต้องการศึกษาหรือตัวแปรตามมาพิจารณาด้วย โดยใช้หลักการของการเพิ่มตัวแปรอิสระ (adding independent variable) ซึ่งจะได้เวกเตอร์ที่จะไปแทนที่ค่าสังเกตที่สูญหาย ดังขั้นตอนต่อไปนี้

1. จากการประมาณค่าข้อมูลสูญหายด้วยวิธี FOR แล้ว เราจะพิจารณาเพิ่มตัวแปรตามในการประมาณค่าสูญหาย ดังสมการต่อไปนี้

(2.13) $$\underset{\sim c}{x} = Z_c \underset{\sim 1}{\beta^*} + y \underset{\sim c}{\beta_2^*} + \underset{\sim}{\varepsilon}$$

2. ประมาณค่าสัมประสิทธิ์การถดถอย $\underset{\sim 1}{b}$ และ b_2 ของ $\underset{\sim\ 1}{\beta^*}$ และ β_2^* ตามลำดับในสมการ (2.13) ดังขั้นตอนต่อไปนี้

2.1 $\underset{\sim 1}{b} = \underset{\sim 1}{b^*} - Lb_2$ ซึ่ง $\underset{\sim 1}{b^*} = \underset{\sim\ OLS}{\hat{\beta}}$ ในสมการ (2.11)

2.2 $b_2 = N \underset{\sim c}{y^{*\prime}} \underset{\sim c}{x}$

เมื่อ $M_1 = I - Z_c(Z_c'Z_c)^{-1}Z_c'$

$L = -(Z_c'Z_c)^{-1}Z_c' \underset{\sim c}{y}$

$N = \dfrac{1}{\underset{\sim c}{y}' M_1 \underset{\sim c}{y}}$

และ $\underset{\sim c}{y^*} = M_1 \underset{\sim c}{y}$

ดังนั้นค่าประมาณของข้อมูลสูญหายสามารถเขียนได้ ดังนี้

(2.14) $\underset{\sim *}{\hat{x}} = Z_* \underset{\sim 1}{b} + \underset{\sim *}{y}\, b_2$

วิธีประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR มีข้อดีตรงที่เป็นการใช้ความสัมพันธ์ของตัวแปรทั้งตัวแปรอิสระและการเพิ่มตัวแปรตามในการประมาณค่า ซึ่งถ้าค่าตัวแปรตามที่เพิ่มเข้าไปมีลักษณะคล้ายคลึงกัน การประมาณค่าข้อมูลสูญหายด้วยวิธีนี้ก็จะให้ผลที่ดี ซึ่งวิธีนี้จะเหมาะสมกับข้อมูลทางเศรษฐศาสตร์เนื่องจากในทางเศรษฐศาสตร์ข้อมูลจะมีความสัมพันธ์เกี่ยวโยงกัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาเพื่อเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณกรณีเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระด้วยวิธี CC วิธี FOR-W วิธี FOR-SR วิธี MFOR-W และวิธี MFOR-SR ซึ่งในการเปรียบเทียบจะพิจารณาจากค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุคูณ (ARMSE) ข้อมูลที่ใช้ในการวิจัยได้จากการจำลองด้วยเทคนิคมอนติคาร์โล (Monte Carlo technique) โดยใช้โปรแกรม MATLAB 7 (Matrix Laboratory Version 7) ซึ่งแผนการทดลองและขั้นตอนในการวิจัยมีรายละเอียดดังนี้

3.1 แผนการทดลอง

ในการวิจัยครั้งนี้ได้กำหนดสถานการณ์ต่างๆ สำหรับเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีต่างๆทั้ง 5 วิธี ดังนี้

1. ค่าความคลาดเคลื่อนมีการแจกแจงปกติที่มีค่าเฉลี่ย (μ) มีค่าเท่ากับ 0 ส่วนเบี่ยงเบนมาตรฐาน (σ) เท่ากับ 0.5 1 และ 3 ตามลำดับ
2. จำนวนตัวแปรอิสระ (p) ที่ศึกษา เท่ากับ 3
3. ขนาดตัวอย่างที่ศึกษามี 4 ขนาด คือ 5p 10p 15p และ 20p[1] ตามลำดับ
4. การสูญหายของข้อมูลกำหนดให้ปรากฏขึ้นโดยสุ่ม และกำหนดให้อัตราการสูญหายของข้อมูลซึ่ง มีค่าเท่ากับ 5% 15 % และ 25 %
5. ระดับพหุสัมพันธ์ของตัวแปรอิสระ แบ่งเป็น 3 ระดับ คือ

   ระดับต่ำ ค่า ρ = (0.1 0.2 0.3)

   ระดับปานกลาง ค่า ρ = (0.4 0.5 0.6)

   ระดับสูง ค่า ρ = (0.7 0.7 0.8)

   โดยค่า ρ ในวงเล็บ คือ ความสัมพันธ์ระหว่าง X_1 กับ X_2, X_1 กับ X_3 และ X_2 กับ X_3 ตามลำดับ
6. กำหนดการประมวลผลในแต่ละสถานการณ์เท่ากับ 1,000 รอบ

---

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

3.2 **ขั้นตอนในการวิจัย** มีดังนี้

1. สร้างข้อมูลของความคลาดเคลื่อนที่มีการแจกแจงปกติ โดยมีพารามิเตอร์ตามที่กำหนด

2. สร้างข้อมูลของตัวแปรอิสระ (X) ให้มีระดับความสัมพันธ์ตามที่กำหนด และสร้างข้อมูลของตัวแปรตาม (y) จากรูปแบบความสัมพันธ์ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ โดยกำหนดให้ $\underset{\sim}{\beta}$ เป็นค่าคงที่ใดๆ คือ β_j = 1 โดยที่ j = 0, 1, …, p

3. สร้างข้อมูลของตัวแปรอิสระให้เกิดการสูญหายโดยเกิดขึ้นอย่างสุ่ม ตามที่กำหนด

4. ประมาณค่าข้อมูลเพื่อแทนที่ข้อมูลที่สูญหายในตัวแปรอิสระ ด้วยวิธี FOR และวิธี MFOR

5. ประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี CC วิธี FOR-W วิธี FOR-SR วิธี MFOR-W และวิธี MFOR-SR

6. คำนวณหาค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) พร้อมทั้งค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เพื่อเปรียบเทียบประสิทธิภาพของวิธีการทั้ง 5 วิธี

<u>รายละเอียดของแต่ละขั้นตอนมีดังนี้</u>

1. **การสร้างการแจกแจงของค่าความคลาดเคลื่อนตามที่กำหนด**

สร้างข้อมูลของความคลาดเคลื่อนที่มีการแจกแจงปกติ พารามิเตอร์ $\mu = 0$ และ $\sigma = 0.5$ 1 และ 3 โดยที่ฟังก์ชันความหนาแน่นของการแจกแจงปกติอยู่ในรูปของ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\}$$

เมื่อ $E(X) = \mu$, $Var(X) = \sigma^2$

2. **การสร้างตัวแปรอิสระ และสร้างตัวแปรตามให้มีความสัมพันธ์ตามที่กำหนด**

ในการวิจัยครั้งนี้ สร้างข้อมูลของตัวแปรอิสระ $X = \left[\underset{\sim}{1} \vdots X_1^{**} \right]$ เมื่อ $X_1^{**} = [\underset{\sim}{X_1^*}, \underset{\sim}{X_2^*}, ..., \underset{\sim}{X_n^*}]$ ซึ่ง $\underset{\sim}{X^*}_{i(1\times p)}, i = 1, 2, \ldots, n$ เป็นข้อมูลชุดที่ i ที่มีตัวแปรอิสระ p ตัว ซึ่งเราสามารถสร้างข้อมูลในแต่ละชุด p ตัว ให้มีการแจกแจงปกติของหลายตัวแปร (multivariate

normal distribution) ที่มีค่าเฉลี่ย $\underset{\sim}{\mu} = \underset{\sim}{0}$ และ $\sum$ เป็นเมทริกซ์ความแปรปรวนร่วม (covariance matrix) ที่แสดงความสัมพันธ์ระหว่างตัวแปรอิสระ สามารถเขียนได้เป็น $\underset{\sim}{X_i^*} \sim N_p\left(\underset{\sim}{\mu}, \Sigma_{p\times p}\right)$

ในปี ค.ศ. 1972 บาร์ (Barr) และเซิลสาค (Slesak) เสนอวิธีการสร้างข้อมูลที่มีการแจกแจงปกติของหลายตัวแปร ดังนี้

กำหนดให้ $\underset{\sim}{X_i^*}' = (X_{i1}, X_{i2}, \ldots, X_{ip})'$ เป็นเวกเตอร์ของตัวแปรอิสระที่มีการแจกแจงแบบปกติซึ่งมีความสัมพันธ์กันโดยมีเวกเตอร์ค่าเฉลี่ย $\underset{\sim}{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)'$ เราสามารถเขียนเมทริกซ์ความแปรปรวนร่วม $\sum$ ได้ดังนี้

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

เมื่อ σ_{ij} คือ ความแปรปรวนร่วมระหว่างตัวแปร X_i และ X_j

โดยที่ $\sigma_{ij} = \rho_{ij}(\sigma_i\sigma_j)\,;\ i, j = 1, 2, \ldots., p$

และ ρ_{ij} คือ สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X_i และ X_j

จากนั้นสร้างข้อมูลของตัวแปรตาม ($\underset{\sim}{y}$) จากรูปแบบความสัมพันธ์ $\underset{\sim}{y} = X\,\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ โดยกำหนดให้ $\underset{\sim}{\beta} = \underset{\sim}{1}$

3. การสร้างข้อมูลของตัวแปรอิสระให้เกิดการสูญหายขึ้นอย่างสุ่ม

จากการสร้างตัวแปรอิสระและตัวแปรตามในข้อ 2 แล้ว ก็จะทำการสุ่มข้อมูลให้เกิดการสูญหายโดยการสูญหายนั้น ผู้วิจัยได้กำหนดให้เกิดขึ้นอย่างสุ่มในตัวแปรอิสระเพียงหนึ่งตัวแปร โดยสัดส่วนข้อมูลสูญหายในตัวแปรอิสระตามที่ได้กำหนดไว้ในขอบเขตการวิจัย

4. ประมาณค่าข้อมูลเพื่อแทนที่ข้อมูลที่สูญหายในตัวแปรอิสระ

จากข้อ 1, 2 และ 3 จากนั้นทำการประมาณค่าข้อมูลสูญหายด้วยวิธี FOR และวิธี MFOR ที่ได้กล่าวไว้แล้วในบทที่ 2 ซึ่งวิธี FOR จะใช้หลักการของการวิเคราะห์สมการถดถอยด้วยความสัมพันธ์ระหว่างตัวแปรอิสระ p ตัว เพื่อประมาณค่าข้อมูลสูญหาย ส่วนวิธี MFOR จะใช้หลักการเช่นเดียวกันกับวิธี FOR แต่พิจารณาข้อมูลของตัวแปรตามเพิ่มเข้าไปเพื่อใช้ในการประมาณค่าข้อมูลสูญหาย

5. การประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี CC วิธี FOR-W วิธี FOR-SR วิธี MFOR-W และวิธี MFOR-SR

จากข้อ 1, 2, 3 และ 4 จากนั้นทำการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี CC วิธี FOR-W วิธี FOR-SR วิธี MFOR-W และวิธี MFOR-SR ตามทฤษฎีที่ได้กล่าวไว้แล้วในบทที่ 2 ซึ่งวิธี CC จะต้องทำการตัดชุดข้อมูลที่มีการสูญหายในตัวแปรอิสระก่อนแล้วจึงคำนวณค่าประมาณจากวิธี CC ส่วนวิธี FOR-W เป็นการประมาณค่าข้อมูลสูญหายด้วยวิธี FOR ก่อนแล้วจึงคำนวณค่าประมาณจากวิธี WM ส่วนวิธี FOR-SR เป็นการประมาณค่าข้อมูลสูญหายด้วยวิธี FOR ก่อนแล้วจึงคำนวณค่าประมาณจากวิธี SRM ส่วนวิธี MFOR-W เป็นการประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR ก่อนแล้วจึงคำนวณค่าประมาณจากวิธี WM ส่วนวิธี MFOR-SR เป็นการประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR ก่อนแล้วจึงคำนวณค่าประมาณจากวิธี SRM

6. การหาค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุคูณ และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

6.1 คำนวณค่า RMSE ของสัมประสิทธิ์การถดถอยพหุคูณแต่ละตัวในแต่ละวิธี เมื่อกระทำซ้ำ 1,000 รอบ

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

เมื่อ β_j แทนสัมประสิทธิ์การถดถอยพหุคูณตัวที่ j

$\hat{\beta}_{ij}$ แทนตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณตัวที่ j จากการประมาณครั้งที่ i

และ $RMSE_i$ แทนค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์การถดถอยพหุคูณจากการประมาณครั้งที่ i

6.2 คำนวณค่า ARMSE แต่ละวิธี โดยการหาค่าเฉลี่ยของ RMSE

$$ARMSE = \frac{1}{1000}\sum_{i=1}^{1000} RMSE_i$$

6.3 คำนวณค่า DIFF เพื่อใช้วัดว่าวิธีที่ให้ค่า ARMSE ต่ำที่สุดจะดีกว่าวิธีอื่นๆกี่เปอร์เซนต์

$$DIFF = \left[\frac{ARMSE_{(k)} - ARMSE_{(\min)}}{ARMSE_{(\min)}} \right] \times 100 \text{ , k = 1, 2, 3, 4, 5}$$

เมื่อ ARMSEแทนค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุคูณ

DIFF แทนเปอร์เซ็นต์ของอัตราส่วนผลต่างค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีต่างๆ

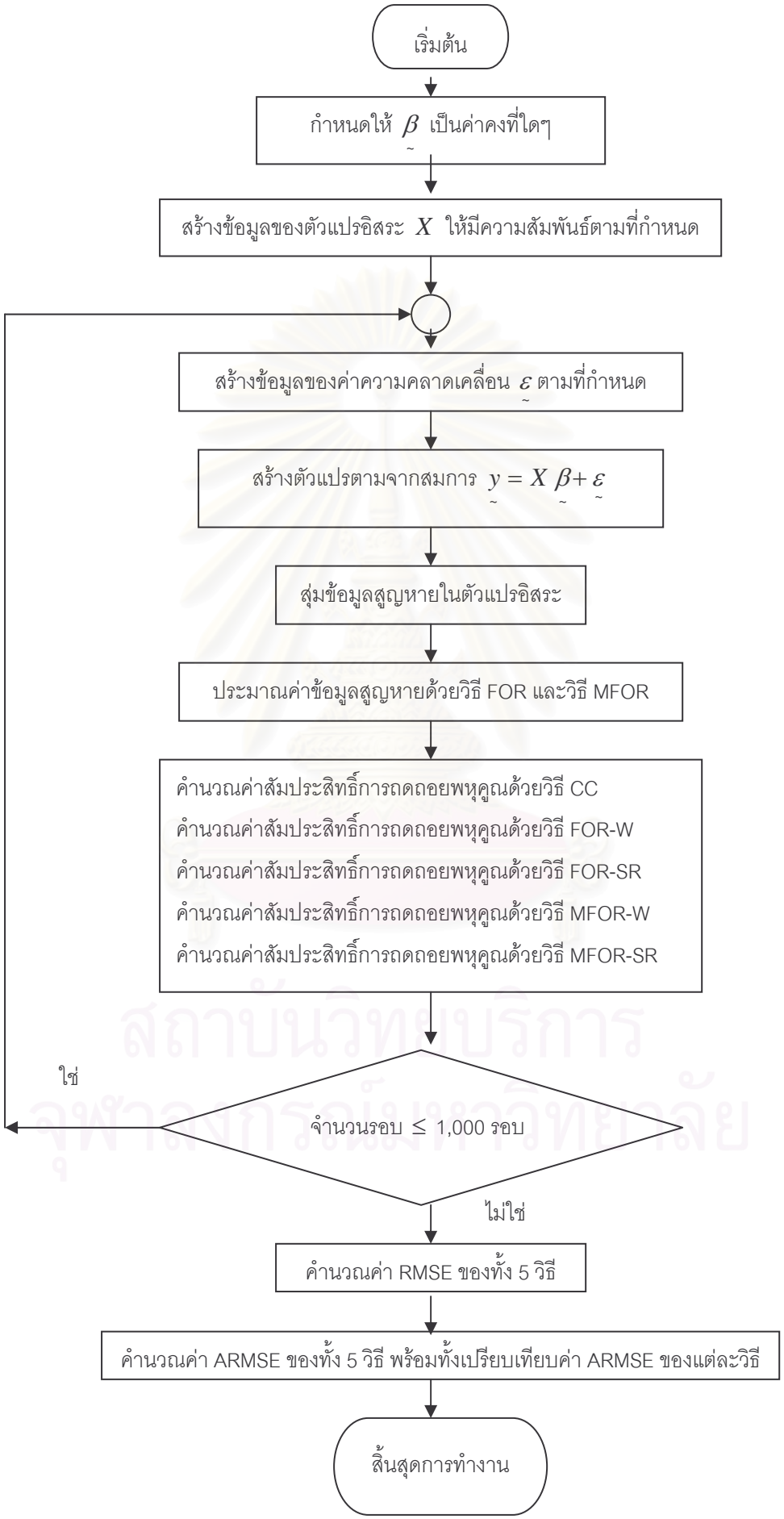
$ARMSE_{min}$แทนค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุคูณจากวิธีที่มีค่าน้อยที่สุด

และ $ARMSE_k$ แทนค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุคูณแต่ละวิธี

ผู้วิจัยได้แสดงตารางแผนผังการเขียนโปรแกรมดังรูปที่ 3.1

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>รูปที่ 3.1</u> แผนผังการเขียนโปรแกรม



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

ผลการวิจัย

ในการวิจัยครั้งนี้มีวัตถุประสงค์ที่จะเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดข้อมูลสูญหายในตัวแปรอิสระด้วยวิธี CC วิธี FOR-W วิธี FOR-SR วิธี MFOR-W และวิธี MFOR-SR โดยเกณฑ์ที่ใช้ในการตัดสินใจว่า วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณวิธีการใดทำให้ได้ค่าประมาณที่ใกล้เคียงกับค่าจริงมากที่สุด จะพิจารณาจากค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอย (ARMSE) และในการเปรียบเทียบประสิทธิภาพของวิธีการทั้ง 5 วิธี จะใช้ค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) มาประกอบการพิจารณาซึ่งวิธีใดให้ค่า ARMSE และค่า DIFF ต่ำที่สุดจะเป็นวิธีที่ดีที่สุด

การนำเสนอผลการวิจัยจะนำเสนอในรูปตารางและรูปกราฟ โดยใช้สัญลักษณ์แทนความหมายต่าง ๆ ดังนี้

n หมายถึง ขนาดตัวอย่าง

p หมายถึง จำนวนตัวแปรอิสระ เท่ากับ 3

CC หมายถึง การประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี CC

FOR-W หมายถึง การประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี FOR-W

FOR-SR หมายถึง การประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี FOR-SR

MFOR-W หมายถึง การประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี MFOR-W

MFOR-SR หมายถึง การประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี MFOR-SR

ARMSE หมายถึง ค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอย

SD หมายถึง ส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอย

DIFF หมายถึง ค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอย

การนำเสนอผลการวิจัยของการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระทั้ง 5 วิธีนั้น ประกอบด้วยตารางและรูปภาพ โดยแบ่งการนำเสนอเป็น 3 ส่วน ซึ่งใช้ระดับของค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเป็นเกณฑ์ในการแบ่งส่วนของการนำเสนอผลการวิจัย ดังนี้

ส่วนที่ 1 ผลการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ ในกรณีส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 0.5

ส่วนที่ 2 ผลการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ ในกรณีส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 1

ส่วนที่ 3 ผลการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ ในกรณีส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 3

รูปแบบการนำเสนอผลการวิจัยในทุกตอนนั้นจะเริ่มจากผลการวิจัยที่ประกอบด้วยตารางและรูปภาพสำหรับแต่ละส่วน ซึ่งเมื่อเปลี่ยนระดับของสัดส่วนข้อมูลสูญหายในตัวแปรอิสระจะมีการอธิบายผลการวิจัยที่ได้ จากนั้นจะทำการอธิบายผลการวิจัยที่ได้อีกครั้งเมื่อเปลี่ยนระดับของส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่ม และทำการอธิบายผลการวิจัยทั้งหมดของส่วนนั้นในตอนท้าย ของบท หลังจากที่นำเสนอผลการวิจัยครบทุกส่วนแล้วจะมีการอธิบายสรุปผลการวิจัยทั้งหมดอีกครั้งหนึ่ง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4.1 ผลการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ ในกรณีส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 0.5

การวิจัยในส่วนนี้ผู้วิจัยศึกษา กรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 0.5 ที่ความสัมพันธ์ระหว่างตัวแปรอิสระอยู่ในระดับต่ำ ปานกลาง และสูง และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 % 15 % และ 25 % ตามลำดับ โดยมีขนาดตัวอย่างเท่ากับ 5p 10p 15p และ 20p ตามลำดับ เมื่อ p คือ จำนวนตัวแปรอิสระซึ่งเท่ากับ 3 ซึ่งผลการวิจัยส่วนนี้นำเสนอในตารางที่ 4.1.1 – 4.1.3

รายละเอียดตารางที่ 4.1.1 – 4.1.3

ตารางที่	ส่วนเบี่ยงเบนมาตรฐาน	สัดส่วนข้อมูลสูญหายในตัวแปรอิสระ
4.1.1	0.5	5 %
4.1.2	0.5	15 %
4.1.3	0.5	25 %

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>ตารางที่ 4.1.1</u> **แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %**

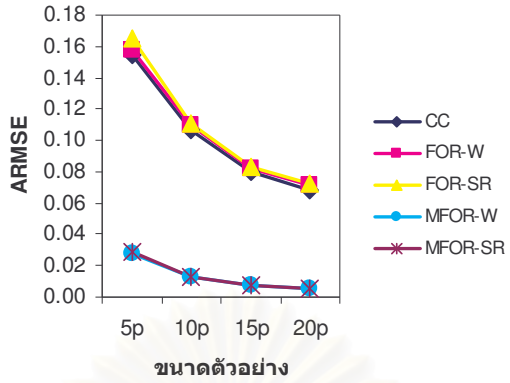
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์ระดับต่ำ	5p	ARMSE	0.1540	0.1654	0.1580	0.0283	0.0274
		SD	(0.0624)	(0.068)	(0.0641)	(0.0115)	(0.0112)
		DIFF	462.52	503.90	476.95	3.20	0.00
	10p	ARMSE	0.1061	0.1108	0.1102	0.0132	0.0131
		SD	(0.0419)	(0.0434)	(0.0434)	(0.0054)	(0.0052)
		DIFF	709.43	745.46	741.03	0.69	0.00
	15p	ARMSE	0.0795	0.0829	0.0823	0.0075	0.0075
		SD	(0.0294)	(0.0303)	(0.0299)	(0.0032)	(0.0032)
		DIFF	962.42	1007.03	999.93	0.53	0.00
	20p	ARMSE	0.0685	0.0725	0.0717	0.0057	0.0056
		SD	(0.0261)	(0.028)	(0.0278)	(0.0022)	(0.0021)
		DIFF	1122.19	1194.99	1180.06	1.34	0.00
ความสัมพันธ์ระดับปานกลาง	5p	ARMSE	0.1754	0.1824	0.1782	0.0362	0.0352
		SD	(0.0737)	(0.0773)	(0.0754)	(0.0152)	(0.0153)
		DIFF	398.81	418.80	407.00	2.93	0.00
	10p	ARMSE	0.1212	0.1238	0.1235	0.0177	0.0176
		SD	(0.0515)	(0.0518)	(0.0516)	(0.0071)	(0.0069)
		DIFF	588.78	603.91	602.26	0.57	0.00
	15p	ARMSE	0.0912	0.0928	0.0927	0.0097	0.0097
		SD	(0.0353)	(0.0362)	(0.0361)	(0.0034)	(0.0033)
		DIFF	839.02	856.02	854.35	0.22	0.00
	20p	ARMSE	0.0764	0.0781	0.0778	0.0070	0.0070
		SD	(0.0301)	(0.0308)	(0.0308)	(0.0027)	(0.0026)
		DIFF	998.28	1023.02	1018.65	0.69	0.00

ตารางที่ 4.1.1(ต่อ) แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %

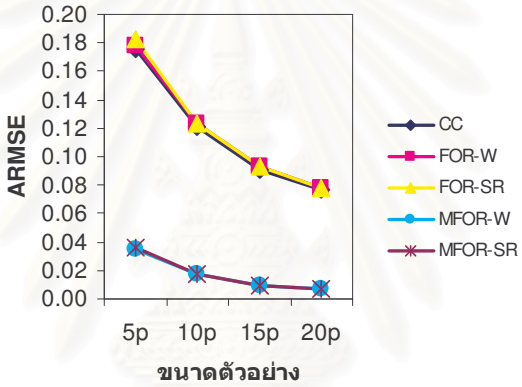
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์เชิงเส้น	5p	ARMSE	0.2810	0.2781	0.2778	0.1004	0.0999
		SD	(0.1514)	(0.1543)	(0.1535)	(0.0558)	(0.0559)
		DIFF	181.23	178.36	178.03	0.50	0.00
	10p	ARMSE	0.2014	0.2010	0.2009	0.0520	0.0519
		SD	(0.106)	(0.106)	(0.106)	(0.0269)	(0.0267)
		DIFF	287.94	287.19	286.82	0.21	0.00
	15p	ARMSE	0.1447	0.1442	0.1441	0.0258	0.0257
		SD	(0.0713)	(0.0705)	(0.0706)	(0.0121)	(0.0120)
		DIFF	462.73	460.48	460.32	0.30	0.00
	20p	ARMSE	0.1211	0.1208	0.1206	0.0184	0.0183
		SD	(0.0603)	(0.0604)	(0.0604)	(0.0082)	(0.0082)
		DIFF	562.18	559.24	559.18	0.62	0.00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

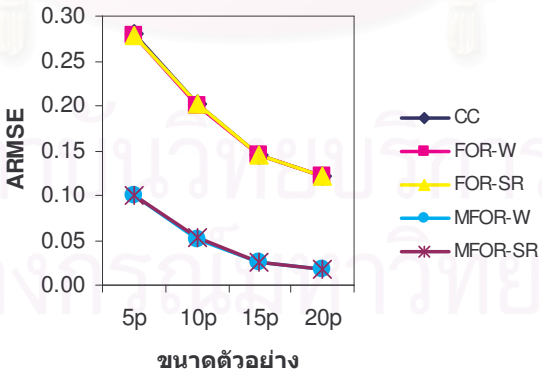
ก) ความสัมพันธ์ระดับต่ำ



ข) ความสัมพันธ์ระดับปานกลาง



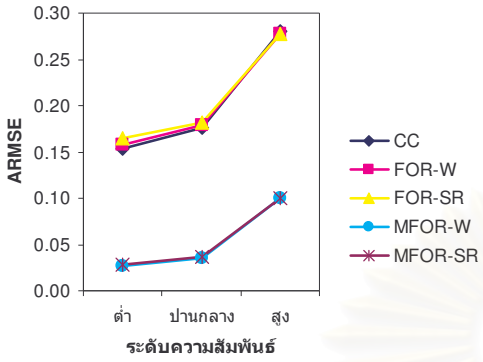
ค) ความสัมพันธ์ระดับสูง



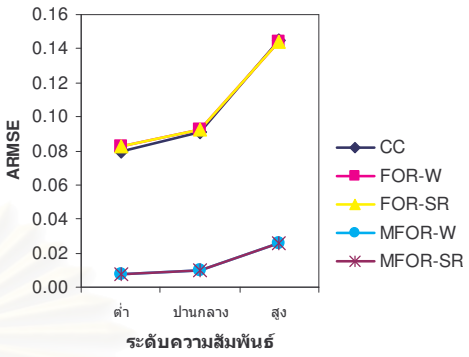
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>รูปที่ 4.1.1</u> แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และขนาดตัวอย่างเมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %

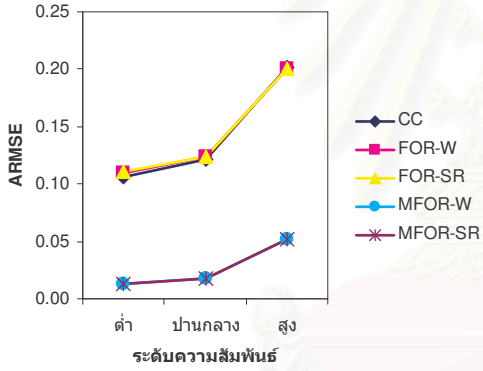
ก) n= 5p



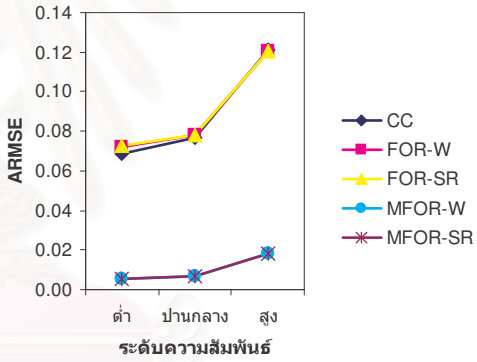
ค) n=15p



ข) n=10p



ง) n=20p



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>รูปที่ 4.1.2</u> แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และระดับความสัมพันธ์ของตัวแปรอิสระ เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %

ผลการวิจัยของการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ กรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 % พบว่า กรณีความสัมพันธ์ระหว่างตัวแปรอิสระระดับต่ำและปานกลาง วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง รองลงมาคือ วิธี MFOR-W วิธี CC วิธี FOR-SR และวิธี FOR-W ตามลำดับ ส่วนกรณีความสัมพันธ์ระหว่างตัวแปรอิสระระดับสูง วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง วิธีที่รองลงมาคือ วิธี MFOR-W วิธี FOR-SR วิธี FOR-W และวิธี CC โดยที่แต่ละวิธีเมื่อขนาดตัวอย่างเพิ่มขึ้น จะทำให้ค่า ARMSE มีแนวโน้มลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นทำให้ความแปรปรวนลดลงส่งผลให้ค่าความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยลดลง ส่วนระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้นจะทำให้ค่า ARMSE ของทุกวิธีเพิ่มขึ้น เนื่องจากระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้น ทำให้ค่าลักษณะเฉพาะของเมทริกซ์ $X'X$ มีค่าลดลงจึงส่งผลทำให้ค่า ARMSE มีแนวโน้มเพิ่มขึ้น

เมื่อพิจารณาค่า DIFF ของแต่ละวิธีการที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยในทุกขนาดตัวอย่าง พบว่า ประสิทธิภาพของวิธี MFOR-SR จะโดดเด่นกว่าอีก 3 วิธี คือ วิธี FOR-SR วิธี FOR-W และวิธี CC อย่างชัดเจน และมีประสิทธิภาพใกล้เคียงกับวิธี MFOR-W เพราะเป็นการประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR เช่นเดียวกัน โดยประสิทธิภาพของวิธี MFOR-SR จะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น กล่าวคือวิธี MFOR-SR มีประสิทธิภาพดีกว่าวิธี CC อย่างมากประมาณ 462.52 % 709.43 % 962.42 % และ 1122.19 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-W อย่างมากประมาณ 503.90 % 745.46 % 1007.03 % และ 1194.99 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-SR อย่างมากประมาณ 476.95 % 741.03 % 999.93 % และ 1180.06 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น และมีประสิทธิภาพดีกว่าวิธี MFOR-W อย่างมากประมาณ 3.2 % 0.69 % 0.53 % และ 1.34 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น แต่วิธี MFOR-SR จะมีประสิทธิภาพลดลงเมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.1.2 แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่ม เท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %

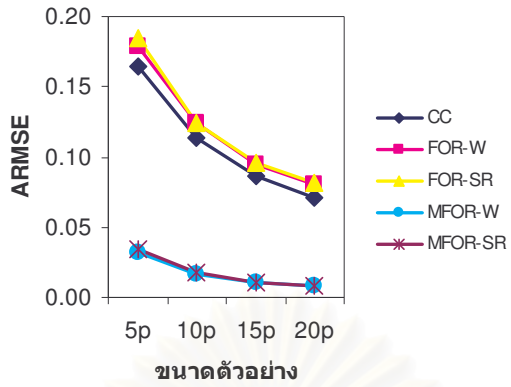
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์ระดับต่ำ	5p	ARMSE	0.1646	0.1849	0.1784	0.0343	0.0324
		SD	(0.066)	(0.0726)	(0.0708)	(0.0146)	(0.0145)
		DIFF	407.45	470.03	450.12	5.60	0.00
	10p	ARMSE	0.1131	0.1249	0.1238	0.0172	0.0169
		SD	(0.0463)	(0.05)	(0.0495)	(0.0067)	(0.0069)
		DIFF	568.54	637.93	631.90	1.67	0.00
	15p	ARMSE	0.0860	0.0962	0.0946	0.0106	0.0103
		SD	(0.0308)	(0.0365)	(0.0357)	(0.0038)	(0.0041)
		DIFF	737.62	837.55	821.42	3.46	0.00
	20p	ARMSE	0.0715	0.0813	0.0806	0.0082	0.0080
		SD	(0.0277)	(0.0317)	(0.0315)	(0.0028)	(0.0027)
		DIFF	793.34	915.84	907.27	2.23	0.00
ความสัมพันธ์ระดับปานกลาง	5p	ARMSE	0.1829	0.1986	0.1917	0.0420	0.0401
		SD	(0.0772)	(0.0842)	(0.0815)	(0.0172)	(0.0173)
		DIFF	356.51	395.82	378.42	4.94	0.00
	10p	ARMSE	0.1307	0.1384	0.1377	0.0225	0.0222
		SD	(0.058)	(0.0587)	(0.0585)	(0.094)	(0.0054)
		DIFF	488.19	522.83	519.72	1.24	0.00
	15p	ARMSE	0.0968	0.1022	0.1013	0.0123	0.0120
		SD	(0.037)	(0.039)	(0.0386)	(0.0061)	(0.0060)
		DIFF	703.55	748.24	740.93	2.14	0.00
	20p	ARMSE	0.0820	0.0873	0.0869	0.0091	0.0090
		SD	(0.0323)	(0.0346)	(0.0345)	(0.0035)	(0.0035)
		DIFF	807.20	866.41	861.63	1.24	0.00

ตารางที่ 4.1.2(ต่อ) แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %

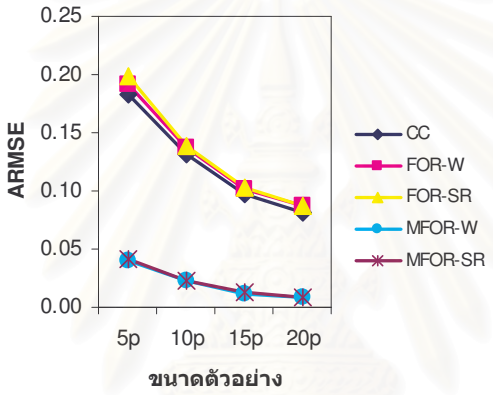
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์เชิงเส้นตรง	5p	ARMSE	0.2940	0.2860	0.2848	0.1141	0.1120
		SD	(0.158)	(0.1554)	(0.1541)	(0.0633)	(0.0626)
		DIFF	162.55	155.40	154.29	1.85	0.00
	10p	ARMSE	0.2125	0.2096	0.2094	0.0610	0.0606
		SD	(0.1151)	(0.1145)	(0.1147)	(0.0328)	(0.0323)
		DIFF	250.80	245.91	245.66	0.62	0.00
	15p	ARMSE	0.1552	0.1533	0.1533	0.0299	0.0298
		SD	(0.0761)	(0.0764)	(0.0764)	(0.0150)	(0.0151)
		DIFF	420.60	414.22	414.09	0.23	0.00
	20p	ARMSE	0.1269	0.1252	0.1251	0.0203	0.0203
		SD	(0.0647)	(0.064)	(0.0641)	(0.0095)	(0.0093)
		DIFF	526.30	517.62	517.47	0.14	0.00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

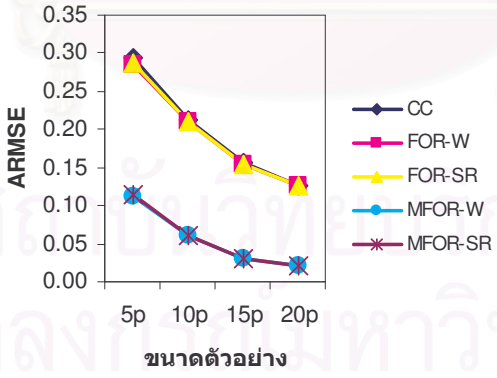
ก) ความสัมพันธ์ระดับต่ำ



ข) ความสัมพันธ์ระดับปานกลาง



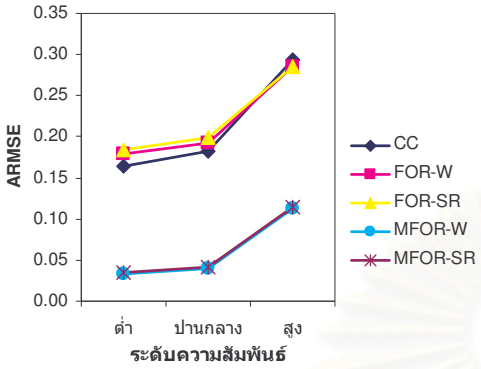
ค) ความสัมพันธ์ระดับสูง



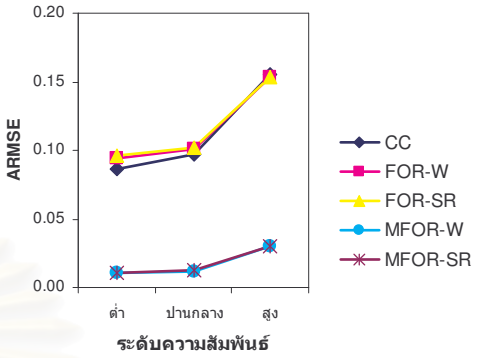
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 4.1.3 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และขนาดตัวอย่าง เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %

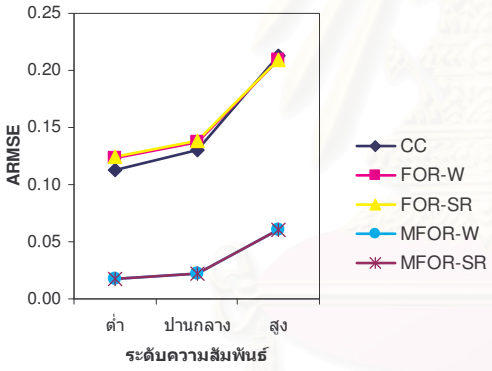
ก) n= 5p



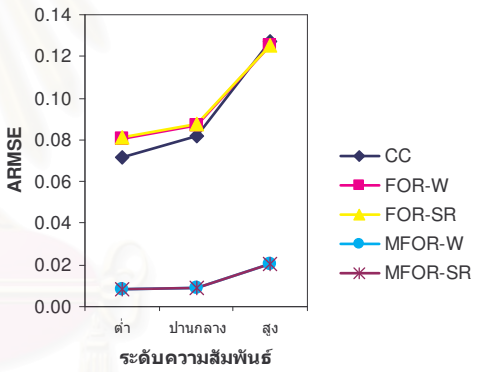
ค) n=15p



ข) n=10p



ง) n=20p



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>รูปที่ 4.1.4</u> แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และระดับความสัมพันธ์ของตัวแปรอิสระ เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %

ผลการวิจัยของการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ กรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 % พบว่า กรณีความสัมพันธ์ระหว่างตัวแปรอิสระระดับต่ำและปานกลาง วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง รองลงมาคือ วิธี MFOR-W วิธี CC วิธี FOR-SR และวิธี FOR-W ตามลำดับ ส่วนกรณีความสัมพันธ์ระหว่างตัวแปรอิสระระดับสูง วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง วิธีที่รองลงมาคือ วิธี MFOR-W วิธี FOR-SR วิธี FOR-W และวิธี CC โดยที่แต่ละวิธีเมื่อขนาดตัวอย่างเพิ่มขึ้น จะทำให้ค่า ARMSE มีแนวโน้มลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นทำให้ความแปรปรวนลดลงส่งผลให้ค่าความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยลดลง ส่วนระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้นจะทำให้ค่า ARMSE ของทุกวิธีเพิ่มขึ้น เนื่องจากระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้น ทำให้ค่าลักษณะเฉพาะของเมทริกซ์ $X'X$ มีค่าลดลงจึงส่งผลทำให้ค่า ARMSE มีแนวโน้มเพิ่มขึ้น

เมื่อพิจารณาค่า DIFF ของแต่ละวิธีการที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอย ในทุกขนาดตัวอย่าง พบว่า ประสิทธิภาพของวิธี MFOR-SR จะโดดเด่นกว่าอีก 3 วิธี คือ วิธี FOR-SR วิธี FOR-W และวิธี CC อย่างชัดเจน และมีประสิทธิภาพใกล้เคียงกับวิธี MFOR-W เพราะเป็นการประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR เช่นเดียวกัน โดยประสิทธิภาพของวิธี MFOR-SR จะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น กล่าวคือวิธี MFOR-SR มีประสิทธิภาพดีกว่าวิธี CC อย่างมากประมาณ 407.45 % 568.54 % 737.62 % และ 793.34 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-W อย่างมากประมาณ 470.03 % 637.93 % 837.55 % และ 915.84 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-SR อย่างมากประมาณ 450.12 % 631.90 % 821.42 % และ 907.27 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น และมีประสิทธิภาพดีกว่าวิธี MFOR-W อย่างมากประมาณ 5.60 % 1.67 % 3.46 % และ 2.23 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น แต่วิธี MFOR-SR จะมีประสิทธิภาพลดลงเมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.1.3 แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %

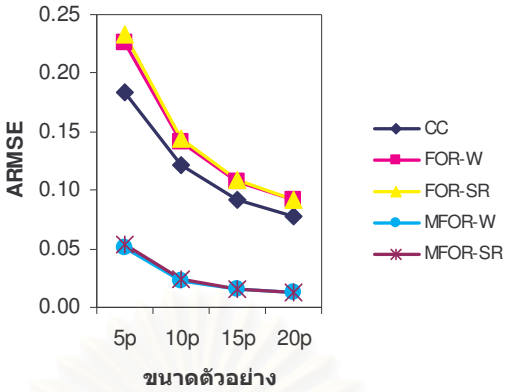
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์ระดับต่ำ	5p	ARMSE	0.1841	0.2327	0.2262	0.0534	0.0510
		SD	(0.0792)	(0.0965)	(0.096)	(0.0242)	(0.0222)
		DIFF	261.02	356.32	343.59	4.73	0.00
	10p	ARMSE	0.1209	0.1436	0.1407	0.0242	0.0230
		SD	(0.0494)	(0.0583)	(0.0575)	(0.0106)	(0.0106)
		DIFF	425.80	524.67	512.01	5.15	0.00
	15p	ARMSE	0.0915	0.1083	0.1072	0.0159	0.0155
		SD	(0.0355)	(0.0437)	(0.0432)	(0.0056)	(0.0057)
		DIFF	490.98	599.35	592.32	2.79	0.00
	20p	ARMSE	0.0782	0.0924	0.0914	0.0126	0.0122
		SD	(0.0291)	(0.0347)	(0.0344)	(0.0050)	(0.0051)
		DIFF	541.67	658.15	650.14	3.54	0.00
ความสัมพันธ์ระดับปานกลาง	5p	ARMSE	0.2112	0.2307	0.2257	0.0704	0.0678
		SD	(0.098)	(0.1064)	(0.1059)	(0.0308)	(0.0306)
		DIFF	211.34	240.17	232.83	3.77	0.00
	10p	ARMSE	0.1413	0.1483	0.1464	0.0298	0.0286
		SD	(0.0614)	(0.0652)	(0.0647)	(0.0078)	(0.0133)
		DIFF	394.50	419.11	412.39	4.35	0.00
	15p	ARMSE	0.1027	0.1105	0.1099	0.0166	0.0163
		SD	(0.0423)	(0.0461)	(0.0459)	(0.0069)	(0.0067)
		DIFF	528.37	576.36	572.44	1.80	0.00
	20p	ARMSE	0.0886	0.0952	0.0946	0.0127	0.0124
		SD	(0.0343)	(0.0387)	(0.0384)	(0.0054)	(0.0053)
		DIFF	614.01	667.42	662.96	2.24	0.00

ตารางที่ 4.1.3(ต่อ) แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %

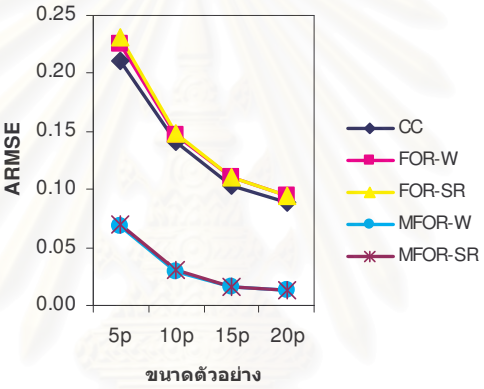
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์ระดับสูง	5p	ARMSE	0.3346	0.3192	0.3183	0.1671	0.1618
		SD	(0.1812)	(0.1759)	(0.1764)	(0.0923)	(0.0921)
		DIFF	106.78	97.27	96.75	3.28	0.00
	10p	ARMSE	0.2291	0.2214	0.2213	0.0786	0.0772
		SD	(0.1243)	(0.123)	(0.123)	(0.0430)	(0.0429)
		DIFF	196.81	186.92	186.78	1.84	0.00
	15p	ARMSE	0.1664	0.1623	0.1622	0.0356	0.0354
		SD	(0.0833)	(0.0825)	(0.0825)	(0.0174)	(0.0172)
		DIFF	370.05	358.58	358.24	0.49	0.00
	20p	ARMSE	0.1393	0.1359	0.1358	0.0249	0.0248
		SD	(0.0686)	(0.0684)	(0.0684)	(0.0126)	(0.0124)
		DIFF	462.22	448.54	448.14	0.51	0.00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

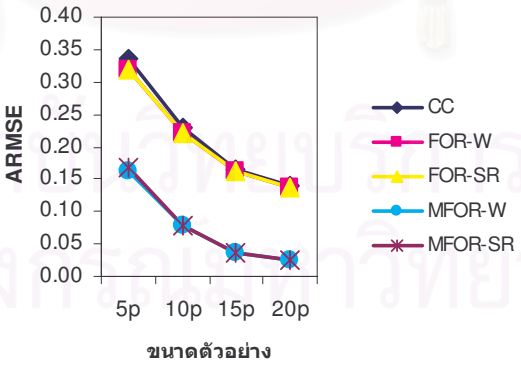
ก) ความสัมพันธ์ระดับต่ำ



ข) ความสัมพันธ์ระดับปานกลาง



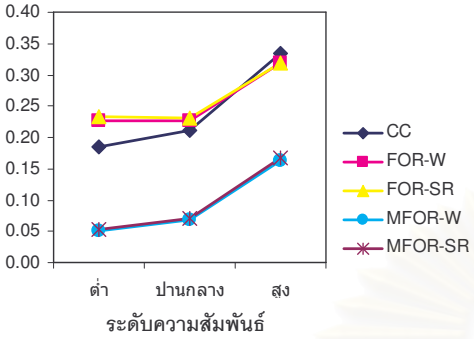
ค) ความสัมพันธ์ระดับสูง



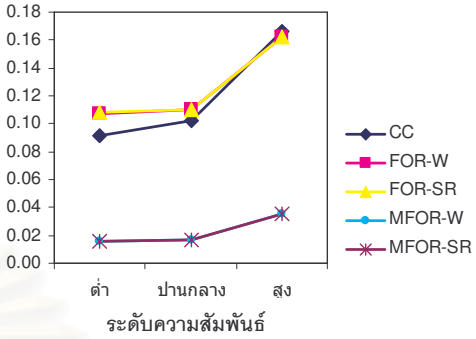
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 4.1.5 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และขนาดตัวอย่าง เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %

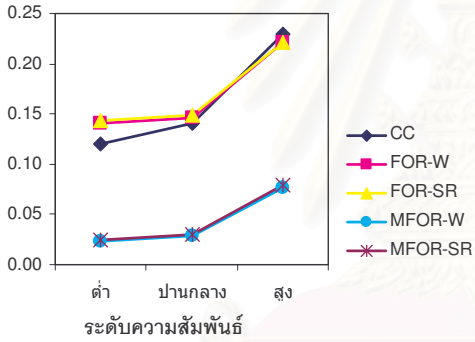
ก) n= 5p



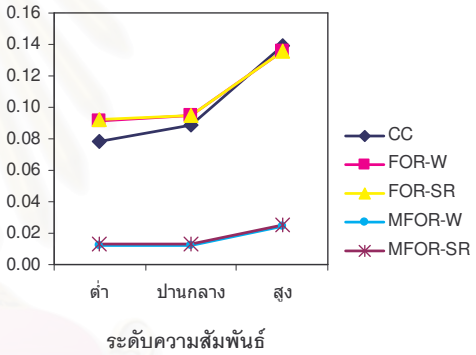
ค) n=15p



ข) n=10p



ง) n=20p



<u>รูปที่ 4.1.6</u> แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และระดับความสัมพันธ์ของตัวแปรอิสระ เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %

ผลการวิจัยของการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ กรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 0.5 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 % พบว่า กรณีความสัมพันธ์ระหว่างตัวแปรอิสระระดับต่ำและปานกลาง วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง รองลงมาคือ วิธี MFOR-W วิธี CC วิธี FOR-SR และวิธี FOR-W ตามลำดับ ส่วนกรณีความสัมพันธ์ระหว่างตัวแปรอิสระระดับสูง วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง วิธีที่รองลงมาคือ วิธี MFOR-W วิธี FOR-SR วิธี FOR-W และวิธี CCโดยที่แต่ละวิธีเมื่อขนาดตัวอย่างเพิ่มขึ้น จะทำให้ค่า ARMSE มีแนวโน้มลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นทำให้ความแปรปรวนลดลงส่งผลให้ค่าความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยลดลง ส่วนระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้นจะทำให้ค่า ARMSE ของทุกวิธีเพิ่มขึ้น เนื่องจากระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้น ทำให้ค่าลักษณะเฉพาะของเมทริกซ์ $X'X$ มีค่าลดลงจึงส่งผลทำให้ค่า ARMSE มีแนวโน้มเพิ่มขึ้น

เมื่อพิจารณาค่า DIFF ของแต่ละวิธีการที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยในทุกขนาดตัวอย่าง พบว่า ประสิทธิภาพของวิธี MFOR-SR จะโดดเด่นกว่าอีก 3 วิธี คือ วิธี FOR-SR วิธี FOR-W และวิธี CC อย่างชัดเจน และมีประสิทธิภาพใกล้เคียงกับวิธี MFOR-W เพราะเป็นการประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR เช่นเดียวกัน โดยประสิทธิภาพของวิธี MFOR-SR จะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น กล่าวคือวิธี MFOR-SR มีประสิทธิภาพดีกว่าวิธี CC อย่างมากประมาณ 261.02 % 425.80 % 490.98 % และ 541.67 % ตามลำดับ ทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-W อย่างมากประมาณ 356.32 % 524.67 % 599.35 % และ 658.15 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-SR อย่างมากประมาณ 343.59 % 512.01 % 592.32 % และ 650.14 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น และมีประสิทธิภาพดีกว่าวิธี MFOR-W อย่างมากประมาณ 4.73 % 5.15 % 2.79 % และ 3.54 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น แต่วิธี MFOR-SR จะมีประสิทธิภาพลดลงเมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น

สรุปส่วนที่ 4.1 ผลการวิจัยของการเปรียบเทียบการประมาณสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ กรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 0.5

ทุกๆสัดส่วนข้อมูลสุญหายในตัวแปรอิสระ พบว่า เมื่อความสัมพันธ์ระหว่างตัวแปรอิสระระดับต่ำและปานกลาง วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง รองลงมาคือ วิธี MFOR-W วิธี CC วิธี FOR-SR และวิธี FOR-W ตามลำดับ และเมื่อกรณีความสัมพันธ์ระหว่างตัวแปรอิสระระดับสูง วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง รองลงมาคือ วิธี MFOR-W วิธี FOR-SR วิธี FOR-W และวิธีCC เพราะวิธีประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีของ SRM เป็นวิธีที่ลดความคลาดเคลื่อนของการประมาณค่าที่เกิดจากความสัมพันธ์กันระหว่างตัวแปรอิสระ แต่วิธีของ WM นั่นเป็นการถ่วงน้ำหนักชุดของข้อมูลที่มีการสูญหายภายหลังจากที่ทำการประมาณค่าแล้ว ส่วนวิธีประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR จะให้ค่าที่ใกล้เคียงค่าจริงมากกว่าวิธี FOR เนื่องจากมีการเพิ่มตัวแปรตามในการประมาณค่า โดยที่ข้อมูลของตัวแปรตามมีความแปรปรวนน้อย และวิธี CC การที่สัดส่วนข้อมูลสูญหายเพิ่มขึ้นจะส่งผลให้การตัดชุดข้อมูลที่มีการสูญหายนั่นจะทำให้เกิดความคลาดเคลื่อนในการประมาณค่าไปจากความเป็นจริงเพิ่มขึ้นเนื่องจากขนาดของข้อมูลที่ลดลง

จากผลการวิจัยในส่วนที่ 4.1 สามารถสรุปเกี่ยวกับปัจจัยที่มีผลต่อค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) ได้ดังนี้

1. **เมื่อขนาดตัวอย่างเพิ่มขึ้น** วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระทั้ง 5 วิธี จะให้ค่า ARMSE ลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นจะช่วยลดความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยลดลง

2. **เมื่อสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเพิ่มขึ้น** วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระทั้ง 5 วิธี จะทำให้ค่า ARMSE เพิ่มขึ้น เนื่องจากในวิธี CC เป็นการประมาณค่าสัมประสิทธิ์การถดถอยโดยการตัดชุดข้อมูลสูญหาย จึงทำให้ขนาดตัวอย่างลดลง มีผลให้ในการประมาณค่าสัมประสิทธิ์การถดถอยมีความคลาดเคลื่อนไปจากความเป็นจริง ส่วนค่า ARMSE ของวิธี MFOR-SR วิธี MFOR-W วิธี FOR-W และวิธี FOR-SR ที่เพิ่มขึ้น เนื่องจากค่าความคลาดเคลื่อนในการประมาณค่าข้อมูลสูญหาย จึงส่งผลให้ค่าความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยเพิ่มขึ้น

3. **เมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น** จะส่งผลให้ค่า ARMSE มีแนวโน้มสูงขึ้น เนื่องจากระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้น ทำให้ค่าลักษณะเฉพาะของเมทริกซ์ $X'X$ มีค่าลดลงจึงส่งผลทำให้ค่า ARMSE มีแนวโน้มเพิ่มขึ้น

4.2 ผลการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ ในกรณีส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 1

การวิจัยในส่วนนี้ผู้วิจัยศึกษา กรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 1 ที่ความสัมพันธ์ระหว่างตัวแปรอิสระอยู่ในระดับต่ำ ปานกลาง และสูง และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 % 15 % และ 25 % ตามลำดับ โดยมีขนาดตัวอย่างเท่ากับ 5p 10p 15p และ 20p ตามลำดับ เมื่อ p คือ จำนวนตัวแปรอิสระซึ่งเท่ากับ 3 ซึ่งผลการวิจัยส่วนนี้นำเสนอในตารางที่ 4.2.1 – 4.2.3

รายละเอียดตารางที่ 4.2.1 – 4.2.3

ตารางที่	ส่วนเบี่ยงเบนมาตรฐาน	สัดส่วนข้อมูลสูญหายในตัวแปรอิสระ
4.2.1	1	5 %
4.2.2	1	15 %
4.2.3	1	25 %

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.2.1 แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %

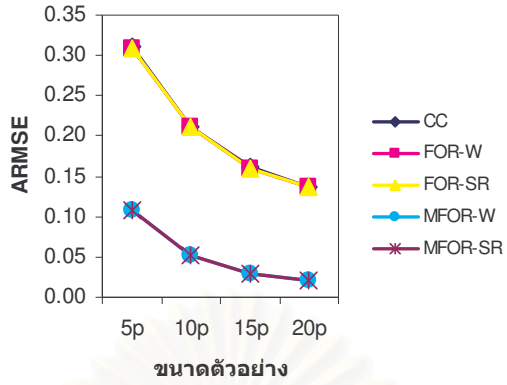
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์ระดับเส้นตรง	5p	ARMSE	0.3102	0.3084	0.3079	0.1080	0.1071
		SD	(0.1233)	(0.1214)	(0.1217)	(0.0896)	(0.0886)
		DIFF	189.52	187.87	187.39	0.80	0.00
	10p	ARMSE	0.2120	0.2114	0.2117	0.0513	0.0513
		SD	(0.0839)	(0.082)	(0.0818)	(0.0403)	(0.0403)
		DIFF	313.61	313.42	312.12	0.14	0.00
	15p	ARMSE	0.1606	0.1594	0.1593	0.0288	0.0287
		SD	(0.0585)	(0.0595)	(0.0594)	(0.0205)	(0.0204)
		DIFF	458.97	454.86	454.38	0.29	0.00
	20p	ARMSE	0.1367	0.1368	0.1365	0.0210	0.0210
		SD	(0.0513)	(0.0514)	(0.0515)	(0.0158)	(0.0158)
		DIFF	552.48	552.32	551.57	0.02	0.00
ความสัมพันธ์ระดับโค้งกลาง	5p	ARMSE	0.3525	0.3445	0.3468	0.1392	0.1381
		SD	(0.1505)	(0.1498)	(0.1498)	(0.1249)	(0.1261)
		DIFF	155.21	149.42	151.11	0.80	0.00
	10p	ARMSE	0.2419	0.2389	0.2393	0.0686	0.0684
		SD	(0.1044)	(0.1017)	(0.102)	(0.0618)	(0.0616)
		DIFF	253.52	249.12	249.69	0.25	0.00
	15p	ARMSE	0.1820	0.1803	0.1805	0.0375	0.0374
		SD	(0.0711)	(0.0703)	(0.0704)	(0.0288)	(0.0287)
		DIFF	386.01	381.49	382.13	0.28	0.00
	20p	ARMSE	0.1526	0.1524	0.1525	0.0269	0.0269
		SD	(0.0603)	(0.0605)	(0.0606)	(0.0219)	(0.0218)
		DIFF	467.40	466.76	466.95	0.06	0.00

ตารางที่ 4.2.1(ต่อ) แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %

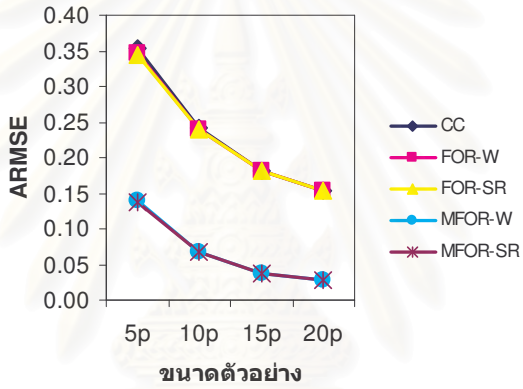
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์ระดับปานกลาง	5p	ARMSE	0.5582	0.5439	0.5490	0.4018	0.4017
		SD	(0.3036)	(0.2991)	(0.3028)	(0.4819)	(0.48)
		DIFF	38.95	35.39	36.67	0.02	0.00
	10p	ARMSE	0.3977	0.3913	0.3924	0.2088	0.2085
		SD	(0.2133)	(0.2118)	(0.2125)	(0.2488)	(0.2491)
		DIFF	90.77	87.70	88.26	0.16	0.00
	15p	ARMSE	0.2876	0.2840	0.2846	0.1026	0.1021
		SD	(0.1399)	(0.1394)	(0.1398)	(0.1065)	(0.1061)
		DIFF	181.67	178.11	178.72	0.43	0.00
	20p	ARMSE	0.2407	0.2384	0.2388	0.0729	0.0728
		SD	(0.1174)	(0.117)	(0.1173)	(0.0751)	(0.075)
		DIFF	230.50	227.28	227.85	0.12	0.00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

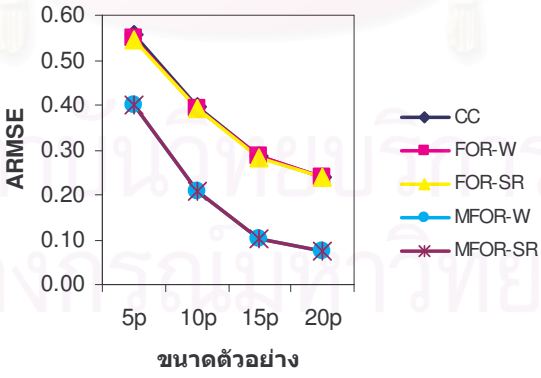
ก) ความสัมพันธ์ระดับต่ำ



ข) ความสัมพันธ์ระดับปานกลาง



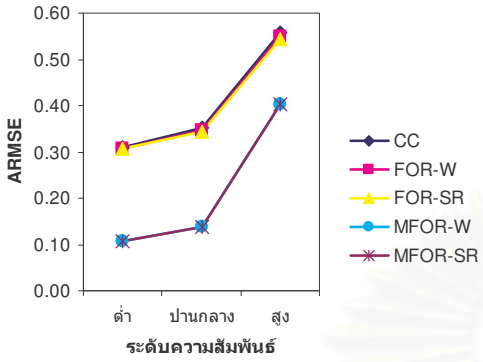
ค) ความสัมพันธ์ระดับสูง



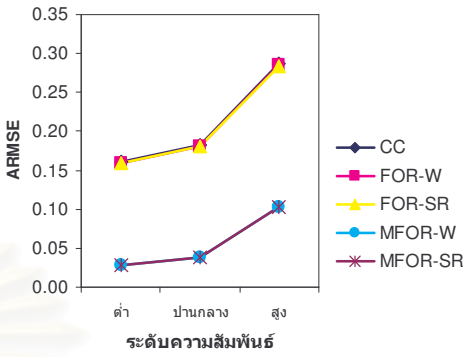
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>รูปที่ 4.2.1</u> แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และขนาดตัวอย่าง เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %

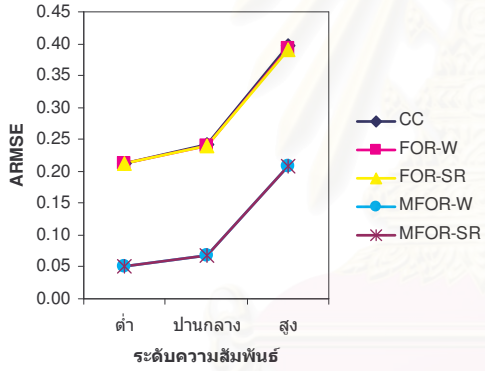
ก) n= 5p



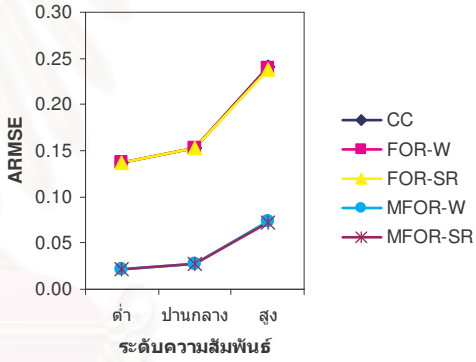
ค) n=15p



ข) n=10p



ง) n=20p



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 4.2.2 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และระดับความสัมพันธ์ของตัวแปรอิสระ เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %

ผลการวิจัยของการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ กรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 % พบว่า กรณีความสัมพันธ์ระหว่างตัวแปรอิสระระดับต่ำ วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง รองลงมาคือ วิธี MFOR-W วิธี FOR-W วิธี FOR-SR และวิธี CC ตามลำดับ ส่วนกรณีความสัมพันธ์ระหว่างตัวแปรอิสระระดับปานกลางสูง วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง วิธีที่รองลงมาคือ วิธี MFOR-W วิธี FOR-W วิธี FOR-SR และ วิธี CC ตามลำดับ โดยที่แต่ละวิธีเมื่อขนาดตัวอย่างเพิ่มขึ้น จะทำให้ค่า ARMSE มีแนวโน้มลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นทำให้ความแปรปรวนลดลงส่งผลให้ค่าความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยลดลง ส่วนระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้นจะทำให้ค่า ARMSE ของทุกวิธีเพิ่มขึ้น เนื่องจากระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้น ทำให้ค่าลักษณะเฉพาะของเมทริกซ์ $X'X$ มีค่าลดลงจึงส่งผลทำให้ค่า ARMSE มีแนวโน้มเพิ่มขึ้น

เมื่อพิจารณาค่า DIFF ของแต่ละวิธีการที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยในทุกขนาดตัวอย่าง พบว่า ประสิทธิภาพของวิธี MFOR-SR จะโดดเด่นกว่าอีก 3 วิธี คือ วิธี FOR-SR วิธี FOR-W และวิธี CC อย่างชัดเจน และมีประสิทธิภาพใกล้เคียงกับวิธี MFOR-W เพราะเป็นการประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR เช่นเดียวกัน โดยประสิทธิภาพของวิธี MFOR-SR จะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น กล่าวคือวิธี MFOR-SR มีประสิทธิภาพดีกว่าวิธี CC อย่างมากประมาณ 189.52 % 313.61 % 458.97 % และ 552.48 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-W อย่างมากประมาณ 187.87 % 313.12 % 454.86 % และ 551.52 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-SR อย่างมากประมาณ 187.87 % 313.12 % 454.86 % และ 551.52 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น และมีประสิทธิภาพดีกว่าวิธี MFOR-W อย่างมากประมาณ 0.80 % 0.25 % 0.43 % และ 0.12 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น แต่วิธี MFOR-SR จะมีประสิทธิภาพลดลงเมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น

<u>ตารางที่ 4.2.2</u> **แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %**

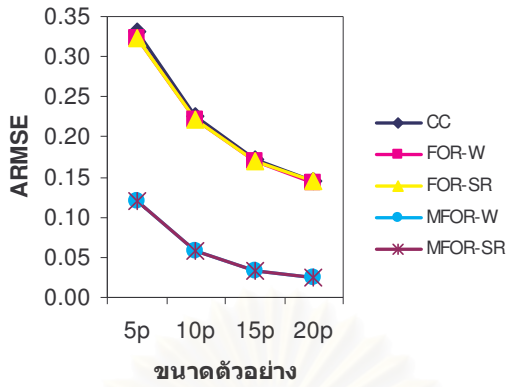
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์ระดับต่ำ	5p	ARMSE	0.3309	0.3232	0.3230	0.1211	0.1206
		SD	(0.1355)	(0.133)	(0.1329)	(0.1046)	(0.1042)
		DIFF	174.34	167.95	167.85	0.42	0.00
	10p	ARMSE	0.2262	0.2218	0.2215	0.0586	0.0585
		SD	(0.0903)	(0.0883)	(0.0884)	(0.048)	(0.0478)
		DIFF	286.87	279.33	278.85	0.17	0.00
	15p	ARMSE	0.1713	0.1698	0.1692	0.0334	0.0330
		SD	(0.0628)	(0.0624)	(0.0621)	(0.024)	(0.0238)
		DIFF	418.72	414.17	412.54	1.05	0.00
	20p	ARMSE	0.1444	0.1441	0.1439	0.0240	0.0239
		SD	(0.055)	(0.0532)	(0.0532)	(0.0172)	(0.0172)
		DIFF	505.40	504.18	503.35	0.45	0.00
ความสัมพันธ์ระดับปานกลาง	5p	ARMSE	0.3649	0.3541	0.3545	0.1511	0.1507
		SD	(0.1591)	(0.1531)	(0.1542)	(0.1397)	(0.141)
		DIFF	142.07	134.93	135.22	0.27	0.00
	10p	ARMSE	0.2576	0.2489	0.2494	0.0803	0.0801
		SD	(0.111)	(0.1095)	(0.1097)	(0.0753)	(0.0749)
		DIFF	221.48	210.62	211.26	0.21	0.00
	15p	ARMSE	0.1959	0.1895	0.1896	0.0431	0.0428
		SD	(0.0788)	(0.0754)	(0.0756)	(0.0345)	(0.034)
		DIFF	357.55	342.71	342.79	0.56	0.00
	20p	ARMSE	0.1628	0.1589	0.1590	0.0306	0.0305
		SD	(0.0651)	(0.0634)	(0.0636)	(0.025)	(0.025)
		DIFF	433.08	420.28	420.41	0.27	0.00

<u>ตารางที่ 4.2.2(ต่อ)</u> แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %

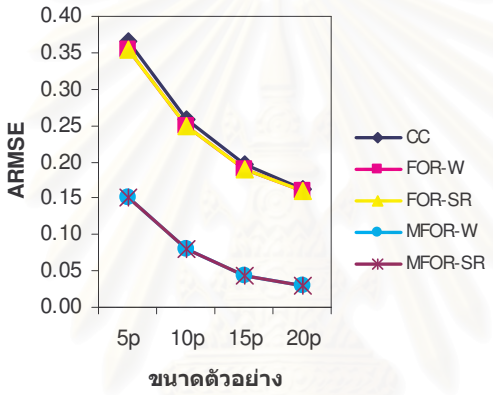
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์แบบเส้นตรง	5p	ARMSE	0.5916	0.5627	0.5691	0.4599	0.4578
		SD	(0.3229)	(0.3095)	(0.314)	(0.5444)	(0.5412)
		DIFF	29.23	22.92	24.32	0.47	0.00
	10p	ARMSE	0.4254	0.4104	0.4113	0.2623	0.2605
		SD	(0.2315)	(0.2321)	(0.2326)	(0.3133)	(0.3097)
		DIFF	63.32	57.57	57.92	0.69	0.00
	15p	ARMSE	0.3107	0.3018	0.3024	0.1280	0.1272
		SD	(0.1536)	(0.1529)	(0.1532)	(0.1426)	(0.1399)
		DIFF	144.18	137.20	137.69	0.61	0.00
	20p	ARMSE	0.2517	0.2412	0.2416	0.0821	0.0821
		SD	(0.1223)	(0.1225)	(0.1226)	(0.0915)	(0.0913)
		DIFF	206.75	193.89	194.36	0.07	0.00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

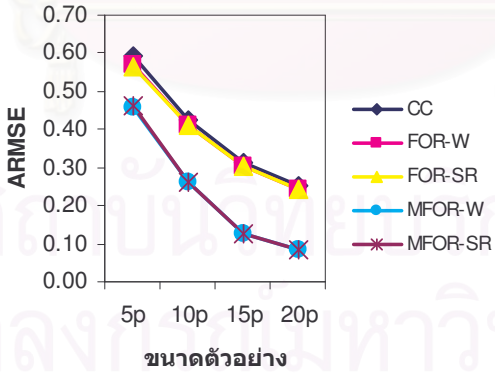
ก) ความสัมพันธ์ระดับต่ำ



ข) ความสัมพันธ์ระดับปานกลาง



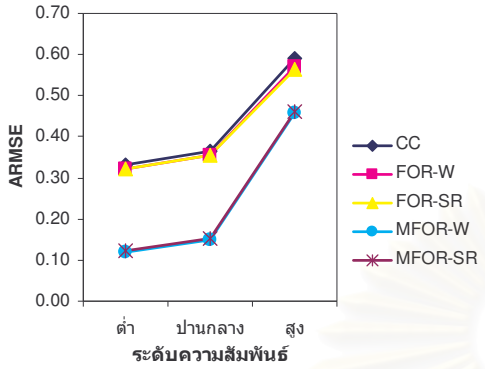
ค) ความสัมพันธ์ระดับสูง



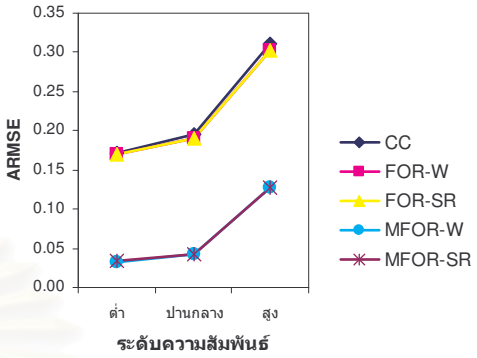
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 4.2.3 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และขนาดตัวอย่าง เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %

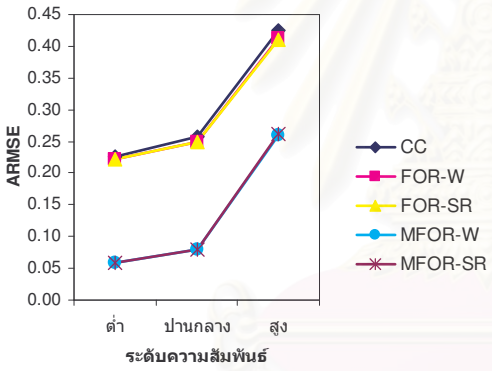
ก) n= 5p



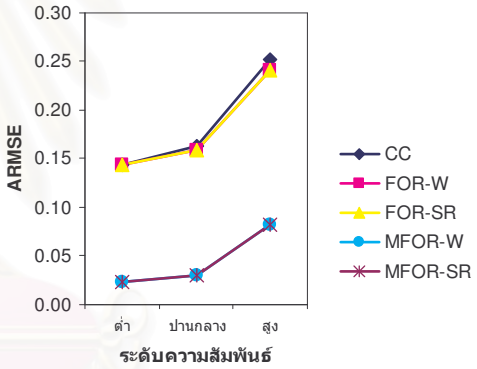
ค) n=15p



ข) n=10p



ง) n=20p



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>รูปที่ 4.2.4</u> แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และระดับความสัมพันธ์ของตัวแปรอิสระ เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %

ผลการวิจัยของการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ กรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 % พบว่า กรณีความสัมพันธ์ระหว่างตัวแปรอิสระระดับต่ำ วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง รองลงมาคือวิธี MFOR-W วิธี FOR-W วิธี FOR-SR และวิธี CC ตามลำดับ ส่วนกรณีความสัมพันธ์ระหว่างตัวแปรอิสระระดับปานกลางสูง วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง วิธีที่รองลงมาคือวิธี MFOR-W วิธี FOR-W วิธี FOR-SR และวิธี CC ตามลำดับ โดยที่แต่ละวิธีเมื่อขนาดตัวอย่างเพิ่มขึ้น จะทำให้ค่า ARMSE มีแนวโน้มลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นทำให้ความแปรปรวนลดลงส่งผลให้ค่าความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยลดลง ส่วนระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้นจะทำให้ค่า ARMSE ของทุกวิธีเพิ่มขึ้น เนื่องจากระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้น ทำให้ค่าลักษณะเฉพาะของเมทริกซ์ $X'X$ มีค่าลดลงจึงส่งผลทำให้ค่า ARMSE มีแนวโน้มเพิ่มขึ้น

เมื่อพิจารณาค่า DIFF ของแต่ละวิธีการที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยในทุกขนาดตัวอย่าง พบว่า ประสิทธิภาพของวิธี MFOR-SR จะโดดเด่นกว่าอีก 3 วิธี คือ วิธี FOR-SR วิธี FOR-W และวิธี CC อย่างชัดเจน และมีประสิทธิภาพใกล้เคียงกับวิธี MFOR-W เพราะเป็นการประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR เช่นเดียวกัน โดยประสิทธิภาพของวิธี MFOR-SR จะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น กล่าวคือวิธี MFOR-SR มีประสิทธิภาพดีกว่าวิธี CC อย่างมากประมาณ 174.34 % 286.87 % 418.72 % และ 505.40 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-W อย่างมากประมาณ 167.95 % 279.33 % 414.17 % และ 504.18 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-SR อย่างมากประมาณ 167.85 % 278.85 % 412.54 % และ 503.35 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น และมีประสิทธิภาพดีกว่าวิธี MFOR-W อย่างมากประมาณ 0.47 % 0.69 % 1.05 % และ 0.45 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น แต่วิธี MFOR-SR จะมีประสิทธิภาพลดลงเมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น

<u>ตารางที่ 4.2.3</u> **แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %**

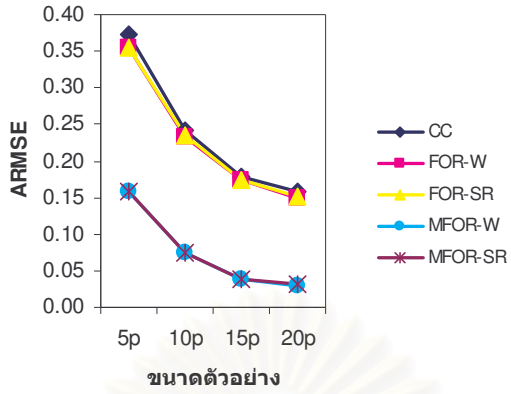
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์ระดับต่ำ	5p	ARMSE	0.3721	0.3553	0.3542	0.1590	0.1574
		SD	(0.1588)	(0.1527)	(0.1524)	(0.147)	(0.1475)
		DIFF	136.45	125.77	125.07	1.00	0.00
	10p	ARMSE	0.2429	0.2341	0.2335	0.0753	0.0739
		SD	(0.0966)	(0.0939)	(0.0938)	(0.0663)	(0.0651)
		DIFF	228.84	216.90	216.14	1.91	0.00
	15p	ARMSE	0.1795	0.1740	0.1738	0.0389	0.0384
		SD	(0.0675)	(0.0655)	(0.0654)	(0.0284)	(0.0279)
		DIFF	366.96	352.52	352.06	1.15	0.00
	20p	ARMSE	0.1572	0.1505	0.1501	0.0306	0.0301
		SD	(0.061)	(0.0587)	(0.0585)	(0.0235)	(0.0232)
		DIFF	422.67	400.20	398.90	1.61	0.00
ความสัมพันธ์ระดับปานกลาง	5p	ARMSE	0.4202	0.3946	0.3950	0.2160	0.2129
		SD	(0.1961)	(0.1823)	(0.1833)	(0.2424)	(0.2416)
		DIFF	97.36	85.33	85.53	1.46	0.00
	10p	ARMSE	0.2793	0.2590	0.2592	0.1028	0.1004
		SD	(0.1232)	(0.1214)	(0.121)	(0.1064)	(0.1032)
		DIFF	178.07	157.93	158.10	2.37	0.00
	15p	ARMSE	0.2097	0.1988	0.1989	0.0528	0.0524
		SD	(0.0849)	(0.0821)	(0.0822)	(0.0469)	(0.0466)
		DIFF	300.19	279.48	27957	0.78	0.00
	20p	ARMSE	0.1748	0.1683	0.1684	0.0390	0.0386
		SD	(0.0738)	(0.0723)	(0.0723)	(0.0341)	(0.0338)
		DIFF	353.27	336.46	336.49	1.14	0.00

<u>ตารางที่ 4.2.3(ต่อ)</u> **แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %**

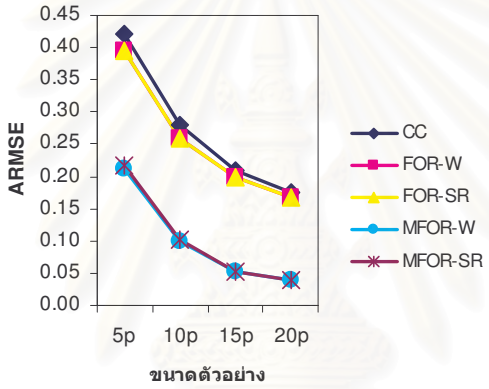
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์เชิงเส้น	5p	ARMSE	0.6947	0.6704	0.6797	0.6178	0.6123
		SD	(0.8833)	(0.387)	(0.8726)	(0.3663)	(0.3637)
		DIFF	13.46	9.50	11.01	0.90	0.00
	10p	ARMSE	0.4566	0.4373	0.4378	0.3424	0.3341
		SD	(0.2469)	(0.244)	(0.2448)	(0.4032)	(0.3931)
		DIFF	36.65	30.88	31.04	2.49	0.00
	15p	ARMSE	0.3260	0.3091	0.3097	0.1527	0.1518
		SD	(0.161)	(0.1613)	(0.1615)	(0.1696)	(0.1684)
		DIFF	114.78	103.62	104.01	0.59	0.00
	20p	ARMSE	0.2749	0.2627	0.2629	0.1098	0.1090
		SD	(0.1365)	(0.1357)	(0.1359)	(0.1293)	(0.1284)
		DIFF	152.20	141.00	141.25	0.73	0.00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

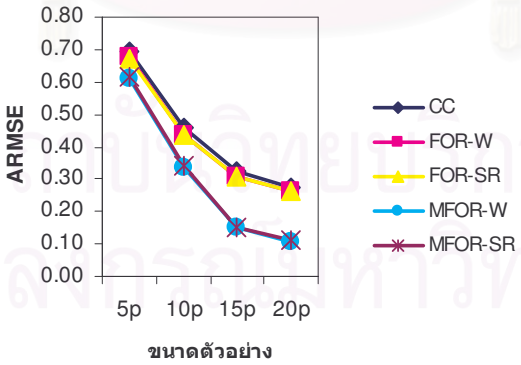
ก) ความสัมพันธ์ระดับต่ำ



ข) ความสัมพันธ์ระดับปานกลาง



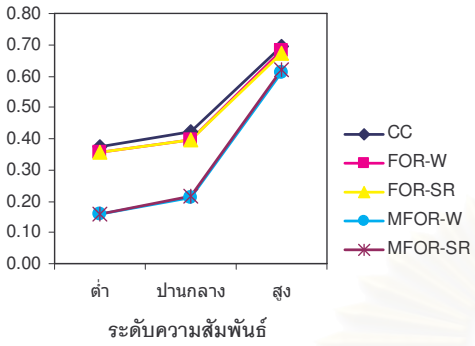
ค) ความสัมพันธ์ระดับสูง



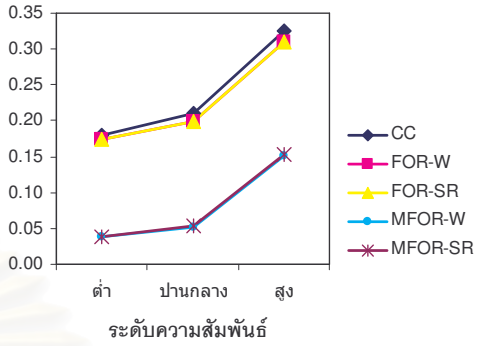
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 4.2.5 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และขนาดตัวอย่าง เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %

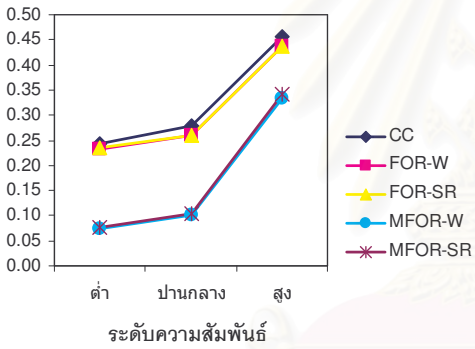
ก) n= 5p



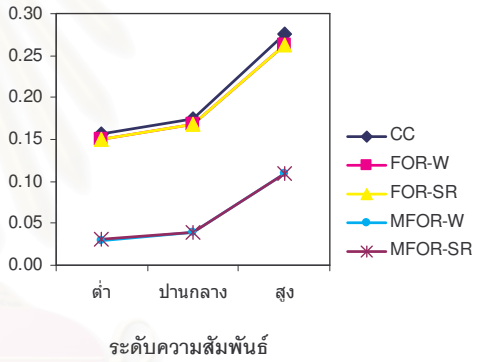
ค) n=15p



ข) n=10p



ง) n=20p



รูปที่ 4.2.6 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และระดับความสัมพันธ์ของตัวแปรอิสระ เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %

ผลการวิจัยของการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ กรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 1 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 % พบว่า กรณีความสัมพันธ์ระหว่างตัวแปรอิสระระดับต่ำ วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง รองลงมาคือวิธี MFOR-W วิธี FOR-W วิธี FOR-SR และวิธี CC ตามลำดับ ส่วนกรณีความสัมพันธ์ระหว่างตัวแปรอิสระระดับปานกลางสูง วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง วิธีที่รองลงมาคือวิธี MFOR-W วิธี FOR-W วิธี FOR-SR และวิธี CC ตามลำดับ โดยที่แต่ละวิธีเมื่อขนาดตัวอย่างเพิ่มขึ้น จะทำให้ค่า ARMSE มีแนวโน้มลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นทำให้ความแปรปรวนลดลงส่งผลให้ค่าความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยลดลง ส่วนระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้นจะทำให้ค่า ARMSE ของทุกวิธีเพิ่มขึ้น เนื่องจากระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้น ทำให้ค่าลักษณะเฉพาะของเมทริกซ์ $X'X$ มีค่าลดลงจึงส่งผลทำให้ค่า ARMSE มีแนวโน้มเพิ่มขึ้น

เมื่อพิจารณาค่า DIFF ของแต่ละวิธีการที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยในทุกขนาดตัวอย่าง พบว่า ประสิทธิภาพของวิธี MFOR-SR จะโดดเด่นกว่าอีก 3 วิธี คือ วิธี FOR-SR วิธี FOR-W และวิธี CC อย่างชัดเจน และมีประสิทธิภาพใกล้เคียงกับวิธี MFOR-W เพราะเป็นการประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR เช่นเดียวกัน โดยประสิทธิภาพของวิธี MFOR-SR จะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น กล่าวคือวิธี MFOR-SR มีประสิทธิภาพดีกว่าวิธี CC อย่างมากประมาณ 136.45 % 228.84 % 366.96 % และ 422.67 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-W อย่างมากประมาณ 125.77 % 216.90 % 352.52 % และ 400.20 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-SR อย่างมากประมาณ 125.07 % 216.14 % 352.06 % และ 398.90 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น และมีประสิทธิภาพดีกว่าวิธี MFOR-W อย่างมากประมาณ 1.46 % 2.49 % 1.15 % และ 1.61 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น แต่วิธี MFOR-SR จะมีประสิทธิภาพลดลงเมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น

สรุปส่วนที่ 4.2 ผลการวิจัยของการเปรียบเทียบการประมาณสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ กรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 1

กรณีที่สัดส่วนข้อมูลสุญหายในตัวแปรอิสระเท่ากับ 5 % พบว่า ทุกระดับความสัมพันธ์ระหว่างตัวแปรอิสระ วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง รองลงมาคือ วิธี MFOR-W วิธี FOR-W วิธี FOR-SR และวิธี CC ตามลำดับ ส่วนกรณีที่สัดส่วนข้อมูลสุญหายในตัวแปรอิสระเท่ากับ 15 % และ 25 % พบว่า เมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระต่ำ วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง รองลงมาคือวิธี MFOR-W วิธี FOR-SR วิธี FOR-W และวิธี CC ตามลำดับ และกรณีระดับความสัมพันธ์ระหว่างตัวแปรอิสระปานกลางและสูง วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง รองลงมาคือวิธี MFOR-W วิธี FOR-W วิธี FOR-SR และวิธี CC ตามลำดับ เพราะวิธีประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีของ SRM เป็นวิธีที่ลดความคลาดเคลื่อนของการประมาณค่าที่เกิดจากความสัมพันธ์กันระหว่างตัวแปรอิสระ แต่วิธีของ WM นั่นเป็นการถ่วงน้ำหนักชุดของข้อมูลที่มีการสูญหายภายหลังจากที่ทำการประมาณค่าแล้ว ส่วนวิธีประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR จะให้ค่าที่ใกล้เคียงค่าจริงมากกว่าวิธี FOR เนื่องจากมีการเพิ่มตัวแปรตามในการประมาณค่า โดยที่ข้อมูลของตัวแปรตามมีความแปรปรวนน้อย และการที่สัดส่วนข้อมูลสูญหายเพิ่มขึ้นจะส่งผลให้การตัดชุดข้อมูลที่มีการสูญหายนั่นจะทำให้เกิดความคลาดเคลื่อนในการประมาณค่าไปจากความเป็นจริงเพิ่มขึ้น เนื่องจากขนาดของข้อมูลที่ลดลง

จากผลการวิจัยในส่วนที่ 4.2 สามารถสรุปเกี่ยวกับปัจจัยที่มีผลต่อค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) ได้ดังนี้

1. **เมื่อขนาดตัวอย่างเพิ่มขึ้น** วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระทั้ง 5 วิธี จะให้ค่า ARMSE ลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นจะช่วยลดความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยลดลง

2. **เมื่อสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเพิ่มขึ้น** วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระทั้ง 5 วิธี จะทำให้ค่า ARMSE เพิ่มขึ้น เนื่องจากในวิธี CC เป็นการประมาณค่าสัมประสิทธิ์การถดถอยโดยการตัดชุดข้อมูลสูญหาย จึงทำให้ขนาดตัวอย่างลดลง มีผลให้ในการประมาณค่าสัมประสิทธิ์การถดถอยมีความคลาดเคลื่อนไปจากความเป็นจริง ส่วนค่า ARMSE ของวิธี MFOR-SR วิธี MFOR-W

วิธี FOR-W และวิธี FOR-SR ที่เพิ่มขึ้น เนื่องจากค่าความคลาดเคลื่อนในการประมาณค่าข้อมูลสูญหาย จึงส่งผลให้ค่าความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยเพิ่มขึ้น

3. **เมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น** จะส่งผลให้ค่า ARMSE มีแนวโน้มสูงขึ้น เนื่องจากระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้น ทำให้ค่าลักษณะเฉพาะของเมทริกซ์ $X'X$ มีค่าลดลงจึงส่งผลทำให้ค่า ARMSE มีแนวโน้มเพิ่มขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4.3 ผลการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ ในกรณีส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 3

การวิจัยในส่วนนี้ผู้วิจัยศึกษา กรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 3 ที่ความสัมพันธ์ระหว่างตัวแปรอิสระอยู่ในระดับต่ำ ปานกลาง และสูง และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 % 15 % และ 25 % ตามลำดับ โดยมีขนาดตัวอย่างเท่ากับ 5p 10p 15p และ 20p ตามลำดับ เมื่อ p คือ จำนวนตัวแปรอิสระซึ่งเท่ากับ 3 ซึ่งผลการวิจัยส่วนนี้นำเสนอในตารางที่ 4.3.1 – 4.3.3

รายละเอียดตารางที่ 4.3.1 – 4.3.3

ตารางที่	ส่วนเบี่ยงเบนมาตรฐาน	สัดส่วนข้อมูลสูญหายในตัวแปรอิสระ
4.3.1	3	5 %
4.3.2	3	15 %
4.3.3	3	25 %

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>ตารางที่ 4.3.1</u> **แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %**

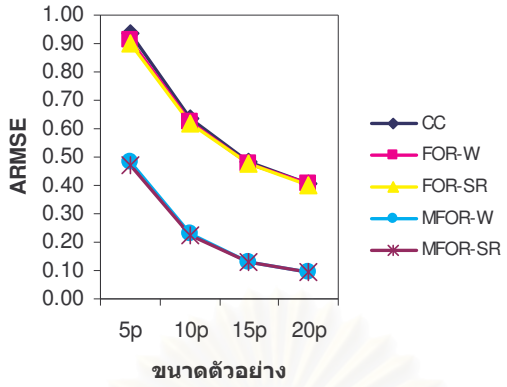
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์ระดับต่ำ	5p	ARMSE	0.9379	0.9139	0.9026	0.4851	0.4729
		SD	(0.382)	(0.3603)	(0.3421)	(0.1939)	(0.1883)
		DIFF	98.31	93.25	90.86	2.57	0.00
	10p	ARMSE	0.6361	0.6253	0.6206	0.2290	0.2258
		SD	(0.2486)	(0.2464)	(0.2378)	(0.0909)	(0.0895)
		DIFF	181.74	176.98	174.88	1.44	0.00
	15p	ARMSE	0.4807	0.4782	0.4760	0.1289	0.1277
		SD	(0.1739)	(0.17)	(0.1657)	(0.0448)	(0.0444)
		DIFF	276.50	274.51	272.81	0.95	0.00
	20p	ARMSE	0.4067	0.4035	0.4020	0.0927	0.0920
		SD	(0.1535)	(0.1514)	(0.1483)	(0.0348)	(0.0344)
		DIFF	342.11	338.63	337.09	0.80	0.00
ความสัมพันธ์ระดับปานกลาง	5p	ARMSE	1.0455	1.0251	1.0115	0.6319	0.6160
		SD	(0.442)	(0.4297)	(0.4127)	(0.2963)	(0.2887)
		DIFF	69.72	66.40	64.19	2.57	0.00
	10p	ARMSE	0.7313	0.7152	0.7104	0.3115	0.3075
		SD	(0.3164)	(0.309)	(0.2999)	(0.1388)	(0.1368)
		DIFF	137.83	132.59	131.01	1.31	0.00
	15p	ARMSE	0.5415	0.5392	0.5367	0.1680	0.1664
		SD	(0.2128)	(0.2097)	(0.2042)	(0.0651)	(0.0643)
		DIFF	225.42	224.08	222.56	0.99	0.00
	20p	ARMSE	0.4625	0.4551	0.4534	0.1214	0.1205
		SD	(0.1843)	(0.1805)	(0.1768)	(0.0507)	(0.0502)
		DIFF	283.92	277.78	276.37	0.78	0.00

<u>ตารางที่ 4.3.1(ต่อ)</u> **แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %**

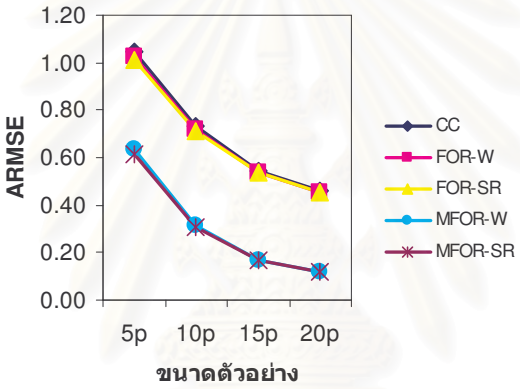
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์แบบเส้นตรง	5p	ARMSE	1.7810	1.7665	1.7608	1.5932	1.5492
		SD	(0.9056)	(0.8977)	(0.8651)	(1.0445)	(1.0173)
		DIFF	14.96	14.03	13.66	2.84	0.00
	10p	ARMSE	1.1921	1.1783	1.1708	0.9431	0.9316
		SD	(0.6409)	(0.6374)	(0.6217)	(0.523)	(0.5168)
		DIFF	27.96	26.48	25.68	1.23	0.00
	15p	ARMSE	0.8740	0.8688	0.8644	0.4856	0.4806
		SD	(0.4266)	(0.4247)	(0.4148)	(0.2507)	(0.248)
		DIFF	81.84	80.76	79.86	1.03	0.00
	20p	ARMSE	0.7267	0.7195	0.7167	0.3368	0.3340
		SD	(0.362)	(0.3602)	(0.3539)	(0.1738)	(0.1723)
		DIFF	117.56	115.39	114.54	0.81	0.00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

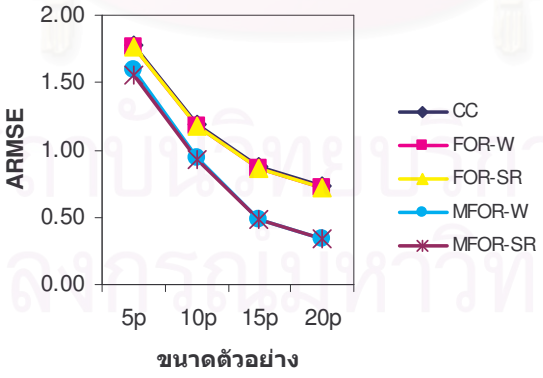
ก) ความสัมพันธ์ระดับต่ำ



ข) ความสัมพันธ์ระดับปานกลาง



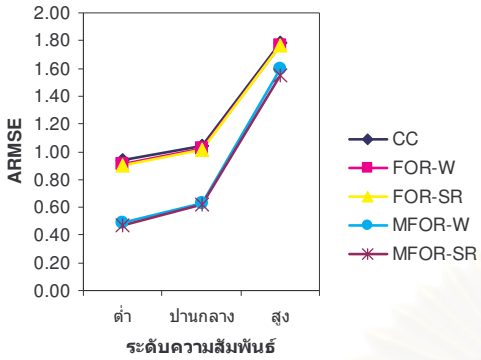
ค) ความสัมพันธ์ระดับสูง



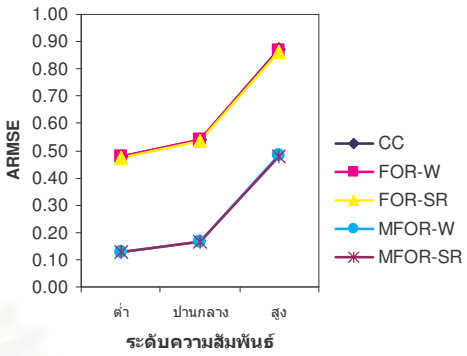
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>รูปที่ 4.3.1</u> แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และขนาดตัวอย่าง เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %

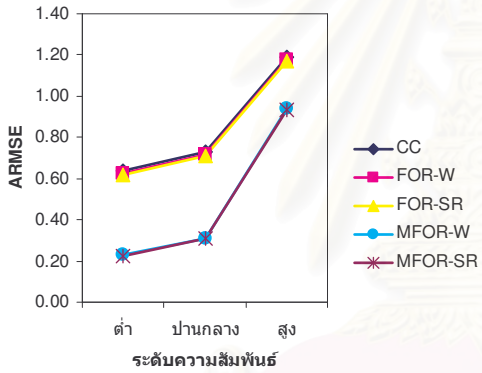
ก) n= 5p



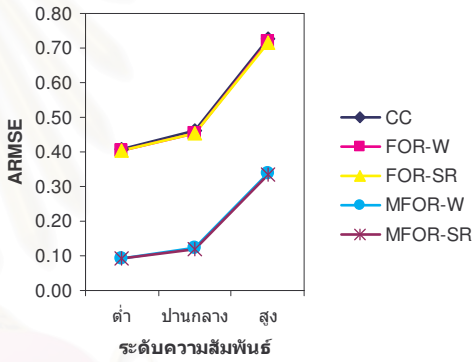
ค) n=15p



ข) n=10p



ง) n=20p



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 4.3.2 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และระดับความสัมพันธ์ของตัวแปรอิสระ เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 %

ผลการวิจัยของการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ กรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 5 % พบว่า วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกๆกรณี รองลงมาคือ วิธี MFOR-W วิธี FOR-SR วิธี FOR-W และวิธี CC โดยที่แต่ละวิธีเมื่อขนาดตัวอย่างเพิ่มขึ้น จะทำให้ค่า ARMSE มีแนวโน้มลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นทำให้ความแปรปรวนลดลงส่งผลให้ค่าความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยลดลง ส่วนระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้นจะทำให้ค่า ARMSE ของทุกวิธีเพิ่มขึ้น เนื่องจากระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้นทำให้ค่าลักษณะเฉพาะของเมทริกซ์ $X'X$ มีค่าลดลงจึงส่งผลทำให้ค่า ARMSE มีแนวโน้มเพิ่มขึ้น

เมื่อพิจารณาค่า DIFF ของแต่ละวิธีการที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยในทุกขนาดตัวอย่าง พบว่า ประสิทธิภาพของวิธี MFOR-SR จะโดดเด่นกว่าอีก 3 วิธี คือ วิธี FOR-SR วิธี FOR-W และวิธี CC อย่างชัดเจน และมีประสิทธิภาพใกล้เคียงกับวิธี MFOR-W เพราะเป็นการประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR เช่นเดียวกัน โดยประสิทธิภาพของวิธี MFOR-SR จะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น กล่าวคือวิธี MFOR-SR มีประสิทธิภาพดีกว่าวิธี CC อย่างมากประมาณ 98.31 % 181.74 % 276.50 % และ 342.11 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-W อย่างมากประมาณ 93.25 % 176.98 % 274.51 % และ 338.63 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-SR อย่างมากประมาณ 90.86 % 174.88 % 272.81 % และ 337.09 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น และมีประสิทธิภาพดีกว่าวิธี MFOR-W อย่างมากประมาณ 2.57 % 1.44 % 0.95 % และ 0.80 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น แต่วิธี MFOR-SR จะมีประสิทธิภาพลดลงเมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.3.2 แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %

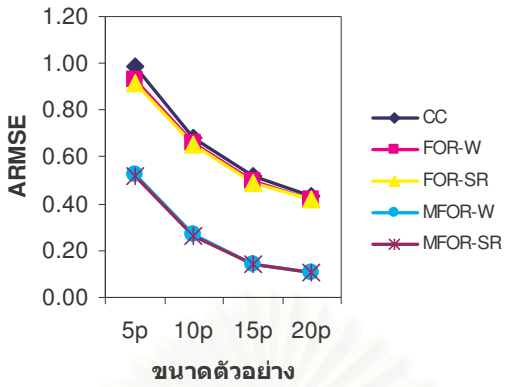
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์ระดับต่ำ	5p	ARMSE	0.9862	0.9296	0.9176	0.5271	0.5154
		SD	(0.4156)	(0.3856)	(0.3694)	(0.2219)	(0.2168)
		DIFF	91.35	80.37	78.03	2.26	0.00
	10p	ARMSE	0.6799	0.6572	0.6524	0.2690	0.2660
		SD	(0.2717)	(0.2551)	(0.2472)	(0.1095)	(0.1084)
		DIFF	155.62	147.09	145.28	1.14	0.00
	15p	ARMSE	0.5204	0.4938	0.4917	0.1441	0.1430
		SD	(0.1934)	(0.1825)	(0.1779)	(0.0506)	(0.0502)
		DIFF	264.06	245.45	243.98	0.81	0.00
	20p	ARMSE	0.4323	0.4186	0.4173	0.1041	0.1033
		SD	(0.1617)	(0.1532)	(0.1499)	(0.038)	(0.0377)
		DIFF	318.42	305.14	303.83	0.72	0.00
ความสัมพันธ์ระดับปานกลาง	5p	ARMSE	1.0925	1.0466	1.0339	0.6841	0.6705
		SD	(0.4858)	(0.4598)	(0.4396)	(0.3083)	(0.3021)
		DIFF	62.93	56.09	54.19	2.02	0.00
	10p	ARMSE	0.7758	0.7493	0.7445	0.3761	0.3726
		SD	(0.3348)	(0.3195)	(0.3116)	(0.182)	(0.1804)
		DIFF	108.20	101.08	99.81	0.94	0.00
	15p	ARMSE	0.5852	0.5596	0.5572	0.1935	0.1922
		SD	(0.2346)	(0.2266)	(0.221)	(0.0764)	(0.0759)
		DIFF	204.47	191.17	189.91	0.70	0.00
	20p	ARMSE	0.4887	0.4698	0.4683	0.1383	0.1375
		SD	(0.1957)	(0.1891)	(0.1855)	(0.057)	(0.0566)
		DIFF	255.32	241.58	240.47	0.59	0.00

<u>ตารางที่ 4.3.2(ต่อ)</u> **แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %**

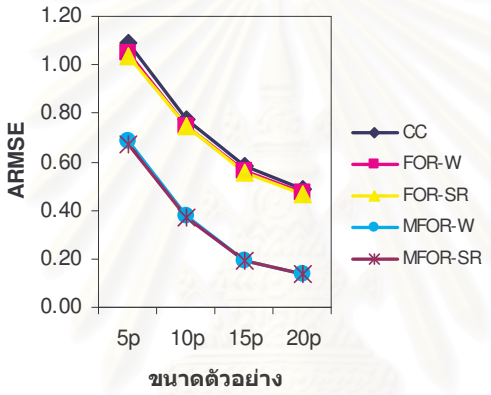
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์แบบเส้นตรง	5p	ARMSE	1.8077	1.7702	1.7562	1.6238	1.6002
		SD	(0.9758)	(0.9556)	(0.9178)	(1.3299)	(1.3021)
		DIFF	12.97	10.62	9.75	1.47	0.00
	10p	ARMSE	1.2603	1.2374	1.2301	1.0011	0.9901
		SD	(0.6796)	(0.6753)	(0.6604)	(0.7131)	(0.7076)
		DIFF	27.30	24.98	24.25	1.11	0.00
	15p	ARMSE	0.9307	0.9087	0.9047	0.6132	0.6087
		SD	(0.4458)	(0.4489)	(0.4409)	(0.3327)	(0.3305)
		DIFF	52.89	49.29	48.63	0.74	0.00
	20p	ARMSE	0.7660	0.7427	0.7401	0.4090	0.4064
		SD	(0.3833)	(0.3839)	(0.3777)	(0.2272)	(0.2257)
		DIFF	88.50	82.77	82.13	0.64	0.00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

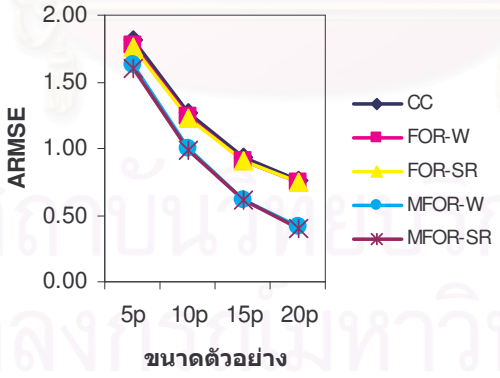
ก) ความสัมพันธ์ระดับต่ำ



ข) ความสัมพันธ์ระดับปานกลาง



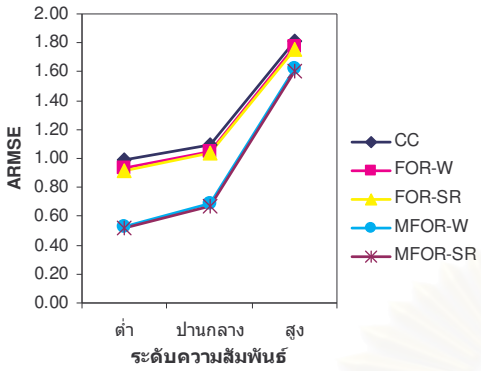
ค) ความสัมพันธ์ระดับสูง



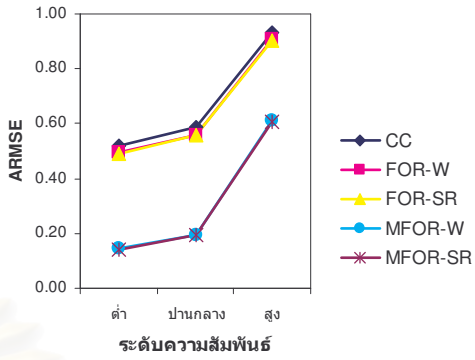
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>รูปที่ 4.3.3</u> แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และขนาดตัวอย่าง เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %

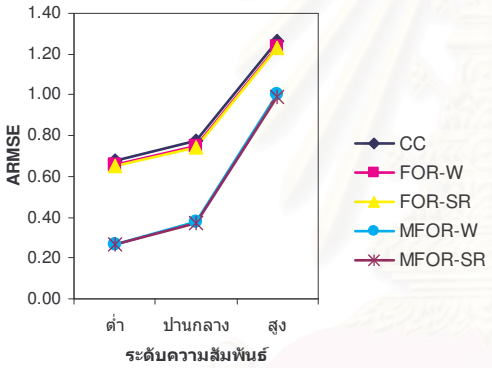
ก) n= 5p



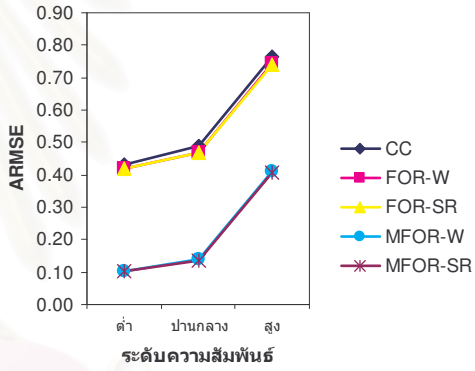
ค) n=15p



ข) n=10p



ง) n=20p



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>รูปที่ 4.3.4</u> แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และระดับความสัมพันธ์ของตัวแปรอิสระ เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 %

ผลการวิจัยของการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ กรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 15 % พบว่า วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกๆกรณี รองลงมาคือวิธี MFOR-W วิธี FOR-SR วิธี FOR-W และวิธี CC โดยที่แต่ละวิธีเมื่อขนาดตัวอย่างเพิ่มขึ้น จะทำให้ค่า ARMSE มีแนวโน้มลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นทำให้ความแปรปรวนลดลงส่งผลให้ค่าความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยลดลง ส่วนระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้นจะทำให้ค่า ARMSE ของทุกวิธีเพิ่มขึ้น เนื่องจากระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้น ทำให้ค่าลักษณะเฉพาะของเมทริกซ์ $X'X$ มีค่าลดลงจึงส่งผลทำให้ค่า ARMSE มีแนวโน้มเพิ่มขึ้น

เมื่อพิจารณาค่า DIFF ของแต่ละวิธีการที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยในทุกขนาดตัวอย่าง พบว่า ประสิทธิภาพของวิธี MFOR-SR จะโดดเด่นกว่าอีก 3 วิธี คือ วิธี FOR-SR วิธี FOR-W และวิธี CC อย่างชัดเจน และมีประสิทธิภาพใกล้เคียงกับวิธี MFOR-W เพราะเป็นการประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR เช่นเดียวกัน โดยประสิทธิภาพของวิธี MFOR-SR จะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น กล่าวคือวิธี MFOR-SR มีประสิทธิภาพดีกว่าวิธี CC อย่างมากประมาณ 91.35 % 155.62 % 264.06 % และ 318.42 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-W อย่างมากประมาณ 80.37 % 147.09 % 245.45 % และ 305.14 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-SR อย่างมากประมาณ 78.03 % 145.28 % 243.98 % และ 303.83 % ตามลำดับ ทุกขนาดตัวอย่างที่เพิ่มขึ้น และมีประสิทธิภาพดีกว่าวิธี MFOR-W อย่างมากประมาณ 2.26 % 1.14 % 0.81 % และ 0.72 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น แต่วิธี MFOR-SR จะมีประสิทธิภาพลดลงเมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>ตารางที่ 4.3.3</u> **แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %**

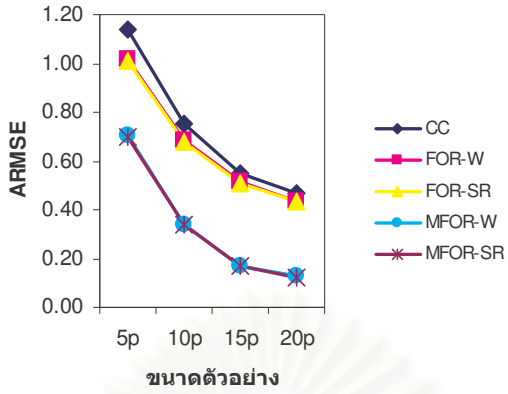
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์ระดับต่ำ	5p	ARMSE	1.1388	1.0202	1.0081	0.7055	0.6967
		SD	(0.4639)	(0.3997)	(0.3851)	(0.3781)	(0.3713)
		DIFF	63.47	46.44	44.70	1.27	0.00
	10p	ARMSE	0.7549	0.6854	0.6809	0.3374	0.3360
		SD	(0.3001)	(0.2727)	(0.266)	(0.1495)	(0.1488)
		DIFF	124.66	103.98	102.64	0.43	0.00
	15p	ARMSE	0.5467	0.5130	0.5111	0.1688	0.1677
		SD	(0.2021)	(0.1849)	(0.1807)	(0.0619)	(0.0616)
		DIFF	225.99	205.88	204.75	0.63	0.00
	20p	ARMSE	0.4666	0.4325	0.4313	0.1259	0.1254
		SD	(0.1839)	(0.1686)	(0.1651)	(0.0494)	(0.0492)
		DIFF	272.06	244.90	243.94	0.38	0.00
ความสัมพันธ์ระดับปานกลาง	5p	ARMSE	1.2695	1.1490	1.1349	0.9691	0.9600
		SD	(0.5843)	(0.5157)	(0.4989)	(0.5319)	(0.5256)
		DIFF	32.25	19.70	18.22	0.95	0.00
	10p	ARMSE	0.8366	0.7719	0.7680	0.4575	0.4561
		SD	(0.353)	(0.3401)	(0.3324)	(0.2263)	(0.2273)
		DIFF	83.43	69.24	68.39	0.32	0.00
	15p	ARMSE	0.6274	0.5872	0.5848	0.2383	0.2372
		SD	(0.255)	(0.2409)	(0.236)	(0.1037)	(0.1032)
		DIFF	164.52	147.58	146.57	0.46	0.00
	20p	ARMSE	0.5183	0.4822	0.4808	0.1643	0.1640
		SD	(0.2147)	(0.2066)	(0.2035)	(0.0674)	(0.0672)
		DIFF	216.11	194.08	193.24	0.20	0.00

<u>ตารางที่ 4.3.3(ต่อ)</u> **แสดงค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เมื่อส่วนเบี่ยงเบนมาตรฐานของคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %**

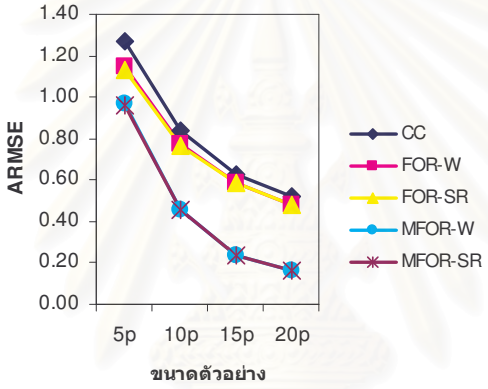
	ขนาดตัวอย่าง		CC	FOR-W	FOR-SR	MFOR-W	MFOR-SR
ความสัมพันธ์เชิงเส้นตรง	5p	ARMSE	1.9948	1.9716	1.9585	1.8284	1.8031
		SD	(1.1228)	(1.0727)	(1.0356)	(2.1276)	(2.1459)
		DIFF	10.63	9.35	8.62	1.40	0.00
	10p	ARMSE	1.3842	1.3226	1.3160	1.2062	1.1902
		SD	(0.7362)	(0.7306)	(0.7156)	(1.0386)	(1.0382)
		DIFF	16.31	11.13	10.57	1.35	0.00
	15p	ARMSE	1.0064	0.9588	0.9550	0.7782	0.7751
		SD	(0.5067)	(0.5042)	(0.4955)	(0.4573)	(0.4558)
		DIFF	29.85	23.71	23.22	0.41	0.00
	20p	ARMSE	0.8146	0.7662	0.7639	0.5033	0.5021
		SD	(0.4029)	(0.4008)	(0.3953)	(0.2805)	(0.2799)
		DIFF	62.24	52.60	52.14	0.23	0.00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

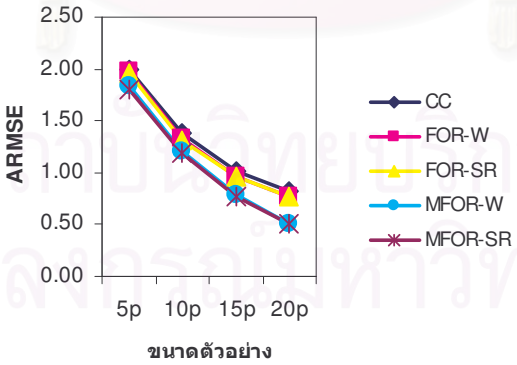
ก) ความสัมพันธ์ระดับต่ำ



ข) ความสัมพันธ์ระดับปานกลาง



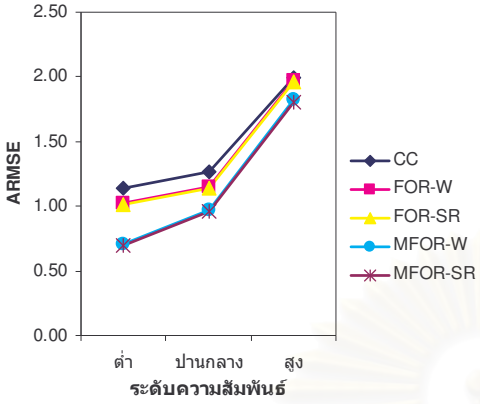
ค) ความสัมพันธ์ระดับสูง



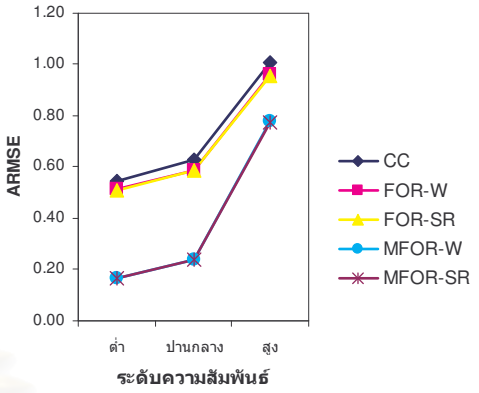
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 4.3.5 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และขนาดตัวอย่าง เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %

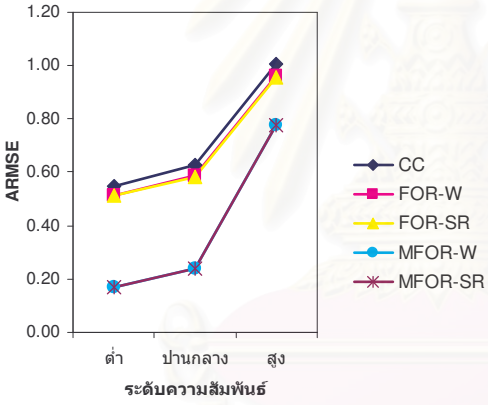
ก) n= 5p



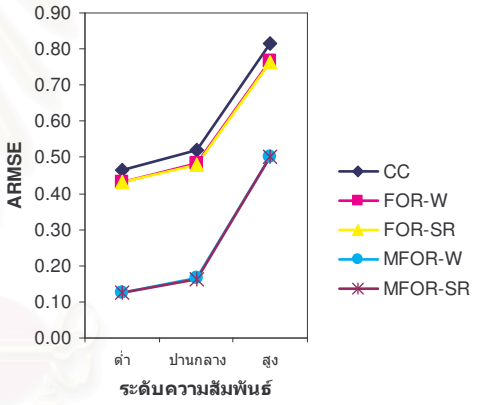
ค) n=15p



ข) n=10p



ง) n=20p



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 4.3.6 แสดงการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระระหว่างค่า ARMSE และระดับความสัมพันธ์ของตัวแปรอิสระ เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 %

ผลการวิจัยของการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ กรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 3 และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเท่ากับ 25 % พบว่า วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกๆกรณี รองลงมาคือ วิธี MFOR-W วิธี FOR-SR วิธี FOR-W และวิธี CC โดยที่แต่ละวิธีเมื่อขนาดตัวอย่างเพิ่มขึ้น จะทำให้ค่า ARMSE มีแนวโน้มลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นทำให้ความแปรปรวนลดลงส่งผลให้ค่าความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยลดลง ส่วนระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้นจะทำให้ค่า ARMSE ของทุกวิธีเพิ่มขึ้น เนื่องจากระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้น ทำให้ค่าลักษณะเฉพาะของเมทริกซ์ $X'X$ มีค่าลดลงจึงส่งผลทำให้ค่า ARMSE มีแนวโน้มเพิ่มขึ้น

เมื่อพิจารณาค่า DIFF ของแต่ละวิธีการที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยในทุกขนาดตัวอย่าง พบว่า ประสิทธิภาพของวิธี MFOR-SR จะโดดเด่นกว่าอีก 3 วิธี คือ วิธี FOR-SR วิธี FOR-W และวิธี CC อย่างชัดเจน และมีประสิทธิภาพใกล้เคียงกับวิธี MFOR-W เพราะเป็นการประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR เช่นเดียวกัน โดยประสิทธิภาพของวิธี MFOR-SR จะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น กล่าวคือวิธี MFOR-SR มีประสิทธิภาพดีกว่าวิธี CC อย่างมากประมาณ 63.47 % 124.66 % 225.99 % และ 272.06 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-W อย่างมากประมาณ 46.44 % 103.98 % 205.88 % และ 244.90 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น มีประสิทธิภาพดีกว่าวิธี FOR-SR อย่างมากประมาณ 44.70 % 102.64 % 204.75 % และ 243.94 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น และมีประสิทธิภาพดีกว่าวิธี MFOR-W อย่างมากประมาณ 1.27 % 0.43 % 0.63 % และ 0.38 % ตามลำดับทุกขนาดตัวอย่างที่เพิ่มขึ้น แต่วิธี MFOR-SR จะมีประสิทธิภาพลดลงเมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สรุปส่วนที่ 4.3 ผลการวิจัยของการเปรียบเทียบการประมาณสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ กรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเท่ากับ 3

ทุกระดับสัดส่วนข้อมูลสุญหายในตัวแปรอิสระและระดับความสัมพันธ์ระหว่างตัวแปรอิสระ พบว่า วิธีที่ดีที่สุดได้แก่ วิธี MFOR-SR ทุกขนาดตัวอย่าง รองลงมาคือ วิธี MFOR-W วิธี FOR-SR วิธี FOR-W และวิธี CC ตามลำดับ เพราะวิธีประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีของ SRM เป็นวิธีที่ลดความคลาดเคลื่อนของการประมาณค่าที่เกิดจากความสัมพันธ์กันระหว่างตัวแปรอิสระ แต่วิธีของ WM นั่นเป็นการถ่วงน้ำหนักชุดของข้อมูลที่มีการสูญหายภายหลังจากที่ทำการประมาณค่าแล้ว ส่วนวิธีประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR จะให้ค่าที่ใกล้เคียงค่าจริงมากกว่าวิธี FOR เนื่องจากมีการเพิ่มตัวแปรตามในการประมาณค่า โดยที่ข้อมูลของตัวแปรตามมีความแปรปรวนน้อย และการที่สัดส่วนข้อมูลสูญหายเพิ่มขึ้นจะส่งผลให้การตัดชุดข้อมูลที่มีการสูญหายนั่นจะทำให้เกิดความคลาดเคลื่อนในการประมาณค่าไปจากความเป็นจริงเพิ่มขึ้นเนื่องจากขนาดของข้อมูลที่ลดลง

จากผลการวิจัยในส่วนที่ 4.3 สามารถสรุปเกี่ยวกับปัจจัยที่มีผลต่อค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) ได้ดังนี้

1. **เมื่อขนาดตัวอย่างเพิ่มขึ้น** วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระทั้ง 5 วิธี จะให้ค่า ARMSE ลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นจะช่วยลดความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยลดลง

2. **เมื่อสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเพิ่มขึ้น** วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระทั้ง 5 วิธี จะทำให้ค่า ARMSE เพิ่มขึ้น เนื่องจากในวิธี CC เป็นการประมาณค่าสัมประสิทธิ์การถดถอยโดยการตัดชุดข้อมูลสูญหาย จึงทำให้ขนาดตัวอย่างลดลง มีผลให้ในการประมาณค่าสัมประสิทธิ์การถดถอยมีความคลาดเคลื่อนไปจากความเป็นจริง ส่วนค่า ARMSE ของวิธี MFOR-SR วิธี MFOR-W วิธี FOR-W และวิธี FOR-SR ที่เพิ่มขึ้น เนื่องจากค่าความคลาดเคลื่อนในการประมาณค่าข้อมูลสูญหายจึงส่งผลให้ค่าความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยเพิ่มขึ้น

3. **เมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น** จะส่งผลให้ค่า ARMSE มีแนวโน้มสูงขึ้น เนื่องจากระดับความสัมพันธ์ระหว่างตัวแปรอิสระที่เพิ่มขึ้น ทำให้ค่าลักษณะเฉพาะของเมทริกซ์ $X'X$ มีค่าลดลงจึงส่งผลทำให้ค่า ARMSE มีแนวโน้มเพิ่มขึ้น

จากผลการวิจัยในส่วนที่ 4.1-4.3 **สามารถสรุปเกี่ยวกับปัจจัยที่มีผลต่อค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง** (ARMSE) ได้ดังนี้

1. **เมื่อขนาดตัวอย่างเพิ่มขึ้น** วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระทั้ง 5 วิธี จะให้ค่า ARMSE ลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นจะช่วยลดความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยลง

2. **เมื่อสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเพิ่มขึ้น** ค่า ARMSE มีแนวโน้มเพิ่มขึ้นเมื่อสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเพิ่มขึ้น เนื่องจากค่าความคลาดเคลื่อนที่เกิดจากการประมาณค่าข้อมูลสูญหาย และในกรณีที่ตัดชุดข้อมูลสูญหาย ทำให้ขนาดตัวอย่างลดลงมีผลให้ในการประมาณค่าสัมประสิทธิ์การถดถอยมีความคลาดเคลื่อนไปจากความเป็นจริง

3. **เมื่อระดับสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น** ค่า ARMSE มีแนวโน้มเพิ่มขึ้นเมื่อระดับความสัมพันธ์เพิ่มขึ้น เพราะเมื่อระดับความสัมพันธ์เพิ่มขึ้นทำให้ค่าลักษณะเฉพาะของเมทริกซ์ $X'X$ มีค่าลดลง จึงทำให้ค่า ARMSE มีค่าเพิ่มขึ้นและมีผลทำให้ความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณลดลง

4. **เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเพิ่มขึ้น** ค่า ARMSE มีแนวโน้มเพิ่มขึ้นเมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเพิ่มขึ้น เพราะเมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเพิ่มขึ้น หมายความว่าความคลาดเคลื่อนมีการกระจายมากขึ้น จึงทำให้ค่า ARMSE มีค่าเพิ่มขึ้นและมีผลทำให้ความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณลดลง

ผู้วิจัยสามารถสรุปผลเกี่ยวกับการเปลี่ยนแปลงของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (ARMSE) ได้ดังนี้

1. **ขนาดตัวอย่าง** ค่า ARMSE จะ**แปรผกผัน**กับขนาดตัวอย่าง คือขนาดตัวอย่างเพิ่มขึ้นจะทำให้ค่า ARMSE ลดลง

2. **ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่ม** ค่า ARMSE จะ**แปรผันตาม**ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่ม คือส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเพิ่มขึ้นจะทำให้ค่า ARMSE เพิ่มขึ้น

3. **ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ** ค่า ARMSE จะ**แปรผันตาม**ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ คือ ระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มจะทำให้ค่า ARMSE เพิ่มขึ้น

4. **สัดส่วนข้อมูลสูญหายในตัวแปรอิสระ** ค่า ARMSE จะ**แปรผันตาม**สัดส่วนข้อมูลสูญหายในตัวแปรอิสระ คือ สัดส่วนข้อมูลสูญหายในตัวแปรอิสระเพิ่มขึ้นจะทำให้ค่า ARMSE เพิ่มขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

การวิจัยครั้งนี้เป็นการวิจัยเพื่อเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระด้วยวิธี CC วิธี FOR-W วิธี FOR-SR วิธี MFOR-W และวิธี MFOR-SR โดยใช้เกณฑ์ค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุคูณ (ARMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (DIFF) เป็นส่วนประกอบในการเปรียบเทียบ โดยมีสถานการณ์ที่ศึกษา ดังนี้

1. ค่าความคลาดเคลื่อนมีการแจกแจงปกติที่มีค่าเฉลี่ย (μ) มีค่าเท่ากับศูนย์ และส่วนเบี่ยงเบนมาตรฐาน (σ) เท่ากับ 0.5, 1 และ 3 ตามลำดับ
2. จำนวนตัวแปรอิสระ (p) ที่ศึกษาเท่ากับ 3
3. ขนาดตัวอย่างที่ศึกษามี 4 ขนาด คือ 5p 10p 15p และ 20p ตามลำดับ
4. การสูญหายของข้อมูลกำหนดให้ปรากฏขึ้นโดยสุ่ม และกำหนดให้อัตราการสูญหายของข้อมูลซึ่งมีค่าเท่ากับ 5 % 15% และ 25 %
5. ระดับพหุสัมพันธ์ของตัวแปรอิสระแบ่งเป็น 3 ระดับ คือ

   ระดับต่ำ ค่า ρ = (0.1 0.2 0.3)

   ระดับปานกลาง ค่า ρ = (0.4 0.5 0.6)

   ระดับสูง ค่า ρ = (0.7 0.7 0.8)

   เมื่อค่า ρ ในวงเล็บ คือ ความสัมพันธ์ระหว่าง X_1 กับ X_2, X_1 กับ X_3 และ X_2 กับ X_3 ตามลำดับ

วิธีการดำเนินการวิจัยครั้งนี้ ใช้วิธีการจำลองด้วยเทคนิคมอนติคาร์โล (Monte Carlo technique) โดยใช้โปรแกรม MATLAB 7 เพื่อสร้างข้อมูลตามสถานการณ์ที่กำหนดโดยกระทำซ้ำ 1,000 ครั้งในแต่ละสถานการณ์

5.1 ผลการเปรียบเทียบค่าเฉลี่ยรากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย

จากการเปรียบเทียบค่า ARMSE ของทั้ง 5 วิธี พบว่า เมื่อสัดส่วนข้อมูลสูญหายในตัวแปรอิสระทุกระดับ คือ 5 %, 15 % และ 25 % วิธีที่ดีที่สุดคือวิธี MFOR-SR ทุกขนาดตัวอย่าง ทุกระดับความสัมพันธ์ระหว่างตัวแปรอิสระ และทุกส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่ม เพราะวิธีประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีของ SRM เป็นวิธีที่ลดความคลาดเคลื่อนของ

การประมาณค่าที่เกิดจากความสัมพันธ์กันระหว่างตัวแปรอิสระ ส่วนวิธีประมาณค่าข้อมูลสูญหายด้วยวิธี MFOR จะให้ค่าที่ใกล้เคียงค่าจริงมากกว่าวิธี FOR เนื่องจากมีการเพิ่มตัวแปรตามในการประมาณค่า โดยที่ข้อมูลของตัวแปรตามมีความแปรปรวนน้อยก็จะทำให้ได้ค่าประมาณที่ใกล้เคียงค่าจริง

สำหรับการวิเคราะห์ความถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระในกรณีที่ขนาดตัวอย่างน้อยแต่มีสัดส่วนข้อมูลสูญหายมากจะทำให้ค่าประมาณค่าสัมประสิทธิ์ความถดถอยพหุคูณที่ได้เกิดความคลาดเคลื่อนมากขึ้น

5.2 ปัจจัยที่มีผลต่อค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของแต่ละวิธี

1. ขนาดตัวอย่าง

เมื่อขนาดตัวอย่างเพิ่มขึ้น ค่า ARMSE ของทุกวิธีมีแนวโน้มลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นจะส่งผลให้ค่าความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยลดลง

2. สัดส่วนข้อมูลสูญหายในตัวแปรอิสระ

เมื่อสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเพิ่มขึ้น ค่า ARMSE ของทุกวิธีมีแนวโน้มเพิ่มขึ้น เนื่องจากมีการประมาณค่าข้อมูลสูญหายจึงทำให้มีค่าความคลาดเคลื่อนจากการประมาณค่าเพิ่มขึ้น ส่วนวิธีที่ทำการตัดชุดข้อมูลที่สูญหายทิ้งไปนั้น ทำให้ขนาดตัวอย่างลดลง มีผลให้ในการประมาณค่าสัมประสิทธิ์การถดถอยคลาดเคลื่อนไปจากความเป็นจริง

3. ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่ม

เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มเพิ่มขึ้น ค่า ARMSE ของทุกวิธีมีแนวโน้มเพิ่มขึ้น เพราะเมื่อส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้น หมายความว่าความคลาดเคลื่อนมีการกระจายมากขึ้น จึงทำให้ค่า ARMSE มีค่าเพิ่มขึ้นและมีผลทำให้ในการประมาณค่าสัมประสิทธิ์การถดถอยคลาดเคลื่อนไปจากความเป็นจริง

4. ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ

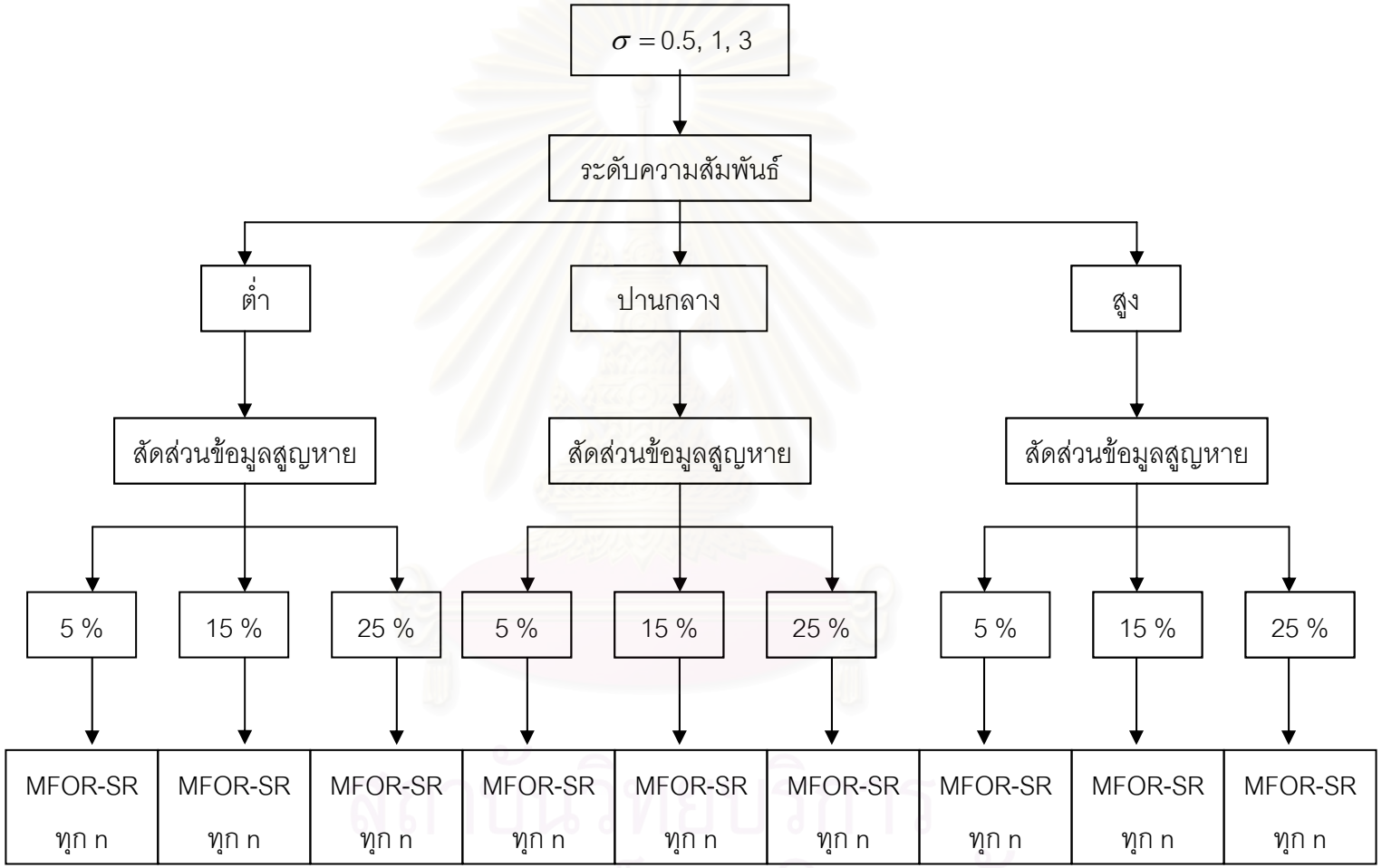
วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระทั้ง 5 วิธี จะให้ค่า ARMSE เพิ่มขึ้น เมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น เพราะระดับความสัมพันธ์ที่เพิ่มขึ้นจะทำให้ค่าลักษณะเฉพาะของเมทริกซ์ $X'X$ มีค่าลดลง จึงส่งผลให้ค่า ARMSE มีแนวโน้มเพิ่มขึ้น

5.3 ผลสรุปการเลือกวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ

ในการวิจัยครั้งนี้พบว่าการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี MFOR-SR จะให้ผลดีทุกกรณี เพราะวิธีนี้เป็นการประมาณค่าข้อมูลสูญหายโดยใช้วิธี MFOR ซึ่งเป็นการใช้ข้อมูลทั้งหมดซึ่งได้แก่ตัวแปรอิสระและตัวแปรตาม และทำการประมาณค่าสัมประสิทธิ์การถดถอยด้วยตัวประมาณผสม (mixed estimator) ที่ปรับค่าโดยวิธีของสไตน์ (stein rule method) เพื่อลดความคลาดเคลื่อนที่เกิดจากการที่ตัวแปรอิสระมีความสัมพันธ์กัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>รูปที่ 5.1</u> แผนผังสรุปผลวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระเชิงทฤษฎี



5.4 ข้อเสนอแนะ

ผลการวิจัยครั้งนี้จะเสนอแนะเป็น 2 ด้าน คือ

5.4.1 ด้านการนำไปใช้ประโยชน์

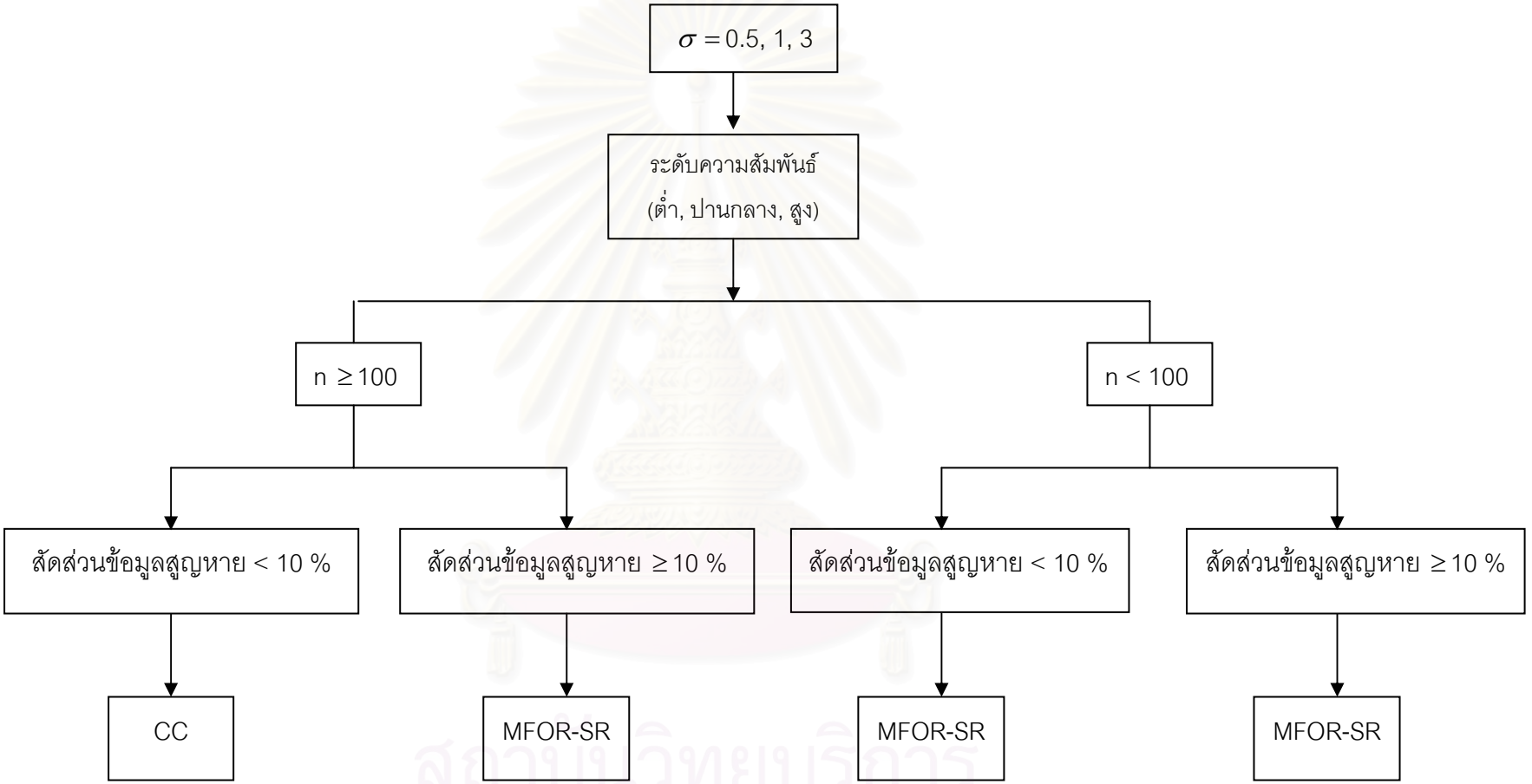
เพื่อเป็นแนวทางในการเลือกวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ เมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระให้เหมาะสมในแต่ละสถานการณ์

เมื่อพิจารณาจากขนาดตัวอย่างกับสัดส่วนข้อมูลสูญหายในทางปฏิบัติจะพบว่า ถ้าขนาดตัวอย่างที่ทำการศึกษามีขนาดใหญ่และสัดส่วนข้อมูลสูญหายน้อยการตัดชุดข้อมูลที่เกิดการสูญหายจะส่งผลกระทบต่อค่าความคลาดเคลื่อนในการประมาณค่าไม่มาก โดยทำการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด ซึ่งในงานวิจัยนี้เรียกว่าวิธี CC

ในกรณีที่ขนาดตัวอย่างน้อยและมีสัดส่วนข้อมูลสูญหายมาก การตัดชุดข้อมูลที่เกิดการสูญหาย อาจส่งผลต่อการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเพราะข้อมูลที่มีอยู่อาจไม่เพียงพอทำให้ค่าความคลาดเคลื่อนในการประมาณค่ามีค่าสูงเกินจริง ดังนั้นจึงต้องประมาณค่าข้อมูลที่เกิดการสูญหายด้วยวิธี MFOR ซึ่งเป็นการใช้รายละเอียดของข้อมูลทั้งตัวแปรตามและตัวแปรอิสระ แล้วประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี SRM เป็นตัวประมาณที่ลดความคลาดเคลื่อนที่เกิดจากการสัมพันธ์กันของตัวแปร จากที่กล่าวมาสามารถแสดงการเลือกใช้ตัวประมาณที่เหมาะสมในเชิงปฏิบัติ ได้ดังนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>รูปที่ 5.2</u> แผนผังแสดงการเลือกวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระเชิงปฏิบัติ



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

5.4.2 ด้านการศึกษาวิจัย

เพื่อเป็นแนวทางให้ผู้ที่สนใจได้ศึกษาเพิ่มเติมและเป็นการขยายผลการวิจัยออกไปให้เกิดประโยชน์มากขึ้น โดยทำการศึกษาดังต่อไปนี้

1. การวิจัยครั้งนี้ผู้วิจัยศึกษากรณีความคลาดเคลื่อนมีการแจกแจงแบบปกติ ผู้ทำการวิจัยต่อไปควรศึกษาเพิ่มเติมเมื่อความคลาดเคลื่อนมีการแจกแจงแบบอื่นๆ เช่น การแจกแจกลอกนอร์มอล เพราะเมื่อการแจกแจงของความคลาดเคลื่อนที่ใช้ในการประมาณค่าข้อมูลสูญหายเปลี่ยนไปอาจทำให้ผลการวิจัยที่ออกมาเปลี่ยนแปลงไปจากเดิม

2. ศึกษาเพิ่มเติมเมื่อขยายขอบเขตให้กว้างขึ้นกว่าที่ผู้วิจัยกำหนด เช่น กำหนดสัดส่วนข้อมูลสูญหายของตัวแปรอิสระมากขึ้นอาจทำให้ประสิทธิภาพของการประมาณค่าสูญหายเปลี่ยนแปลงไปจากเดิมและเมื่อกำหนดให้ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมากขึ้นอาจทำให้ผลการวิจัยที่ออกมาเปลี่ยนแปลงไปจากเดิม

3. ควรศึกษาเพิ่มเติมในกรณีที่ข้อมูลเป็นแบบอนุกรมเวลา เพราะข้อมูลแบบอนุกรมเวลาจะมีค่าเพิ่มขึ้นหรือลดลงเมื่อเวลาเปลี่ยนไป ซึ่งการสูญหายของข้อมูล ณ ช่วงเวลาหนึ่งอาจส่งผลต่อการประมาณค่าต่างๆ

4. ศึกษาเพิ่มเติมเมื่อวิธีประมาณค่าข้อมูลสูญหายแบบอื่นๆ และวิธีประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณวิธีอื่นๆ ที่เหมาะสมกับการวิเคราะห์ความถดถอยเมื่อเกิดปัญหาข้อมูลสูญหายเพื่อเปรียบเทียบกับวิธีที่ผู้วิจัยทำการศึกษา

5. ข้อสังเกตสำหรับวิธีการประมาณค่าข้อมูลสูญหายด้วยการเพิ่มตัวแปรตามนั้นเป็นวิธีการที่เหมาะสมกับข้อมูลทางเศรษฐศาสตร์ เพราะได้ใช้ความสัมพันธ์ของข้อมูล ซึ่งวิธีนี้จะเป็นวิธีที่น่าจะเด่นกว่าวิธีการประมาณค่าข้อมูลสูญหายที่ไม่ได้พิจารณาการเพิ่มตัวแปรตาม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

ภาษาไทย

ชุติมา ชัยมุสิก. การวิเคราะห์การถดถอยเชิงซ้อนเมื่อข้อมูลของตัวแปรอิสระสูญหาย . วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2533.

วิวัฒน์ สกลสนธิเศรษฐ์. การประมาณค่าแบบริดจ์เมื่อค่าสังเกตสูญหาย. วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2532.

ภาษาอังกฤษ

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

Shaalabh and Alan,T. K. Wan. Stein-Rule Estimation in Mixed Regression Models. Biometrical Journal 42, (2000): 203 – 214.

Toutenburg, H., Srivastava, V.K. and Fieger, A. Estimation of Parameters in Multiple Regression with Missing X-Observation using First Order Regression Procedure. University of Munich and Lucknow University, 1996.

Toutenburg, H. and Srivastava, V.K. Improving the Estimation of Coefficients in Linear Regression Model with Some Missing Observations on Some Explanatory Variables. University of Munich and Lucknow University, 1998.

Toutenburg, H., Srivastava, V.K. and Fieger, A. Weighted Modified First Order Regression Procedures for Estimation in Linear Models with Missing X-Observations. University of Munich and Lucknow University, 1998.

บรรณานุกรม

ภาษาไทย

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

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

ภาษาอังกฤษ

Judge, G. G., Griffiths, W.E., Hill, R. C., Lütkepohl, H. and Lee, T. C. The Theory and Practice of Econometrics, 2nd ed. New York: John Wiley & Sons, 1982.

Little, R. J. A. Regression with missing X's: a review, Journal of the American Statistical Association 87, (1992): 1227-1237

Little, R. J. A. and Rubin, D. B. Statistical Analysis with Missing Data. New York: John Wiley & Sons, 1987.

Rao, C.R. and Toutenburg, H. Linear Models: Least Squares and Alternatives. New York: Springer, 1995.

Toutenburg, H. and Srivastava, V.K. On the First Order Regression Procedures of Estimation for Incomplete Regression Models. University of Munich and Lucknow University, 1999.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายละเอียดของโปรแกรมที่ใช้ในงานวิจัย

ในงานวิจัยครั้งนี้ได้ใช้โปรแกรม MATLAB สำหรับหาค่าประมาณสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระด้วยวิธีข้อมูลสมบูรณ์ (CC) วิธีการถดถอยอันดับที่หนึ่ง-วิธีผสมแบบถ่วงน้ำหนัก (FOR-W) วิธีการถดถอยอันดับที่หนึ่ง-วิธีผสมแบบสไตน์ (FOR-SR) วิธีประยุกต์การถดถอยอันดับที่หนึ่ง-วิธีผสมแบบถ่วงน้ำหนัก (MFOR-W) และวิธีประยุกต์การถดถอยอันดับที่หนึ่ง-วิธีผสมแบบสไตน์ (MFOR-SR) ซึ่งเป็นโปรแกรมที่ใช้การพัฒนาโปรแกรมบน Windows ส่งผลให้การพัฒนาโปรแกรมสามารถทำได้ง่ายและสะดวก รายละเอียดทั้งหมดของโปรแกรมที่ใช้ในงานวิจัยมีดังนี้

ตารางแสดงลักษณะการทำงานของโปรแกรมที่ใช้ในงานวิจัย

ลำดับที่	ชื่อโปรแกรม	การทำงานของโปรแกรม	ชื่อโปรแกรมย่อย (ฟังก์ชันที่เรียกใช้)
โปรแกรมหลัก			
1	Main Program	- รวบรวมค่า ARMSE SDและ DIFFเพื่อแสดงค่าในตาราง	- product
ฟังก์ชัน			
1	random	- สร้างตัวแปรสุ่ม	
2	normal	- สร้างข้อมูลที่มีการแจกแจงปกติ	- random
3	generate	- สร้างค่าคลาดเคลื่อน $\underset{\sim}{\varepsilon}$ - กำหนดค่าพารามิเตอร์ $\underset{\sim}{\beta}$ - สร้างตัวแปรตาม	- normal
4	product	- กำหนดจำนวนรอบ - สร้างข้อมูลตัวแปรอิสระ - ประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี CC วิธี FOR-W วิธี FOR-SR วิธี MFOR-W และวิธี MFOR-SR - คำนวณค่า ARMSE SD และ DIFF	- normal - CC - FOR-W - FOR-SR - MFOR-W - MFOR-SR

ลำดับที่	ชื่อโปรแกรม	การทำงานของโปรแกรม	ชื่อโปรแกรมย่อย (ฟังก์ชันที่เรียกใช้)
5	CC	- ประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี CC - คำนวณค่า MSE เมื่อประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี CC	
6	FOR	- ประมาณค่าข้อมูลสูญหายในตัวแปรอิสระด้วยวิธี FOR	
7	MFOR	- ประมาณค่าข้อมูลสูญหายในตัวแปรอิสระด้วยวิธี MFOR	
8	MIXED	- ประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี mixed	
9	FOR_W	- ประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี FOR-W - คำนวณค่า MSE เมื่อประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี FOR-W	- FOR
10	MFOR_W	- ประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี MFOR-W - คำนวณค่า MSE เมื่อประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี MFOR-W	- MFOR
11	FOR_SR	- ประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี FOR-SR - คำนวณค่า MSE เมื่อประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี FOR-SR	- FOR

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ลำดับที่	ชื่อโปรแกรม	การทำงานของโปรแกรม	ชื่อโปรแกรมย่อย (ฟังก์ชันที่เรียกใช้)
12	MFOR_SR	- ประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี MFOR-SR - คำนวณค่า MSE เมื่อประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี MFOR-SR	- MFOR

 # การแสดงผลในรูปตารางตามความแปรปรวนของค่าคลาดเคลื่อนที่เปลี่ยนไป #

```
%  Main program
sigmaey=0.5; %sigmaey= 0.5, 1, 3%
p=3;
%================covariance matrix==============================
cov3l=[1 0.1 0.2;0.1 1 0.3;0.2 0.3 1];
cov3m=[1 0.4 0.5;0.4 1 0.6;0.5 0.6 1];
cov3h=[1 0.7 0.7;0.7 1 0.8;0.7 0.8 1];
%================================================================
mull=cov3l;
per1=5;
[mullper1n1,MSETL1n1]=product(p,per1,mull,5*p,sigmaey);
[mullper1n2,MSETL1n2]=product(p,per1,mull,10*p,sigmaey);
[mullper1n3,MSETL1n3]=product(p,per1,mull,15*p,sigmaey);
[mullper1n4,MSETL1n4]=product(p,per1,mull,20*p,sigmaey);
TL5=[mullper1n1;mullper1n2;mullper1n3;mullper1n4];
per2=15;
[mullper2n1,MSETL2n1]=product(p,per2,mull,5*p,sigmaey);
[mullper2n2,MSETL2n2]=product(p,per2,mull,10*p,sigmaey);
[mullper2n3,MSETL2n3]=product(p,per2,mull,15*p,sigmaey);
[mullper2n4,MSETL2n4]=product(p,per2,mull,20*p,sigmaey);
TL15=[mullper2n1;mullper2n2;mullper2n3;mullper2n4];
per3=25;
[mullper3n1,MSETL3n1]=product(p,per3,mull,5*p,sigmaey);
[mullper3n2,MSETL3n2]=product(p,per3,mull,10*p,sigmaey);
[mullper3n3,MSETL3n3]=product(p,per3,mull,15*p,sigmaey);
[mullper3n4,MSETL3n4]=product(p,per3,mull,20*p,sigmaey);
TL25=[mullper3n1;mullper3n2;mullper3n3;mullper3n4];
```

```
mulm=cov3m;
per1=5;
[mulmper1n1,MSETM1n1]=product(p,per1,mulm,5*p,sigmaey);
[mulmper1n2,MSETM1n2]=product(p,per1,mulm,10*p,sigmaey);
[mulmper1n3,MSETM1n3]=product(p,per1,mulm,15*p,sigmaey);
[mulmper1n4,MSETM1n4]=product(p,per1,mulm,20*p,sigmaey);
TM5=[ mulmper1n1;mulmper1n2;mulmper1n3;mulmper1n4];
per2=15;
[mulmper2n1,MSETM2n1]=product(p,per2,mulm,5*p,sigmaey);
[mulmper2n2,MSETM2n2]=product(p,per2,mulm,10*p,sigmaey);
[mulmper2n3,MSETM2n3]=product(p,per2,mulm,15*p,sigmaey);
[mulmper2n4,MSETM2n4]=product(p,per2,mulm,20*p,sigmaey);
TM15=[ mulmper2n1;mulmper2n2;mulmper2n3;mulmper2n4];
per3=25;
[mulmper3n1,MSETM3n1]=product(p,per3,mulm,5*p,sigmaey);
[mulmper3n2,MSETM3n2]=product(p,per3,mulm,10*p,sigmaey);
[mulmper3n3,MSETM3n3]=product(p,per3,mulm,15*p,sigmaey);
[mulmper3n4,MSETM3n4]=product(p,per3,mulm,20*p,sigmaey);
TM25=[ mulmper3n1;mulmper3n2;mulmper3n3;mulmper3n4];
mulh=cov3h;
per1=5;
[mulhper1n1,MSETH1n1]=product(p,per1,mulh,5*p,sigmaey);
[mulhper1n2,MSETH1n2]=product(p,per1,mulh,10*p,sigmaey);
[mulhper1n3,MSETH1n3]=product(p,per1,mulh,15*p,sigmaey);
[mulhper1n4,MSETH1n4]=product(p,per1,mulh,20*p,sigmaey);
TH5=[ mulhper1n1;mulhper1n2;mulhper1n3;mulhper1n4];
per2=15;
[mulhper2n1,MSETH2n1]=product(p,per2,mulh,5*p,sigmaey);
[mulhper2n2,MSETH2n2]=product(p,per2,mulh,10*p,sigmaey);
[mulhper2n3,MSETH2n3]=product(p,per2,mulh,15*p,sigmaey);
[mulhper2n4,MSETH2n4]=product(p,per2,mulh,20*p,sigmaey);
TH15=[ mulhper2n1;mulhper2n2;mulhper2n3;mulhper2n4];
per3=25;
[mulhper3n1,MSETH3n1]=product(p,per3,mulh,5*p,sigmaey);
[mulhper3n2,MSETH3n2]=product(p,per3,mulh,10*p,sigmaey);
[mulhper3n3,MSETH3n3]=product(p,per3,mulh,15*p,sigmaey);
[mulhper3n4,MSETH3n4]=product(p,per3,mulh,20*p,sigmaey);
TH25=[ mulhper3n1;mulhper3n2;mulhper3n3;mulhper3n4];
Table=[TL5 TL15 TL25;TM5 TM15 TM25;TH5 TH15 TH25];
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

 # การสร้างเลขสุ่ม #

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

 # การสร้างข้อมูลให้มีการแจกแจงปกติ #

```
function[temp,ix,mm,z2]=normal(mean,sigma,ix,mm,z2)
if mm~=1
   [r1,ix]=rand_number(ix);
   [r2,ix]=rand_number(ix);
   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
```

 # การสร้างข้อมูล #

```
function[y,ix,mm,z2]=generate(X,sigmaey,ix,mm,z2,p,n,col)
b=ones(col,1);
for i=1:n
   [ey(i),ix,mm,z2]=normal(0,sigmaey,ix,mm,z2);
end
ey=ey';
y=(X*b)+ey;
```

 # การกำหนดค่าตัวแปร จำนวนรอบ และผลลัพธ์ที่ต้องการแสดงค่าในตาราง #

```
function[RESULT,MSE]=product(p, per, covx, n,sigmaey)
tt=1000;
mean=0;
sigma=sqrt(1);
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
col=p+1;
ix=65539;
mm=0;
z2=0;
%========== generate X matrix=====================
for i = 1:n
   for j = 1:p
      [XS(i,j),ix,mm,z2]= normal(mean,sigma,ix,mm,z2);
   end
end

%=====compute multicollinearity by lomer triangular matrix and cholesky======
r=Cholesky(covx);
C=r';
cov=C*r;
XX=C*XS';
XX=XX';
X=[ones(n,1) XX];

%============ get y==================
for t=1:tt
   [y,ix,mm,z2]=generate(X,sigmaey,ix,mm,z2,p,n,col);

% ========defind missing in matrix X==========
x=X;
cul=round(n*per/100);
missing1=randperm(n);
missing2=missing1(1:cul);
for i = 1:cul
  x(missing2(1,i),col)=nan;
end
data_old=[y x];
data_new = sortrows(data_old,col+1);
nmm=cul;
nmc=n-nmm;
%=========separate data for analysis==========
data_CC=excise(data_old);
y_CC=data_CC(:,1);
X_CC=data_CC(:,2:col+1);
ZCCC=X_CC(:,2:col-1);
ZCCM=X_CC(:,col);
```

สำนักวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
data_MC=data_new(nmc+1:n,:);
y_MC=data_MC(:,1);
X_MC=data_MC(:,2:col+1);
ZMMC=X_MC(:,2:col-1);
ZMMM=X_MC(:,col);

 %========Estimate Parameter by CC Method==========
[bhcc,msecc]=CC(X_CC,y_CC,p,nmc,col);
   MSECC(t)=msecc;
   BHCC(t,:)=bhcc';

%========Estimate Parameter Missing by FOR Method==========
 [B_LS,est_MMMfor]=FOR(ZCCC,ZCCM,ZMMC,ZMMM,p,nmc,col);
 X_estMCFOR=[ones(nmm,1) ZMMC est_MMMfor];
  XX_FOR=[X_estMCFOR' X_CC' ];
 X_FOR=XX_FOR';
[bhfor,msefor]=MIXED(X_FOR,X_CC,X_estMCFOR,y_CC,y_MC,p,nmm,nmc,col);

%========Estimate Parameter by FOR-SR Method=============
[bhfor_sr,msefor_sr]=FOR_SR(bhfor,X_FOR,X_CC,X_estMCFOR,y_CC,y_MC,p,nmm,nmc,col);
   MSEFOR_SR(t)=msefor_sr;
   BHFOR_SR(t,:)=bhfor_sr';

%========Estimate Parameter by FOR-W Method===============
[bhfor_w,msefor_w]=FOR_W(y,X_FOR,X_CC,X_estMCFOR,ZMMC,y_CC,y_MC,p,n,nmc,nmm,col);
   MSEFOR_W(t)=msefor_w;
   BHFOR_W(t,:)=bhfor_w';

%=========Estimate Parameter Missing by MFOR Method=========
[B_LS,est_MMMfor]=FOR(ZCCC,ZCCM,ZMMC,ZMMM,p,nmc,col);
[B_2,B_1,est_MMMmfor]=MFOR(B_LS,y_CC,y_MC,ZCCC,ZCCM,ZMMC,ZMMM,p,nmc,col);
X_estMCMFOR=[ones(nmm,1) ZMMC est_MMMmfor];
XX_MFOR=[X_estMCMFOR' X_CC' ];
 X_MFOR=XX_MFOR';
[bhmfor,msemfor]=MIXED(X_MFOR,X_CC,X_estMCMFOR,ZMMC,y_CC,y_MC,p,nmc,nmm,col);
   MSEMFOR(t)=msemfor;
   BHMFOR(t,:)=bhmfor';
```

```
%==========Estimate Parameter by MFOR-W Method==========
[bhmfor_w,msemfor_w]=MFOR_W(y,X_MFOR,X_CC,X_estMCMFOR,ZMMC,y_CC,y_MC,p,n,nmc,nmm,col);
   MSEMFOR_W(t)=msemfor_w;
   BHMFOR_W(t,:)=bhmfor_w';

%=========Estimate Parameter by MFOR-SR Method==========
[bhmfor_sr,msemfor_sr]= MFOR_SR (bhmfor,X_MFOR,X_CC,X_estMCMFOR,y_CC,y_MC,p,nmm,nmc,col);
   MSEMFOR_SR(t)=msemfor_sr;
   BHWMFOR_SR(t,:)=bhmfor_sr';
end

%================== Criterion=====================
MSE=[MSECC;MSEFOR_W;MSEFOR_SR;MSEMFOR_W;MSEMFOR_SR];

amsecc=mean2(MSECC);
amsewfor=mean2(MSEFOR_W);
amsefor_sr=mean2(MSEFOR_SR);
amsemfor_sr=mean2(MSEMFOR_SR);
amsewmfor=mean2(MSEMFOR_W);
AMSE=[amsecc amsefor_w amsefor_sr amsemfor_w amsemfor_sr];

stdcc=std(MSECC);
stdwfor=std(MSEFOR_W);
stdfor_sr=std(MSEFOR_SR);
stdwmfor=std(MSEMFOR_W);
stdmfor_sr=std(MSEMFOR_SR);
SD=[stdcc stdfor_w stdfor_sr stdmfor_w stdmfor_sr];

amsem=min(AMSE);
diffcc=100*(amsecc-amsem)/amsem;
diffwfor=100*(amsefor_w-amsem)/amsem;
difffor_sr=100*(amsefor_sr-amsem)/amsem;
diffwmfor=100*(amsemfor_w-amsem)/amsem;
diffmfor_sr=100*(amsemfor_sr-amsem)/amsem;
DIFF=[diffcc difffor_w difffor_sr diffmfor_w diffmfor_sr];

RESULT=[AMSE;SD;DIFF];
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี CC

```
function[bhcc,msecc]=CC(X_CC,y_CC,p,nmc,col)
bhcc=(inv(X_CC'*X_CC))*(X_CC'*y_CC);
yhcc=X_CC*bhcc;
b=ones(col,1);
sscc=bhcc-b;
sscc=sscc.^2;
msecc=sqrt(sum(sscc)/col);
```

การประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี FOR-W

```
function[bhwfor,msewfor]=FOR_W(y,X_FOR,X_CC,X_estMCFOR,ZMMC,y_CC,y_MC,p,n,nmc,nmm,col)
P3=X_FOR(:,1:col)*inv(X_FOR(:,1:col)'*X_FOR(:,1:col))*X_FOR(:,1:col)';
forvarresidualyxp=std((eye(n)-)*y);
P4=X_FOR(:,1:col-1)*inv(X_FOR(:,1:col-1)'*X_FOR(:,1:col-1))*X_FOR(:,1:col-1)';
forvarresidualyxp_1=std((eye(n)-P4)*y);
ww=(forvarresidualyxp)^2/(forvarresidualyxp_1)^2;
bhfor_w=(inv((X_CC'*X_CC)+((ww^2)*(X_estMCFOR'*X_estMCFOR))))*((X_CC'*y_CC)+((ww^2)*(X_estMCFOR'*y_MC
)));
yhfor_w=X_FOR*bhfor_w;
ssfor_w=bhfor_w-ones(col,1);
ssfor_w=ssfor_w.^2;
ssfor_w=ssfor_w.^2;
msefor_w=sqrt(ssfor_w/col);
```

การประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี FOR-SR

```
function[bhfor_sr,msefor_sr]=FOR_SR(bhfor,X_FOR,X_CC,X_estMCFOR,y_CC,y_MC,p,nmm,nmc,col)
hh=(p-2)*inv((1+(nmm/(nmc-p+2)))*(1+(2/(nmc+nmm-p+2))));
h=hh/(nmc+nmm-p+2);
eccfor=y_CC-X_CC*bhfor;
emsfor=y_MC-X_estMCFOR*bhfor;
bhfor_sr=(1-h*(((eccfor'*eccfor)+(emsfor'*emsfor))/(bhfor'*(X_CC'*X_CC+X_estMCFOR'*
X_estMCFOR)*bhfor)))*bhfor;
yhfor_sr=X_FOR*bhfor_sr;
ssfor_sr=bhfor_sr-ones(col,1);
ssfor_sr=ssfor_sr.^2;
msefor_sr=sqrt(sum(ssfor_sr)/col);
```

การประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี MFOR-W

```
function[bhmfor_w,msemfor_w]=MFOR_W(y,X_MFOR,X_CC,X_estMCMFOR,ZMMC,y_CC,y_MC,p,n,nmc,nmm,col)
P=X_MFOR(:,1:col)*inv(X_MFOR(:,1:col)'*X_MFOR(:,1:col))*X_MFOR(:,1:col)';
varresidualyxp=std((eye(n)-P)*y);
P2=X_MFOR(:,1:col-1)*inv(X_MFOR(:,1:col-1)'*X_MFOR(:,1:col-1))*X_MFOR(:,1:col-1)';
varresidualyxp_1=std((eye(n)-P2)*y);
w=(varresidualyxp)^2/(varresidualyxp_1)^2;
bhmfor_w=(inv((X_CC'*X_CC)+((w^2)*(X_estMCMFOR'*X_estMCMFOR))))*((X_CC'*y_CC)+((w^2)*(X_estMCMFOR'*y
_MC)));
yhmfor_w=X_MFOR*bhmfor_w;
ssmfor_w=bhmfor_w-ones(col,1);
ssmfor_w=ssmfor_w.^2;
ssmfor_w=ssmfor_w.^2;
msemfor_w=ssmfor_w/col;
```

การประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี MFOR-SR

```
function[bhmfor_sr,msemfor_sr]=MFOR_SR(bhmfor,X_MFOR,X_CC,X_estMCMFOR,y_CC,y_MC,p,nmm,nmc,col)
hh=(p-2)*inv((1+(nmm/(nmc-p+2)))*(1+(2/(nmc+nmm-p+2))));
h=hh/(nmc+nmm-p+2);
eccmfor=y_CC-X_CC*bhmfor;
emsmfor=y_MC-X_estMCMFOR*bhmfor;
bhmfor_sr=(1-h*(((eccmfor'*eccmfor)+(emsmfor'*emsmfor))/(bhmfor'*(X_CC'*X_CC+X_estMCMFOR'*X_estMCMFOR)*
bhmfor)))*bhmfor;
yhmfor_sr=X_MFOR*bhmfor_sr;
ssmfor_sr=bhmfor_sr-ones(col,1);
ssmfor_sr=ssmfor_sr.^2;
msemfor_sr=sum(ssmfor_sr)/col;
```

การประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีแบบผสม

```
function[bhfor,msemixed]=MIXED(X_MIXED,X_CC,X_estMCMIXED,y_CC,y_MC,p,nmm,nmc,col)
bhmixed=(inv((X_CC'*X_CC)+(X_estMCMIXED'*X_estMCMIXED)))*((X_CC'*y_CC)+(X_estMCMIXED'*y_MC));
yhmixed=X_MIXED*bhmixed;
ssmixed=bhmixed-ones(col,1);
ssmixed=ssmixed.^2;
msemixed=sum(ssmixed)/col;
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การประมาณค่าข้อมูลสูญหายในตัวแปรอิสระด้วยวิธี FOR

```
function[B_LS,est_MMMfor]=FOR(ZCCC,ZCCM,ZMMC,ZMMM,p,nmc,col)
B_LS=(inv(ZCCC'*ZCCC))*(ZCCC'*ZCCM);
est_MMMfor=ZMMC*B_LS;
```

การประมาณค่าข้อมูลสูญหายในตัวแปรอิสระด้วยวิธี MFOR

```
function[B_2,B_1,est_MMMmfor]=MFOR(B_LS,y_CC,y_MC,ZCCC,ZCCM,ZMMC,ZMMM,p,nmc,col)
M1=eye(nmc)-(ZCCC*(inv(ZCCC'*ZCCC))*ZCCC');
L=inv(ZCCC'*ZCCC)*(ZCCC'*y_CC);
y_CCP=M1*y_CC;
N=inv(y_CCP'*y_CCP);
B_2=N*y_CCP'*ZCCM;
B_1=B_LS-(L*B_2);
est_MMMmfor=(ZMMC*B_1)+(y_MC*B_2);
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวสิริกัญญา สุขวิเสส เกิดวันพฤหัสบดีที่ 12 สิงหาคม พ.ศ. 2525 สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร ในปีการศึกษา 2546 และเข้าศึกษาต่อในหลักสูตรสถิติศาสตรมหาบัณฑิต(สต.ม.) สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปี พ.ศ. 2547

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย