การประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นเมื่อมีค่าผิดปกติในตัวแปรตาม

นางสาวกัญญารัตน์ โพธิสุทธิ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2547

ISBN 974-53-1934-1

ESTIMATION OF PARAMETERS IN LINEAR REGRESSION MODEL HAVING OUTLIERS IN DEPENDENT VARIABLE

Miss Kanyarat Potisut

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2004

ISBN 974-53-1934-1

หัวข้อวิทยานิพนธ์	การประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นเมื่อมีค่าผิดปกติในตัวแปรตาม
โดย	นางสาวกัญญารัตน์ โพธิสุทธิ์
ภาควิชา	สถิติ
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ร้อยเอกมานพ วราภักดิ์

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

...................................คณบดีคณะพาณิชยศาสตร์และการบัญชี

(ผู้ช่วยศาสตราจารย์ ดร.ดนุชา คุณพนิชกิจ)

คณะกรรมการสอบวิทยานิพนธ์

....ประธานกรรมการ

(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

....อาจารย์ที่ปรึกษา

(รองศาสตราจารย์ ร้อยเอกมานพ วราภักดิ์)

....กรรมการ

(รองศาสตราจารย์ ดร.ธีระพร วีระถาวร)

....กรรมการ

(รองศาสตราจารย์ ผกาวดี ศิริรังษี)

กัญญารัตน์ โพธิสุทธิ์ : การประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นเมื่อมีค่าผิดปกติในตัวแปรตาม (ESTIMATION OF PARAMETERS IN LINEAR REGRESSION MODEL HAVING OUTLIERS IN DEPENDENT VARIABLE). อ. ที่ปรึกษา : รศ.ร.อ.มานพ วราภักดิ์, 150 หน้า. ISBN 974-53-1934-1

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณในการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้น เมื่อมีค่าผิดปกติในตัวแปรตาม โดยทำการเปรียบเทียบตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (OLSE- Ordinary Least Squares Estimator) ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ (AWLSE - Adaptive Weighted Least Squares Estimator) และตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ (REWLSE - Robust and Efficient Weighted Least Squares Estimator) ซึ่งเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของตัวประมาณคือ ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) ของพารามิเตอร์ สถานการณ์ที่ศึกษาคือกำหนดการแจกแจงความคลาดเคลื่อนสุ่ม($\underline{\varepsilon}$) สองการแจกแจง คือการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ N(0,10C^2)โดยกำหนดให้สเกลแฟกเตอร์ (C) มีค่าเท่ากับ 3 สำหรับข้อมูลที่มีค่าผิดปกติในระดับไม่รุนแรง และสเกลแฟกเตอร์เท่ากับ 12 สำหรับข้อมูลที่มีค่าผิดปกติในระดับรุนแรง และจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,β) โดยกำหนดให้ β =8 เมื่อข้อมูลมีค่าผิดปกติในระดับไม่รุนแรง และ $\beta = 25$ สำหรับข้อมูลที่มีค่าผิดปกติในระดับรุนแรง กำหนดค่าพารามิเตอร์ $\underline{\beta} = (5, 1, 1)^T$ ตัวแปรอิสระ x_1 จำลองมาจากการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ 20 และความแปรปรวนเท่ากับ 10 ตัวแปรอิสระ x_2 จำลองมาจากการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ 30 และความแปรปรวนเท่ากับ 25 โดยแต่ละระดับความรุนแรงของค่าผิดปกติจะกำหนดให้มีขนาดตัวอย่าง (n) เท่ากับ 20, 30, 40, 50, 60, 70, 80, 90 และ 100 และสัดส่วนการปลอมปน (p) เท่ากับ 0.05, 0.10, 0.15 และ 0.20 จำลองสถานการณ์การทดลองด้วยเทคนิคมอนติคาร์โลซึ่งทำซ้ำ 500 ครั้งในแต่ละสถานการณ์

ผลการวิจัยปรากฏว่าระดับค่าผิดปกติ สัดส่วนการปลอมปน และขนาดตัวอย่าง ต่างมีผลต่อตัวประมาณค่าพารามิเตอร์ของทั้ง 3 ตัว โดยค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของพารามิเตอร์จะเพิ่มขึ้นเมื่อระดับค่าผิดปกติหรือสัดส่วนการปลอมปนเพิ่มขึ้น แต่จะมีค่าลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

กรณีที่ไม่มีค่าผิดปกติในตัวแปรตามและในตัวแปรอิสระ

ในทุกขนาดตัวอย่างและทุกสัดส่วนการปลอมปน ตัวประมาณ OLS ให้ประสิทธิภาพในการประมาณสูงที่สุด และเมื่อขนาดตัวอย่างตั้งแต่ 60 ขึ้นไป ตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS จะมีค่า AMSE ใกล้เคียงกัน

กรณีที่ตัวแปรตามมีค่าผิดปกติในระดับไม่รุนแรง

กรณีที่สัดส่วนการปลอมปนน้อย ($p \in [0.05,0.10]$) และขนาดตัวอย่างมีขนาดเล็ก ($n \in [20,30]$) ตัวประมาณ REWLS ให้ประสิทธิภาพในการประมาณสูงที่สุด ในขณะที่สัดส่วนการปลอมปนน้อย ($p \in [0.05,0.10]$) และขนาดตัวอย่างเพิ่มขึ้น ($n \in (30,100]$) ตัวประมาณ AWLS ให้ประสิทธิภาพในการประมาณสูงที่สุด สำหรับกรณีที่สัดส่วนการปลอมปนเพิ่มขึ้น ($p \in (0.10,0.20]$) ในทุกๆขนาดตัวอย่าง ($n \in [20,100]$) ตัวประมาณ AWLS ให้ประสิทธิภาพในการประมาณสูงที่สุด

กรณีที่ตัวแปรตามมีค่าผิดปกติในระดับรุนแรง

ในทุกขนาดตัวอย่างและทุกสัดส่วนการปลอมปน ตัวประมาณ REWLS ให้ประสิทธิภาพในการประมาณสูงที่สุด และเมื่อสัดส่วนการปลอมปนน้อย (p = 0.05) และขนาดตัวอย่างตั้งแต่ 40 ขึ้นไป พบว่าตัวประมาณ AWLS และตัวประมาณ REWLS มีประสิทธิภาพในการประมาณใกล้เคียงกัน

ภาควิชา……………… สถิติ……………… ลายมือชื่อนิสิต..

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

ปีการศึกษา………….. 2547 ………………

4582161426 : MAJOR STATISTICS
KEY WORD : LINEAR REGRESSION / OUTLIERS / ADAPTIVE WEIGHTED LEAST SQUARES ESTIMATOR / ROBUST AND EFFICIENT WEIGHTED LEAST SQUARES ESTIMATOR

KANYARAT POTISUT : ESTIMATION OF PARAMETERS IN LINEAR REGRESSION MODEL HAVING OUTLIERS IN DEPENDENT VARIABLE. THESIS ADVISOR : ASSOC. PROF. CAPT. MANOP VARAPHAKDI , M.S. 150 pp. ISBN 974-53-1934-1

The objective of this research is to compare the efficiency of estimators for parametor estimation in linear regression model when the dependent variable has outliers. The estimators are Ordinary Least Squares Estimator (OLSE), Adaptive Weighted Least squares Estimator (AWLSE), and Robust and Efficient Weighted Least Squares Estimator (REWLSE). The measurement for the efficiency of estimators is the Average Mean Square Error (AMSE). Random Errors ($\underline{\varepsilon}$) are independent and identically distributed normal that are generated from two distributions and was done under mild and extreme outliers. The contaminated normal distribution is mixture of the normal distribution having mean of zero and variance of 10, and the normal distribution having mean of zero and variance of $10C^2$ where C is a scale factor that is 3 for mild level and 10 for extreme level. And the contaminated normal distribution is mixture of the normal distribution having mean of zero and variance of 10 and the Laplace distribution having mean of zero and variance of $2\beta^2$ where β is 8 for mild level and 25 for extreme level. This research specified the parameter $\underline{\beta}$ = (5, 1, 1 $)^T$. The observations of independent variable X_1 are generated from the normal distribution with mean of 20 and variance of 10. The observations of independent variable X_2 are generated from the normal distribution with mean of 30 and variance of 25. The sample sizes (n) are 20, 30, 40, 50, 60, 70, 80, 90 and 100. The proportions of contamination (p) are 0.05, 0.10, 0.15 and 0.20. The AMSE of the estimators are computed through the Monte Carlo Simulation method. This simulation is repeated 500 times in each situation.

The results of this research show that the level of outliers, proportions of contamination and sample sizes have effected on the parameter estimations. The average values of mean square error of parameters increase when level of outliers or proportions of contamination increase but they decrease when the sample sizes increase.

In case of no outliers in dependent variable and independent variables

For all sample sizes and proportions of contamination, OLSE is the most efficient. Whereas n $\geq$ 60, the AMSE of OLSE, AWLSE and REWLSE are nearly the same.

In case of dependent variable has mild outliers

For small proportions of contamination (p $\in$ [0.05,0.10]) and sample sizes (n $\in$ [20,30]), REWLSE is the most efficient. Whereas AWLSE is the most efficient when sample size increases (n $\in$ (30,100]). For large proportions of contamination (p $\in$ 0.10,0.20]) and for all n (n $\in$ [20,100]), AWLSE is the most efficient.

In case of dependent variable has extreme outliers

For all sample sizes and proportions of contamination, REWLSE is the most efficient. But the AMSE of AWLSE and REWLSE are a nearly efficiency at p = 0.05 for n $\geq$ 40

Department……….Statistics……… Student's signature…………..………………………...
Field of study……..Statistics……… Advisor's signature……………..................................
Academic year……..2004…………

กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้สามารถสำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของ รองศาสตราจารย์ ร้อยเอก มานพ วราภักดิ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำแนะนำสั่งสอน และให้ข้อคิดเห็นต่าง ๆ ตลอดจนช่วยเหลือแก้ไขข้อบกพร่องต่าง ๆ จนกระทั่งวิทยานิพนธ์เสร็จสมบูรณ์ ผู้วิจัยใคร่ขอกราบขอบพระคุณเป็นอย่างสูง

ผู้วิจัยขอกราบขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ซึ่งประกอบไปด้วย รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา ประธานกรรมการ รองศาสตราจารย์ ผกาวดี ศิริรังษี และรองศาสตราจารย์ ดร.ธีระพร วีระถาวร ผู้เป็นกรรมการ ที่ช่วยตรวจสอบและแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น ขอกราบขอบพระคุณคณาจารย์ทุกท่านที่ได้ให้ความรู้ให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษา

ท้ายนี้ ผู้วิจัยใคร่ขอน้อมรำลึกถึงพระคุณครูอาจารย์ทุกท่านที่ได้ให้ความรู้แก่ผู้วิจัย และกราบขอบพระคุณ บิดา มารดา ตลอดจนทุกคนในครอบครัวซึ่งสนับสนุนและให้กำลังใจเสมอมาจนสำเร็จการศึกษา และขอขอบคุณเพื่อน ๆ ทุกคนสำหรับมิตรภาพและกำลังใจที่มีให้ตลอดมา

สารบัญ

สารบัญ (ต่อ)

หน้า

สารบัญตาราง

ตาราง หน้า

สารบัญภาพ

รูปที่ หน้า

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันนี้ได้มีการนำความรู้ทางด้านสถิติไปประยุกต์ใช้กับงานด้านต่างๆเป็นอันมาก โดยเฉพาะอย่างยิ่งการศึกษารูปแบบความสัมพันธ์ที่เรียกว่ารูปแบบการถดถอย โดยกำหนดให้ตัวแปรที่สนใจศึกษาเป็นตัวแปรตาม ส่วนตัวแปรที่เหลือแทนอิทธิพลหรือปัจจัยที่มีผลต่อตัวแปรตาม เรียกว่า ตัวแปรอิสระ

รูปแบบการถดถอยเชิงเส้นแบบพหุเป็นกรณีหนึ่งของการวิเคราะห์การถดถอยเชิงเส้น เมื่อข้อมูลที่สนใจศึกษามีความสัมพันธ์กับปัจจัยมากกว่า 1 ปัจจัย ซึ่งสามารถเขียนอยู่ในรูปทั่วไปได้ดังนี้

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$$

โดยที่

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \qquad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

และ
$$\underline{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ 1 & X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{pn} \end{bmatrix}$$

เมื่อ $\underline{Y}$ แทนเวกเตอร์ของตัวแปรตาม

$\underline{\beta}$ แทนเวกเตอร์ของสัมประสิทธิ์การถดถอย

$\underline{\varepsilon}$ แทนเวกเตอร์ของความคลาดเคลื่อน

$\underline{X}$ แทนเมทริกซ์ของตัวแปรอิสระ

n แทนขนาดตัวอย่าง

และ p แทนจำนวนตัวแปรอิสระ

ในการประมาณค่าสัมประสิทธิ์การถดถอย β_j (j = 1, 2, … , p) ตัวประมาณที่นิยมนำมาใช้กันโดยทั่วไป คือตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (OLSE - Ordinary Least Squares Estimator) โดยตัวประมาณกำลังสองน้อยที่สุดแบบสามัญของ $\underline{\beta}$ คือ $\underline{\hat{\beta}}_{OLS}$ ที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อน $(\underline{Y}-\underline{X}\underline{\beta})^T(\underline{Y}-\underline{X}\underline{\beta})$ มีค่าต่ำสุดคือ

$$\underline{\hat{\beta}}_{OLS} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y}$$

โดยที่

$$\underline{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ 1 & X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{pn} \end{bmatrix} \qquad \underline{X}^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ X_{p1} & X_{p2} & \cdots & X_{pn} \end{bmatrix}$$

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \text{และ} \quad \underline{\hat{\beta}}_{OLS} = \begin{bmatrix} \underline{\hat{\beta}}_0 \\ \underline{\hat{\beta}}_1 \\ \vdots \\ \underline{\hat{\beta}}_p \end{bmatrix}$$

การประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยโดยใช้ตัวประมาณ OLS จะเป็นตัวประมาณไม่เอนเอียงเชิงเส้นที่ดีที่สุด (BLUE - Best Linear Unbiased Estimator) ในการประมาณภายใต้เงื่อนไขเบื้องต้นเกี่ยวกับความคลาดเคลื่อนสุ่ม ดังนี้

1. ความคลาดเคลื่อน ε_i (i = 1, 2, … , n) มีค่าเฉลี่ยเป็นศูนย์ และมีค่าความแปรปรวนคงที่ให้แทนด้วย σ^2
2. ε_i และ ε_j ไม่มีความแปรปรวนร่วมกัน $COV(\varepsilon_i, \varepsilon_j) = 0$ สำหรับ $i \neq j$

อย่างไรก็ตาม ในกรณีที่ค่าสังเกตมีค่าผิดปกติ (Outliers) รวมอยู่ด้วย ตัวประมาณกำลังสองน้อยที่สุดแบบสามัญอาจจะไม่เหมาะสม เนื่องจากวิธีการนี้มีความไวต่อข้อมูลที่ผิดปกติ ทำให้เส้นการถดถอยที่ได้เบี่ยงเบนไปจากข้อมูลกลุ่มใหญ่ เนื่องจากเส้นสมการถดถอยที่ได้ถูกปรับทิศทางไปตามข้อมูลที่ผิดปกติ ส่งผลให้ความแปรปรวนของ $\underline{\hat{\beta}}_{OLS}$ มีค่าสูงขึ้น ซึ่งทำให้ค่าสัมประสิทธิ์การถดถอยที่ได้มีค่าที่ไม่เหมาะสม

ผลงานวิจัยที่เกี่ยวข้องกับการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นเมื่อมีค่าผิดปกติในตัวแปรตาม มีดังนี้

ในปี ค.ศ. 2001 Thomas W.O'Gorman ได้เสนอตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ (AWLSE - Adaptive Weighted Least Squares Estimator) เพื่อใช้ในการหาค่าเฉลี่ยของตัวแปรตามในตัวแบบการถดถอยเชิงเส้นอย่างง่ายเมื่อข้อมูลมีค่าผิดปกติหรือมีการแจกแจงแบบสมมาตร และใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นอย่างง่ายเมื่อการแจกแจงของความคลาดเคลื่อนเป็นการแจกแจงแบบหางยาว ในกรณีที่ข้อมูลมีค่าผิดปกติในตัวแปรตาม พบว่าตัวประมาณ AWLS มีประสิทธิภาพในการประมาณดีกว่าตัวประมาณที่ได้รับการปรับของ Han และ Hawkins (Han and Hawkins's adaptive estimator)

ในปี ค.ศ. 2002 Daniel Gervini และ Victor J. Yohai ได้เสนอตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ (REWLSE - Robust and Efficient Weighted Least Squares Estimator) เพื่อใช้ในการหาค่าสัมประสิทธิ์การถดถอยเชิงเส้นในกรณีที่ไม่มีค่าผิดปกติและมีค่าผิดปกติในตัวแปรตาม ซึ่งตัวประมาณ REWLS เป็นตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก โดยใช้ฟังก์ชันการแจกแจงเชิงสังเกต (Empirical Distribution Function) ในการหาค่า Cut-off เพื่อใช้ในการกำหนดค่าถ่วงน้ำหนักให้กับค่าสังเกต โดยค่าสังเกต X_{ji} และ Y_i ที่ไม่ผิดปกติจะกำหนดค่าถ่วงน้ำหนักให้มีค่าเท่ากับ 1 ในขณะที่ค่าสังเกต X_{ji} และ Y_i ที่เป็นค่าที่ผิดปกติจะถูกตัดออกไป จากการศึกษาพบว่าในกรณีที่ไม่มีค่าผิดปกติ ตัวประมาณ REWLS จะมีประสิทธิภาพในการประมาณค่าใกล้เคียงกับตัวประมาณ OLS เมื่อขนาดตัวอย่างใหญ่ สำหรับในกรณีที่มีค่าผิดปกติในตัวแปรตาม พบว่าตัวประมาณ REWLS มีประสิทธิภาพในการประมาณค่าพารามิเตอร์สูงกว่าตัวประมาณ OLS และตัวประมาณกำลังสองค่ามัธยฐานต่ำสุด (LMSE - Least Median Squares Estimator)

จิตรวี วีระประดิษฐ์ (2539) ได้ทำวิทยานิพนธ์เรื่องการประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นพหุเมื่อข้อมูลมีค่าผิดปกติ โดยทำการศึกษาวิธีประมาณ 3 วิธีคือ วิธีกำลังสองน้อยที่สุด วิธีตัวประมาณ M และวิธีตัวประมาณ Bounded-Influence เมื่อใช้เกณฑ์ความแกร่งของ Huber และของ Tukey ผลการวิจัยสรุปได้ว่า กรณีข้อมูลเกิดค่าผิดปกติในตัวแปรตามและกรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระและตัวแปรตาม โดยทั่วไปวิธีตัวประมาณ M มีค่า MSE (Mean Square Error) ต่ำกว่าวิธีอื่น แต่บางสถานการณ์วิธีตัวประมาณ M และวิธีตัวประมาณ Bounded-Influence จะมีค่า MSE ใกล้เคียงกัน และเมื่อขนาดของค่าผิดปกติของตัวแปรตามมีขนาดใหญ่ โดยทั่วไปวิธีตัวประมาณ M เมื่อใช้เกณฑ์ความแกร่งของ Tukey จะมีค่า MSE ต่ำที่สุด

ปัทมวดี นันทนาเนตร์ (2544) ได้ทำวิทยานิพนธ์เรื่องการเปรียบเทียบตัวประมาณการถดถอยเมื่อมีพหุสัมพันธ์และ/หรือมีค่าผิดปกติ ซึ่งศึกษาตัวแปรประมาณการถดถอย 5 ตัว คือ ตัวประมาณกำลังสองน้อยสุด ตัวประมาณค่าสัมบูรณ์น้อยสุด ตัวประมาณริดจ์ ตัวประมาณริดจ์ที่มี

ค่าสัมบูรณ์น้อยสุด และตัวประมาณริดจ์ที่มีการถ่วงน้ำหนัก ในกรณีที่ตัวแปรตามมีค่าผิดปกติ สรุปได้ว่า ตัวประมาณค่าสัมบูรณ์น้อยสุดจะให้ค่า RMSE (Square Root Mean Square Error) ต่ำที่สุดในทุกสถานการณ์

จากการศึกษาข้างต้นพบว่า มีตัวประมาณที่ยังไม่ได้ทำศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นเมื่อมีค่าผิดปกติในตัวแปรตาม นั่นคือ ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ (AWLSE) และตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ (REWLSE) จึงทำให้ผู้วิจัยมีความสนใจที่จะทำการศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้น โดยประกอบไปด้วย

1. ตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (OLSE - Ordinary Least Squares Estimator)
2. ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ (AWLSE - Adaptive Weighted Least Squares Estimator)
3. ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ (REWLSE - Robust and Efficient Weighted Least Squares Estimator)

เพื่อศึกษาว่าการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นวิธีใดมีประสิทธิภาพมากที่สุด เมื่อมีค่าผิดปกติในตัวแปรตามโดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE - Average Mean Square Error) เป็นเกณฑ์ในการพิจารณา

1.2 วัตถุประสงค์ของการวิจัย

ในการวิจัยนี้มีวัตถุประสงค์ดังนี้

1. เพื่อศึกษาประสิทธิภาพของตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (OLSE) ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ (AWLSE) และตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ (REWLSE) ในการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นเมื่อมีค่าผิดปกติในตัวแปรตาม ภายใต้กรณีต่าง ๆ

2. เพื่อหาข้อเสนอแนะวิธีการประมาณที่เหมาะสม ในการประมาณค่าพารามิเตอร์ของตัวแบบการถดถอยเชิงเส้นเมื่อมีค่าผิดปกติในตัวแปรตาม

1.3 สมมติฐานของการวิจัย

ในการวิจัยนี้มีสมมติฐานดังนี้

การประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นเมื่อมีค่าผิดปกติในตัวแปรตาม ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ (AWLSE) มีประสิทธิภาพในการประมาณสูงกว่าตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (OLSE) และตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ (REWLSE)

1.4 ข้อตกลงเบื้องต้น

การวิจัยครั้งนี้มีข้อตกลงเบื้องต้นสำหรับการดำเนินงานวิจัย ดังนี้

1. การศึกษาในครั้งนี้ใช้รูปแบบการถดถอยเชิงเส้นที่มีตัวแปรอิสระ 2 ตัว ซึ่งมีรูปแบบดังนี้

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, \quad i = 1,2,...,n$$

โดยที่ Y_i แทนค่าสังเกตที่ i ของตัวแปรตาม

X_{1i} แทนค่าสังเกตที่ i ของตัวแปรอิสระตัวที่ 1

X_{2i} แทนค่าสังเกตที่ i ของตัวแปรอิสระตัวที่ 2

$\beta_0,\ \beta_1,\ \beta_2$ แทนสัมประสิทธิ์การถดถอยที่ต้องการประมาณค่า

ε_i แทนความคลาดเคลื่อนตัวที่ i

และ n แทนขนาดตัวอย่าง

2. กำหนดให้มีค่าผิดปกติในตัวแปรตาม ในที่นี้จะกำหนดให้ตัวแปรตามมีค่าผิดปกติ 2 ระดับ คือระดับไม่รุนแรง (Mild Outliers) และระดับรุนแรง (Extreme Outliers) ตามเงื่อนไขของการตรวจสอบค่าผิดปกติโดยใช้กราฟแบบ Box และ Whisker คือ

2.1 ข้อมูลที่มีค่าผิดปกติในระดับไม่รุนแรง คือข้อมูลที่มีค่าตกอยู่ในช่วง

$(Q_1 - 3(IQR), Q_1 - 1.5(IQR))$ หรือตกอยู่ในช่วง $(Q_3 + 1.5(IQR), Q_3 + 3(IQR))$

2.2 ข้อมูลที่มีค่าผิดปกติในระดับรุนแรง คือข้อมูลที่มีค่าตกอยู่ในช่วง

$(-\infty,\ Q_1 - 3(IQR))$ หรือตกอยู่ในช่วง $(Q_3 + 3(IQR), \infty)$

เมื่อ Q_1 คือค่าควอไทล์ที่ 1 (The First Quartile) ของข้อมูล

Q_3คือค่าควอไทล์ที่ 3 (The Third Quartile) ของข้อมูล

IQR คือค่าพิสัยระหว่างควอไทล์ (The Interquartile Range) ซึ่งมีค่าเท่ากับ Q_3- Q_1

1.5 ขอบเขตของการวิจัย

ในการวิจัยครั้งนี้ ทำการศึกษาในกรณีที่ตัวแปรตามมีค่าผิดปกติ โดยกำหนดขอบเขตของการวิจัย ดังนี้

1. กำหนดลักษณะความคลาดเคลื่อนตามการแจกแจงของข้อมูลดังนี้

1.1 การแจกแจงแบบปกติ (Normal Distribution)

ฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(\varepsilon) \quad = \frac{1}{\sqrt{2\pi}\sigma}\exp\{\frac{-(\varepsilon-\mu)^2}{2\sigma^2}\} \quad ; -\infty < \varepsilon < \infty,\ \sigma > 0,\ -\infty < \mu < \infty$$

เมื่อ $E(\varepsilon) = \mu$ และ $Var(\varepsilon) = \sigma^2$

ในที่นี้กำหนดให้ $\mu = 0$ และ $\sigma^2 = 10$

1.2 การแจกแจงแบบปกติปลอมปน

1.2.1 ระหว่างการแจกแจงแบบปกติกับการแจกแจงแบบปกติ

ฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(\varepsilon) \quad = \ (1-p).N(0,\sigma^2) + p.N(0, C^2\sigma^2) \ ; -\infty < \varepsilon < \infty$$

กำหนดให้ $\sigma^2 = 10$ และ C คือสเกลแฟกเตอร์ (Scale Factor) โดยกำหนดให้สเกลแฟกเตอร์มีค่าเท่ากับ 3 สำหรับข้อมูลที่มีค่าผิดปกติในระดับไม่รุนแรง และสเกลแฟกเตอร์เท่ากับ 12 สำหรับข้อมูลที่มีค่าผิดปกติในระดับรุนแรง

1.2.2 ระหว่างการแจกแจงแบบปกติกับการแจกแจงแบบลาปลาซ

ฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(\varepsilon) \quad = \ (1-p).N(0,\sigma^2) + p.L(0,\beta) \ ; -\infty < \varepsilon < \infty \ ; \beta > 0$$

ในที่นี้กำหนดให้ $\sigma^2 = 10$, $\beta = 8$ เมื่อข้อมูลมีค่าผิดปกติในระดับไม่รุนแรง และ $\sigma^2 = 10$, $\beta = 25$ สำหรับข้อมูลที่มีค่าผิดปกติในระดับรุนแรง

p คือสัดส่วนการปลอมปนจากข้อมูลทั้งหมดซึ่งมีขนาดเท่ากับ n โดยในที่นี้จะศึกษาเมื่อสัดส่วนการปลอมปนมีค่าเท่ากับ 0.05, 0.1, 0.15 และ 0.2

$L(\theta, \beta)$ คือการแจกแจงแบบลาปลาซ ซึ่งมีฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(\varepsilon) \quad = \quad \frac{1}{2\beta}\exp\{-\frac{|\varepsilon - \theta|}{\beta}\} \quad ; -\infty < \varepsilon < \infty \quad ; -\infty < \theta < \infty \quad ; \beta > 0$$

เมื่อ $E(\varepsilon) = \theta$ และ $Var(\varepsilon) = 2\beta^2$

ในที่นี้กำหนดให้ $\theta = 0,\ \beta = 8$ เมื่อข้อมูลมีค่าผิดปกติในระดับไม่รุนแรง และ $\theta = 0,\ \beta = 25$ สำหรับข้อมูลที่มีค่าผิดปกติในระดับรุนแรง

2. กำหนดขนาดตัวอย่าง n ที่ใช้ในการวิจัยเท่ากับ 20, 30, 40, 50, 60, 70, 80, 90 และ 100

3. จำนวนค่าผิดปกติเมื่อสัดส่วนการปลอมปนเท่ากับ 0.05, 0.1, 0.15 และ 0.2 ในแต่ละขนาดตัวอย่างที่ศึกษา กรณีที่จำนวนที่ได้เป็นเลขไม่ลงตัว ให้ปัดเป็นเลขจำนวนเต็มที่มีค่าใกล้เคียงมากที่สุด

4. กำหนดลักษณะข้อมูลของตัวแปรอิสระ X_1 และ X_2 จากการแจกแจงแบบปกติที่อิสระซึ่งกันและกัน นั่นคือไม่เกิดปัญหาความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระ (Multicolinearlity)

โดยกำหนดข้อมูลของ X_1 มาจากการแจกแจงแบบปกติ $N(\mu_1, \sigma_1^2)$ และข้อมูลของ X_2 มาจากการแจกแจงแบบปกติ $N(\mu_2, \sigma_2^2)$ ซึ่งค่าของพารามิเตอร์ μ_1, σ_1^2, μ_2 และ σ_2^2 สามารถเป็นค่าใดๆ สำหรับในงานวิจัยนี้เลือกกำหนด $\mu_1 = 20,\ \sigma_1^2 = 10,\ \mu_2 = 30,$ และ $\sigma_2^2 = 25$

5. สร้างตัวแปรตาม Y จากสมการ $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i,\ \ i = 1,2,\ldots,n$ โดยที่พารามิเตอร์ $\beta_0, \beta_1, \beta_2$ เป็นค่าใดๆ (ผลสรุปไม่ขึ้นอยู่กับค่าพารามิเตอร์) สำหรับในงานวิจัยนี้กำหนดให้ $\beta_0 = 5,\ \beta_1 = 1$ และ $\beta_2 = 1$

6. ทำการประมาณค่าสัมประสิทธิ์การถดถอยด้วยตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ และตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ

7. ทำการทดลองหาค่าประมาณ AMSE โดยใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) ซึ่งเขียนด้วยโปรแกรมภาษาฟอร์แทน (Fortran) จำลองซ้ำ 500 รอบในแต่ละสถานการณ์ของการทดลอง

1.6 เกณฑ์ที่ใช้ในการประเมินประสิทธิภาพของตัวประมาณ

เกณฑ์ที่ใช้ในการประเมินว่าตัวประมาณที่ใช้ในการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นตัวใดมีประสิทธิภาพมากที่สุดจะพิจารณาจากเกณฑ์ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) โดยตัวประมาณที่ให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดจะเป็นตัวประมาณที่มีประสิทธิภาพมากที่สุด โดยมีวิธีในการคำนวณดังนี้

$$\text{AMSE} = \frac{\sum_{i=0}^{2}\sum_{j=1}^{500}\frac{(\beta_i - \hat{\beta}_{ij})^2}{500}}{3}$$

เมื่อ β_i แทนค่าพารามิเตอร์ที่ต้องการประมาณ โดยที่ i = 0, 1, 2

$\hat{\beta}_{ij}$ แทนค่าประมาณของพารามิเตอร์ β_i ในรอบที่ j

และ AMSE แทนค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์จำนวน 3 ตัว

1.7 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัยในครั้งนี้ คือ

1. เพื่อทราบประสิทธิภาพของตัวประมาณที่ศึกษาในแต่ละสถานการณ์ของการทดลอง
2. เพื่อเป็นแนวทางในการตัดสินใจเลือกตัวประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้น เมื่อมีค่าผิดปกติในตัวแปรตามได้อย่างเหมาะสม
3. เพื่อใช้เป็นแนวทางในการศึกษาเปรียบเทียบกับตัวประมาณอื่น ๆ ต่อไป

บทที่ 2

ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความเที่ยงตรงของตัวประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้น เมื่อมีค่าผิดปกติในตัวแปรตาม โดยการศึกษาในครั้งนี้ได้เสนอตัวประมาณค่าพารามิเตอร์ 3 ตัว คือ ตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (OLSE - Ordinary Least Squares Estimator) ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ (AWLSE - Adaptive Weighted Least Squares Estimator) และตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ (REWLSE - Robust and Efficient Weighted Least Squares Estimator) ซึ่งแต่ละตัวประมาณมีรายละเอียดดังนี้

2.1 ตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (OLSE - Ordinary Least Squares Estimator)

การประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นด้วยตัวประมาณกำลังสองน้อยที่สุดแบบสามัญมีหลักในการประมาณค่าสัมประสิทธิ์คือ ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนมีค่าต่ำที่สุด

จากตัวแบบการถดถอยเชิงเส้นพหุคูณ $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$ ตัวประมาณกำลังสองน้อยที่สุดของ $\underline{\beta}$ คือ $\hat{\underline{\beta}}$ ที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนมีค่าต่ำสุด ซึ่งมีรายละเอียดดังนี้

กำหนด

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \qquad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

และ $$\underline{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ 1 & X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{pn} \end{bmatrix}$$

เมื่อ $\underline{Y}$ แทนเวกเตอร์ของตัวแปรตาม

$\underline{\beta}$ แทนเวกเตอร์ของสัมประสิทธิ์การถดถอย

$\underline{\varepsilon}$ แทนเวกเตอร์ของความคลาดเคลื่อน

และ $\underline{X}$ แทนเมทริกซ์ของตัวแปรอิสระ

ต้องการหาตัวประมาณ $\hat{\underline{\beta}}$ ของ $\underline{\beta}$ ที่ทำให้ $(\underline{Y}-\underline{X}\underline{\beta})^T(\underline{Y}-\underline{X}\underline{\beta})$ มีค่าต่ำสุด

ให้ $L(\underline{\beta}) = (\underline{Y}-\underline{X}\underline{\beta})^T(\underline{Y}-\underline{X}\underline{\beta})$

$$= \underline{Y}^T\underline{Y} - \underline{\beta}^T\underline{X}^T\underline{Y} - \underline{Y}^T\underline{X}\underline{\beta} + \underline{\beta}^T\underline{X}^T\underline{X}\underline{\beta}$$

$$= \underline{Y}^T\underline{Y} - 2\underline{\beta}^T\underline{X}^T\underline{Y} + \underline{\beta}^T\underline{X}^T\underline{X}\underline{\beta}$$

(เพราะว่า $\underline{Y}^T\underline{X}\underline{\beta} = (\underline{Y}^T\underline{X}\underline{\beta})^T = \underline{\beta}^T\underline{X}^T\underline{Y}$)

แทนค่า $\underline{\beta}$ ด้วย $\hat{\underline{\beta}}$ ในสมการ

$$\left.\frac{\partial}{\partial\underline{\beta}}L(\underline{\beta})\right|_{\underline{\beta}=\hat{\underline{\beta}}} = \underline{0}$$

ได้

$$-2\underline{X}^T\underline{Y} + 2\underline{X}^T\underline{X}\hat{\underline{\beta}} = \underline{0}$$

ดังนั้น

$$\hat{\underline{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = (\underline{X}^T\underline{X})^{-1}(\underline{X}^T\underline{Y})$$

โดยที่ $\underline{X}^T\underline{X}$ ต้องไม่เป็นเมทริกซ์เอกฐาน (Singular Matrix) นั่นคือสามารถหาเมทริกซ์ผกผันของ $\underline{X}^T\underline{X}$ ได้

2.2 ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ (AWLSE - Adaptive Weighted Least Squares Estimator)

ในปี ค.ศ. 2001 Thomas W.O' Gorman ได้เสนอตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ (AWLSE - Adaptive Weighted Least Squares Estimator) ในการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้น ซึ่งมีหลักในการประมาณคือ ใช้ความยาวหางของความคลาดเคลื่อน $\underline{\hat{\varepsilon}}$ ในการกำหนดค่าถ่วงน้ำหนักให้กับค่าสังเกตของ $\underline{X}$, $\underline{Y}$ และ $\underline{\varepsilon}$ ในตัวแบบการถดถอยเชิงเส้น โดยความคลาดเคลื่อน $\underline{\hat{\varepsilon}} = \underline{Y} - \underline{X}\underline{\hat{\beta}}$ หาได้จากตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (OLSE) ของ $\underline{\beta}$ จากนั้นใช้ตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (OLSE) อีกครั้งในการประมาณค่าพารามิเตอร์ $\underline{\beta}$ คือ $\underline{\hat{\beta}}$ ที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนมีค่าต่ำสุด

จากตัวแบบการถดถอยเชิงเส้นพหุคูณ $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$ ใช้ตัวประมาณกำลังสองน้อยที่สุดแบบสามัญในการประมาณค่า $\underline{\beta}$ คือ $\underline{\hat{\beta}}$ ที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนมีค่าต่ำสุด

ได้ $$\underline{\hat{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = (\underline{X}^T \underline{X})^{-1} (\underline{X}^T \underline{Y})$$

นำค่าประมาณ $\underline{\hat{\beta}}$ ที่ได้ไปคำนวณหาความคลาดเคลื่อน $\underline{\hat{\varepsilon}}$ ดังนี้

$$\underline{Y} - \underline{X}\underline{\hat{\beta}} = \underline{\hat{\varepsilon}}$$

กำหนดให้ $$\underline{\hat{\varepsilon}} = \begin{bmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_n \end{bmatrix}$$

จากนั้นนำค่าความคลาดเคลื่อน $\underline{\hat{\varepsilon}}$ ที่ได้ไปกำหนดค่าถ่วงน้ำหนักให้กับค่าสังเกตของ $\underline{X}$, $\underline{Y}$ และ $\underline{\varepsilon}$

เมื่อได้ค่าถ่วงน้ำหนักให้กับค่าสังเกตของ $\underline{X}$, $\underline{Y}$ และ $\underline{\varepsilon}$ แล้ว จากนั้นทำการสร้างเมทริกซ์ค่าถ่วงน้ำหนัก ดังนี้

$$\underline{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix}$$

โดยที่ w_i คือค่าถ่วงน้ำหนักของค่าสังเกตของ $\underline{X}$, $\underline{Y}$ และ $\underline{\varepsilon}$ ที่อยู่ตำแหน่งเดียงกับ $\hat{\varepsilon}_i$ เมื่อ $i = 1, 2, \ldots, n$

จากตัวแบบการถดถอยเชิงเส้นพหุคูณ $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$ กำหนดให้

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \qquad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

และ $$\underline{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ 1 & X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{pn} \end{bmatrix}$$

เมื่อ $\underline{Y}$ แทนเวกเตอร์ของตัวแปรตาม

$\underline{\beta}$ แทนเวกเตอร์ของสัมประสิทธิ์การถดถอย

$\underline{\varepsilon}$ แทนเวกเตอร์ของความคลาดเคลื่อน

และ $\underline{X}$ แทนเมทริกซ์ของตัวแปรอิสระ

กำหนดเมทริกซ์ค่าถ่วงน้ำหนัก $\underline{W}$ ให้กับ $\underline{X}$, $\underline{Y}$ และ $\underline{\varepsilon}$ ในตัวแบบการถดถอยเชิงเส้นดังนี้

$$\underline{W}\underline{Y} = \underline{W}\underline{X}\underline{\beta} + \underline{W}\underline{\varepsilon}$$

เพราะฉะนั้น

$$\underline{W}\underline{Y} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} w_1 Y_1 \\ w_2 Y_2 \\ \vdots \\ w_n Y_n \end{bmatrix}$$

$$\underline{W}\underline{X}\underline{\beta} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix} \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ 1 & X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{pn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

$$= \begin{bmatrix} \beta_0 w_1 + \beta_1 w_1 X_{11} + \beta_2 w_1 X_{21} + \ldots + \beta_p w_1 X_{p1} \\ \beta_0 w_2 + \beta_1 w_2 X_{12} + \beta_2 w_2 X_{22} + \ldots + \beta_p w_2 X_{p2} \\ \vdots \\ \beta_0 w_n + \beta_1 w_n X_{1n} + \beta_2 w_n X_{2n} + \ldots + \beta_p w_n X_{pn} \end{bmatrix}$$

และ $$\underline{W}\underline{\varepsilon} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} w_1\,\varepsilon_1 \\ w_2\,\varepsilon_2 \\ \vdots \\ w_n\,\varepsilon_n \end{bmatrix}$$

กำหนดให้ $\underline{Y}_w = \underline{W}\underline{Y},\ \ \underline{X}_w = \underline{W}\underline{X}$ และ $\underline{\varepsilon}_w = \underline{W}\underline{\varepsilon}$ จากนั้นใช้ตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (OLSE) ในการหาตัวประมาณ $\hat{\underline{\beta}}$ ของ $\underline{\beta}$ ที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อน $(\underline{Y}_w - \underline{X}_w\underline{\beta})^T(\underline{Y}_w - \underline{X}_w\underline{\beta})$ มีค่าต่ำสุด

ให้ $$L(\underline{\beta}) = (\underline{Y}_w - \underline{X}_w\underline{\beta})^T(\underline{Y}_w - \underline{X}_w\underline{\beta})$$

$$= \underline{Y}_w^T\underline{Y}_w - \underline{\beta}^T\underline{X}_w^T\underline{Y}_w - \underline{Y}_w^T\underline{X}_w\underline{\beta} + \underline{\beta}^T\underline{X}_w^T\underline{X}_w\underline{\beta}$$

$$= \underline{Y}_w^T\underline{Y}_w - 2\underline{\beta}^T\underline{X}_w^T\underline{Y}_w + \underline{\beta}^T\underline{X}_w^T\underline{X}_w\underline{\beta}$$

(เพราะว่า $\underline{Y}_w^T\underline{X}_w\underline{\beta} = (\underline{Y}_w^T\underline{X}_w\underline{\beta})^T = \underline{\beta}^T\underline{X}_w^T\underline{Y}_w$)

จากนั้นแทนค่า $\underline{\beta}$ ด้วย $\hat{\underline{\beta}}$

จากสมการ $$\frac{\partial}{\partial\underline{\beta}}\underline{Y}_w^T\underline{Y}_w - 2\underline{\beta}^T\underline{X}_w^T\underline{Y}_w + \underline{\beta}^T\underline{X}_w^T\underline{X}_w\underline{\beta} = \underline{0}$$

จะได้ว่า $$\frac{\partial}{\partial\hat{\underline{\beta}}}\underline{Y}_w^T\underline{Y}_w - 2\hat{\underline{\beta}}^T\underline{X}_w^T\underline{Y}_w + \hat{\underline{\beta}}^T\underline{X}_w^T\underline{X}_w\hat{\underline{\beta}} = \underline{0}$$

$$-2\underline{X}_w^T\underline{Y}_w + 2\underline{X}_w^T\underline{X}_w\hat{\underline{\beta}} = \underline{0}$$

ดังนั้น

$$\hat{\underline{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = (\underline{X}_w^T\underline{X}_w)^{-1}(\underline{X}_w^T\underline{Y}_w)$$

โดยที่ $\underline{X}_w^T\underline{X}_w$ ต้องไม่เป็นเมทริกซ์เอกฐาน (Singular Matrix) นั่นคือสามารถหา เมทริกซ์ผกผัน (Inverse Matrix) ของ $\underline{X}_w^T\underline{X}_w$ ได้

ในส่วนต่อไปจะเป็นวิธีการหาค่าถ่วงน้ำหนัก w_i ให้กับเมทริกซ์ W ซึ่งมีรายละเอียดดังนี้

การหาค่าถ่วงน้ำหนัก w_i ให้กับเมทริกซ์[1]

พิจารณาความคลาดเคลื่อน $\hat{\underline{\varepsilon}} = \underline{Y} - \underline{X}\hat{\underline{\beta}}$ ซึ่งหาได้จากตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (OLSE) ของ $\underline{\beta}$ ที่กล่าวไว้แล้วในข้างต้น

คำนวณหาค่าถ่วงน้ำหนักโดยการกำหนดความยาวของหางสำหรับการแจกแจงของค่าความคลาดเคลื่อน $\hat{\underline{\varepsilon}}$ โดยที่ค่าความยาวของหางทางซ้าย (Left tail length measure-t_L) และค่าความยาวของหางทางขวา (Right tail length measure-t_R) สามารถหาได้จาก

$$t_L = \frac{\hat{\varepsilon}_{P25} - \hat{\varepsilon}_{P2}}{\hat{\varepsilon}_{P75} - \hat{\varepsilon}_{P25}} \qquad \text{และ} \qquad t_R = \frac{\hat{\varepsilon}_{P98} - \hat{\varepsilon}_{P75}}{\hat{\varepsilon}_{P75} - \hat{\varepsilon}_{P25}}$$

เมื่อ t_L คือความยาวหางทางซ้ายของความคลาดเคลื่อน $\hat{\underline{\varepsilon}}$

t_R คือความยาวหางทางขวาของความคลาดเคลื่อน $\hat{\underline{\varepsilon}}$

$\hat{\varepsilon}_{P2}$ คือค่าความคลาดเคลื่อน $\hat{\underline{\varepsilon}}$ ที่ตำแหน่งเปอร์เซ็นไทล์ที่ 2

$\hat{\varepsilon}_{P25}$ คือค่าความคลาดเคลื่อน $\hat{\underline{\varepsilon}}$ ที่ตำแหน่งเปอร์เซ็นไทล์ที่ 25

$\hat{\varepsilon}_{P75}$ คือค่าความคลาดเคลื่อน $\hat{\underline{\varepsilon}}$ ที่ตำแหน่งเปอร์เซ็นไทล์ที่ 75

และ $\hat{\varepsilon}_{P98}$ คือค่าความคลาดเคลื่อน $\hat{\underline{\varepsilon}}$ ที่ตำแหน่งเปอร์เซ็นไทล์ที่ 98

(รายละเอียดของการคำนวณหาค่าเปอร์เซ็นไทล์แสดงไว้ในภาคผนวกหน้าที่ 122)

---

[1] ที่มา : Thomas W. O'Gorman, Adaptive Estimation Using Weighted Least Squares, Australia New Zealand Journal of Statistics, Northern Illinois University, DeKalb, Illinois, USA. Vol.43(3)(2001) 287-297.

ในทำนองเดียวกัน ค่าความยาวหางทางซ้ายและขวาของความคลาดเคลื่อนที่มีการแจกแจงแบบปกติ สามารถหาได้จาก

$$t_{NL} = \frac{\Phi^{-1}(0.25)-\Phi^{-1}(0.02)}{\Phi^{-1}(0.75)-\Phi^{-1}(0.25)}$$

$$= \frac{(-0.67449)-(-2.05375)}{(0.67449)-(-0.67449)} = 1.022$$

และ $$t_{NR} = \frac{\Phi^{-1}(0.98)-\Phi^{-1}(0.75)}{\Phi^{-1}(0.75)-\Phi^{-1}(0.25)}$$

$$= \frac{(2.05375)-(0.67449)}{(0.67449)-(-0.67449)} = 1.022$$

โดยที่ Φ เป็นฟังก์ชันการแจกแจงสะสมของการแจกแจงแบบปกติมาตรฐาน

สำหรับกรณี $n < 50$

สูตรของ t_L และ t_R จะปรับโดยแทนค่า $\hat{\varepsilon}_{P2}$ ด้วยค่าความคลาดเคลื่อนของ $\underline{\hat{\varepsilon}}$ ที่มีค่าน้อยที่สุด และแทนค่า $\hat{\varepsilon}_{P98}$ ด้วยค่าความคลาดเคลื่อนของ $\underline{\hat{\varepsilon}}$ ที่มีค่ามากที่สุด

สำหรับสูตรของ t_{NL} และ t_{NR} จะปรับโดยแทนค่า $\Phi^{-1}(0.02)$ ด้วย $\Phi^{-1}(h)$ เมื่อ $h = 1/(n+1)$ และแทนค่า $\Phi^{-1}(0.98)$ ด้วย $\Phi^{-1}(1-h)$

จากนั้นทำการตรวจสอบค่าผิดปกติโดยเปรียบเทียบค่าความยาวหางทางซ้ายและขวาของความคลาดเคลื่อน $\underline{\hat{\varepsilon}}$ (t_L และ t_R) กับค่าความยาวหางทางซ้ายและขวาที่ได้จากการแจกแจงแบบปกติ (t_{NL} และ t_{NR}) จะได้

$$w_R = t_{NR}/t_R \qquad \text{และ} \qquad w_L = t_{NL}/t_L$$

พิจารณาจากสูตรจะเห็นว่า ถ้าการแจกแจงของ $\underline{\hat{\varepsilon}}$ เป็นการแจกแจงแบบปกติแล้ว ค่า w_R และ w_L จะมีค่าประมาณ 1 ในขณะที่ค่า w_R และ w_L จะมีค่าน้อยกว่า 1 ถ้า $\underline{\hat{\varepsilon}}$ มีค่าผิดปกติ

จากนั้นนำค่า w_R และ w_L ที่ได้ไปกำหนดค่าถ่วงน้ำหนักให้กับ $\underline{\hat{\varepsilon}}$ ซึ่งสามารถแบ่งออกได้เป็น 4 กรณีดังนี้

1. กรณีที่ $\underline{\hat{\varepsilon}}$ มีค่าอยู่ที่ตำแหน่งเปอร์เซ็นไทล์ที่ 2 และเปอร์เซ็นไทล์ที่ 98

ค่าถ่วงน้ำหนักของ $\hat{\varepsilon}_{P2}$ มีค่าเท่ากับ w_L

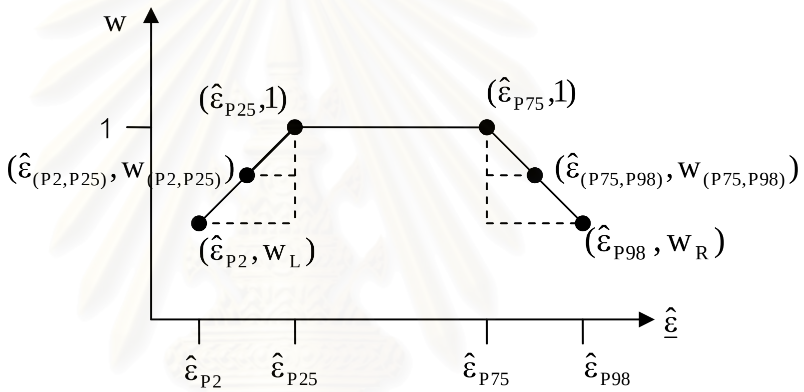
และ ค่าถ่วงน้ำหนักของ $\hat{\varepsilon}_{P98}$ มีค่าเท่ากับ w_R

2. กรณีที่ $\underline{\hat{\varepsilon}}$ มีค่าอยู่ตั้งแต่เปอร์เซ็นไทล์ที่ 25 ถึงเปอร์เซ็นไทล์ที่ 75

ค่าถ่วงน้ำหนักของ $\underline{\hat{\varepsilon}} \in [\, P25, P75 \,]$ มีค่าเท่ากับ 1

3. กรณีที่ $\underline{\hat{\varepsilon}}$ มีค่าอยู่ระหว่างเปอร์เซ็นไทล์ที่ 2 กับเปอร์เซ็นไทล์ที่ 25 และระหว่างเปอร์เซ็นไทล์ที่ 75 กับเปอร์เซ็นไทล์ที่ 98

ใช้วิธีการประมาณค่าในช่วงแบบเชิงเส้น (Linear Interpolation Method) ในการกำหนดค่าถ่วงน้ำหนักให้กับ $\underline{\hat{\varepsilon}}$



3.1 กรณีที่ $\underline{\hat{\varepsilon}}$ มีค่าอยู่ระหว่างเปอร์เซ็นไทล์ที่ 2 กับเปอร์เซ็นไทล์ที่ 25 แทนด้วย $\hat{\varepsilon}_{(P2,\, P25)}$ จะหาค่า $w_{(P2,\, P25)}$ ได้จากสมการต่อไปนี้ [โดยใช้คุณสมบัติสามเหลี่ยมคล้าย (ดูรูปภาพข้างต้นประกอบ)]

$$\frac{1 - w_{(P2,\, P25)}}{\hat{\varepsilon}_{P25} - \hat{\varepsilon}_{(P2,\, P25)}} = \frac{1 - w_L}{\hat{\varepsilon}_{P25} - \hat{\varepsilon}_{P2}}$$

$$w_{(P2,\, P25)} = 1 - \frac{1 - w_L}{\hat{\varepsilon}_{P25} - \hat{\varepsilon}_{P2}} (\hat{\varepsilon}_{P25} - \hat{\varepsilon}_{(P2,P25)})$$

เมื่อ $\hat{\varepsilon}_{(P2,\, P25)}$ คือค่าความคลาดเคลื่อน $\underline{\hat{\varepsilon}}$ ที่มีค่าอยู่ระหว่างเปอร์เซ็นไทล์ที่ 2 กับเปอร์เซ็นไทล์ที่ 25

และ $w_{(P2,\, P25)}$ คือค่าถ่วงน้ำหนักของ $\underline{\hat{\varepsilon}}$ ที่มีค่าอยู่ระหว่างเปอร์เซ็นไทล์ที่ 2 กับเปอร์เซ็นไทล์ที่ 25

ดังนั้น ค่าถ่วงน้ำหนักของ $\underline{\hat{\varepsilon}}$ $\in (\mathrm{P2, P25})$ คือ $1 - \dfrac{1 - w_L}{\hat{\varepsilon}_{P25} - \hat{\varepsilon}_{P2}} (\hat{\varepsilon}_{P25} - \hat{\varepsilon}_{(P2,P25)})$

3.2 กรณีที่ $\underline{\hat{\varepsilon}}$ มีค่าอยู่ระหว่างเปอร์เซ็นไทล์ที่ 75 กับเปอร์เซ็นไทล์ที่ 98 แทนด้วย $\hat{\varepsilon}_{(P75,\,P98)}$ จะหาค่า $w_{(P75,\,P98)}$ ได้จากสมการต่อไปนี้ [โดยใช้คุณสมบัติสามเหลี่ยมคล้าย (ดูรูปภาพข้างต้นประกอบ)]

$$\frac{1 - w_{(P75,\,P98)}}{\hat{\varepsilon}_{(P75,\,P98)} - \hat{\varepsilon}_{P75}} = \frac{1 - w_R}{\hat{\varepsilon}_{P98} - \hat{\varepsilon}_{P75}}$$

$$w_{(P75,\,P98)} = 1 - \frac{1 - w_R}{\hat{\varepsilon}_{P98} - \hat{\varepsilon}_{P75}} (\hat{\varepsilon}_{(P75,P98)} - \hat{\varepsilon}_{P75})$$

เมื่อ $\hat{\varepsilon}_{(P75,\,P98)}$ คือค่าความคลาดเคลื่อน $\underline{\hat{\varepsilon}}$ ที่มีค่าอยู่ระหว่างเปอร์เซ็นไทล์ที่ 75 กับเปอร์เซ็นไทล์ที่ 98

และ $w_{(P75,\,P98)}$ คือค่าถ่วงน้ำหนักของ $\underline{\hat{\varepsilon}}$ ที่มีค่าอยู่ระหว่างเปอร์เซ็นไทล์ที่ 75 กับเปอร์เซ็นไทล์ที่ 98

ดังนั้น ค่าถ่วงน้ำหนักของ $\underline{\hat{\varepsilon}}$ $\in (\mathrm{P75, P98})$ คือ $1 - \dfrac{1 - w_R}{\hat{\varepsilon}_{P98} - \hat{\varepsilon}_{P75}} (\hat{\varepsilon}_{(P75,P98)} - \hat{\varepsilon}_{P75})$

4. กรณีที่ $\underline{\hat{\varepsilon}}$ มีค่าอยู่ต่ำกว่าเปอร์เซ็นไทล์ที่ 2 หรือสูงกว่าเปอร์เซ็นไทล์ที่ 98

สามารถกำหนดค่าถ่วงน้ำหนักได้เช่นเดียวกับกรณีที่ $\underline{\hat{\varepsilon}}$ มีค่าอยู่ระหว่างเปอร์เซ็นไทล์ที่ 2 กับเปอร์เซ็นไทล์ที่ 25 และระหว่างเปอร์เซ็นไทล์ที่ 75 กับเปอร์เซ็นไทล์ที่ 98 โดยค่าถ่วงน้ำหนักของ $\underline{\hat{\varepsilon}}$ ที่มีค่าต่ำกว่าเปอร์เซ็นไทล์ที่ 2 จะมีค่าน้อยกว่า w_L ในขณะที่ค่าถ่วงน้ำหนักของ $\underline{\hat{\varepsilon}}$ ที่มีค่าสูงกว่าเปอร์เซ็นไทล์ที่ 98 จะมีค่าน้อยกว่า w_R

ดังนั้น ค่าถ่วงน้ำหนักของ $\underline{\hat{\varepsilon}}$ $\in [\mathrm{P1, P2})$ คือ $1 - \dfrac{1 - w_L}{\hat{\varepsilon}_{P25} - \hat{\varepsilon}_{P2}} (\hat{\varepsilon}_{P25} - \hat{\varepsilon}_{[P1,P2)})$

และ ค่าถ่วงน้ำหนักของ $\underline{\hat{\varepsilon}}$ $\in (\mathrm{P98, P100}]$ คือ $1 - \dfrac{1 - w_R}{\hat{\varepsilon}_{P98} - \hat{\varepsilon}_{P75}} (\hat{\varepsilon}_{(P98,P100]} - \hat{\varepsilon}_{P75})$

เมื่อ $\hat{\varepsilon}_{[P1,\,P2)}$ คือค่าความคลาดเคลื่อน $\underline{\hat{\varepsilon}}$ ที่มีค่าต่ำกว่าเปอร์เซ็นไทล์ที่ 2

และ $\hat{\varepsilon}_{(P98,\,P100]}$ คือค่าความคลาดเคลื่อน $\underline{\hat{\varepsilon}}$ ที่มีค่าสูงกว่าเปอร์เซ็นไทล์ที่ 98

2.3 ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ (REWLSE - Robust and Efficient Weighted Least Squares Estimator)

ในปี ค.ศ. 2002 Daniel Gervini และ Victor J. Yohai ได้เสนอตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ (Robust and Efficient Weighted Least Squares Estimator - REWLSE) ในการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้น โดยมีแนวคิดมาจากเกณฑ์ในการพิจารณาค่าผิดปกติของ Rousseeuw และ Leroy[2] ที่กำหนดให้ค่าตัดขอบ (Cut-off Value) ของความคลาดเคลื่อนมาตรฐานสัมบูรณ์มีค่าเท่ากับ 2.5 แต่ในทางปฏิบัติมีความเป็นไปได้ที่ความคลาดเคลื่อนมาตรฐานสัมบูรณ์ที่มีค่ามากกว่า 2.5 อาจเป็นข้อมูลที่ปกติได้ ดังนั้น Daniel Gervini และ Victor J. Yohai จึงมีแนวคิดในการกำหนดเกณฑ์ในการพิจารณาค่าผิดปกติที่มีความแกร่งและมีประสิทธิภาพ โดยมีหลักในการประมาณคือ การใช้ฟังก์ชันการแจกแจงเชิงสังเกต (Empirical Distribution Function) ในการหาค่าตัดขอบ (Cut-off Value) ที่มีประสิทธิภาพในการตัดข้อมูลที่ผิดปกติออก แล้วใช้ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักในการประมาณค่า $\underline{\beta}$ คือ $\hat{\underline{\beta}}$ ที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนมีค่าต่ำสุด

การหาค่าตัดขอบที่มีประสิทธิภาพในการตัดข้อมูลที่ผิดปกติของตัวประมาณนี้ จะใช้ฟังก์ชันการแจกแจงเชิงสังเกตของความคลาดเคลื่อนมาตรฐาน (Standardized Residuals) ซึ่งหาได้จาก $\underline{X}$, $\underline{Y}$, $\hat{\underline{\beta}}$ และค่าประมาณแสดงสเกลของตัวประมาณดั้งเดิม ดังนี้

$$\underline{R} = \frac{\underline{Y} - \underline{X}\hat{\underline{\beta}}_{OLD}}{S}$$

กำหนดให้

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \hat{\underline{\beta}}_{OLD} = \begin{bmatrix} \hat{\beta}_{OLD\_0} \\ \hat{\beta}_{OLD\_1} \\ \vdots \\ \hat{\beta}_{OLD\_p} \end{bmatrix} \qquad \underline{R} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}$$

และ $$\underline{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ 1 & X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{pn} \end{bmatrix}$$

---

[2] ที่มา : Rousseeuw, P.J. and Leroy, A.(1987). Robust Regression and Outlier Detection. Wiley, New York. p.77

เมื่อ $\underline{Y}$ แทนเวกเตอร์ของตัวแปรตาม

$\underline{\hat{\beta}}_{OLD}$ แทนเวกเตอร์ค่าประมาณของสัมประสิทธิ์การถดถอยที่ได้จากตัวประมาณดั้งเดิม

$\underline{R}$ แทนเวกเตอร์ค่าสังเกตของความคลาดเคลื่อนมาตรฐาน

r_i แทนความคลาดเคลื่อนมาตรฐานตัวที่ i , $i = 1, 2, \ldots, n$

$\underline{X}$ แทนเมทริกซ์ของตัวแปรอิสระ

และ S แทนค่าประมาณแสดงสเกลที่คำนวณได้จากตัวประมาณดั้งเดิม

โดยการศึกษาในครั้งนี้ใช้ตัวประมาณมัธยฐานกำลังสองน้อยที่สุด (LMSE – Least Median Squares Estimator) เป็นตัวประมาณดั้งเดิม

จากนั้นใช้ฟังก์ชันการแจกแจงเชิงสังเกต (Empirical Distribution Function) ของค่าสัมบูรณ์ของความคลาดเคลื่อนมาตรฐาน $|r_i|$ ในการหาค่าตัดขอบที่มีประสิทธิภาพในการตัดข้อมูลที่ผิดปกติออก

เมื่อได้ค่าตัดขอบแล้ว ทำการกำหนดค่าถ่วงน้ำหนักให้กับค่าสังเกต โดยที่

$$w_i = w(|r_i| / t_n)$$

ซึ่งกำหนดให้ $w(|r_i| / t_n) = I(|r_i| / t_n < 1)$

นั่นคือ

$$w_i = \begin{cases} 1 & , |r_i| / t_n < 1 \\ 0 & , |r_i| / t_n \geq 1 \end{cases}$$

เมื่อ w_i คือค่าถ่วงน้ำหนักของค่าสังเกตตัวที่ i ; $i = 1, 2, \ldots, n$

$|r_i|$ ค่าสัมบูรณ์ของความคลาดเคลื่อนมาตรฐานตัวที่ i ; $i = 1, 2, \ldots, n$

และ t_n คือค่าตัดขอบ

เมื่อได้ค่าถ่วงน้ำหนักแล้ว จากนั้นทำการสร้างเมทริกซ์ค่าถ่วงน้ำหนัก $\underline{W}$ ดังนี้

$$\underline{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix}$$

จากตัวแบบการถดถอยเชิงเส้นพหุคูณ $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$ กำหนดให้

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \qquad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

และ $$\underline{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ 1 & X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{pn} \end{bmatrix}$$

เมื่อ $\underline{Y}$ แทนเวกเตอร์ของตัวแปรตาม

$\underline{\beta}$ แทนเวกเตอร์ของสัมประสิทธิ์การถดถอย

$\underline{\varepsilon}$ แทนเวกเตอร์ของความคลาดเคลื่อน

และ $\underline{X}$ แทนเมทริกซ์ของตัวแปรอิสระ

จากนั้นหาตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (WLSE – Weighted Least Squares Estimator) ของ $\underline{\beta}$ คือ $\hat{\underline{\beta}}_{WLS}$ ที่ทำให้ผลบวกของค่าคลาดเคลื่อนกำลังสองแบบถ่วงน้ำหนัก $(\underline{Y} - \underline{X}\underline{\beta})^T W(Y - \underline{X}\underline{\beta})$ มีค่าต่ำที่สุด

ให้ $$L(\underline{\beta}) = (\underline{Y} - \underline{X}\underline{\beta})^T \underline{W}(\underline{Y} - \underline{X}\underline{\beta})$$

$$= \underline{Y}^T \underline{W}\underline{Y} - \underline{\beta}^T \underline{X}^T \underline{W}\underline{Y} - \underline{Y}^T \underline{W}\underline{X}\underline{\beta} + \underline{\beta}^T \underline{X}^T \underline{W}\underline{X}\underline{\beta}$$

$$= \underline{Y}^T \underline{W}\underline{Y} - 2\underline{\beta}^T \underline{X}^T \underline{W}\underline{Y} + \underline{\beta}^T \underline{X}^T \underline{W}\underline{X}\underline{\beta}$$

จากนั้นแทนค่า $\underline{\beta}$ ด้วย $\hat{\underline{\beta}}$

จากสมการ $$\frac{\partial}{\partial \underline{\beta}} (\underline{Y}^T \underline{W}\underline{Y} - 2\underline{\beta}^T \underline{X}^T \underline{W}\underline{Y} + \underline{\beta}^T \underline{X}^T \underline{W}\underline{X}\underline{\beta}) = \underline{0}$$

จะได้ว่า $$\frac{\partial}{\partial \hat{\underline{\beta}}} (\underline{Y}^T \underline{W}\underline{Y} - 2\hat{\underline{\beta}}^T \underline{X}^T \underline{W}\underline{Y} + \hat{\underline{\beta}}^T \underline{X}^T \underline{W}\underline{X}\hat{\underline{\beta}}) = \underline{0}$$

$$-2\underline{X}^T\underline{W}\underline{Y}+2\underline{X}^T\underline{W}\underline{X}\hat{\underline{\beta}} \quad = \quad \underline{0}$$

ดังนั้น
$$\hat{\underline{\beta}}_{WLS} = \begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \vdots \\ \hat{\beta}_p\end{bmatrix} = (\underline{X}^T\underline{W}\underline{X})^{-1}(\underline{X}^T\underline{W}\underline{Y})$$

โดยที่ $\underline{X}^T\underline{W}\underline{X}$ ต้องไม่เป็นเมทริกซ์เอกฐาน (Singular Matrix)

ดังนั้นตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพของ $\underline{\beta}_{REWLS}$ คือ

$$\hat{\underline{\beta}}_{REWLSE} = \begin{cases}(\underline{X}^T\underline{W}\underline{X})^{-1}\underline{X}^T\underline{W}\underline{Y} & , \quad S>0\\ \hat{\underline{\beta}}_{LMSE} & , \quad S=0\end{cases}$$

เมื่อ $\hat{\underline{\beta}}_{LMSE}$ คือค่าประมาณสัมประสิทธิ์การถอยของตัวประมาณ LMSE

S คือค่าประมาณแสดงสเกลที่คำนวณได้จากตัวประมาณ LMSE

และ $\hat{\underline{\beta}}_{REWLSE}$ คือค่าประมาณสัมประสิทธิ์การถอยของตัวประมาณ REWLS

ในส่วนต่อไปจะเป็นวิธีการหาค่าตัดขอบ (Cut-off Value) ซึ่งมีรายละเอียดดังนี้

การหาค่าตัดขอบ (Cut-off Value)

พิจารณาฟังก์ชันการแจกแจงเชิงสังเกตของค่าสัมบูรณ์ของความคลาดเคลื่อนมาตรฐาน $|r_i|$ ในการหาฟังก์ชันความน่าจะเป็นสะสมของเหตุการณ์ที่ $|r_i| \le t$ โดยกำหนดให้

$$F_n^+(t) = \frac{1}{n}\sum_{i=1}^{n} I(|r_i| \le t)$$

เมื่อ $I(|r_i| \le t)$ คือฟังก์ชันดัชนี (Indicator Function) ซึ่ง

$$I(|r_i| \le t) = \begin{cases}1 & , |r_i| \le t\\ 0 & , |r_i| > t\end{cases}$$

โดยที่ t คือค่า $|r_i|$ ค่าใดค่าหนึ่ง

ทำการเรียงลำดับค่า $|r_i|$ จากค่าน้อยไปมาก แล้วกำหนดให้ $|r|_{(i)}$ คือค่าสัมบูรณ์ของความคลาดเคลื่อนมาตรฐานอันดับที่ i

จากนั้นใช้เกณฑ์พิจารณาค่าผิดปกติของ Rousseeuw และ Leroy ในการพิจารณาค่าผิดปกติในเบื้องต้น โดยทำการเปรียบเทียบค่า $|r|_{(i)}$ กับ η ซึ่งกำหนดให้มีค่าเท่ากับ 2.5 ดังนี้

$$i_0 \quad = \quad \max\{i : |r|_{(i)} < \eta\}, \quad i = 1, 2, \dots, n$$

โดยค่า i_0 ที่ได้คือจำนวน $|r_i|$ ขั้นต่ำที่เป็นข้อมูลที่ปกติ และ $|r|_{(i_0+1)}, |r|_{(i_0+2)}, \dots, |r|_{(n)}$ คือข้อมูลที่คาดว่าจะเป็นข้อมูลที่ผิดปกติ

ตัวอย่างเช่น มีค่าความคลาดเคลื่อนมาตรฐาน (r_i) จำนวน 10 ค่า ดังนี้

-1.6 0.7 3.4 -1.5 1.3 1.2 -0.2 2.9 2.1 -0.8

ซึ่งจะได้ค่าสัมบูรณ์ของความคลาดเคลื่อนมาตรฐาน $|r_i|$ ดังนี้

1.6 0.7 3.4 1.5 1.3 1.2 0.2 2.9 2.1 0.8

พิจารณาค่าสัมบูรณ์ของความคลาดเคลื่อนมาตรฐานแล้วเรียงลำดับค่าจากน้อยไปมาก จะได้ $|r|_{(i)}$ คือ

0.2 0.7 0.8 1.2 1.3 1.5 1.6 2.1 2.9 3.4

เปรียบเทียบค่า $|r|_{(i)}$ กับ η ซึ่งกำหนดให้มีค่าเท่ากับ 2.5 จะได้ $i_0 =$ 8 แสดงว่ามีจำนวน $|r_i|$ อย่างน้อย 8 ค่าที่เป็นค่าปกติ และ $|r|_{(9)} = 2.9$ และ $|r|_{(10)} = 3.4$ เป็นค่าที่คาดว่าจะเป็นค่าที่ผิดปกติ

จากนั้นพิจารณาหาค่าสัดส่วนการปลอมปนของค่าผิดปกติ จากผลต่างฟังก์ชันการแจกแจงสะสมของการแจกแจงแบบปกติมาตรฐานของ $|r|_{(i_0+1)}, |r|_{(i_0+2)}, \dots, |r|_{(n)}$ กับฟังก์ชันความน่าจะเป็นสะสมของ $|r|_{(i_0)}, |r|_{(i_0+1)}, \dots, |r|_{(n)}$ ดังนี้

$$d_{(i)} \quad = \quad \{F^{+}(|r|_{(i)}) - \frac{j}{n}\}^{+} \qquad , \ i > i_0 \ ; \ j = i - 1$$

เมื่อ $d_{(i)}$ คือค่าสัดส่วนการปลอมปนของค่าผิดปกติเมื่อ $|r|_{(i_0+1)}, |r|_{(i_0+2)}, \dots, |r|_{(n)}$ เป็นค่าที่คาดว่าจะเป็นค่าผิดปกติที่เป็นค่า Cut-off

$F^{+}(|r|_{(i)})$ คือฟังก์ชันความน่าจะเป็นสะสมของเหตุการณ์ที่ $|r|_{(i_0+1)}, |r|_{(i_0+2)}, \dots, |r|_{(n)}$ เป็นค่าที่คาดว่าจะเป็นค่าผิดปกติที่เป็นค่า Cut-off

และ $\frac{j}{n}$ คือฟังก์ชันความน่าจะเป็นสะสมของเหตุการณ์ที่ $|r|_{(j)}$ เป็นค่าที่คาดว่าจะเป็นค่าปกติ เมื่อ $|r|_{(i)}, i > i_0$ เป็นค่าที่คาดว่าจะเป็นค่าผิดปกติ โดยที่ $\frac{j}{n}$ หาได้จากฟังก์ชันการแจกแจงเชิงสังเกต $F_n^{+}(|r|_{(j)})$

ดังนั้นค่าสัดส่วนการปลอมปนของค่าผิดปกติ (d_n) คือค่าสัดส่วนการปลอมปนของค่าผิดปกติ $(d_{(i)})$ ที่มีค่าสูงสุด นั่นคือ

$$d_n = \max_{i>i_0}\{d_{(i)}\}^+$$

<u>หมายเหตุ</u> $\{.\}^+$ หมายถึงพิจารณาเฉพาะค่าของฟังก์ชันที่มีค่าเป็นค่าบวก (Positive Part) สำหรับกรณีที่มีค่าเป็นค่าลบจะกำหนดให้มีค่าเท่ากับ 0

เพราะฉะนั้น จำนวนข้อมูลที่เป็นค่าผิดปกติมีค่าเท่ากับ $\lfloor nd_n \rfloor$ (คือจำนวนเต็มที่มีค่ามากที่สุดที่น้อยกว่าหรือเท่ากับ nd_n)

ดังนั้นค่า Cut-off ที่เหมาะสม คือ

$$t_n = |r|_{(n-\lfloor nd_n \rfloor)}$$

เมื่อ $|r|_{(n-\lfloor nd_n \rfloor)}$ คือค่าสัมบูรณ์ของความคลาดเคลื่อนมาตรฐานอันดับที่ $n-\lfloor nd_n \rfloor$

และ t_n คือค่า Cut-off ที่ต้องการ

ในส่วนต่อไปจะเป็นรายละเอียดของตัวประมาณมัธยฐานกำลังสองน้อยที่สุด (LMSE – Least Median Squares Estimator) และค่าประมาณแสดงสเกลของตัวประมาณ LMSE ซึ่งมีรายละเอียดดังนี้

ตัวประมาณมัธยฐานกำลังสองน้อยที่สุด (LMSE – Least Median Squares Estimator)[3]

หลักในการประมาณของตัวประมาณมัธยฐานกำลังสองน้อยที่สุด (LMSE) คือ การใช้กลุ่มตัวอย่างย่อย (Subsampling Algorithm) ในการหาค่าประมาณของ $\underline{\beta}$ คือ $\underline{\hat{\beta}}$ ที่ทำให้ค่ามัธยฐานของความคลาดเคลื่อนกำลังสองมีค่าต่ำสุด

จากตัวแบบการถดถอยเชิงเส้นพหุคูณ $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$ กำหนดให้

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \qquad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

---

[3] ที่มา : Rousseeuw, P.J. and Leroy, A.(1987). Robust Regression and Outlier Detection. Wiley, New York. p.197

และ $\begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ 1 & X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{pn} \end{bmatrix}$

เมื่อ $\underline{Y}$ แทนเวกเตอร์ของตัวแปรตาม

$\underline{\beta}$ แทนเวกเตอร์ของสัมประสิทธิ์การถดถอย

$\underline{\varepsilon}$ แทนเวกเตอร์ของความคลาดเคลื่อน

และ $\underline{X}$ แทนเมทริกซ์ของตัวแปรอิสระ

ทำการสุ่มกลุ่มตัวอย่างย่อยของตัวแปรอิสระ $\underline{X}$ และตัวแปรตาม $\underline{Y}$ ชุดเดียวกันจำนวน 1000 ชุดๆ ละ p+1 ค่า (เท่ากับจำนวนพารามิเตอร์) ดังนี้

กำหนด

$$\underline{X}_j = \begin{bmatrix} 1 & X_{j[1,(1)]} & X_{j[2,(1)]} & \cdots & X_{j[p,(1)]} \\ 1 & X_{j[1,(2)]} & X_{j[2,(2)]} & \cdots & X_{j[p,(2)]} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{j[1,(p+1)]} & X_{j[2,(p+1)]} & \cdots & X_{j[p,(p+1)]} \end{bmatrix}$$

และ $$\underline{Y}_j = \begin{bmatrix} Y_{j(1)} \\ Y_{j(2)} \\ \vdots \\ Y_{j(p+1)} \end{bmatrix}$$

เมื่อ $X_{j[k,(m)]}$ คือค่าสังเกตชุดที่ j ของตัวแปรอิสระตัวที่ k และสุ่มมาเป็นอันดับที่ m

$Y_{j(m)}$ คือค่าสังเกตชุดที่ j ของตัวแปรตามที่สุ่มมาเป็นอันดับที่ m

โดยที่ $j = 1, 2, \ldots, 1000$; $k = 1, 2, \ldots, p$ และ $m=1, 2, \ldots, p+1$

ใช้ตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (OLSE) ในการหาตัวประมาณ $\hat{\underline{\beta}}_j$ ของ $\underline{\beta}_j$ ในตัวแบบการถดถอยเชิงเส้นที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนแต่ละชุดมีค่าต่ำสุด

ได้ $$\hat{\underline{\beta}}_j = \begin{bmatrix} \hat{\beta}_{j0} \\ \hat{\beta}_{j1} \\ \vdots \\ \hat{\beta}_{jp} \end{bmatrix}$$

เมื่อ $\underline{\hat{\beta}}_j$ แทนเวกเตอร์ค่าประมาณสัมประสิทธิ์การถดถอยของกลุ่มตัวอย่างย่อยที่ j

จากนั้นนำค่าประมาณสัมประสิทธิ์การถดถอย $\underline{\hat{\beta}}_j$ ของ $\underline{\beta}_j$ ที่ได้ทั้ง1000 ชุดมาใช้ในการหาค่าความคลาดเคลื่อน $\underline{\hat{\varepsilon}}_j$ ในตัวแบบการถดถอยเชิงเส้นพหุคูณ $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$ ได้

$$\underline{\hat{\varepsilon}}_j = \underline{Y} - \underline{X}\underline{\hat{\beta}}_j$$

โดยที่ $$\underline{\hat{\varepsilon}}_j = \begin{bmatrix} \hat{\varepsilon}_{j1} \\ \hat{\varepsilon}_{j2} \\ \vdots \\ \hat{\varepsilon}_{jn} \end{bmatrix}$$

และ $\hat{\varepsilon}_{ji}$ คือความคลาดเคลื่อนตัวที่ i ที่ได้จากกลุ่มตัวอย่างย่อยกลุ่มที่ j ซึ่งมีค่าเท่ากับ

$$\hat{\varepsilon}_{ji} = Y_i - (\hat{\beta}_{j0} + \hat{\beta}_{j1}X_{1i} + \hat{\beta}_{j2}X_{2i} + ... + \hat{\beta}_{jp}X_{pi}) \text{ , } i = 1, 2, \dots , n \text{ ; } j = 1, 2, \dots , 1000$$

จากนั้นพิจารณาค่ามัธยฐานของความคลาดเคลื่อนกำลังสอง ($\hat{\varepsilon}_{ji}^2$) ของกลุ่มตัวอย่างทั้ง 1000 ชุด

$$med_j = \text{median } \hat{\varepsilon}_{ji}^2, \quad i = 1, 2, \dots , n ; \quad j = 1, 2, \dots , 1000$$

เมื่อ med_j คือค่ามัธยฐานของความคลาดเคลื่อนกำลังสอง($\hat{\varepsilon}_{ji}^2$) ของกลุ่มตัวอย่างย่อยที่ j

เรียงลำดับค่า med_j จากน้อยไปมาก โดยค่าประมาณ $\underline{\hat{\beta}}_j$ ชุดใดที่ทำให้ med_j มีค่าน้อยที่สุดจะเป็นค่าประมาณของ $\underline{\beta}$ คือ $\underline{\hat{\beta}}$ ที่ทำให้ค่ามัธยฐานของความคลาดเคลื่อนกำลังสองมีค่าต่ำสุด นั่นคือ

$$\begin{aligned} \underline{\hat{\beta}} &= \min \underset{i,j}{med} \, (Y_i - (\hat{\beta}_{j0} + \hat{\beta}_{j1}X_{1i} + \hat{\beta}_{j2}X_{2i} + ... + \hat{\beta}_{jp}X_{pi})^2 \\ &= \min \underset{i,j}{med} \, \hat{\varepsilon}_{ji}^2 \end{aligned}$$

ค่าประมาณแสดงสเกลของตัวประมาณ LMSE

จากตัวแบบการถดถอยเชิงเส้นพหุคูณ $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$ กำหนดให้

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \qquad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

และ $$\begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ 1 & X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{pn} \end{bmatrix}$$

เมื่อ $\underline{Y}$ แทนเวกเตอร์ของตัวแปรตาม

$\underline{\beta}$ แทนเวกเตอร์ของสัมประสิทธิ์การถดถอย

$\underline{\varepsilon}$ แทนเวกเตอร์ของความคลาดเคลื่อน

และ $\underline{X}$ แทนเมทริกซ์ของตัวแปรอิสระ

ใช้ค่าประมาณของ $\underline{\beta}$ คือ $\underline{\hat{\beta}}_{LMSE}$ ซึ่งได้จากตัวประมาณ LMSE มาใช้ในการหาค่าความคลาดเคลื่อน $\underline{\hat{\varepsilon}}$ ดังนี้

$$\underline{\hat{\varepsilon}} \quad = \quad \underline{Y} - \underline{X}\underline{\hat{\beta}}_{LMSE}$$

จากนั้นนำค่า $\underline{\hat{\varepsilon}}$ ที่ได้ไปคำนวณหาค่าประมาณแสดงสเกลได้ดังนี้

$$S = \text{median}\ (|\hat{\varepsilon}_1|, |\hat{\varepsilon}_2|, \ldots, |\hat{\varepsilon}_n|) / \Phi^{-1}(3/4)$$

โดยที่ Φ เป็นฟังก์ชันการแจกแจงสะสมของการแจกแจงแบบปกติมาตรฐาน

บทที่ 3

วิธีดำเนินการวิจัย

การศึกษาในครั้งนี้เป็นการศึกษาประสิทธิภาพของตัวประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นเมื่อมีค่าผิดปกติในตัวแปรตาม โดยตัวประมาณที่นำมาใช้ในการศึกษาในครั้งนี้ได้แก่

1. ตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (OLSE - Ordinary Least Squares Estimator)
2. ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ (AWLSE - Adaptive Weighted Least Squares Estimator)
3. ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ (REWLSE - Robust and Efficient Weighted Least Squares Estimator)

การดำเนินการวิจัยทำโดยการนำตัวประมาณที่ต้องการศึกษามาทำการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นเมื่อมีค่าผิดปกติในตัวแปรตาม ภายใต้ข้อมูลที่มีการแจกแจงที่มีค่าผิดปกติอยู่ในระดับไม่รุนแรงและรุนแรง สัดส่วนการปลอมปนเท่ากับ 0.05, 0.10, 0.15 และ 0.20 โดยขนาดตัวอย่างที่ใช้ในการศึกษาคือ 20, 30, 40, 50, 60, 70, 80, 90 และ 100

การวิจัยในครั้งนี้ใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Method) โดยใช้โปรแกรมภาษาฟอร์แทรน (Fortran) ในการจำลองซ้ำ 500 รอบในแต่ละสถานการณ์การทดลอง ในการเปรียบเทียบตัวประมาณจะพิจารณาจากค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE - Average Mean Square Error) ของตัวประมาณทั้ง 3 ตัว โดยตัวประมาณที่มีค่า AMSE ต่ำที่สุดจะถือว่าตัวประมาณนั้นมีประสิทธิภาพในการประมาณสูงที่สุดในกลุ่มตัวประมาณที่เปรียบเทียบ

ในบทนี้จะกล่าวถึงรายละเอียดของการดำเนินการวิจัย ตามลำดับ ส่วนรายละเอียดของโปรแกรมที่ใช้ในการวิจัยจะแสดงไว้ในภาคผนวก

3.1 ข้อกำหนดของการทดลอง

ในการวิจัยครั้งนี้กำหนดสถานการณ์ต่าง ๆ ที่ต้องการศึกษาดังนี้

1. ในกรณีตัวแปรตามมีค่าปกติ ผู้วิจัยสนใจศึกษาเมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติ โดยมีพารามิเตอร์ $\mu = 0$ และ $\sigma^2 = 10$

2. ขนาดตัวอย่างที่ทำการศึกษาเท่ากับ 20, 30, 40, 50, 60, 70, 80, 90 และ 100

3. ในกรณีตัวแปรตามมีค่าผิดปกติ ผู้วิจัยจึงกำหนดการมีค่าผิดปกติ ดังนี้

3.1 กรณีที่ความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน ผู้วิจัยสนใจศึกษาเมื่อมีระดับค่าผิดปกติ 2 ระดับ คือ ระดับไม่รุนแรง และระดับรุนแรง โดยกำหนดให้สเกลแฟกเตอร์มีค่าเท่ากับ 3 สำหรับข้อมูลที่มีค่าผิดปกติในระดับไม่รุนแรง และสเกลแฟกเตอร์เท่ากับ 12 สำหรับข้อมูลที่มีค่าผิดปกติในระดับรุนแรง ซึ่งแต่ละระดับมีสัดส่วนการปลอมปนเท่ากับ 0.05, 0.10, 0.15 และ 0.20

3.2 กรณีที่ความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปนแบบลาปลาซ ผู้วิจัยสนใจศึกษาเมื่อมีระดับค่าผิดปกติ 2 ระดับ คือ ระดับไม่รุนแรง และระดับรุนแรง โดยกำหนดพารามิเตอร์ $\beta = 8$ สำหรับข้อมูลที่มีค่าผิดปกติในระดับไม่รุนแรง และพารามิเตอร์ $\beta = 25$ สำหรับข้อมูลที่มีค่าผิดปกติในระดับรุนแรง ซึ่งแต่ละระดับมีสัดส่วนการปลอมปนเท่ากับ 0.05, 0.10, 0.15 และ 0.20

3.2 ขั้นตอนการดำเนินการวิจัย

ขั้นตอนดำเนินการวิจัยมีดังนี้

1. สร้างข้อมูลให้มีการแจกแจงตามลักษณะที่ต้องการศึกษา และกำหนดขนาดตัวอย่างตามที่ต้องการศึกษา

2. ประมาณค่าพารามิเตอร์ที่ต้องการศึกษาโดยใช้ตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ และตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ

3. คำนวณค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณค่า

4. เปรียบเทียบค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณค่าในแต่ละสถานการณ์ของการทดลอง

5. สรุปผลการวิจัยในแต่ละสถานการณ์

3.2.1 การสร้างตัวแปรอิสระ ความคลาดเคลื่อน และตัวแปรตาม

การสร้างตัวแปรอิสระ

การศึกษาในครั้งนี้กำหนดให้ตัวแปรอิสระ X_1 และ X_2 มีการแจกแจงแบบปกติที่มีฟังก์ชันการแจกแจงอยู่ในรูป

$$f(\varepsilon) \quad = \frac{1}{\sqrt{2\pi}\sigma}\exp\{\frac{-(\varepsilon-\mu)^2}{2\sigma^2}\}$$

โดยกำหนดให้ ตัวแปรอิสระ X_1 มีพารามิเตอร์ $\mu = 20$ และ $\sigma^2 = 10$

ตัวแปรอิสระ X_2 มีพารามิเตอร์ $\mu = 30$ และ $\sigma^2 = 25$

การสร้างความคลาดเคลื่อนที่มีการแจกแจงแบบปกติปลอมปน

การสร้างความคลาดเคลื่อนให้มีลักษณะการแจกแจงตามที่ต้องการศึกษานั้นใช้โปรแกรมภาษาฟอร์แทน 77 กับ PC Computer โดยการสร้างลักษณะการแจกแจงแบบปกติปลอมปนที่มีค่าเฉลี่ยและความแปรปรวนตามที่กำหนด จะใช้วิธีของ Ramsay (1977) ได้เสนอไว้โดยพิจารณาการแจกแจง ซึ่งแปลงมาจากการแจกแจงแบบปกติที่มีฟังก์ชันการแจกแจงอยู่ในรูป

$$f(\varepsilon) = (1-p)\, N(\mu,\sigma^2) + p\, N(\mu, C^2\sigma^2)$$

หมายความว่า ตัวแปรสุ่ม ε มาจากการแจกแจง $N(\mu,\sigma^2)$ ด้วยความน่าจะเป็น $(1-p)$ และจากการแจกแจง $N(\mu,C^2\sigma^2)$ ด้วยความน่าจะเป็น p โดยที่

μ และ σ^2 เป็นค่าเฉลี่ยและความแปรปรวนของการแจกแจงปกติ

p และ C เป็นค่าที่กำหนดสัดส่วนการปลอมปนและสเกลแฟกเตอร์

ในการวิจัยครั้งนี้ กำหนดให้ ความคลาดเคลื่อนมีการแจกแจงปกติ โดยมีพารามิเตอร์ $\mu = 0$ และ $\sigma^2 = 10$ และกำหนดระดับค่าผิดปกติ 2 ระดับ คือ ระดับไม่รุนแรง (C=3) และระดับรุนแรง (C=12) ตามเกณฑ์การกำหนดขนาดค่าผิดปกติด้วย Box Plot โดยแต่ละระดับจะกำหนดให้มีสัดส่วนการปลอมปน (p) เท่ากับ 0.05, 0.10, 0.15 และ 0.20

การสร้างความคลาดเคลื่อนที่มีการแจกแจงแบบปกติปลอมปนแบบลาปลาซ

การแจกแจงแบบปกติปลอมปนแบบลาปลาซสามารถสร้างได้โดยแปลงมาจากการแจกแจงแบบปกติที่มีฟังก์ชันการแจกแจงอยู่ในรูป

$$f(\varepsilon) \quad = \quad (1-p).N(\mu,\sigma^2) + p.L(\theta,\beta)$$

หมายความว่า ตัวแปรสุ่ม ε มาจากการแจกแจง $N(\mu,\sigma^2)$ ด้วยความน่าจะเป็น $(1-p)$ และจากการแจกแจง $L(\theta,\beta)$ ด้วยความน่าจะเป็น p โดยที่

μ และ σ^2 เป็นค่าเฉลี่ยและความแปรปรวนของการแจกแจงปกติ

θ และ β เป็นพารามิเตอร์ของการแจกแจงแบบลาปลาซ

p เป็นค่าที่กำหนดสัดส่วนการปลอมปน

ในการวิจัยครั้งนี้ กำหนดให้การแจกแจงปกติมีพารามิเตอร์ $\mu = 0$ และ $\sigma^2 = 10$ และการแจกแจงแบบลาปลาซมีพารามิเตอร์ $\theta = 0,\ \beta = 8$ เมื่อข้อมูลมีค่าผิดปกติในระดับไม่รุนแรง และ $\beta = 25$ สำหรับข้อมูลที่มีค่าผิดปกติในระดับรุนแรง ตามเกณฑ์การกำหนดขนาดค่าผิดปกติด้วย Box Plot โดยแต่ละระดับจะกำหนดให้มีสัดส่วนการปลอมปน (p) เท่ากับ 0.05, 0.10, 0.15 และ 0.20

การสร้างตัวแปรอิสระ และความคลาดเคลื่อนให้มีค่าผิดปกติจะต้องอาศัยการสร้างลักษณะการแจกแจงแบบปกติและลักษณะการแจกแจงแบบลาปลาซ ซึ่งต้องใช้เลขสุ่ม (random number) ซึ่งมีการแจกแจงเอกรูป (uniform distribution) ในช่วง [0,1] เป็นองค์ประกอบหลัก

การสร้างเลขสุ่มที่เป็นอิสระกันและมีการแจกแจงเอกรูปในช่วง [0,1]*

วิธีการคณิตศาสตร์ในการจำลองเลขสุ่ม(เทียม) มีหลายวิธีการ สำหรับวิธีการที่ได้รับความนิยมใช้กันมากในปัจจุบัน คือ วิธีสมภาค (Congruential Method) ซึ่งมีสูตรหรือตัวแบบหนึ่งที่ใช้กันมาก คือ

$$X_i \quad = \quad (c + aX_{i-1}) \bmod m\ , \qquad i = 1, 2, \ldots$$

โดยที่ค่า c , a และ m เป็นค่าคงที่จำนวนเต็มค่าไม่เป็นลบ และความหมายของตัวแบบ คือ X_i เป็นเศษเหลือที่เป็นจำนวนเต็มที่ได้จากการหาร $(c + aX_{i-1})$ ด้วย m นั่นคือ $X_i = c + aX_{i-1} - mk_i$ ซึ่ง $k_i = \lfloor (c + aX_{i-1})/m \rfloor$ (หมายถึง จำนวนเต็มใหญ่ที่สุดที่น้อยกว่าหรือเท่ากับผลหาร $(c + aX_{i-1})/m$) ดังนั้นค่าเป็นไปได้ของ X_i คือ 0 , 1 , … , m-1 และก่อนที่จะ

---

* ที่มา : มานพ วราภักดิ์, การจำลองเบื้องต้น (กรุงเทพฯ: ศูนย์ผลิตตำราเรียนสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2547), หน้า 43

ได้ค่าของ $X_1, X_2, \ldots$ ต้องกำหนดค่าของ c , a , m และ X_0 เราเรียก X_0 ว่า ซีด (seed) หรือค่าเริ่มต้น (starting value) จาก X_i ที่ได้จากการคำนวณนำมาหาค่า R_i ซึ่ง

$$R_i = \frac{X_i}{m} \quad , i = 1, 2, \ldots$$

จะได้ R_i มีค่าอยู่ในช่วง [0,1) เรียก $R_1, R_2, \ldots$ ว่า เลขสุ่มเทียม หรือ เลขสุ่มคล้าย

ตัวแบบจำลองสมภาคแบบผลคูณที่ใช้กันมากตัวแบบหนึ่ง ซึ่งได้ผ่านการตรวจสอบคุณสมบัติแล้วหลายประการ คือ กำหนด c = 0, m = 2^{31}-1 = 2147483647, a = 7^5 = 16807 และ X_0 เป็นจำนวนเต็มบวกที่เป็นเลขคี่ไม่เกิน m ฟังก์ชันการจำลองเลขสุ่มที่เป็นอิสระกันและมีการแจกแจงเอกรูปในช่วง [0,1] คือ subroutine random

การจำลองตัวแปรสุ่มปกติด้วยวิธีบอกซ์-มูลเลอร์*

George E.P.Box และ Mervin E.Muller (1958) ได้คิดค้นวิธีการจำลองตัวแปรสุ่มปกติมาตรฐาน Z_1 และ Z_2 ที่เป็นอิสระกัน ได้ตัวแบบการจำลอง

$$Z_1 = \sqrt{-2\ln R_1}\cos(2\pi R_2)$$
$$Z_2 = \sqrt{-2\ln R_1}\sin(2\pi R_2) \qquad \ldots\ldots\ldots\ldots(1)$$

โดยที่ $R_1, R_2 \sim U(0,1)$ และเป็นอิสระกัน ซึ่ง R_1 และ R_2 เป็นเลขสุ่มที่สร้างจาก subroutine random (รายละเอียดของการจำลองตัวแปรสุ่มแบบปกติด้วยวิธีบอกซ์-มูลเลอร์แสดงไว้ในภาคผนวกหน้าที่ 120)

จากตัวแบบที่ (1) จะเห็นว่าต้องจำลองเลขสุ่มสองตัว ทำให้เราได้ค่าของตัวแปรสุ่ม $N(0,1)$ สองค่าที่เป็นอิสระกัน ซึ่งในทางปฏิบัติสามารถเลือกใช้เฉพาะสูตรใดสูตรหนึ่งก็ได้

เมื่อเราได้เลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานแล้ว จากนั้นจึงแปลงค่าเลขสุ่มดังกล่าวให้มีการแจกแจงแบบปกติที่มีค่าพารามิเตอร์ μ และ σ ที่ต้องการ โดยใช้สมการ

$$NORMAL_1 = \mu + \sigma Z_1$$

หรือ

$$NORMAL_2 = \mu + \sigma Z_2$$

รายละเอียดของโปรแกรมแสดงไว้ใน subroutine normal

การจำลองตัวแปรสุ่มลาปลาซ*

ตัวแปรสุ่ม X มีการแจกแจงลาปลาซ (Laplace Distribution) หรือการแจกแจงแบบเลขชี้กำลังสองด้าน (Double-Exponential Distribution) ด้วยพารามิเตอร์ θ และ λ ซึ่งเขียนแทนด้วย $X \sim La(\theta,\lambda)$ และมีฟังก์ชันความหนาแน่นดังนี้

---

* ที่มา : มานพ วราภักดิ์, การจำลองเบื้องต้น (กรุงเทพฯ: ศูนย์ผลิตตำราเรียนสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2547), หน้า 142

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x-\theta|} \qquad , \; -\infty < x < \infty ; -\infty < \theta < \infty \; , \; \lambda > 0$$

โดยที่ $E(X) = \theta$ และ $Var(X) = \frac{2}{\lambda^2}$

การแจกแจงลาปลาซเป็นการแจกแจงที่นิยมใช้ในการจำลองข้อมูลให้มีค่าผิดปกติ และใช้เป็นการแจกแจงของความคลาดเคลื่อนสุ่ม

การจำลองตัวแปรสุ่มลาปลาซด้วยวิธีการแปลงผกผัน (Inverse Transformation Method) ได้ตัวแบบการจำลอง X ดังนี้

$$X = \begin{cases} \theta + \frac{1}{\lambda}\ln(2R) & , 0 < R < \frac{1}{2} \\ \theta - \frac{1}{\lambda}\ln(2(1-R)) & , \frac{1}{2} \le R < 1 \end{cases}$$

โดยที่ $R_1, R_2 \sim U(0,1)$ และเป็นอิสระกัน

รายละเอียดของโปรแกรมการจำลองตัวแปรสุ่มแบบลาปลาซศึกษาจาก subroutine laplace

การสร้างตัวแปรตาม

การสร้างตัวแปรตามจะสร้างจากรูปแบบความสัมพันธ์

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i , \qquad i = 1,2,\ldots,n$$

โดยกำหนดให้ $\beta_0 = 5$, $\beta_1 = 1$ และ $\beta_2 = 1$ ตัวแปรอิสระ X_1 มีพารามิเตอร์ $\mu = 20$ และ $\sigma^2 = 10$ ตัวแปรอิสระ X_2 มีพารามิเตอร์ $\mu = 30$ และ $\sigma^2 = 25$ และความคลาดเคลื่อนมีการแจกแจงตามที่กำหนดไว้ข้างต้น

3.2.2 ประมาณค่าพารามิเตอร์ของตัวแบบการถดถอยเชิงเส้นด้วยตัวประมาณกำลังสองน้อยที่สุดแบบธรรมดา ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ และตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ

จากหัวข้อ (3.2.1) เราสามารถสร้างความคลาดเคลื่อน ตัวแปรอิสระ และตัวแปรตามให้มีค่าผิดปกติตรงตามที่กำหนด จากนั้นจึงทำการประมาณค่าพารามิเตอร์ด้วยตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ และตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ โดยมีวิธีการดังที่อธิบายไว้ในบทที่ 2 และได้แสดงโปรแกรมไว้ใน subroutine OLS, subroutine AWLS และ subroutine REWLS ตามลำดับ

3.2.3 การหาค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์ มีวิธีการคำนวณดังนี้

1. คำนวณค่าความคลาดเคลื่อนกำลังสองของพารามิเตอร์แต่ละตัวเมื่อกระทำซ้ำ 500 ครั้ง

$$\mathrm{MSE}_i = \frac{1}{500}\sum_{j=1}^{500}(\beta_i - \hat{\beta}_{ij})^2$$

เมื่อ β_i แทนค่าพารามิเตอร์ที่ต้องการประมาณ โดยที่ $i = 0, 1, 2$

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

MSE_i แทนค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสำหรับ β_i

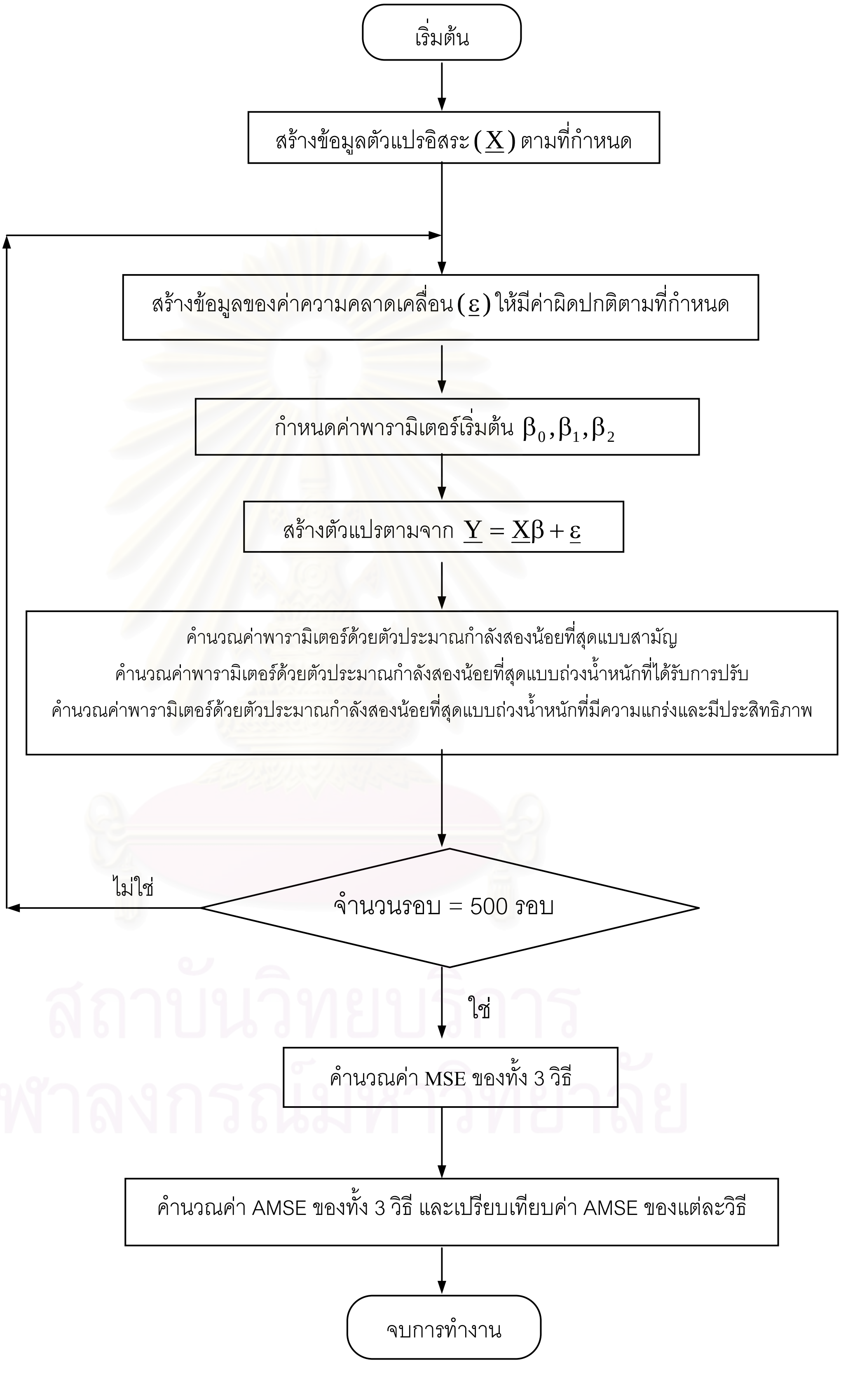
2. คำนวณหาค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์

$$\mathrm{AMSE} = \frac{1}{k}\sum_{i=0}^{2}\mathrm{MSE}_i$$

โดย AMSE แทนค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์จำนวน k ตัว ซึ่งในที่นี้กำหนดให้ k=3

นอกจากนี้ผู้วิจัยได้แสดงแผนผังขั้นตอนการทำงานที่ใช้ในการวิจัยรวมทั้งตารางโปรแกรมดังต่อไปนี้

แผนผังขั้นตอนการทำงาน



ตารางที่ 3.1 โปรแกรมทั้งหมดที่ใช้ในการวิจัย

อันดับที่	ชื่อโปรแกรม	การทำงานของโปรแกรม	ชื่อโปรแกรมย่อย ที่เรียกใช้
โปรแกรมหลัก	MAIN	-อ่านค่าพารามิเตอร์ที่กำหนด	
		-สร้างตัวแปรอิสระ X1 , X2	- gen_x
		-สร้างความคลาดเคลื่อนที่มีค่าผิดปกติ	- cn_residual หรือ cl_residual
		-สร้างตัวแปรตาม	- gen_y
		-คำนวณหาค่าพารามิเตอร์ของตัวประมาณ OLS	- OLS
		-คำนวณหาค่าพารามิเตอร์ของตัวประมาณ AWLS	- AWLS
		-คำนวณหาค่าพารามิเตอร์ของตัวประมาณ REWLS	- REWLS
		-คำนวณหาค่า AMSE	- MSE
โปรแกรมย่อย			
1	gen_x	-สร้างตัวแปรอิสระ X1 , X2	- normal
2	cn_residual	-สร้างความคลาดเคลื่อนที่มีค่าผิดปกติ	- cnormal
	cl_residual		- lnormal
3	OLS	-คำนวณหาค่าพารามิเตอร์ของตัวประมาณ OLS	- inverse3
4	AWLS	-คำนวณหาค่าพารามิเตอร์ของตัวประมาณ AWLS	- OLS - est_y - E_OLS - weight2 - inverse3

ตารางที่ 3.1 (ต่อ)

อันดับที่	ชื่อโปรแกรม	การทำงานของโปรแกรม	ชื่อโปรแกรมย่อย ที่เรียกใช้
5	REWLS	-คำนวณหาค่าพารามิเตอร์ของตัวประมาณ REWLS	- beta_p - LMS - SN - st_residual - weight3 - inverse3
6	weight2	-คำนวณหาค่าถ่วงน้ำหนักของตัวประมาณ AWLS	- rank_error
7	beta_p	- หาพารามิเตอร์ของกลุ่มตัวอย่างย่อยใน LMSE	- grouping
8	LMS	-คำนวณหาค่าพารามิเตอร์ของตัวประมาณ LMSE	- rank_error - rank_med
9	SN	-คำนวณหาค่าประมาณสเกลของตัวประมาณ LMSE	- rank_error
10	weight3	-คำนวณหาค่าถ่วงน้ำหนักของตัวประมาณ REWLS	- rank_error - cumulative - rank_d
11	grouping	-สุ่มกลุ่มตัวอย่างย่อยของตัวประมาณ LMSE	- sampling

บทที่ 4

ผลการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นเมื่อมีค่าผิดปกติในตัวแปรตาม โดยผู้วิจัยได้ทำการเปรียบเทียบตัวประมาณ 3 วิธี ซึ่งได้แก่ ตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (OLSE) ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ (AWLSE) และตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ (REWLSE) โดยใช้เกณฑ์การเปรียบเทียบคือ ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์ ซึ่งมีวิธีการคำนวณเกณฑ์ที่ใช้ในการเปรียบเทียบดังนี้

$$\text{AMSE} = \frac{\sum_{i=0}^{2}\sum_{j=1}^{500}\frac{(\beta_i - \hat{\beta}_{ij})^2}{500}}{3}$$

เมื่อ β_i แทนค่าพารามิเตอร์ที่ต้องการประมาณ โดยที่ i = 0, 1, 2

$\hat{\beta}_{ij}$ แทนค่าประมาณของพารามิเตอร์ β_i ในรอบที่ j

และ AMSE แทนค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์จำนวน 3 ตัว

ผู้วิจัยได้เสนอผลการวิจัยโดยแบ่งออกเป็น 3 ส่วน คือ

<u>ส่วนที่ 1</u> ผลการเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่ไม่มีค่าผิดปกติในตัวแปรตามและในตัวแปรอิสระ

<u>ส่วนที่ 2</u> ผลการเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรตามในระดับไม่รุนแรง และไม่มีค่าผิดปกติในตัวแปรอิสระ

<u>ส่วนที่ 3</u> ผลการเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรตามในระดับรุนแรง และไม่มีค่าผิดปกติในตัวแปรอิสระ

สำหรับการนำเสนอผลการวิจัยจะนำเสนอในรูปแบบตารางและรูปกราฟ เพื่อความสะดวกในการอธิบาย จึงใช้สัญลักษณ์ต่อไปนี้ เพื่อแทนความหมายต่างๆ

p หมายถึง สัดส่วนการปลอมปนของตัวแปรตาม

n หมายถึง ขนาดตัวอย่าง

OLS หมายถึง การประมาณค่าพารามิเตอร์ด้วยตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ

AWLS หมายถึง การประมาณค่าพารามิเตอร์ด้วยตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ

REWLS หมายถึง การประมาณค่าพารามิเตอร์ด้วยตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ

S.D. หมายถึง ส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง โดยที่

$$\mathrm{S.D.} = \sqrt{\sum_{i=0}^{2} \frac{(\mathrm{MSE}_i - \mathrm{AMSE})^2}{2}} \quad \text{สำหรับ } i = 0, 1, 2$$

เมื่อ MSE_i แทนค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสำหรับ β_i

และ AMSE หมายถึง ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์

<u>ส่วนที่ 1 ผลการเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่ไม่มีค่าผิดปกติในตัวแปรตามและในตัวแปรอิสระ</u>

4.1 การเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่ไม่มีค่าผิดปกติในตัวแปรตามและในตัวแปรอิสระ

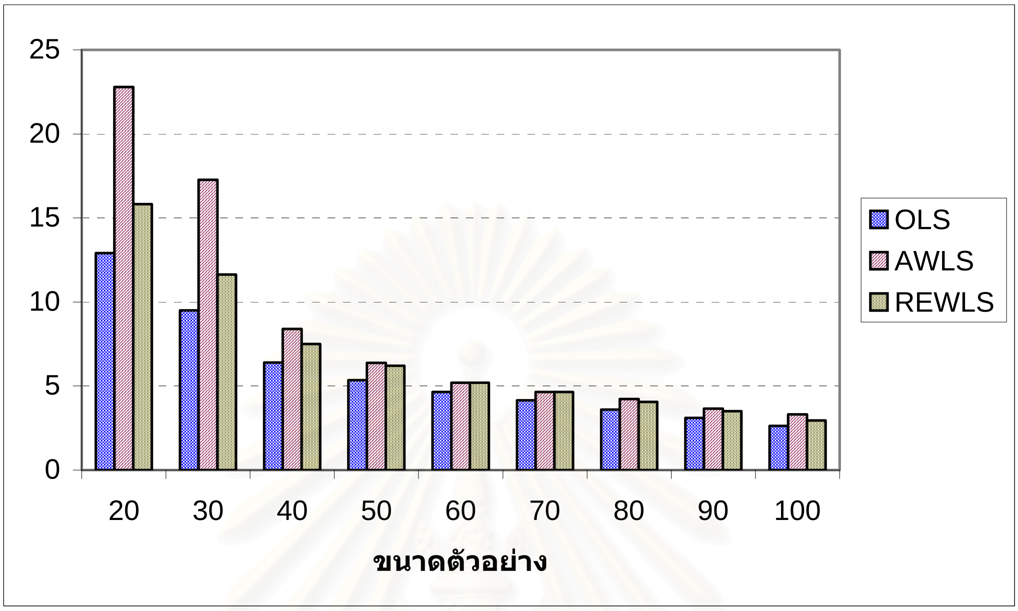
การวิจัยครั้งนี้ ผู้วิจัยทำการศึกษาค่าปกติของตัวแปรตาม ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 4.1 และกราฟรูปที่ 4.1

ตารางที่ 4.1 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามไม่มีค่าผิดปกติ

ขนาดตัวอย่าง n	AMSE		
	OLS (S.D.)	AWLS (S.D.)	REWLS (S.D.)
20	12.907 * (22.300)	22.798 (39.390)	15.829 (27.348)
30	9.509 * (16.429)	17.274 (29.843)	11.633 (20.098)
40	6.405 * (11.0600	8.390 (14.487)	7.493 (12.938)
50	5.353 * (9.242)	6.383 (11.022)	6.202 (10.709)
60	4.646 * (8.025)	5.193 (8.969)	5.191 (8.965)
70	4.150 * (7.170)	4.645 (8.025)	4.642 (8.019)
80	3.597 * (6.214)	4.229 (7.305)	4.058 (7.010)
90	3.108 * (5.368)	3.647 (6.298)	3.512 (6.065)
100	2.631 * (4.544)	3.322 (5.737)	2.960 (5.112)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

รูปที่ 4.1 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามไม่มีค่าผิดปกติ



จากตารางที่ 4.1 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS เมื่อตัวแปรตามไม่มีค่าผิดปกติซึ่งสามารถสรุปผลได้ดังนี้

เมื่อตัวแปรตามไม่มีค่าผิดปกติในทุก ๆ ขนาดตัวอย่าง คือ n = 20, 30, 40, 50, 60, 70, 80, 90 และ100 ตัวประมาณ OLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ REWLS และตัวประมาณ AWLS ตามลำดับ และจะพบว่าเมื่อตัวอย่างมีขนาดใหญ่ (n = 60, 70, 80, 90 และ 100) ตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ให้ค่า AMSE ใกล้เคียงกัน

ข้อสรุป

จากตารางที่4.1 ทุกขนาดตัวอย่างตัวประมาณ OLS ให้ค่า AMSE ต่ำกว่าวิธีอื่น ๆ โดยที่ตัวประมาณ AWLS จะเป็นตัวประมาณที่ให้ค่า AMSE สูงที่สุด และจะพบว่าเมื่อตัวอย่างมีขนาดใหญ่ (n = 60, 70, 80, 90 และ 100) ตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ให้ค่า AMSE ใกล้เคียงกัน และจะสังเกตได้ว่าค่า AMSE ของทุกวิธีมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น เพราะเมื่อขนาดตัวอย่างเพิ่มขึ้นจะทำให้ความแปรปรวนลดลงและค่าประมาณจะเข้าใกล้ค่าจริงมากขึ้นจึงส่งผลให้ค่า AMSE ลดลง

<u>ส่วนที่ 2 ผลการเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรตามในระดับไม่รุนแรง และไม่มีค่าผิดปกติในตัวแปรอิสระ</u>

การวิจัยในส่วนนี้ ผู้วิจัยทำการศึกษาการเกิดค่าผิดปกติของตัวแปรตาม 2 กรณี คือ

1. กรณีเกิดค่าผิดปกติจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ $N(0,10(3)^2)$

2. กรณีเกิดค่าผิดปกติจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,8)

4.2 การเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรตามในระดับไม่รุนแรง กรณีเกิดค่าผิดปกติจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ $N(0,10(3)^2)$

กรณีเกิดค่าผิดปกติจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ $N(0,10C^2)$ เมื่อตัวแปรตามมีระดับค่าผิดปกติระดับไม่รุนแรง (C=3) ทำการศึกษาที่สัดส่วนการปลอมปนเท่ากับ 0.05 , 0.10 , 0.15 และ 0.20 โดยกำหนดขนาดตัวอย่างเท่ากับ 20, 30, 40, 50, 60, 70, 80, 90 และ 100 ซึ่งผลการวิจัยโดยจำแนกตามสัดส่วนการปลอมปนได้นำเสนอในตารางที่ 4.2.1 และกราฟรูปที่ 4.2.1

ตารางที่ 4.2.1 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับไม่รุนแรง จากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ $N(0,10(3)^2)$ โดยจำแนกตามสัดส่วนการปลอมปน

สัดส่วนการปลอมปน p	ขนาดตัวอย่าง n	ตัวประมาณ OLS (S.D.)	ตัวประมาณ AWLS (S.D.)	ตัวประมาณ REWLS (S.D.)
0.05	20	16.598 (28.669)	19.546 (33.771)	15.394 * (26.598)
	30	15.526 (26.835)	13.113 (22.661)	12.443 * (21.505)
	40	8.976 (15.499)	7.364 * (12.713)	7.401 (12.778)
	50	7.318 (12.636)	6.235 * (10.763)	6.697 (11.563)
	60	5.854 (10.108)	5.023 * (8.673)	5.060 (8.737)
	70	5.060 (8.739)	4.410 * (7.618)	5.028 (8.686)
	80	4.767 (8.235)	3.987 * (6.886)	4.224 (7.296)
	90	4.105 (7.090)	3.589 * (6.200)	3.818 (6.595)
	100	3.733 (6.446)	3.096 * (5.347)	3.228 (5.574)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.2.1 (ต่อ)

สัดส่วนการปลอมปน p	ขนาดตัวอย่าง n	ตัวประมาณ		
		OLS (S.D.)	AWLS (S.D.)	REWLS (S.D.)
0.1	20	22.684 (39.187)	21.504 (37.153)	19.966 * (34.503)
	30	20.362 (35.192)	14.419 * (24.916)	16.250 (28.080)
	40	11.015 (19.019)	7.833 * (13.523)	8.960 (15.470)
	50	8.709 (15.034)	7.041 * (12.154)	7.488 (12.927)
	60	7.134 (12.320)	5.421 * (9.362)	6.424 (11.095)
	70	6.902 (11.926)	5.179 * (8.949)	5.790 (10.003)
	80	6.493 (11.218)	4.990 * (8.619)	5.711 (9.865)
	90	5.104 (8.817)	4.195 * (7.247)	4.642 (8.019)
	100	4.285 (7.399)	3.412 * (5.893)	3.826 (6.608)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.2.1 (ต่อ)

สัดส่วนการปลอมปน p	ขนาดตัวอย่าง n	ตัวประมาณ		
		OLS (S.D.)	AWLS (S.D.)	REWLS (S.D.)
0.15	20	42.767 (73.903)	38.514 * (66.553)	38.852 (67.133)
	30	22.9402 (39.6466)	16.414 * (28.359)	20.396 (35.239)
	40	16.926 (29.219)	12.451 * (21.495)	14.765 (25.491)
	50	13.521 (23.346)	10.794 * (18.636)	11.513 (19.879)
	60	12.517 (21.618)	9.029 * (15.593)	9.988 (17.249)
	70	9.909 (17.118)	7.705 * (13.310)	8.523 (14.722)
	80	9.158 (15.818)	6.811 * (11.764)	7.402 (12.786)
	90	8.038 (13.883)	6.376 * (11.014)	7.108 (12.280)
	100	7.219 (12.466)	5.136 * (8.868)	5.421 (9.360)

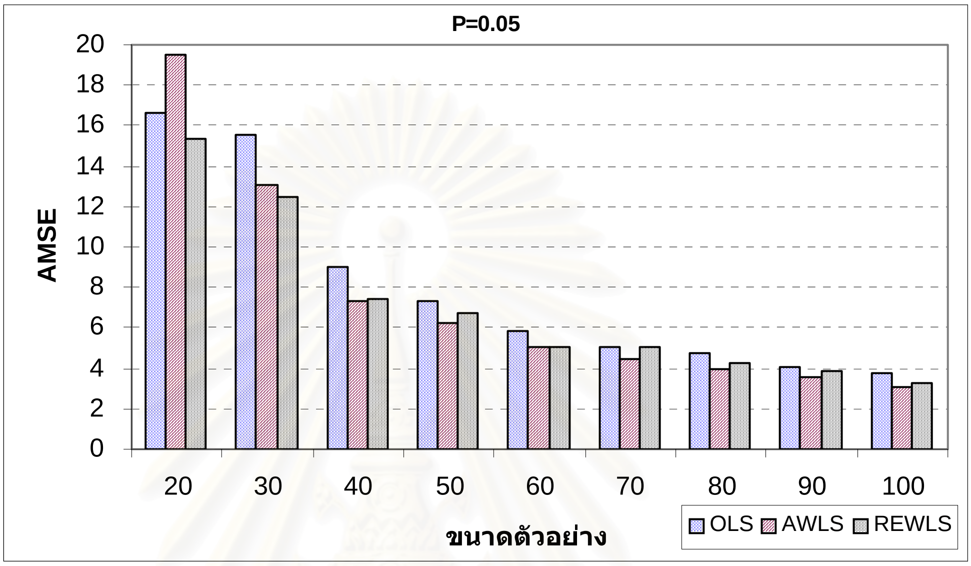
* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

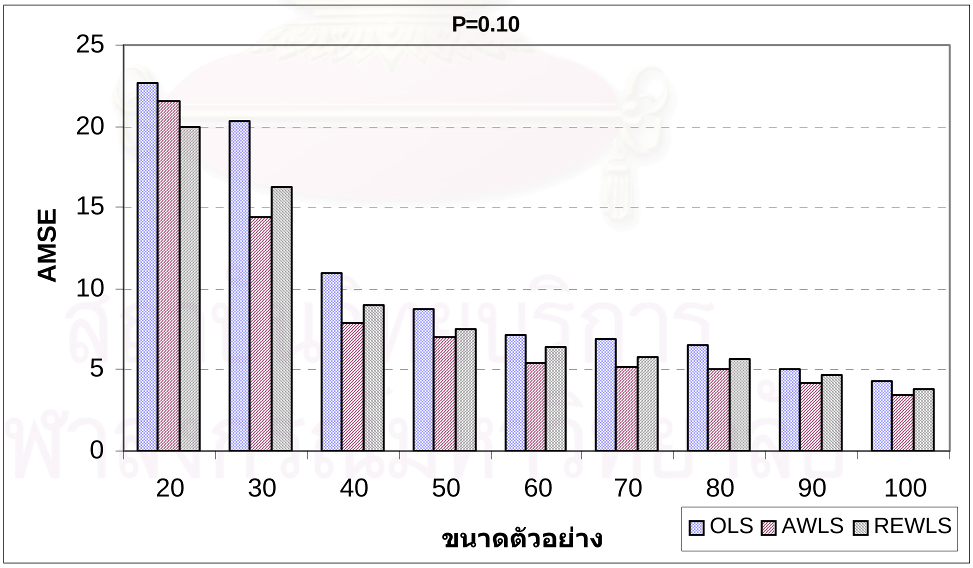
ตารางที่ 4.2.1 (ต่อ)

สัดส่วนการปลอมปน p	ขนาดตัวอย่าง n	ตัวประมาณ		
		OLS (S.D.)	AWLS (S.D.)	REWLS (S.D.)
0.2	20	48.871 (84.424)	39.168 * (67.681)	55.703 (96.243)
	30	36.210 (62.561)	23.437 * (40.483)	39.336 (67.960)
	40	22.739 (39.259)	17.630 * (30.432)	23.656 (40.838)
	50	21.032 (36.325)	16.679 * (28.802)	21.368 (36.901)
	60	15.552 (26.864)	12.457 * (21.520)	16.119 (27.846)
	70	13.997 (24.184)	11.047 * (19.089)	13.937 (24.084)
	80	11.488 (19.842)	9.186 * (15.866)	10.951 (18.915)
	90	9.652 (16.670)	7.594 * (13.118)	9.058 (15.649)
	100	9.356 (16.163)	7.291 * (12.591)	8.552 (14.769)

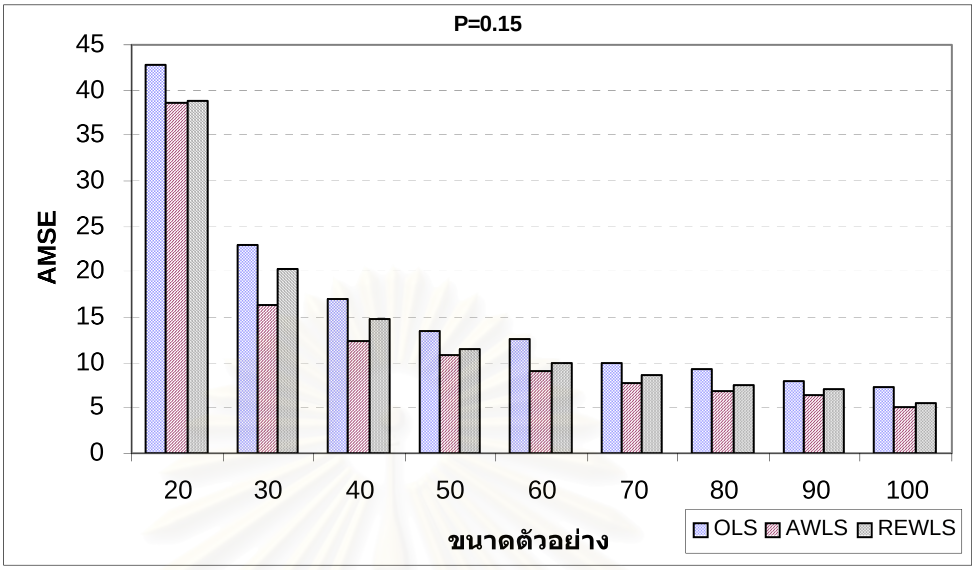
* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

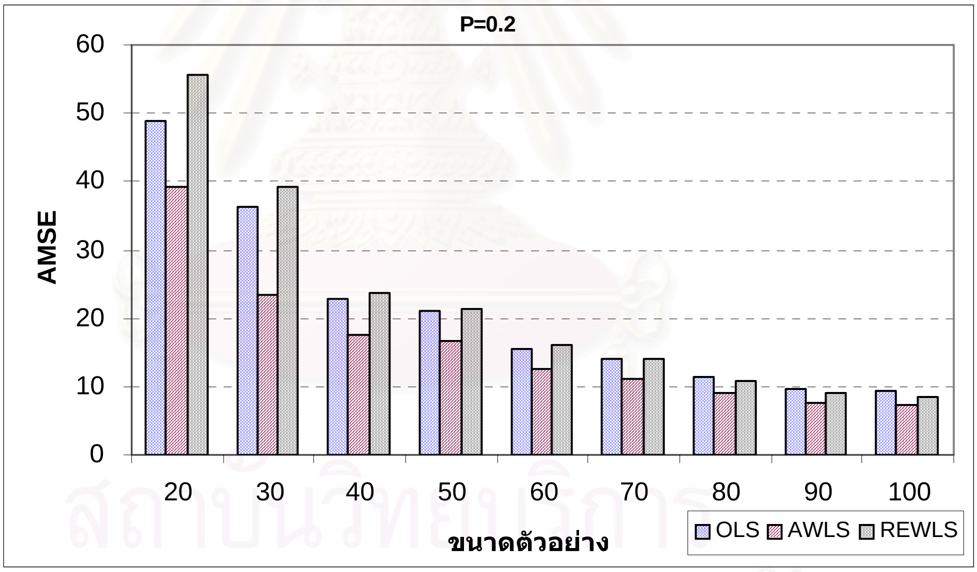
รูปที่ 4.2.1 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับไม่รุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ N(0,10(3)^2) โดยจำแนกตามสัดส่วนการปลอมปน





รูปที่ 4.2.1 (ต่อ)


P=0.15
AMSE
45
40
35
30
25
20
15
10
5
0
20
30
40
50
60
70
80
90
100
ขนาดตัวอย่าง
OLS
AWLS
REWLS


P=0.2
AMSE
60
50
40
30
20
10
0
20
30
40
50
60
70
80
90
100
ขนาดตัวอย่าง
OLS
AWLS
REWLS

จากตารางที่ 4.2.1 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS เมื่อตัวแปรตามมีค่าผิดปกติในระดับไม่รุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ $N(0,10(3)^2)$ โดยจำแนกตามสัดส่วนการปลอมปนได้ดังนี้

กรณีสัดส่วนการปลอมปนเท่ากับ 0.05 เมื่อขนาดตัวอย่างเท่ากับ 20 ตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ OLS และตัวประมาณ AWLS ตามลำดับ ในขณะที่ขนาดตัวอย่างเท่ากับ 30 ตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ AWLS และตัวประมาณ OLS ตามลำดับ ส่วนขนาดตัวอย่างเท่ากับ 40, 50, 60, 70, 80, 90 และ100 พบว่าตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ REWLS และตัวประมาณ OLS ตามลำดับ

กรณีสัดส่วนปลอมปนเท่ากับ 0.1 เมื่อขนาดตัวอย่างเท่ากับ 20 พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ AWLS และตัวประมาณ OLS ตามลำดับ ในขณะที่ขนาดตัวอย่างเท่ากับ 30, 40, 50, 60, 70, 80, 90 และ100 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ REWLS และตัวประมาณ OLS ตามลำดับ

กรณีสัดส่วนการปลอมปนเท่ากับ 0.15 ในทุก ๆ ขนาดตัวอย่าง คือ n = 20, 30, 40, 50, 60, 70, 80, 90 และ100 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ REWLS และตัวประมาณ OLS ตามลำดับ

กรณีสัดส่วนการปลอมปนเท่ากับ 0.2 เมื่อขนาดตัวอย่างเท่ากับ 20, 30, 40, 50 และ 60 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ OLS และตัวประมาณ REWLS ตามลำดับ ในขณะที่ขนาดตัวอย่างเท่ากับ 70, 80, 90 และ 100 พบว่าตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ REWLS และตัวประมาณ OLS ตามลำดับ

ข้อสรุป

ในกรณีที่สัดส่วนการปลอมปนเท่ากับ 0.05 ขนาดตัวอย่างเท่ากับ 20 และ30 และสัดส่วนการปลอมปนเท่ากับ 0.1 ขนาดตัวอย่างเท่ากับ 20 พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด ในขณะที่สัดส่วนการปลอมปนเท่ากับ 0.05 ขนาดตัวอย่างเท่ากับ 40, 50, 60, 70, 80, 90 และ100 สัดส่วนการปลอมปนเท่ากับ 0.1 ขนาดตัวอย่างเท่ากับ 30, 40, 50, 60, 70, 80, 90 และ 100 และสัดส่วนการปลอมปนเท่ากับ 0.15 และ 0.2 ในทุกๆ ขนาดตัวอย่าง พบว่าตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด นอกจากนี้ยังพบว่าเมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกๆ สัดส่วนการปลอมปนค่า AMSE ของทุกตัวประมาณมีแนวโน้มลดลง

กรณีเกิดค่าผิดปกติจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ N(0,10C^2) เมื่อตัวแปรตามมีระดับค่าผิดปกติระดับไม่รุนแรง (C=3) ทำการศึกษาที่สัดส่วนการปลอมปนเท่ากับ 0.05, 0.10, 0.15 และ 0.20 โดยกำหนดขนาดตัวอย่างเท่ากับ 20, 30, 40, 50, 60, 70, 80, 90 และ 100 ซึ่งผลการวิจัยโดยจำแนกตามขนาดตัวอย่างได้นำเสนอในตารางที่ 4.2.2 และกราฟรูปที่ 4.2.2

ตารางที่ 4.2.2 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับไม่รุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ N(0,10$(3)^2$) โดยจำแนกตามขนาดตัวอย่าง

		สัดส่วนการปลอมปน			
n	ตัวประมาณ	0.05	0.10	0.15	0.20
20	OLS	16.598	22.684	42.767	48.871
	(S.D.)	(28.669)	(39.187)	(73.903)	(84.424)
	AWLS	19.546	21.504	38.514 *	39.168 *
	(S.D.)	(33.771)	(37.153)	(66.553)	(67.681)
	REWLS	15.394 *	19.966 *	38.852	55.703
	(S.D.)	(26.598)	(34.503)	(67.133)	(96.243)

		สัดส่วนการปลอมปน			
n	ตัวประมาณ	0.05	0.10	0.15	0.20
30	OLS	15.526	20.362	22.9402	36.210
	(S.D.)	(26.835)	(35.192)	(39.6466)	(62.561)
	AWLS	13.113	14.419 *	16.414 *	23.437 *
	(S.D.)	(22.661)	(24.916)	(28.359)	(40.483)
	REWLS	12.443 *	16.250	20.396	39.336
	(S.D.)	(21.505)	(28.080)	(35.239)	(67.960)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.2.2 (ต่อ)

n	ตัวประมาณ	สัดส่วนการปลอมปน 0.05	0.10	0.15	0.20
40	OLS	8.976	11.015	16.926	22.739
	(S.D.)	(15.499)	(19.019)	(29.219)	(39.259)
	AWLS	7.364 *	7.833 *	12.451 *	17.630 *
	(S.D.)	(12.713)	(13.523)	(21.495)	(30.432)
	REWLS	7.401	8.960	14.765	23.656
	(S.D.)	(12.778)	(15.470)	(25.491)	(40.838)

n	ตัวประมาณ	สัดส่วนการปลอมปน 0.05	0.10	0.15	0.20
50	OLS	7.318	8.709	13.521	21.032
	(S.D.)	(12.636)	(15.034)	(23.346)	(36.325)
	AWLS	6.235 *	7.041 *	10.794 *	16.679 *
	(S.D.)	(10.763)	(12.154)	(18.636)	(28.802)
	REWLS	6.697	7.488	11.513	21.368
	(S.D.)	(11.563)	(12.927)	(19.879)	(36.901)

n	ตัวประมาณ	สัดส่วนการปลอมปน 0.05	0.10	0.15	0.20
60	OLS	5.854	7.134	12.517	15.552
	(S.D.)	(10.108)	(12.320)	(21.618)	(26.864)
	AWLS	5.023 *	5.421 *	9.029 *	12.457 *
	(S.D.)	(8.673)	(9.362)	(15.593)	(21.520)
	REWLS	5.060	6.424	9.988	16.119
	(S.D.)	(8.737)	(11.095)	(17.249)	(27.846)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.2.2 (ต่อ)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
70	OLS	5.060	6.902	9.909	13.997
	(S.D.)	(8.739)	(11.926)	(17.118)	(24.184)
	AWLS	4.410 *	5.179 *	7.705 *	11.047 *
	(S.D.)	(7.618)	(8.949)	(13.310)	(19.089)
	REWLS	5.028	5.790	8.523	13.937
	(S.D.)	(8.686)	(10.003)	(14.722)	(24.084)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
80	OLS	4.767	6.493	9.158	11.488
	(S.D.)	(8.235)	(11.218)	(15.818)	(19.842)
	AWLS	3.987 *	4.990 *	6.811 *	9.186 *
	(S.D.)	(6.886)	(8.619)	(11.764)	(15.866)
	REWLS	4.224	5.711	7.402	10.951
	(S.D.)	(7.296)	(9.865)	(12.786)	(18.915)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
90	OLS	4.105	5.104	8.038	9.652
	(S.D.)	(7.090)	(8.817)	(13.883)	(16.670)
	AWLS	3.589 *	4.195 *	6.376 *	7.594 *
	(S.D.)	(6.200)	(7.247)	(11.014)	(13.118)
	REWLS	3.818	4.642	7.108	9.058
	(S.D.)	(6.595)	(8.019)	(12.280)	(15.649)

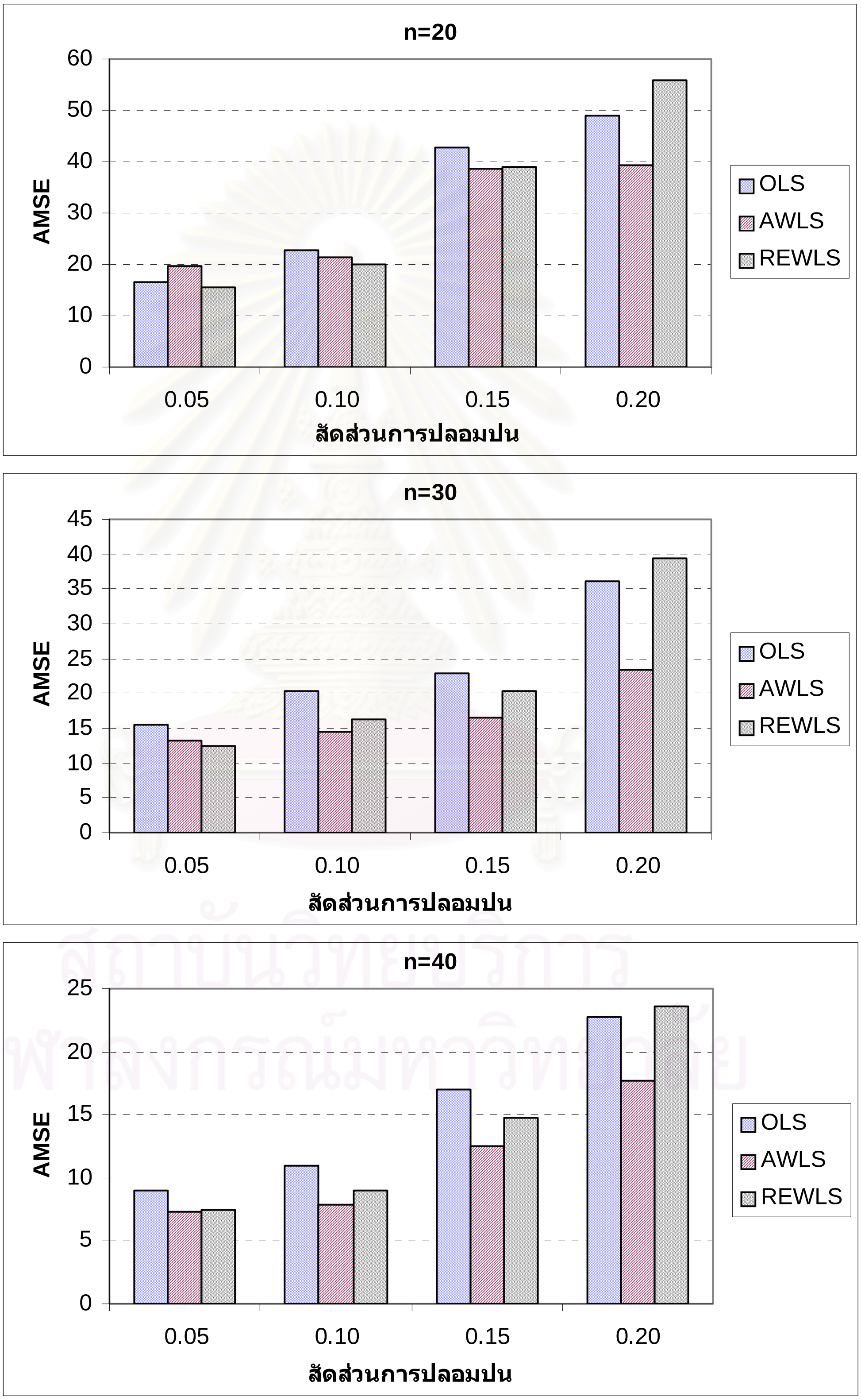
* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.2.2 (ต่อ)

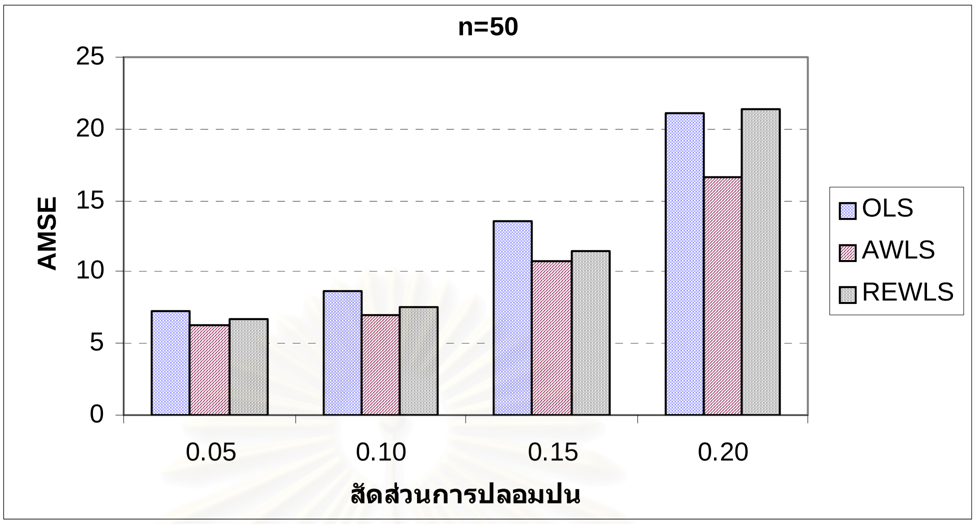
n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
100	OLS	3.733	4.285	7.219	9.356
	(S.D.)	(6.446)	(7.399)	(12.466)	(16.163)
	AWLS	3.096 *	3.412 *	5.136 *	7.291 *
	(S.D.)	(5.347)	(5.893)	(8.868)	(12.591)
	REWLS	3.228	3.826	5.421	8.552
	(S.D.)	(5.574)	(6.608)	(9.360)	(14.769)

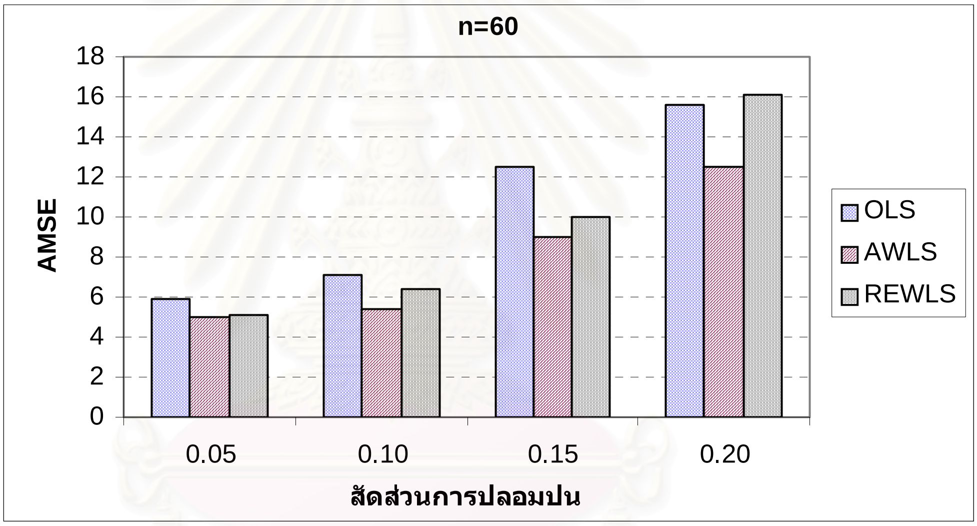
* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

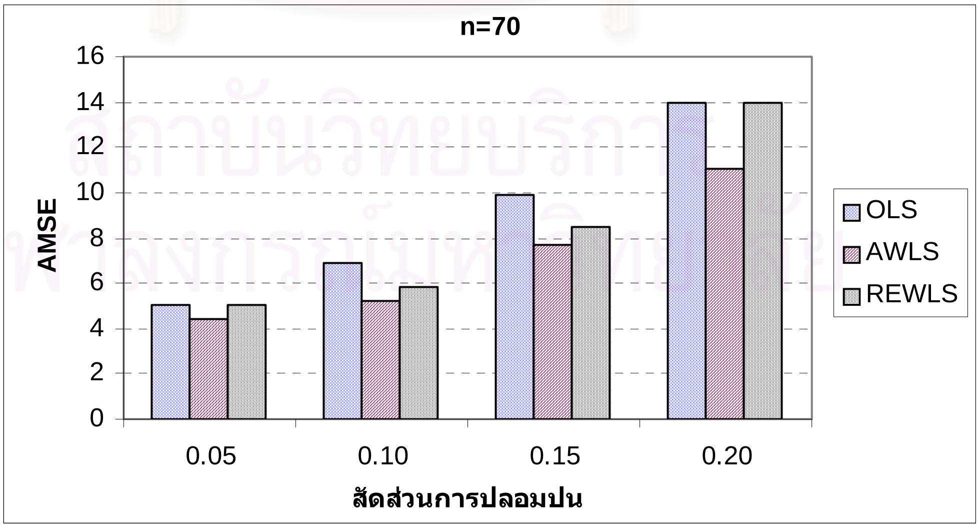
รูปที่ 4.2.2 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับไม่รุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ $N(0,10(3)^2)$ โดยจำแนกตามขนาดตัวอย่าง



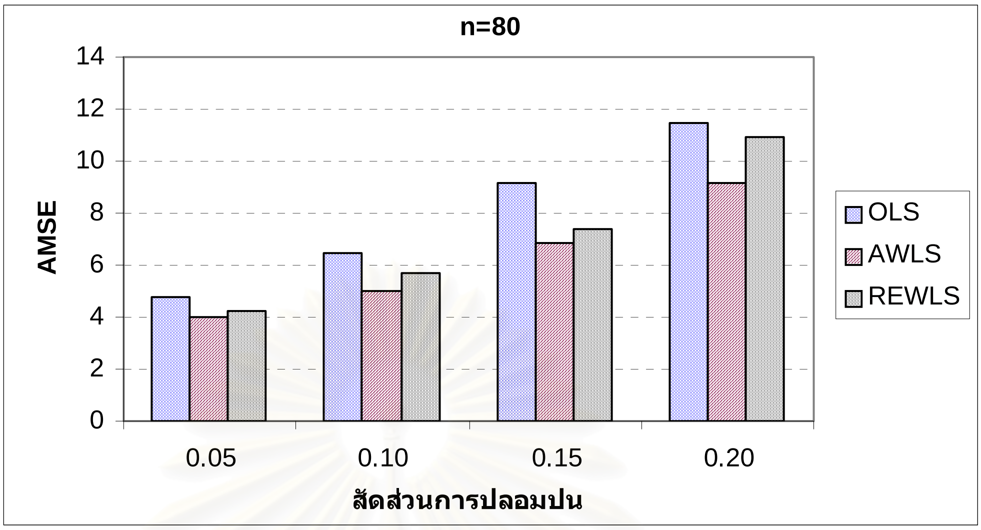
รูปที่ 4.2.2 (ต่อ)

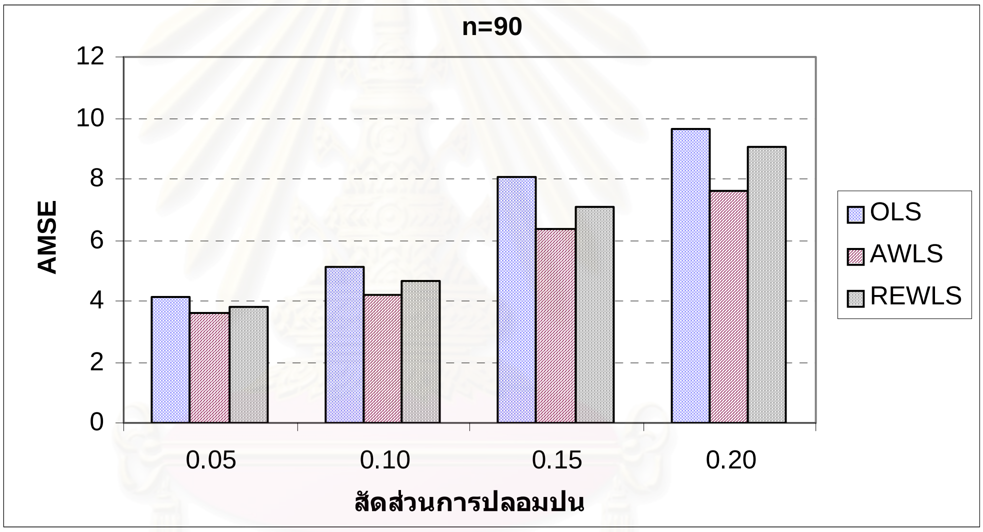

n=50
AMSE
25
20
15
10
5
0
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS

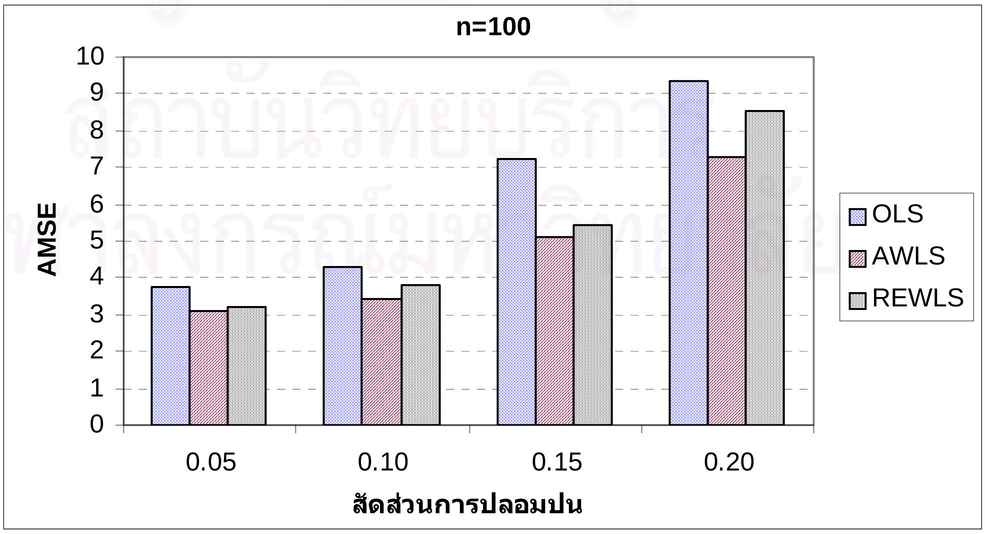

n=60
AMSE
18
16
14
12
10
8
6
4
2
0
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS


n=70
AMSE
16
14
12
10
8
6
4
2
0
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS

รูปที่ 4.2.2 (ต่อ)


n=80
AMSE
0
2
4
6
8
10
12
14
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS


n=90
AMSE
0
2
4
6
8
10
12
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS


n=100
AMSE
0
1
2
3
4
5
6
7
8
9
10
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS

จากตารางที่ 4.2.2 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS เมื่อตัวแปรตามมีค่าผิดปกติในระดับไม่รุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ $N(0,10(3)^2)$ โดยจำแนกตามขนาดตัวอย่างได้ดังนี้

กรณีขนาดตัวอย่างเท่ากับ 20 เมื่อสัดส่วนการปลอมปนเท่ากับ 0.05 และ 0.1 ตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาเมื่อสัดส่วนการปลอมปนเท่ากับ 0.05 พบว่าตัวประมาณ OLS ให้ค่า AMSE ต่ำกว่าตัวประมาณ AWLS ในขณะที่สัดส่วนการปลอมปนเท่ากับ 0.1 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำกว่าตัวประมาณ OLS ส่วนในกรณีที่สัดส่วนการปลอมปนเท่ากับ 0.15 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ REWLS และตัวประมาณ OLS ตามลำดับ และสำหรับกรณีสัดส่วนการปลอมปนเท่ากับ 0.2 พบว่าตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ OLS และ ตัวประมาณ REWLS ตามลำดับ

กรณีขนาดตัวอย่างเท่ากับ 30 เมื่อสัดส่วนการปลอมปนเท่ากับ 0.05 ตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ AWLS และตัวประมาณ OLS ตามลำดับ ในขณะที่สัดส่วนการปลอมปนเท่ากับ 0.1, 0.15 และ 0.2 พบว่าตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาเมื่อสัดส่วนการปลอมปนเท่ากับ 0.1 และ 0.15 ตัวประมาณ REWLS ให้ค่า AMSE ต่ำกว่าตัวประมาณ OLS ในขณะที่สัดส่วนการปลอมปนเท่ากับ 0.2 ตัวประมาณ OLS ให้ค่า AMSE ต่ำกว่าตัวประมาณ REWLS

กรณีขนาดตัวอย่างเท่ากับ 40, 50 และ 60 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุดในทุกๆ สัดส่วนการปลอมปน รองลงมาเมื่อสัดส่วนการปลอมปนเท่ากับ 0.05, 0.1 และ 0.15 พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำกว่าตัวประมาณ OLS ในขณะที่สัดส่วนการปลอมปนเท่ากับ 0.2 ตัวประมาณ OLS ให้ค่า AMSE ต่ำกว่าตัวประมาณ REWLS

กรณีขนาดตัวอย่างเท่ากับ 70, 80, 90 และ 100 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุดในทุกๆ สัดส่วนการปลอมปน รองลงมาคือตัวประมาณ REWLS และตัวประมาณ OLS ตามลำดับ

ข้อสรุป

ในกรณีขนาดตัวอย่างเท่ากับ 20 เมื่อสัดส่วนการปลอมปนเท่ากับ 0.05 และ 0.1 พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด ในขณะที่สัดส่วนการปลอมปนเท่ากับ 0.15 และ 0.2 พบว่าตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด

กรณีขนาดตัวอย่างเท่ากับ 30 สัดส่วนการปลอมปนเท่ากับ 0.05 พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด ในขณะที่สัดส่วนการปลอมปนเท่ากับ 0.1, 0.15 และ 0.2 พบว่าตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด

เมื่อตัวอย่างเท่ากับ 40, 50 และ 60 ในทุกๆ สัดส่วนการปลอมปน ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด และตัวประมาณ OLS ให้ค่า AMSE สูงที่สุดเมื่อสัดส่วนการปลอมปนเท่ากับ 0.05, 0.1 และ 0.15 ในขณะที่ตัวประมาณ REWLS ให้ค่า AMSE สูงที่สุดเมื่อสัดส่วนการปลอมปนเท่ากับ 0.2

ในขณะที่ขนาดตัวอย่างเท่ากับ 70, 80, 90 และ 100 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด และตัวประมาณ OLS ให้ค่า AMSE สูงที่สุดในทุกๆ สัดส่วนการปลอมปน

นอกจากนี้ยังพบว่าเมื่อสัดส่วนการปลอมปนเพิ่มขึ้นในทุกๆ ขนาดตัวอย่างค่า AMSE ของทุกตัวประมาณมีแนวโน้มเพิ่มขึ้น

4.3 การเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรตามในระดับไม่รุนแรง กรณีเกิดค่าผิดปกติจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,8)

กรณีเกิดค่าผิดปกติในระดับไม่รุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,8) ทำการศึกษาที่สัดส่วนการปลอมปนเท่ากับ 0.05, 0.1, 0.15 และ 0.20 โดยกำหนดขนาดตัวอย่างเท่ากับ 20, 30, 40, 50, 60, 70, 80, 90 และ 100 ซึ่งผลการวิจัยโดยจำแนกตามสัดส่วนการปลอมปนได้นำเสนอในตารางที่ 4.3.1 และกราฟรูปที่ 4.3.1

ตารางที่ 4.3.1 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับไม่รุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,8) โดยจำแนกตามสัดส่วนการปลอมปน

สัดส่วนการปลอมปน p	ขนาดตัวอย่าง n	ตัวประมาณ OLS (S.D.)	AWLS (S.D.)	REWLS (S.D.)
0.05	20	18.086 (31.239)	23.069 (39.852)	15.934 * (27.528)
	30	15.225 (26.314)	14.133 (24.421)	12.005 * (20.742)
	40	9.021 (15.577)	7.482 * (12.918)	7.595 (13.114)
	50	7.432 (12.830)	6.126 * (10.578)	6.671 (11.518)
	60	5.859 (10.120)	4.864 * (8.404)	5.231 (9.036)
	70	5.628 (9.723)	4.757 * (8.215)	5.144 (8.887)
	80	5.426 (9.377)	4.440 * (7.672)	4.696 (8.115)
	90	3.690 (6.373)	3.154 * (5.448)	3.440 (5.941)
	100	3.389 (5.852)	3.090 * (5.336)	3.365 (5.810)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.3.1 (ต่อ)

สัดส่วนการปลอมปน p	ขนาดตัวอย่าง n	ตัวประมาณ OLS (S.D.)	AWLS (S.D.)	REWLS (S.D.)
0.1	20	30.009 (51.842)	28.643 (49.494)	26.338 * (45.509)
	30	20.905 (36.111)	15.526 * (26.823)	17.464 (30.166)
	40	14.509 (25.049)	9.361 * (16.157)	10.694 (18.462)
	50	12.660 (21.858)	9.007 * (15.551)	9.502 (16.406)
	60	10.342 (17.862)	7.347 * (12.689)	7.661 (13.233)
	70	9.410 (16.256)	6.999 * (12.093)	7.308 (12.626)
	80	8.711 (15.048)	6.897 * (11.918)	6.983 (12.066)
	90	6.562 (11.333)	4.962 * (8.572)	5.014 (8.662)
	100	6.069 (10.484)	4.412 * (7.620)	4.440 (7.669)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.3.1 (ต่อ)

สัดส่วนการปลอมปน p	ขนาดตัวอย่าง n	ตัวประมาณ OLS (S.D.)	AWLS (S.D.)	REWLS (S.D.)
0.15	20	42.693	33.684 *	40.979
		(73.762)	(58.189)	(70.809)
	30	26.213	16.461 *	22.111
		(45.284)	(28.443)	(38.195)
	40	17.517	12.416 *	15.533
		(30.239)	(21.432)	(26.816)
	50	17.119	11.559 *	13.455
		(29.563)	(19.958)	(23.230)
	60	14.114	10.602 *	10.810
		(24.375)	(18.310)	(18.669)
	70	11.504	8.123 *	9.238
		(19.870)	(14.034)	(15.958)
	80	9.900	7.391 *	7.815
		(17.100)	(12.767)	(13.501)
	90	8.075	5.918 *	6.062
		(13.946)	(10.221)	(10.466)
	100	7.833	5.646 *	5.705
		(13.523)	(9.751)	(9.853)

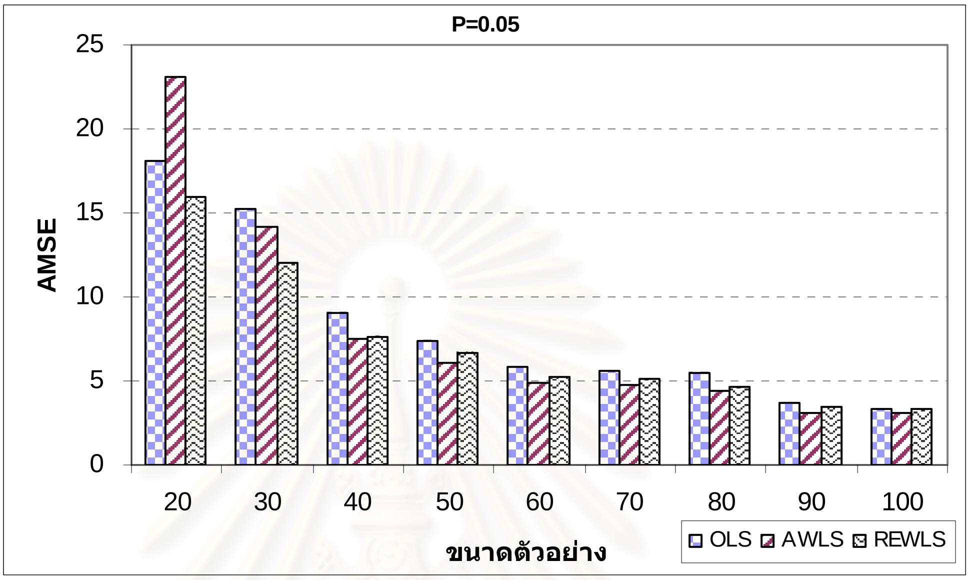
* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

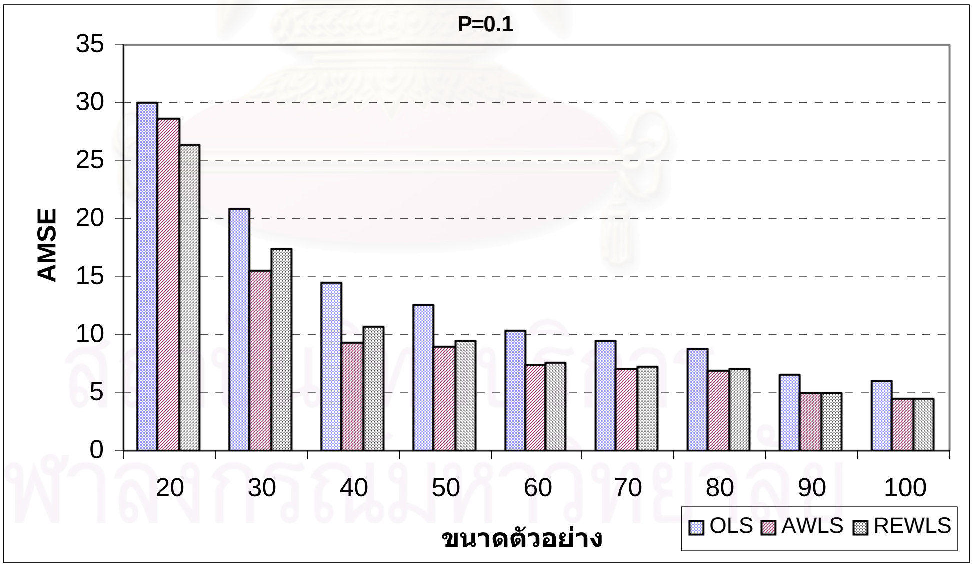
ตารางที่ 4.3.1 (ต่อ)

สัดส่วนการปลอมปน p	ขนาดตัวอย่าง n	ตัวประมาณ		
		OLS (S.D.)	AWLS (S.D.)	REWLS (S.D.)
0.2	20	54.090 (93.469)	42.724 * (73.839)	56.314 (97.318)
	30	36.349 (62.799)	27.080 * (46.791)	36.392 (62.874)
	40	21.445 (37.016)	16.948 * (29.258)	21.649 (37.373)
	50	20.628 (35.620)	16.551 * (28.579)	20.767 (35.862)
	60	14.793 (25.546)	12.080 * (20.857)	14.970 (25.852)
	70	13.164 (22.735)	10.524 * (18.176)	12.232 (21.127)
	80	13.026 (22.507)	9.593 * (16.574)	10.941 (18.892)
	90	10.768 (18.595)	8.721 * (15.061)	10.462 (18.074)
	100	8.111 (14.000)	5.530 * (9.548)	7.092 (12.249)

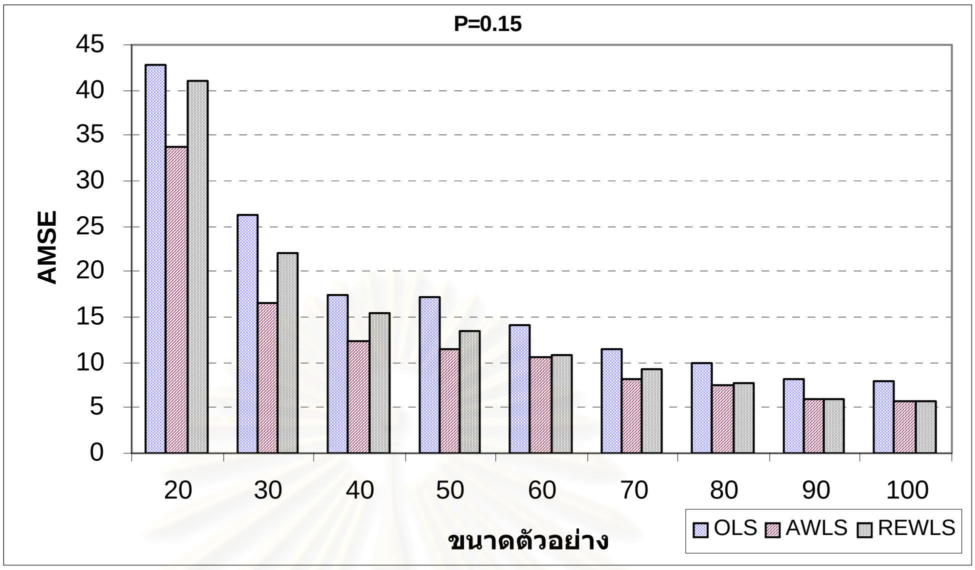
* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

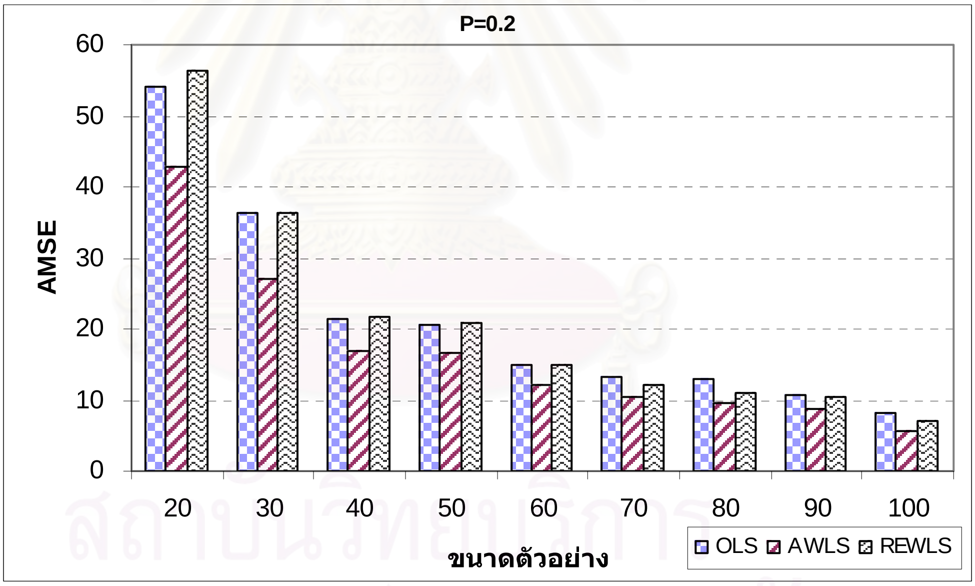
รูปที่ 4.3.1 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับไม่รุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,8) โดยจำแนกตามสัดส่วนการปลอมปน





รูปที่ 4.3.1 (ต่อ)


P=0.15
AMSE
45
40
35
30
25
20
15
10
5
0
20
30
40
50
60
70
80
90
100
ขนาดตัวอย่าง
OLS
AWLS
REWLS


P=0.2
AMSE
60
50
40
30
20
10
0
20
30
40
50
60
70
80
90
100
ขนาดตัวอย่าง
OLS
AWLS
REWLS

จากตารางที่ 4.3.1 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS เมื่อตัวแปรตามมีค่าผิดปกติในระดับไม่รุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,8) โดยจำแนกตามสัดส่วนการปลอมปนได้ดังนี้

กรณีสัดส่วนการปลอมปนเท่ากับ 0.05 เมื่อขนาดตัวอย่างเท่ากับ 20 และ 30 ตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาเมื่อขนาดตัวอย่างเท่ากับ 20 ตัวประมาณ OLS ให้ค่า AMSE ต่ำกว่าตัวประมาณ AWLS ในขณะที่ขนาดตัวอย่างเท่ากับ 30 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำกว่าตัวประมาณ OLS ส่วนขนาดตัวอย่างเท่ากับ 40, 50, 60, 70, 80, 90 และ100 พบว่าตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ REWLS และตัวประมาณ OLS ตามลำดับ

กรณีสัดส่วนปลอมปนเท่ากับ 0.1 เมื่อขนาดตัวอย่างเท่ากับ 20 พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ AWLS และตัวประมาณ OLS ตามลำดับ ในขณะที่ขนาดตัวอย่างเท่ากับ 30, 40, 50, 60, 70, 80, 90 และ100 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ REWLS และตัวประมาณ OLS ตามลำดับ

กรณีสัดส่วนการปลอมปนเท่ากับ 0.15 ในทุก ๆ ขนาดตัวอย่าง คือ n = 20, 30, 40, 50, 60, 70, 80, 90 และ100 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ REWLS และตัวประมาณ OLS ตามลำดับ

กรณีสัดส่วนการปลอมปนเท่ากับ 0.2 เมื่อขนาดตัวอย่างเท่ากับ 20, 30, 40, 50 และ 60 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ OLS และตัวประมาณ REWLS ตามลำดับ ในขณะที่ขนาดตัวอย่างเท่ากับ 70, 80, 90 และ100 พบว่าตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ REWLS และตัวประมาณ OLS ตามลำดับ

ข้อสรุป

ในกรณีที่สัดส่วนการปลอมปนเท่ากับ 0.05 ขนาดตัวอย่างเท่ากับ 20 และ30 และสัดส่วนการปลอมปนเท่ากับ 0.1 ขนาดตัวอย่างเท่ากับ 20 พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด ในขณะที่สัดส่วนการปลอมปนเท่ากับ 0.05 ขนาดตัวอย่างเท่ากับ 40, 50, 60, 70, 80, 90 และ100 สัดส่วนการปลอมปนเท่ากับ 0.1 ขนาดตัวอย่างเท่ากับ 30, 40, 50, 60, 70, 80, 90 และ 100 และสัดส่วนการปลอมปนเท่ากับ 0.15 และ 0.2 ในทุกๆ ขนาดตัวอย่าง พบว่าตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด นอกจากนี้ยังพบว่าเมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกๆ สัดส่วนการปลอมปนค่า AMSE ของทุกตัวประมาณมีแนวโน้มลดลง

กรณีเกิดค่าผิดปกติในระดับไม่รุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,8) ทำการศึกษาที่สัดส่วนการปลอมปนเท่ากับ 0.05, 0.10, 0.15 และ 0.20 โดยกำหนดขนาดตัวอย่างเท่ากับ 20, 30, 40, 50, 60, 70, 80, 90 และ 100 ซึ่งผลการวิจัยโดยจำแนกตามขนาดตัวอย่างได้นำเสนอในตารางที่ 4.3.2 และกราฟรูปที่ 4.3.2

ตารางที่ 4.3.2 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับไม่รุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,8) โดยจำแนกตามขนาดตัวอย่าง

		สัดส่วนการปลอมปน			
n	ตัวประมาณ	0.05	0.10	0.15	0.20
20	OLS	18.086	30.009	42.693	54.090
	(S.D.)	(31.239)	(51.842)	(73.762)	(93.469)
	AWLS	23.069	28.643	33.684 *	42.724 *
	(S.D.)	(39.852)	(49.494)	(58.189)	(73.839)
	REWLS	15.934 *	26.338 *	40.979	56.314
	(S.D.)	(27.528)	(45.509)	(70.809)	(97.318)

		สัดส่วนการปลอมปน			
n	ตัวประมาณ	0.05	0.10	0.15	0.20
30	OLS	15.225	20.905	26.213	36.349
	(S.D.)	(26.314)	(36.111)	(45.284)	(62.799)
	AWLS	14.133	15.526 *	16.461 *	27.080 *
	(S.D.)	(24.421)	(26.823)	(28.443)	(46.791)
	REWLS	12.005 *	17.464	22.111	36.392
	(S.D.)	(20.742)	(30.166)	(38.195)	(62.874)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.3.2 (ต่อ)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
40	OLS	9.021	14.509	17.517	21.445
	(S.D.)	(15.577)	(25.049)	(30.239)	(37.016)
	AWLS	7.482 *	9.361 *	12.416 *	16.948 *
	(S.D.)	(12.918)	(16.157)	(21.432)	(29.258)
	REWLS	7.595	10.694	15.533	21.649
	(S.D.)	(13.114)	(18.462)	(26.816)	(37.373)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
50	OLS	7.432	12.660	17.119	20.628
	(S.D.)	(12.830)	(21.858)	(29.563)	(35.620)
	AWLS	6.126 *	9.007 *	11.559 *	16.551 *
	(S.D.)	(10.578)	(15.551)	(19.958)	(28.579)
	REWLS	6.671	9.502	13.455	20.767
	(S.D.)	(11.518)	(16.406)	(23.230)	(35.862)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
60	OLS	5.859	10.342	14.114	14.793
	(S.D.)	(10.120)	(17.862)	(24.375)	(25.546)
	AWLS	4.864 *	7.347 *	10.602 *	12.080 *
	(S.D.)	(8.404)	(12.689)	(18.310)	(20.857)
	REWLS	5.231	7.661	10.810	14.970
	(S.D.)	(9.036)	(13.233)	(18.669)	(25.852)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.3.2 (ต่อ)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
70	OLS	5.628	9.410	11.504	13.164
	(S.D.)	(9.723)	(16.256)	(19.870)	(22.735)
	AWLS	4.757 *	6.999 *	8.123 *	10.524 *
	(S.D.)	(8.215)	(12.093)	(14.034)	(18.176)
	REWLS	5.144	7.308	9.238	12.232
	(S.D.)	(8.887)	(12.626)	(15.958)	(21.127)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
80	OLS	5.426	8.711	9.900	13.026
	(S.D.)	(9.377)	(15.048)	(17.100)	(22.507)
	AWLS	4.440 *	6.897 *	7.391 *	9.593 *
	(S.D.)	(7.672)	(11.918)	(12.767)	(16.574)
	REWLS	4.696	6.983	7.815	10.941
	(S.D.)	(8.115)	(12.066)	(13.501)	(18.892)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
90	OLS	3.690	6.562	8.075	10.768
	(S.D.)	(6.373)	(11.333)	(13.946)	(18.595)
	AWLS	3.154 *	4.962 *	5.918 *	8.721 *
	(S.D.)	(5.448)	(8.572)	(0.221)	(15.061)
	REWLS	3.440	5.014	6.062	10.462
	(S.D.)	(5.941)	(8.662)	(10.466)	(18.074)

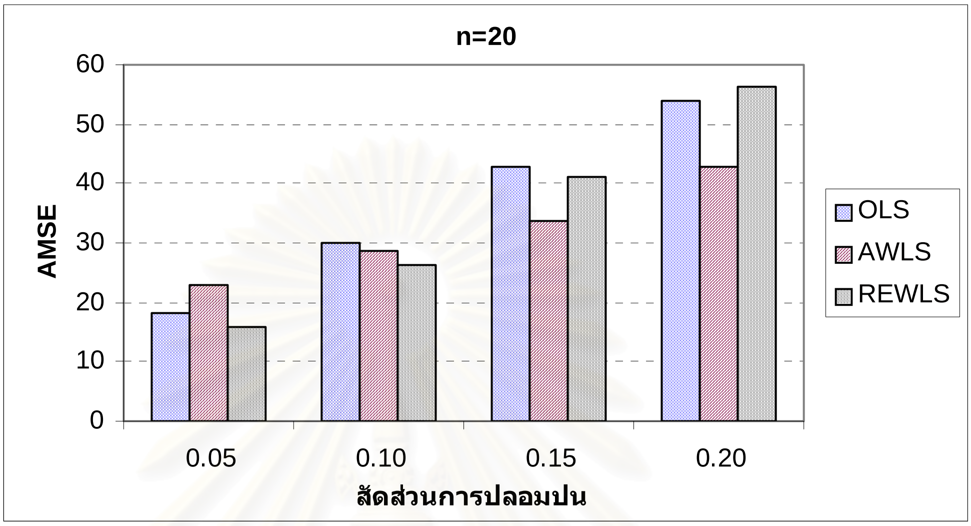
* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

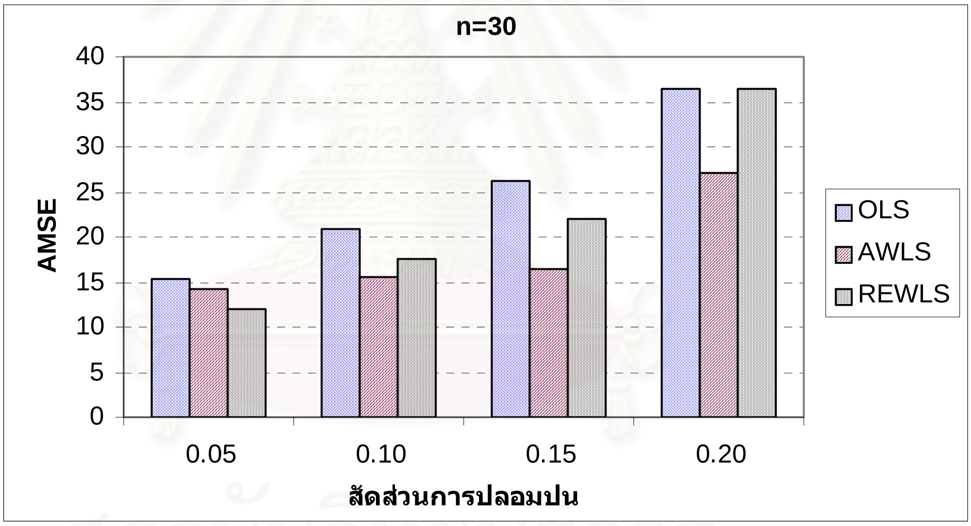
ตารางที่ 4.3.2 (ต่อ)

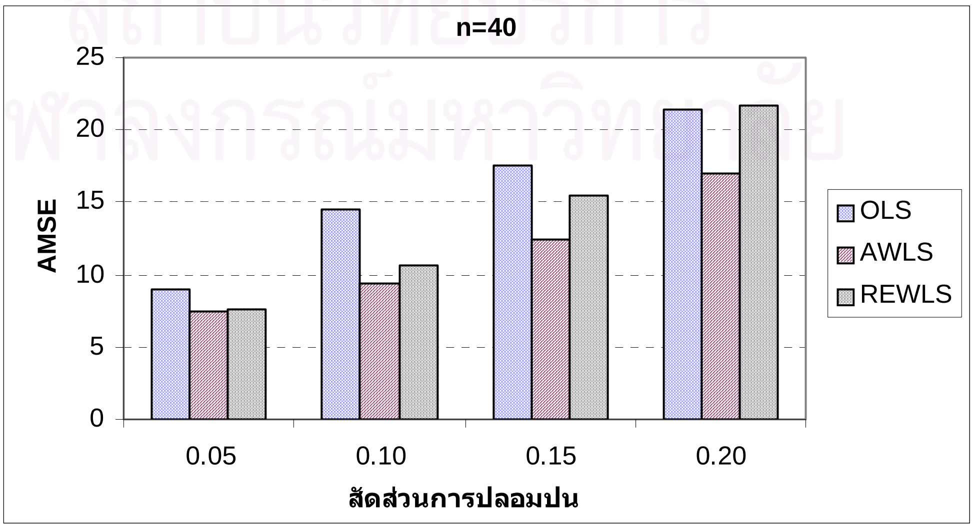
n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
100	OLS	3.389	6.069	7.833	8.111
	(S.D.)	(5.852)	(10.484)	(13.523)	(14.000)
	AWLS	3.090 *	4.412 *	5.646 *	5.530 *
	(S.D.)	(5.336)	(7.620)	(9.751)	(9.548)
	REWLS	3.365	4.440	5.705	7.092
	(S.D.)	(5.810)	(7.669)	(9.853)	(12.249)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

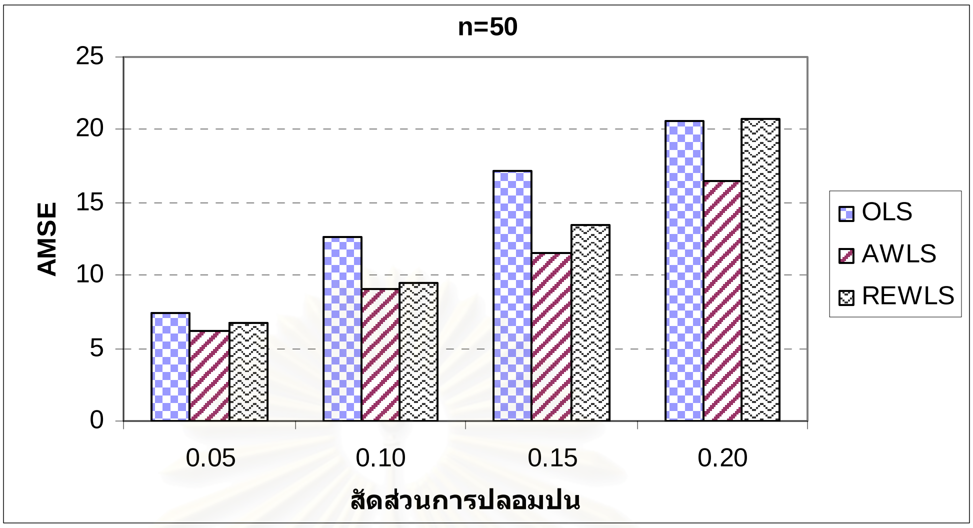
รูปที่ 4.3.2 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับไม่รุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,8) โดยจำแนกตามขนาดตัวอย่าง

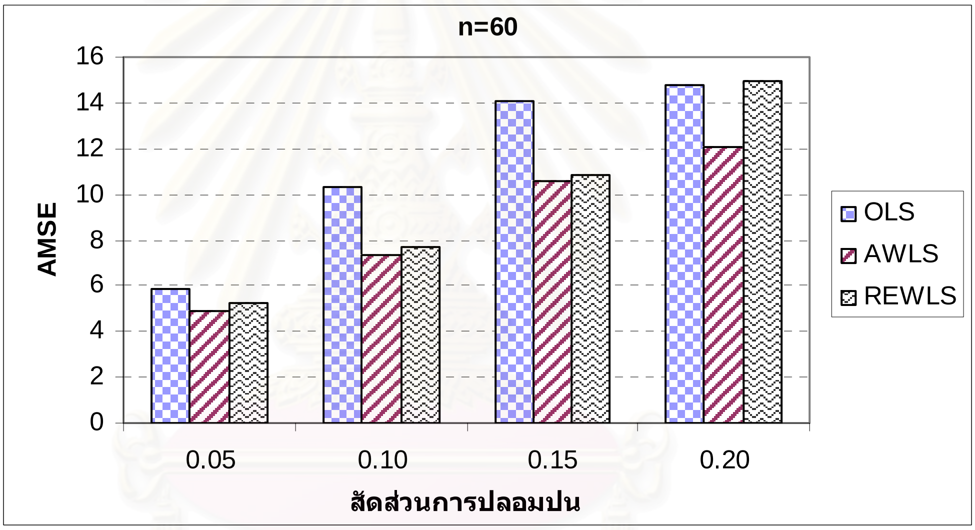


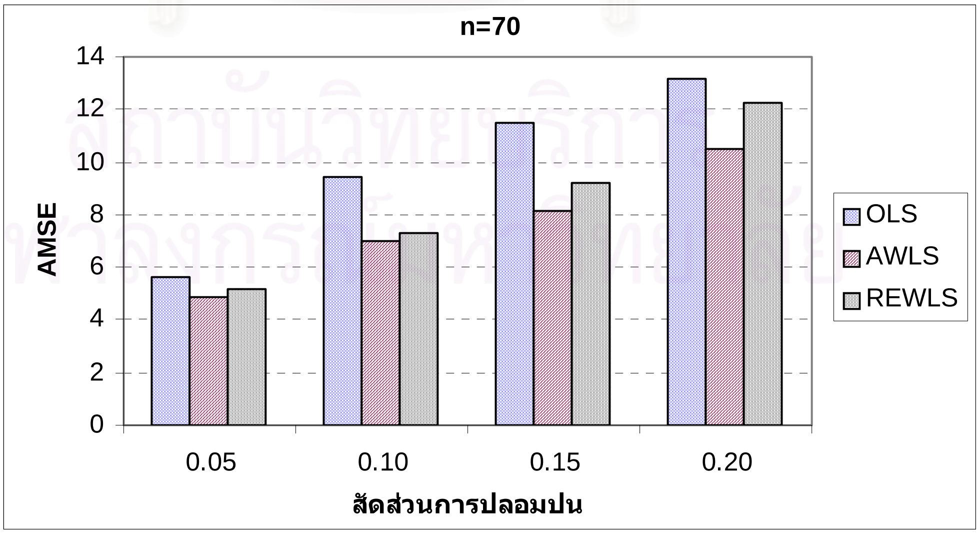




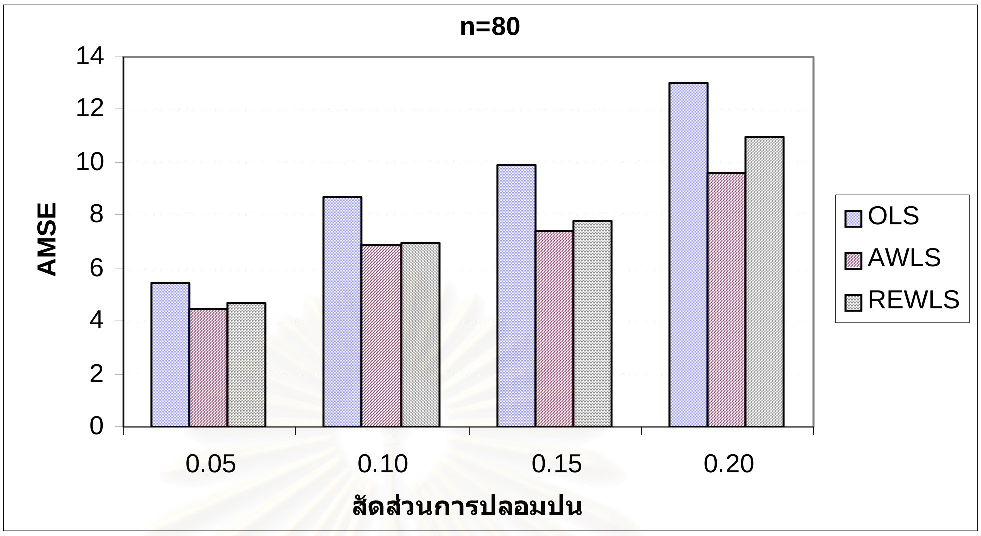
รูปที่ 4.3.2 (ต่อ)

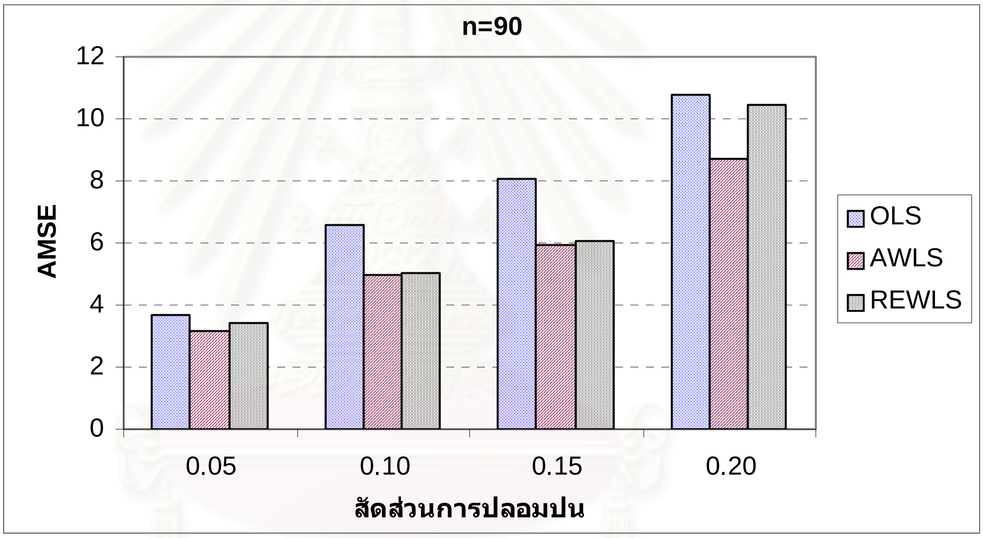

n=50
AMSE
25
20
15
10
5
0
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS

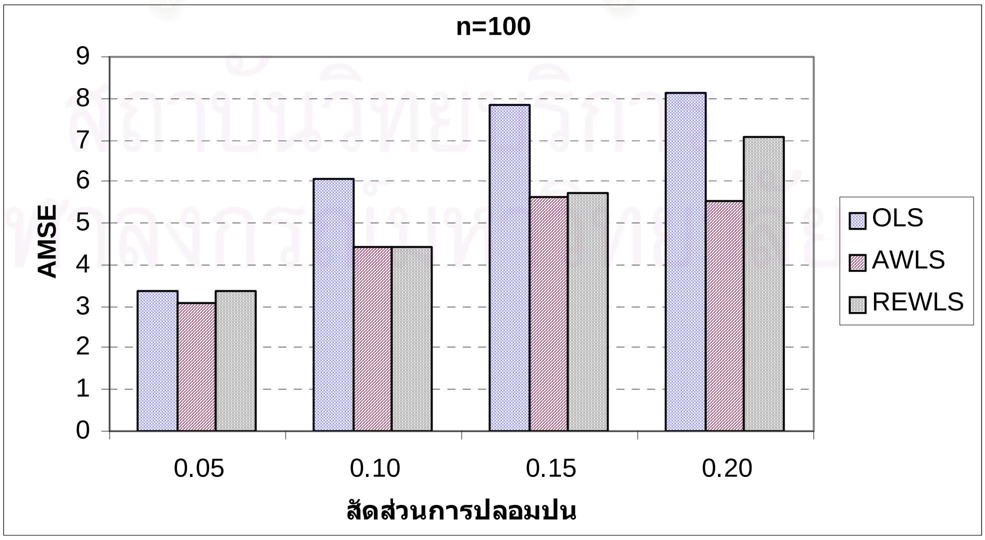

n=60
AMSE
16
14
12
10
8
6
4
2
0
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS


n=70
AMSE
14
12
10
8
6
4
2
0
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS

รูปที่ 4.3.2 (ต่อ)


n=80
AMSE
0
2
4
6
8
10
12
14
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS


n=90
AMSE
0
2
4
6
8
10
12
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS


n=100
AMSE
0
1
2
3
4
5
6
7
8
9
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS

จากตารางที่ 4.3.2 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS เมื่อตัวแปรตามมีค่าผิดปกติในระดับไม่รุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,8) โดยจำแนกตามขนาดตัวอย่างได้ดังนี้

กรณีขนาดตัวอย่างเท่ากับ 20 เมื่อสัดส่วนการปลอมปนเท่ากับ 0.05 พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด และตัวประมาณ AWLS ให้ค่า AMSE สูงที่สุด ในขณะที่สัดส่วนการปลอมปนเท่ากับ 0.1 ตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด และตัวประมาณ OLS ให้ค่า AMSE สูงที่สุด ส่วนในกรณีที่สัดส่วนการปลอมปนเท่ากับ 0.15 และ 0.2 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาเมื่อสัดส่วนการปลอมปนเท่ากับ 0.15 ตัวประมาณ REWLS ให้ค่า AMSE ต่ำกว่าตัวประมาณ OLS ในขณะที่สัดส่วนการปลอมปนเท่ากับ 0.2 พบว่าตัวประมาณ OLS ให้ค่า AMSE ต่ำกว่าตัวประมาณ REWLS

กรณีขนาดตัวอย่างเท่ากับ 30 เมื่อสัดส่วนการปลอมปนเท่ากับ 0.05 ตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ AWLS และตัวประมาณ OLS ตามลำดับ ในขณะที่สัดส่วนการปลอมปนเท่ากับ 0.1, 0.15 และ 0.2 พบว่าตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาเมื่อสัดส่วนการปลอมปนเท่ากับ 0.1 และ 0.15 ตัวประมาณ REWLS ให้ค่า AMSE ต่ำกว่าตัวประมาณ OLS ในขณะที่สัดส่วนการปลอมปนเท่ากับ 0.2 ตัวประมาณ OLS ให้ค่า AMSE ต่ำกว่าตัวประมาณ REWLS

กรณีขนาดตัวอย่างเท่ากับ 40, 50 และ 60 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุดในทุกๆ สัดส่วนการปลอมปน รองลงมาเมื่อสัดส่วนการปลอมปนเท่ากับ 0.05, 0.1 และ 0.15 พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำกว่าตัวประมาณ OLS ในขณะที่สัดส่วนการปลอมปนเท่ากับ 0.2 ตัวประมาณ OLS ให้ค่า AMSE ต่ำกว่าตัวประมาณ REWLS

กรณีขนาดตัวอย่างเท่ากับ 70, 80, 90 และ 100 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุดในทุกๆ สัดส่วนการปลอมปน รองลงมาคือตัวประมาณ REWLS และตัวประมาณ OLS ตามลำดับ

ข้อสรุป

ในกรณีขนาดตัวอย่างเท่ากับ 20 เมื่อสัดส่วนการปลอมปนเท่ากับ 0.05 และ 0.1 พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด ในขณะที่สัดส่วนการปลอมปนเท่ากับ 0.15 และ 0.2 พบว่าตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด

กรณีขนาดตัวอย่างเท่ากับ 30 สัดส่วนการปลอมปนเท่ากับ 0.05 พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด ในขณะที่สัดส่วนการปลอมปนเท่ากับ 0.1, 0.15 และ 0.2 พบว่าตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด

เมื่อตัวอย่างเท่ากับ 40, 50 และ 60 ในทุกๆ สัดส่วนการปลอมปน ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด และตัวประมาณ OLS ให้ค่า AMSE สูงที่สุดเมื่อสัดส่วนการปลอมปนเท่ากับ 0.05, 0.1 และ 0.15 ในขณะที่ตัวประมาณ REWLS ให้ค่า AMSE สูงที่สุดเมื่อสัดส่วนการปลอมปนเท่ากับ 0.2

ในขณะที่ขนาดตัวอย่างเท่ากับ 70, 80, 90 และ 100 ตัวประมาณ AWLS ให้ค่า AMSE ต่ำที่สุด และตัวประมาณ OLS ให้ค่า AMSE สูงที่สุดในทุกๆ สัดส่วนการปลอมปน

นอกจากนี้ยังพบว่าเมื่อสัดส่วนการปลอมปนเพิ่มขึ้นในทุกๆ ขนาดตัวอย่าง ค่า AMSE ของทุกตัวประมาณมีแนวโน้มเพิ่มขึ้น

<u>ส่วนที่ 3</u> ผลการเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรตามในระดับรุนแรง และไม่มีค่าผิดปกติในตัวแปรอิสระ

การวิจัยในส่วนนี้ ผู้วิจัยทำการศึกษาการเกิดค่าผิดปกติของตัวแปรตาม 2 กรณี คือ

1. กรณีเกิดค่าผิดปกติจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ $N(0,10(12)^2)$

2. กรณีเกิดค่าผิดปกติจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,25)

4.4 การเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรตามในระดับรุนแรง กรณีเกิดค่าผิดปกติจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ $N(0,10(12)^2)$

กรณีเกิดค่าผิดปกติจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ $N(0,10C^2)$ เมื่อตัวแปรตามมีระดับค่าผิดปกติระดับรุนแรง (C=12) ทำการศึกษาที่สัดส่วนการปลอมปนเท่ากับ 0.05, 0.10, 0.15 และ 0.20 โดยกำหนดขนาดตัวอย่างเท่ากับ 20, 30, 40, 50, 60, 70, 80, 90 และ 100 ซึ่งผลการวิจัยโดยจำแนกตามสัดส่วนการปลอมปนได้นำเสนอในตารางที่ 4.4.1 และกราฟรูปที่ 4.4.1

ตารางที่ 4.4.1 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ $N(0,10(12)^2)$ โดยจำแนกตามสัดส่วนการปลอมปน

สัดส่วนการปลอมปน p	ขนาดตัวอย่าง n	ตัวประมาณ OLS (S.D.)	ตัวประมาณ AWLS (S.D.)	ตัวประมาณ REWLS (S.D.)
0.05	20	203.195 (351.417)	27.533 (47.572)	14.418 * (24.911)
	30	108.663 (187.572)	26.546 (45.895)	13.580 * (23.469)
	40	75.491 (130.395)	8.891 (15.349)	7.330 * (12.654)
	50	65.008 (112.284)	8.627 (14.890)	6.635 * (11.454)
	60	56.492 (97.576)	6.596 (11.390)	5.308 * (9.164)
	70	45.293 (78.254)	5.785 (9.993)	4.844 * (8.370)
	80	44.640 (77.106)	4.743 (8.193)	4.020 * (6.943)
	90	29.884 (51.587)	3.760 (6.491)	3.288 * (5.678)
	100	27.572 (47.579)	3.306 (5.707)	2.848 * (4.917)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.4.1 (ต่อ)

สัดส่วนการปลอมปน p	ขนาดตัวอย่าง n	ตัวประมาณ		
		OLS (S.D.)	AWLS (S.D.)	REWLS (S.D.)
0.1	20	273.037 (471.640)	75.607 (130.717)	42.227 * (72.989)
	30	241.487 (417.329)	29.215 (50.486)	20.674 * (35.727)
	40	142.295 (245.780)	14.752 (25.450)	9.173 * (15.841)
	50	125.531 (216.739)	14.411 (24.882)	8.253 * (14.241)
	60	103.664 (179.058)	12.401 (21.420)	6.198 * (10.704)
	70	89.366 (154.388)	11.745 (20.290)	5.839 * (10.089)
	80	87.685 (151.486)	10.082 (17.417)	5.088 * (8.790)
	90	62.747 (108.354)	6.552 (11.315)	3.371 * (5.821)
	100	60.908 (105.169)	5.422 (9.361)	2.894 * (4.997)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.4.1 (ต่อ)

สัดส่วนการปลอมปน p	ขนาดตัวอย่าง n	ตัวประมาณ		
		OLS (S.D.)	AWLS (S.D.)	REWLS (S.D.)
0.15	20	423.299 (731.299)	144.664 (250.117)	121.230 * (209.579)
	30	399.142 (689.821)	64.870 (112.106)	54.346 * (93.927)
	40	202.223 (349.196)	26.719 (46.127)	19.363 * (33.425)
	50	194.351 (335.671)	26.536 (45.786)	15.092 * (26.062)
	60	144.854 (250.213)	21.901 (37.823)	8.667 * (14.969)
	70	124.589 (215.269)	21.398 (36.978)	7.936 * (13.713)
	80	122.810 (212.141)	15.583 (26.917)	6.979 * (12.058)
	90	87.840 (151.641)	10.964 (18.924)	4.563 * (7.880)
	100	86.578 (149.532)	10.582 (18.278)	4.403 * (7.605)

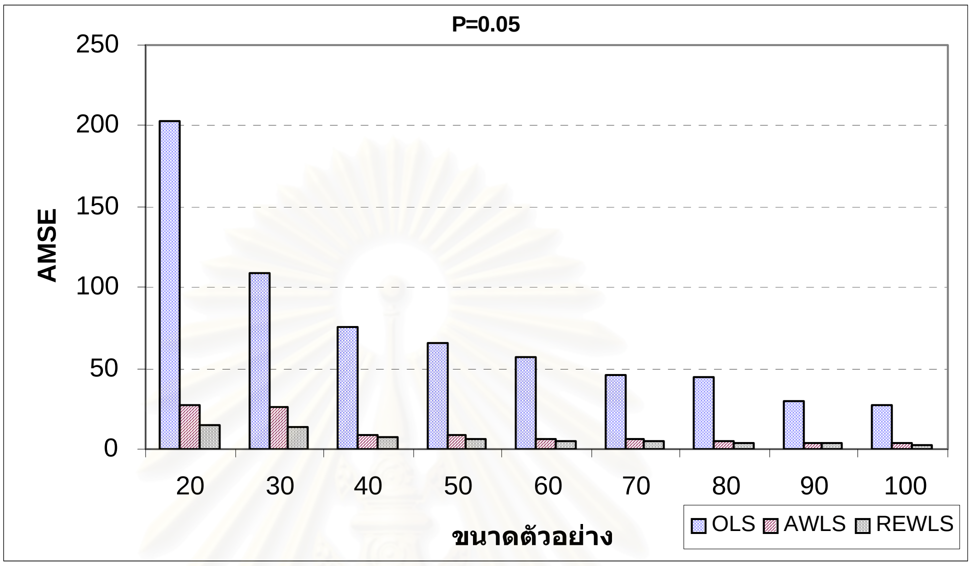
* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

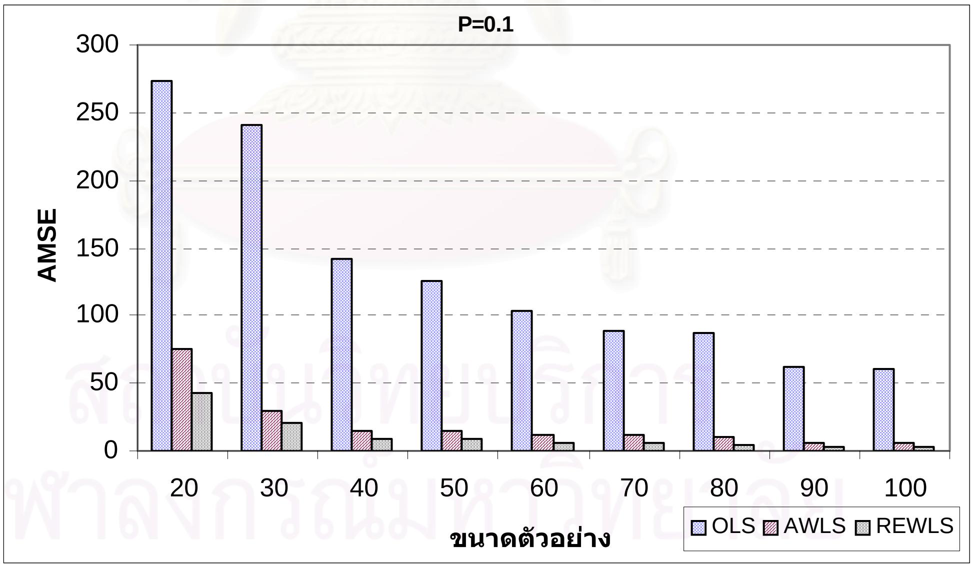
ตารางที่ 4.4.1 (ต่อ)

สัดส่วนการปลอมปน p	ขนาดตัวอย่าง n	ตัวประมาณ		
		OLS (S.D.)	AWLS (S.D.)	REWLS (S.D.)
0.2	20	622.903 (1076.371)	292.400 (505.534)	200.985 * (347.410)
	30	437.621 (756.079)	89.636 (154.851)	74.657 * (128.975)
	40	296.085 (511.266)	55.401 (95.622)	41.053 * (70.871)
	50	285.168 (492.471)	53.452 (92.280)	31.031 * (53.543)
	60	207.669 (358.626)	40.490 (69.906)	21.872 * (37.771)
	70	185.890 (321.118)	35.974 (62.150)	14.489 * (25.033)
	80	155.686 (268.878)	25.203 (43.527)	10.570 * (18.252)
	90	136.151 (235.137)	20.316 (35.081)	7.865 * (13.585)
	100	124.013 (214.170)	18.079 (31.220)	6.471 * (11.174)

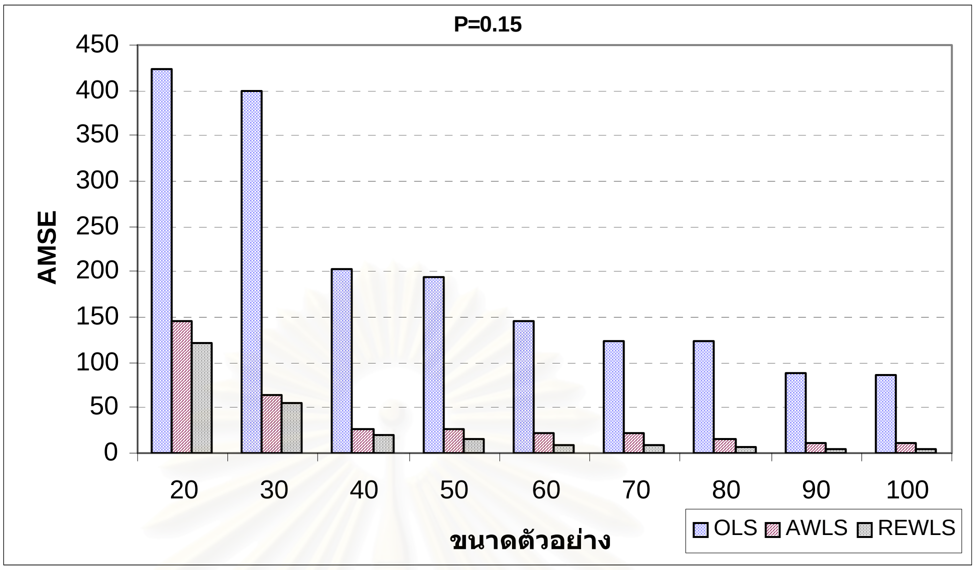
* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

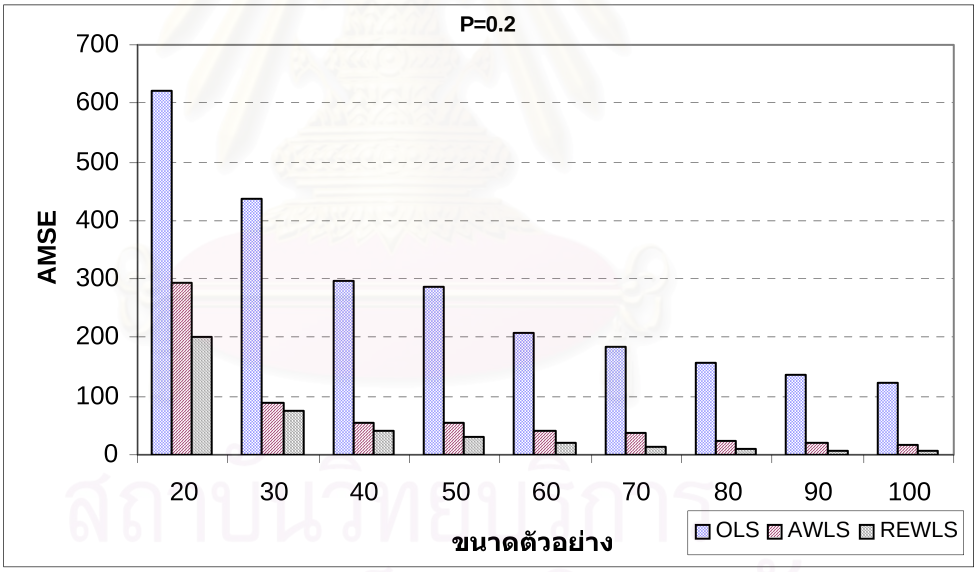
รูปที่ 4.4.1 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ N(0,10(12)^2) โดยจำแนกตามสัดส่วนการปลอมปน





รูปที่ 4.4.1 (ต่อ)


P=0.15
AMSE
450
400
350
300
250
200
150
100
50
0
20
30
40
50
60
70
80
90
100
ขนาดตัวอย่าง
OLS
AWLS
REWLS


P=0.2
AMSE
700
600
500
400
300
200
100
0
20
30
40
50
60
70
80
90
100
ขนาดตัวอย่าง
OLS
AWLS
REWLS

จากตารางที่ 4.4.1 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS เมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ N(0,10$(12)^2$) โดยจำแนกตามสัดส่วนการปลอมปนได้ดังนี้

เมื่อสัดส่วนการปลอมปนเท่ากับ 0.05, 0.1, 0.15 และ 0.2 ในทุกๆ ขนาดตัวอย่าง (n=20, 30, 40, 50, 60, 70, 80, 90 และ 100) พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ AWLS และตัวประมาณ OLS ตามลำดับ และพบว่าเมื่อสัดส่วนการปลอมปนเท่ากับ 0.05 ขนาดตัวอย่างเท่ากับ 40, 50, 60, 70, 80, 90 และ 100 ตัวประมาณ REWLS และตัวประมาณ AWLS มีค่า AMSE ใกล้เคียงกัน

ข้อสรุป

ในทุกๆ สถานการณ์ของการทดลองเมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรง พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ AWLS และตัวประมาณ OLS ตามลำดับ นอกจากนี้ยังพบว่าเมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกๆ สัดส่วนการปลอมปน ค่า AMSE ของทุกตัวประมาณมีแนวโน้มลดลง

กรณีเกิดค่าผิดปกติจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ $N(0,10C^2)$ เมื่อตัวแปรตามมีระดับค่าผิดปกติระดับรุนแรง (C=12) ทำการศึกษาที่สัดส่วนการปลอมปนเท่ากับ 0.05, 0.1, 0.15 และ 0.2 โดยกำหนดขนาดตัวอย่างเท่ากับ 20, 30, 40, 50, 60, 70, 80, 90 และ 100 ซึ่งผลการวิจัยโดยจำแนกตามขนาดตัวอย่างได้นำเสนอในตารางที่ 4.4.2 และ กราฟรูปที่ 4.4.2

ตารางที่ 4.4.2 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ $N(0,10(12)^2)$ โดยจำแนกตามขนาดตัวอย่าง

		สัดส่วนการปลอมปน			
n	ตัวประมาณ	0.05	0.10	0.15	0.20
20	OLS	203.195	273.037	423.299	622.903
	(S.D.)	(351.417)	(471.640)	(731.299)	(1076.371)
	AWLS	27.533	75.607	144.664	292.400
	(S.D.)	(47.572)	(130.717)	(250.117)	(505.534)
	REWLS	14.418 *	42.227 *	121.230 *	200.985 *
	(S.D.)	(24.911)	(72.989)	(209.579)	(347.410)

		สัดส่วนการปลอมปน			
n	ตัวประมาณ	0.05	0.10	0.15	0.20
30	OLS	108.663	241.487	399.142	437.621
	(S.D.)	(187.572)	(417.329)	(689.821)	(756.079)
	AWLS	26.546	29.215	64.870	89.636
	(S.D.)	(45.895)	(50.486)	(112.106)	(154.851)
	REWLS	13.580 *	20.674 *	54.346 *	74.657 *
	(S.D.)	(23.469)	(35.727)	(93.927)	(128.975)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.4.2 **(ต่อ)**

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
40	OLS	75.491	142.295	202.223	296.085
	(S.D.)	(130.395)	(245.780)	(349.196)	(511.266)
	AWLS	8.891	14.752	26.719	55.401
	(S.D.)	(15.349)	(25.450)	(46.127)	(95.622)
	REWLS	7.330 *	9.173 *	19.363 *	41.053 *
	(S.D.)	(12.654)	(15.841)	(33.425)	(70.871)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
50	OLS	65.008	125.531	194.351	285.168
	(S.D.)	112.284	216.739	335.671	492.471
	AWLS	8.627	14.411	26.536	53.452
	(S.D.)	14.890	24.882	45.786	92.280
	REWLS	6.635 *	8.253 *	15.092 *	31.031 *
	(S.D.)	11.454	14.241	26.062	53.543

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
60	OLS	56.492	103.664	144.854	207.669
	(S.D.)	(97.576)	(179.058)	(250.213)	(358.626)
	AWLS	6.596	12.401	21.901	40.490
	(S.D.)	(11.390)	(21.420)	(37.823)	(69.906)
	REWLS	5.308 *	6.198 *	8.667 *	21.872 *
	(S.D.)	(9.164)	(10.704)	(14.969)	(37.771)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.4.2 **(ต่อ)**

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
70	OLS	45.293	89.366	124.589	185.890
	(S.D.)	(78.254)	(154.388)	(215.269)	(321.118)
	AWLS	5.785	11.745	21.398	35.974
	(S.D.)	(9.993)	(20.290)	(36.978)	(62.150)
	REWLS	4.844 *	5.839 *	7.936 *	14.489 *
	(S.D.)	(8.370)	(10.089)	(13.713)	(25.033)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
80	OLS	44.640	87.685	122.810	155.686
	(S.D.)	(77.106)	(151.486)	(212.141)	(268.878)
	AWLS	4.743	10.082	15.583	25.203
	(S.D.)	(8.193)	(17.417)	(26.917)	(43.527)
	REWLS	4.020 *	5.088 *	6.979 *	10.570 *
	(S.D.)	(6.943)	(8.790)	(12.058)	(18.252)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
90	OLS	29.884	62.747	87.840	136.151
	(S.D.)	(51.587)	(108.354)	(151.641)	(235.137)
	AWLS	3.760	6.552	10.964	20.316
	(S.D.)	(6.491)	(11.315)	(18.924)	(35.081)
	REWLS	3.288 *	3.371 *	4.563 *	7.865 *
	(S.D.)	(5.678)	(5.821)	(7.880)	(13.585)

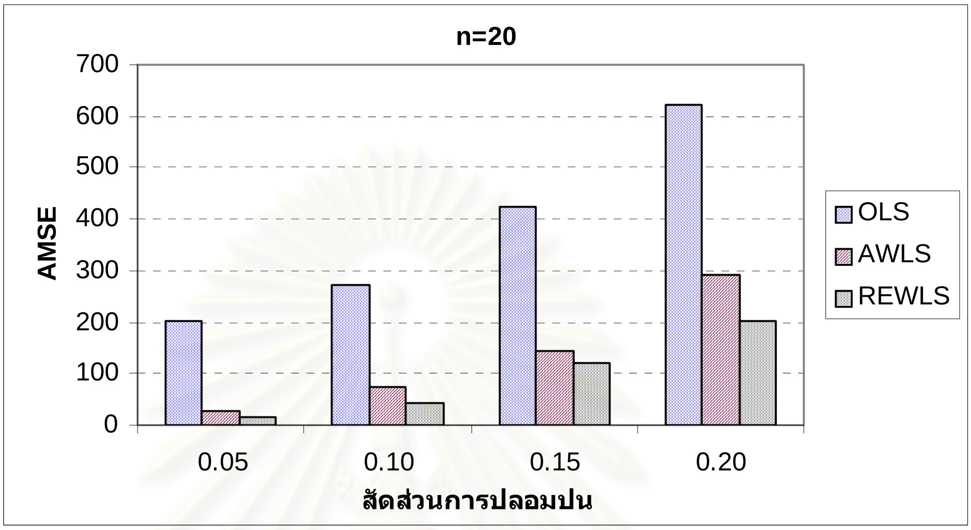
* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

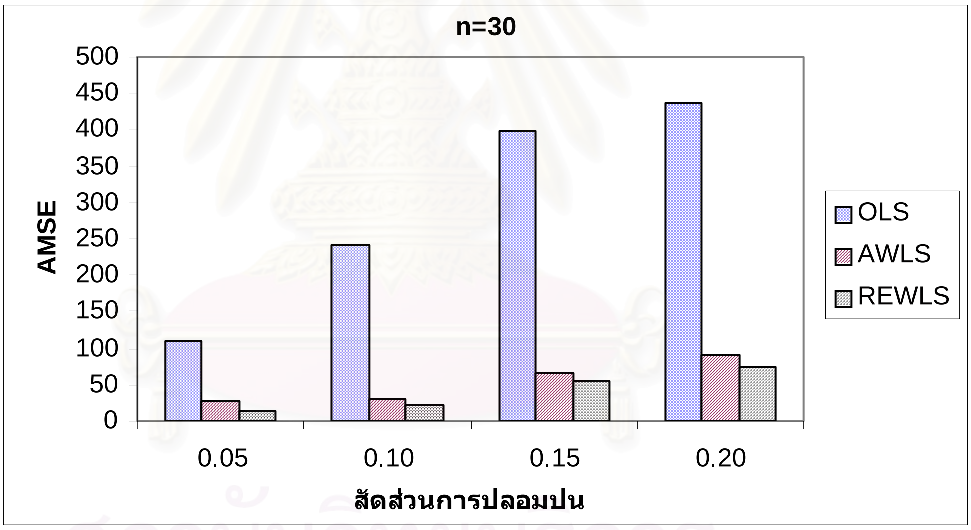
ตารางที่ 4.4.2 **(ต่อ)**

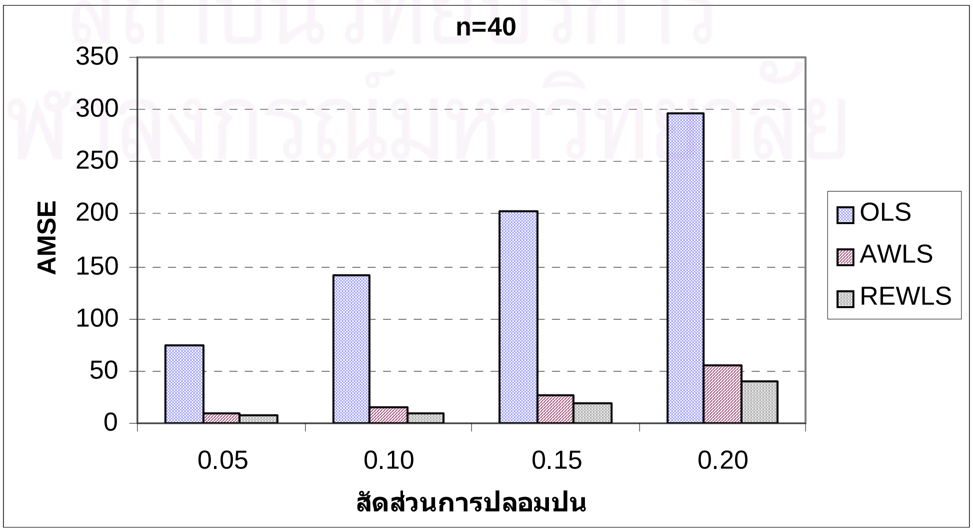
n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
100	OLS	27.572	60.908	86.578	124.013
	(S.D.)	(47.579)	(105.169)	(149.532)	(214.170)
	AWLS	3.306	5.422	10.582	18.079
	(S.D.)	(5.707)	(9.361)	(18.278)	(31.220)
	REWLS	2.848 *	2.894 *	4.403 *	6.471 *
	(S.D.)	(4.917)	(4.997)	(7.605)	(11.174)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

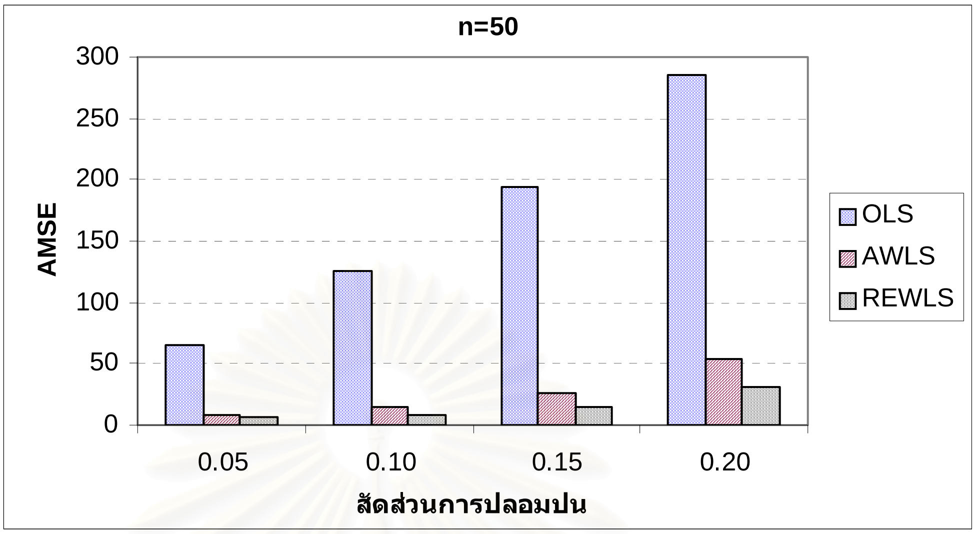
รูปที่ 4.4.2 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ $N(0,10(12)^2)$ โดยจำแนกตามขนาดตัวอย่าง

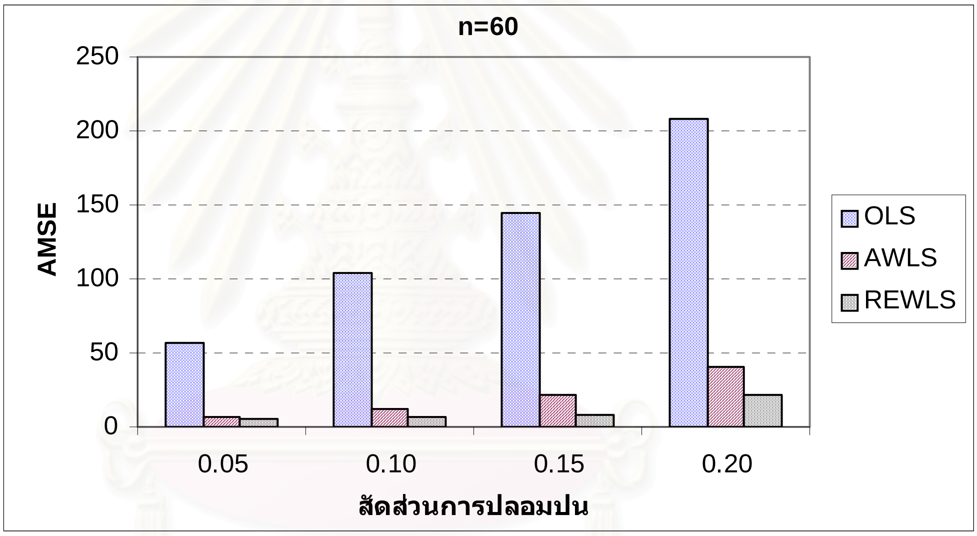


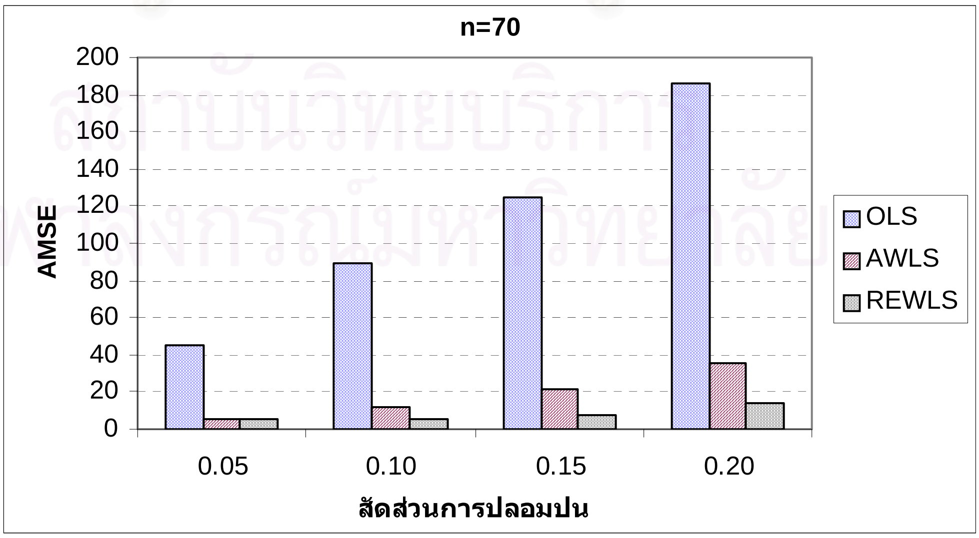




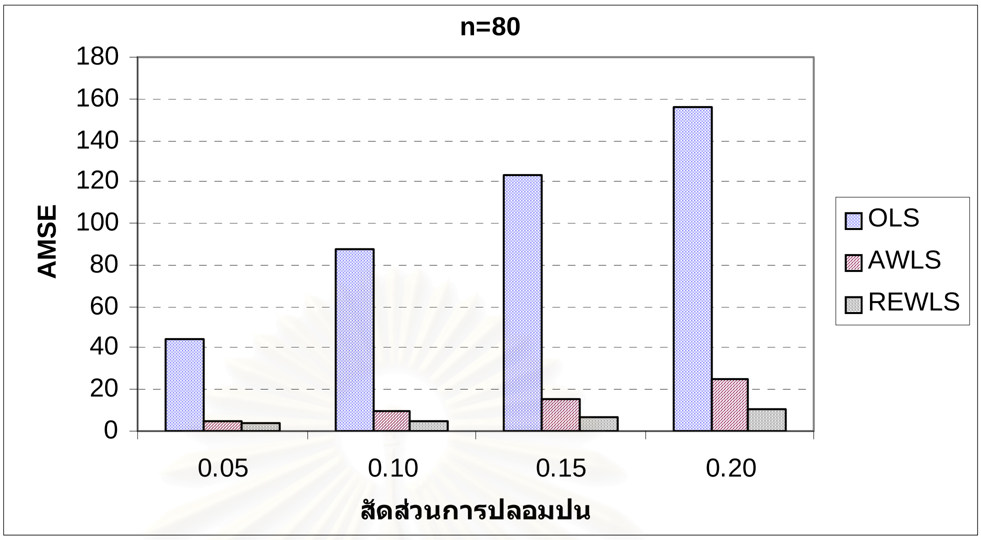
รูปที่ 4.4.2 **(ต่อ)**

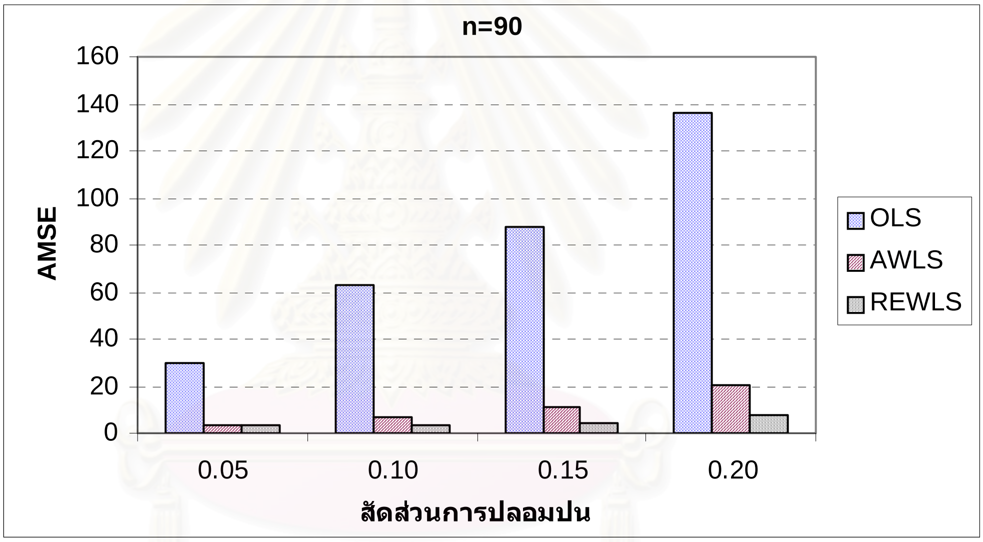

n=50
AMSE
300
250
200
150
100
50
0
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS

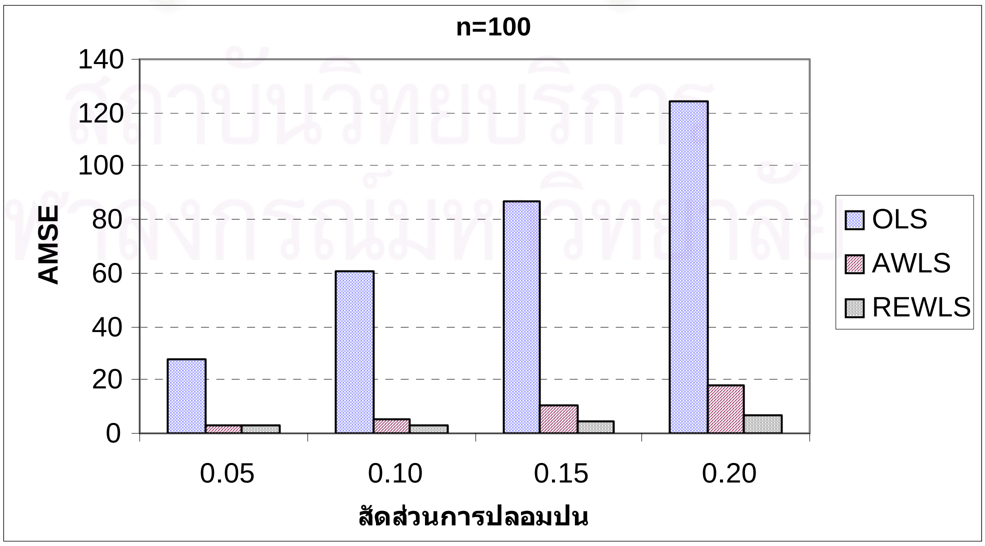

n=60
AMSE
250
200
150
100
50
0
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS


n=70
AMSE
200
180
160
140
120
100
80
60
40
20
0
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS

รูปที่ 4.4.2 (ต่อ)


n=80
180
160
140
120
100
80
60
40
20
0
AMSE
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS


n=90
160
140
120
100
80
60
40
20
0
AMSE
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS


n=100
140
120
100
80
60
40
20
0
AMSE
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS

จากตารางที่ 4.4.2 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS เมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ $N(0,10(12)^2)$ โดยจำแนกตามขนาดตัวอย่างได้ดังนี้

ในทุกๆ ขนาดตัวอย่าง (n = 20, 30, 40, 50, 60, 70, 80, 90 และ 100) และทุกๆ สัดส่วนการปลอมปน พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ AWLS และตัวประมาณ OLS ตามลำดับ นอกจากนี้ยังพบว่า เมื่อขนาดตัวอย่างเท่ากับ 40, 50, 60, 70, 80, 90 และ 100 สัดส่วนการปลอมปนเท่ากับ 0.05 ตัวประมาณ REWLS และตัวประมาณ AWLS ให้ค่า AMSE ใกล้เคียงกัน

ข้อสรุป

ในทุกๆ สถานการณ์ของการทดลองเมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรง พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ AWLS และตัวประมาณ OLS ตามลำดับ นอกจากนี้ยังพบว่าเมื่อสัดส่วนการปลอมปนเพิ่มขึ้นในทุกๆ ขนาดตัวอย่าง ค่า AMSE ของทุกตัวประมาณมีแนวโน้มเพิ่มขึ้น

4.5 การเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรตามในระดับรุนแรง กรณีเกิดค่าผิดปกติจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,25)

กรณีเกิดค่าผิดปกติในระดับรุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,25) ทำการศึกษาที่สัดส่วนการปลอมปนเท่ากับ 0.05, 0.1, 0.15 และ 0.2 โดยกำหนดขนาดตัวอย่างเท่ากับ 20, 30, 40, 50, 60, 70, 80, 90 และ 100 ซึ่งผลการวิจัยโดยจำแนกตามสัดส่วนการปลอมปนได้นำเสนอในตารางที่ 4.5.1 และกราฟรูปที่ 4.5.1

ตารางที่ 4.5.1 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,25) โดยจำแนกตามสัดส่วนการปลอมปน

สัดส่วนการ		ตัวประมาณ		
ปลอมปน	ขนาดตัวอย่าง	OLS	AWLS	REWLS
p	n	(S.D.)	(S.D.)	(S.D.)
0.05	20	186.374 (321.982)	37.311 (64.481)	18.603 * (32.148)
	30	157.171 (271.723)	26.078 (45.083)	14.345 * (24.793)
	40	87.967 (151.933)	9.798 (16.918)	7.760 * (13.400)
	50	63.251 (109.242)	9.474 (16.358)	5.937 * (10.248)
	60	52.216 (90.191)	6.704 (11.578)	5.464 * (9.436)
	70	49.855 (86.144)	6.410 (11.076)	5.167 * (8.928)
	80	46.261 (79.904)	4.681 (8.085)	4.085 * (7.057)
	90	34.572 (59.674)	4.246 (7.333)	3.554 * (6.137)
	100	32.936 (56.858)	3.638 (6.280)	3.310 * (5.716)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.5.1 (ต่อ)

สัดส่วนการปลอมปน p	ขนาดตัวอย่าง n	ตัวประมาณ OLS (S.D.)	AWLS (S.D.)	REWLS (S.D.)
0.1	20	281.589 (486.484)	51.354 (88.724)	44.711 * (77.270)
	30	239.520 (413.939)	29.070 (50.232)	19.304 * (33.355)
	40	141.897 (245.002)	14.653 (25.301)	10.596 * (18.294)
	50	128.831 (222.500)	14.499 (25.033)	8.094 * (13.972)
	60	106.728 (184.365)	11.680 (20.173)	6.258 * (10.808)
	70	105.991 (183.112)	11.200 (19.349)	5.952 * (10.282)
	80	84.363 (145.708)	8.773 (15.155)	4.445 * (7.678)
	90	65.306 (112.781)	7.217 (12.465)	4.247 * (7.336)
	100	63.910 (110.343)	5.654 (9.761)	3.392 * (5.857)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.5.1 (ต่อ)

สัดส่วนการปลอมปน p	ขนาดตัวอย่าง n	ตัวประมาณ OLS (S.D.)	AWLS (S.D.)	REWLS (S.D.)
0.15	20	462.509 (799.093)	111.009 (191.856)	101.710 * (175.799)
	30	440.236 (760.878)	71.331 (123.298)	54.346 * (93.927)
	40	223.159 (385.395)	31.691 (54.718)	18.073 * (31.197)
	50	205.352 (354.700)	28.481 (49.168)	17.133 * (29.577)
	60	156.817 (270.823)	20.094 (34.695)	8.906 * (15.377)
	70	136.211 (235.320)	19.600 (33.857)	8.439 * (14.580)
	80	116.970 (202.048)	13.676 (23.623)	6.359 * (10.984)
	90	92.715 (160.131)	12.057 (20.824)	5.843 * (10.093)
	100	90.266 (155.885)	10.986 (18.973)	4.764 * (8.227)

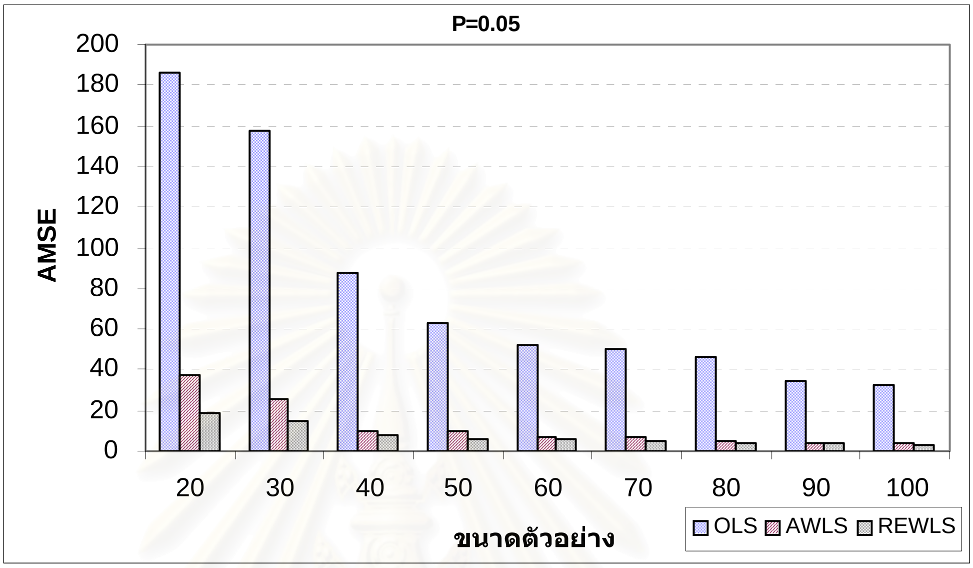
* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

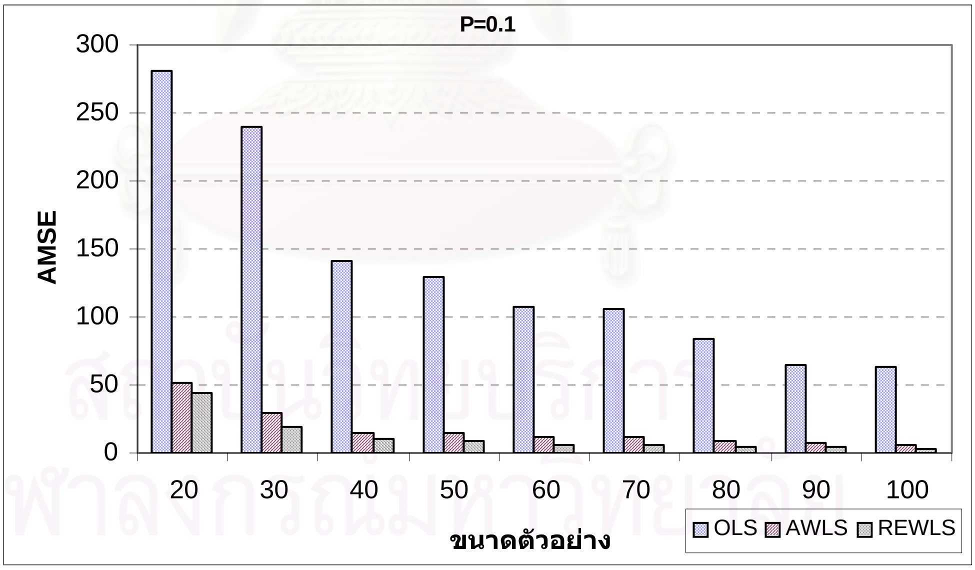
ตารางที่ 4.5.1 (ต่อ)

สัดส่วนการปลอมปน p	ขนาดตัวอย่าง n	ตัวประมาณ OLS (S.D.)	AWLS (S.D.)	REWLS (S.D.)
0.2	20	651.649 (1125.829)	203.464 (351.604)	167.000 * (288.549)
	30	519.751 (898.168)	99.309 (171.645)	92.687 * (160.197)
	40	289.387 (499.669)	58.136 (100.342)	37.949 * (65.521)
	50	262.114 (452.632)	50.588 (87.340)	35.466 * (61.237)
	60	228.890 (395.356)	42.698 (73.752)	23.523 * (40.630)
	70	219.866 (379.856)	40.218 (69.487)	20.278 * (35.033)
	80	175.439 (303.049)	22.367 (38.634)	14.659 * (25.320)
	90	150.921 (260.646)	19.564 (33.790)	11.825 * (20.425)
	100	137.903 (238.164)	18.150 (31.347)	10.249 * (17.700)

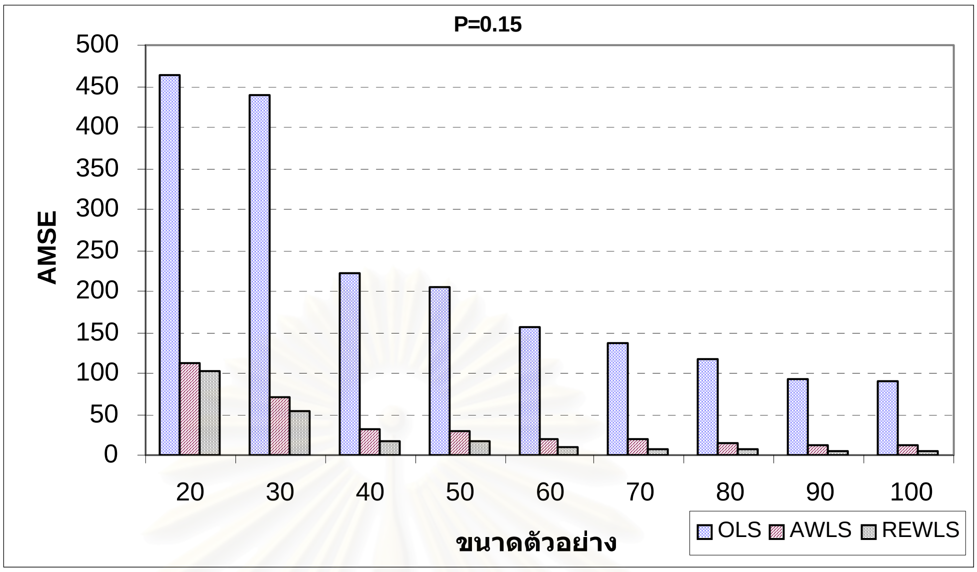
* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

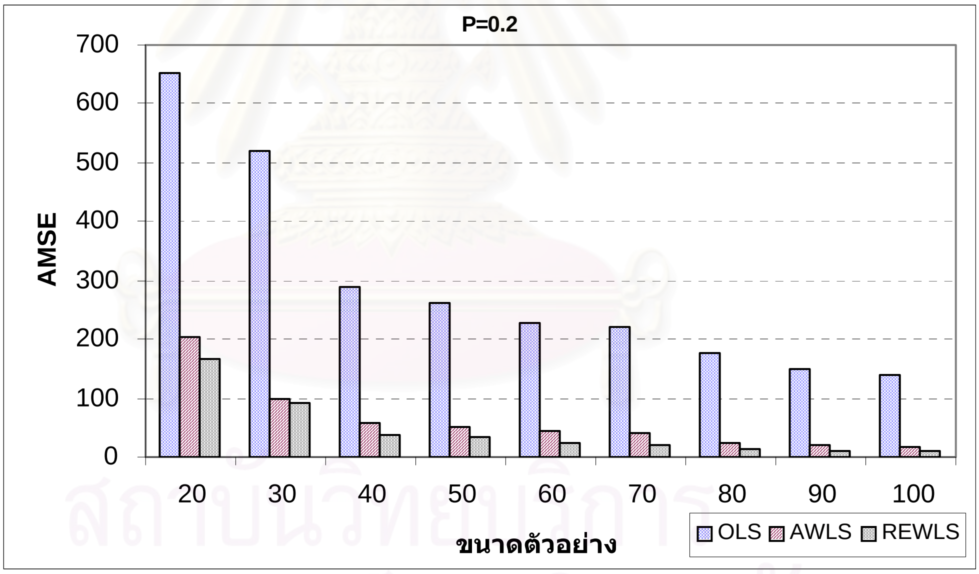
รูปที่ 4.5.1 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,25) โดยจำแนกตามสัดส่วนการปลอมปน





รูปที่ 4.5.1 (ต่อ)


P=0.15
AMSE
500
450
400
350
300
250
200
150
100
50
0
20
30
40
50
60
70
80
90
100
ขนาดตัวอย่าง
OLS
AWLS
REWLS


P=0.2
AMSE
700
600
500
400
300
200
100
0
20
30
40
50
60
70
80
90
100
ขนาดตัวอย่าง
OLS
AWLS
REWLS

จากตารางที่ 4.5.1 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS เมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,25) โดยจำแนกตามสัดส่วนการปลอมปนได้ดังนี้

เมื่อสัดส่วนการปลอมปนเท่ากับ 0.05, 0.1, 0.15 และ 0.2 ในทุกๆ ขนาดตัวอย่าง (n=20, 30, 40, 50, 60, 70, 80, 90 และ 100) พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด และตัวประมาณ OLS ให้ค่า AMSE สูงที่สุด และพบว่าเมื่อสัดส่วนการปลอมปนเท่ากับ 0.05 ขนาดตัวอย่างเท่ากับ 40, 50, 60, 70, 80, 90 และ 100 ตัวประมาณ REWLS และตัวประมาณ AWLS ให้ค่า AMSE ใกล้เคียงกัน

ข้อสรุป

ในทุกๆ สถานการณ์ของการทดลองเมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรง พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ AWLS และตัวประมาณ OLS ตามลำดับ นอกจากนี้ยังพบว่าเมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกๆ สัดส่วนการปลอมปน ค่า AMSE ของทุกตัวประมาณมีแนวโน้มลดลง

กรณีเกิดค่าผิดปกติในระดับรุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,25) ทำการศึกษาที่สัดส่วนการปลอมปนเท่ากับ 0.05, 0.1, 0.15 และ 0.2 โดยกำหนดขนาดตัวอย่างเท่ากับ 20, 30, 40, 50, 60, 70, 80, 90 และ 100 ซึ่งผลการวิจัยโดยจำแนกตามขนาดตัวอย่างได้นำเสนอในตารางที่ 4.5.2 และกราฟรูปที่ 4.5.2

ตารางที่ 4.5.2 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,25) โดยจำแนกตามขนาดตัวอย่าง

		สัดส่วนการปลอมปน			
n	ตัวประมาณ	0.05	0.10	0.15	0.20
20	OLS	186.374	281.589	462.509	651.649
	(S.D.)	(321.982)	(486.484)	(799.093)	(1125.829)
	AWLS	37.311	51.354	111.009	203.464
	(S.D.)	(64.481)	(88.724)	(191.856)	(351.604)
	REWLS	18.603 *	44.711 *	101.710 *	167.000 *
	(S.D.)	(32.148)	(77.270)	(175.799)	(288.549)

		สัดส่วนการปลอมปน			
n	ตัวประมาณ	0.05	0.10	0.15	0.20
30	OLS	157.171	239.520	440.236	519.751
	(S.D.)	(271.723)	(413.939)	(760.878)	(898.168)
	AWLS	26.078	29.070	71.331	99.309
	(S.D.)	(45.083)	(50.232)	(123.298)	(171.645)
	REWLS	14.345 *	19.304 *	54.346 *	92.687 *
	(S.D.)	(24.793)	(33.355)	(93.927)	(160.197)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.5.2 (ต่อ)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
40	OLS	87.967	141.897	223.159	289.387
	(S.D.)	(151.933)	(245.002)	(385.395)	(499.669)
	AWLS	9.798	14.653	31.691	58.136
	(S.D.)	(16.918)	(25.301)	(54.718)	(100.342)
	REWLS	7.760 *	10.596 *	18.073 *	37.949 *
	(S.D.)	(13.400)	(18.294)	(31.197)	(65.521)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
50	OLS	63.251	128.831	205.352	262.114
	(S.D.)	(109.242)	(222.500)	(354.700)	(452.632)
	AWLS	9.474	14.499	28.481	50.588
	(S.D.)	(16.358)	(25.033)	(49.168)	(87.340)
	REWLS	5.937 *	8.094 *	17.133 *	35.466 *
	(S.D.)	(10.248)	(13.972)	(29.577)	(61.237)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
60	OLS	52.216	106.728	156.817	228.890
	(S.D.)	(90.191)	(184.365)	(270.823)	(395.356)
	AWLS	6.704	11.680	20.094	42.698
	(S.D.)	(11.578)	(20.173)	(34.695)	(73.752)
	REWLS	5.464 *	6.258 *	8.906 *	23.523 *
	(S.D.)	(9.436)	(10.808)	(15.377)	(40.630)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่ 4.5.2 (ต่อ)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
70	OLS	49.855	105.991	136.211	219.866
	(S.D.)	(86.144)	(183.112)	(235.320)	(379.856)
	AWLS	6.410	11.200	19.600	40.218
	(S.D.)	(11.076)	(19.349)	(33.857)	(69.487)
	REWLS	5.167 *	5.952 *	8.439 *	20.278 *
	(S.D.)	(8.928)	(10.282)	(14.580)	(35.033)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
80	OLS	46.261	84.363	116.970	175.439
	(S.D.)	(79.904)	(145.708)	(202.048)	(303.049)
	AWLS	4.681	8.773	13.676	22.367
	(S.D.)	(8.085)	(15.155)	(23.623)	(38.634)
	REWLS	4.085 *	4.445 *	6.359 *	14.659 *
	(S.D.)	(7.057)	(7.678)	(10.984)	(25.320)

n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
90	OLS	34.572	65.306	92.715	150.921
	(S.D.)	(59.674)	(112.781)	(160.131)	(260.646)
	AWLS	4.246	7.217	12.057	19.564
	(S.D.)	(7.333)	(12.465)	(20.824)	(33.790)
	REWLS	3.554 *	4.247 *	5.843 *	11.825 *
	(S.D.)	(6.137)	(7.336)	(10.093)	(20.425)

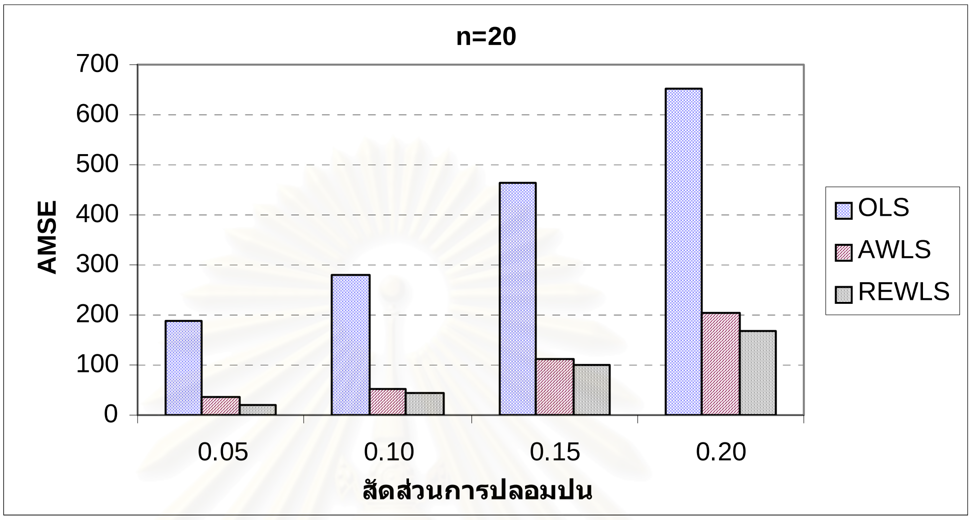
* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

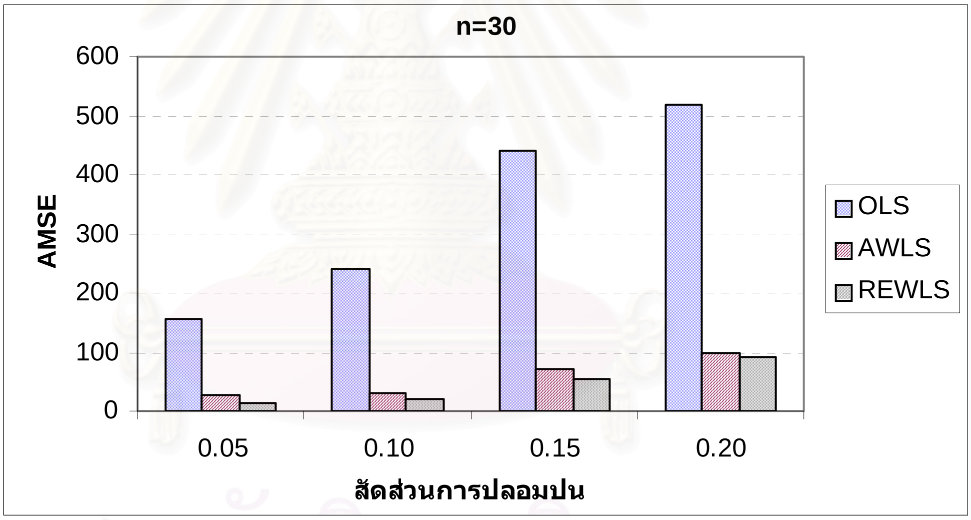
ตารางที่ 4.5.2 (ต่อ)

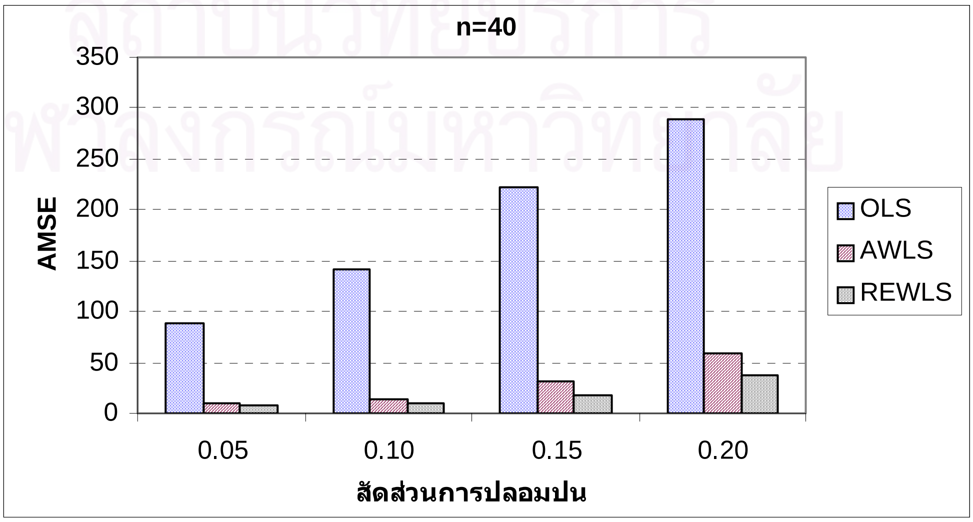
n	ตัวประมาณ	สัดส่วนการปลอมปน			
		0.05	0.10	0.15	0.20
100	OLS	32.936	63.910	90.266	137.903
	(S.D.)	(56.858)	(110.343)	(155.885)	(238.164)
	AWLS	3.638	5.654	10.986	18.150
	(S.D.)	(6.280)	(9.761)	(18.973)	(31.347)
	REWLS	3.310 *	3.392 *	4.764 *	10.249 *
	(S.D.)	(5.716)	(5.857)	(8.227)	(17.700)

* หมายถึง ตัวประมาณที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

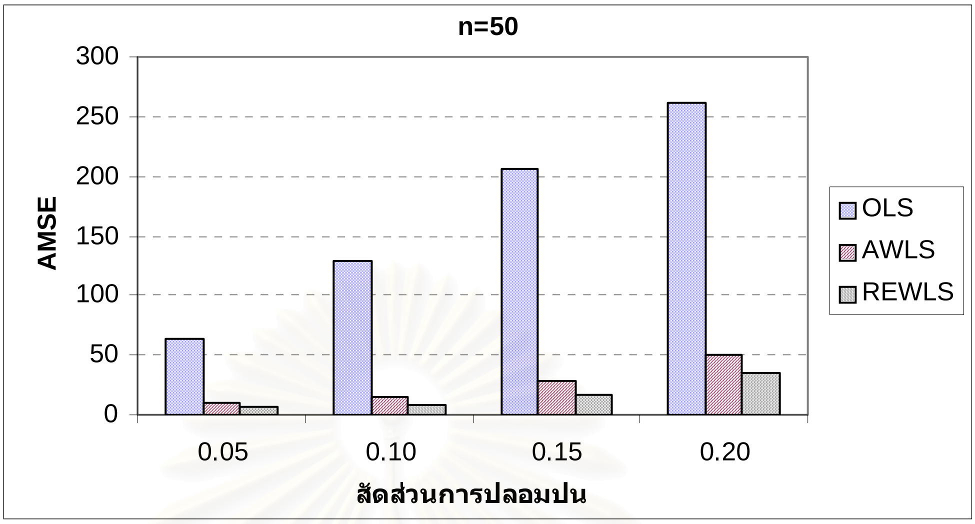
รูปที่ 4.5.2 การเปรียบเทียบค่าประมาณพารามิเตอร์จากตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ด้วยค่า AMSE เมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,25) โดยจำแนกตามขนาดตัวอย่าง

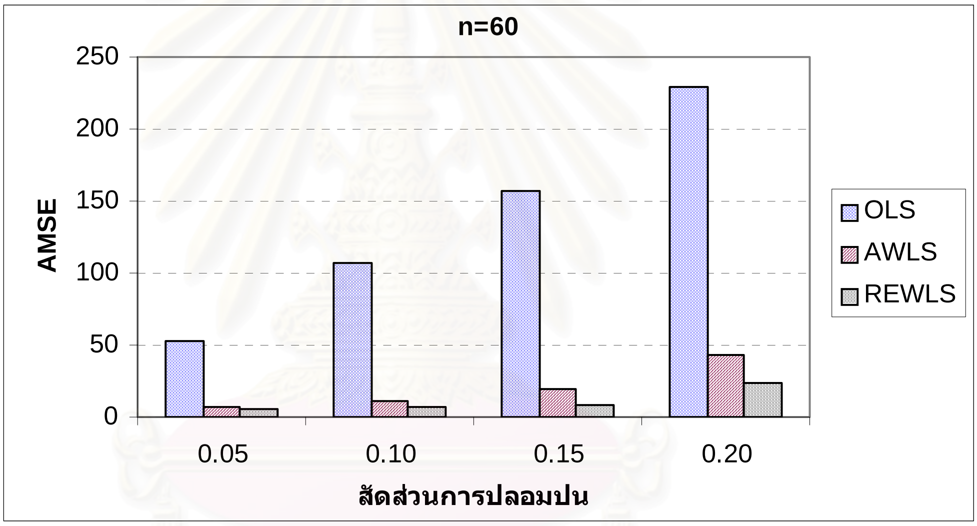


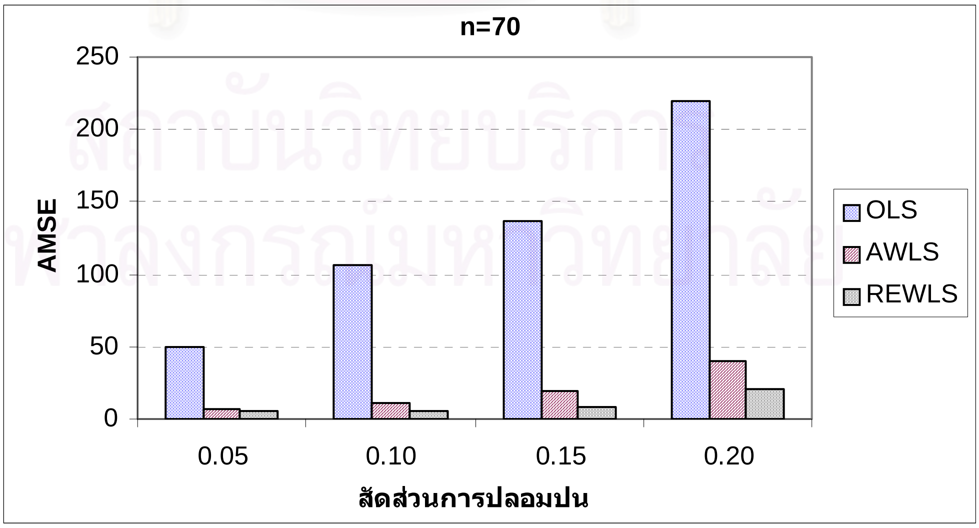




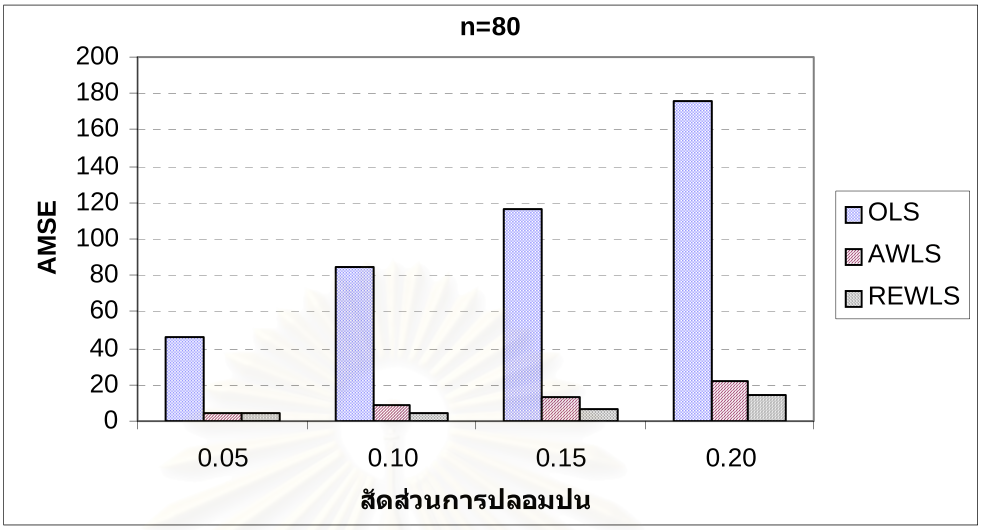
รูปที่ 4.5.2 (ต่อ)

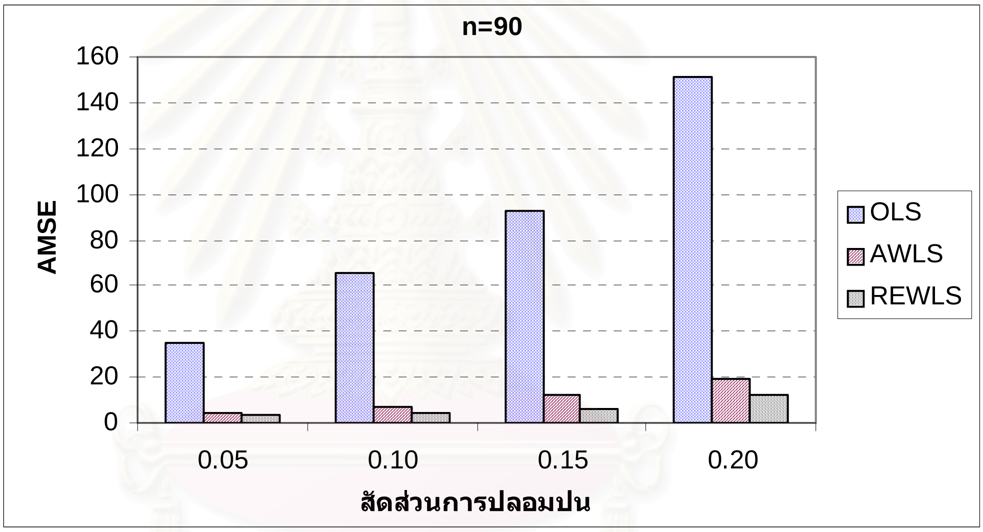

n=50
300
250
200
150
100
50
0
AMSE
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS

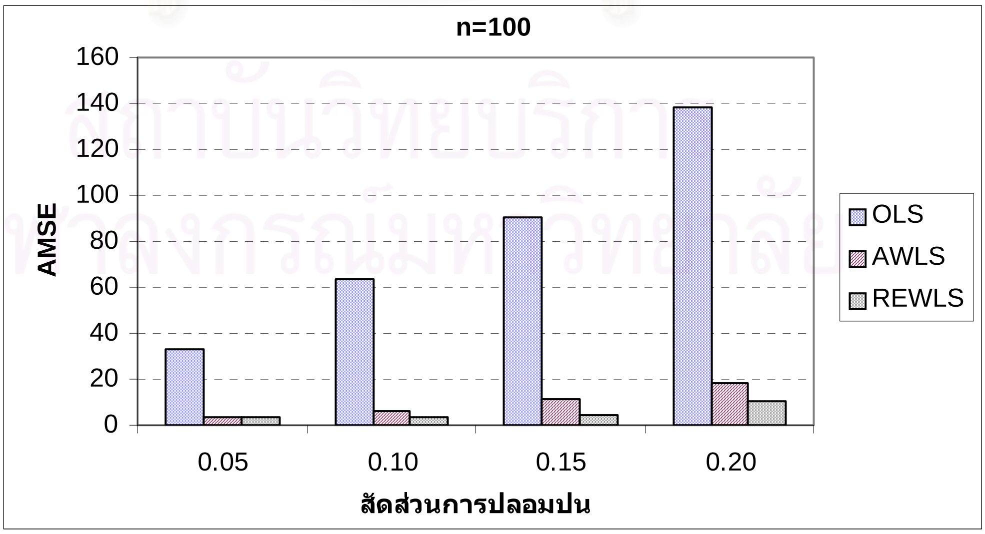

n=60
250
200
150
100
50
0
AMSE
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS


n=70
250
200
150
100
50
0
AMSE
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS

รูปที่ 4.5.2 (ต่อ)


n=80
200
180
160
140
120
100
80
60
40
20
0
AMSE
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS


n=90
160
140
120
100
80
60
40
20
0
AMSE
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS


n=100
160
140
120
100
80
60
40
20
0
AMSE
0.05
0.10
0.15
0.20
สัดส่วนการปลอมปน
OLS
AWLS
REWLS

จากตารางที่ 4.5.2 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS เมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรงจากการแจกแจงแบบปกติปลอมปนระหว่าง N(0,10) กับ L(0,25) โดยจำแนกตามขนาดตัวอย่างได้ดังนี้

ในทุกๆ ขนาดตัวอย่าง (n = 20, 30, 40, 50, 60, 70, 80, 90 และ 100) และทุกๆ สัดส่วนการปลอมปน พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ AWLS และตัวประมาณ OLS ตามลำดับ นอกจากนี้ยังพบว่า เมื่อขนาดตัวอย่างเท่ากับ 40, 50, 60, 70, 80, 90 และ 100 สัดส่วนการปลอมปนเท่ากับ 0.05 ตัวประมาณ REWLS และตัวประมาณ AWLS ให้ค่า AMSE ใกล้เคียงกัน

ข้อสรุป

ในทุกๆ สถานการณ์ของการทดลองเมื่อตัวแปรตามมีค่าผิดปกติในระดับรุนแรง พบว่าตัวประมาณ REWLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือตัวประมาณ AWLS และตัวประมาณ OLS ตามลำดับ นอกจากนี้ยังพบว่าเมื่อสัดส่วนการปลอมปนเพิ่มขึ้นในทุกๆ ขนาดตัวอย่าง ค่า AMSE ของทุกตัวประมาณมีแนวโน้มเพิ่มขึ้น

บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลองเพื่อการศึกษาและเปรียบเทียบประสิทธิภาพของตัวประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้น เมื่อมีค่าผิดปกติในตัวแปรตาม โดยผู้วิจัยได้ทำการเปรียบเทียบตัวประมาณ 3 ตัว ซึ่งได้แก่ ตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (OLSE) ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ (AWLSE) และตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ (REWLSE) โดยสถานการณ์ที่ศึกษามีดังนี้

1. ตัวแปรตามไม่มีค่าผิดปกติและมีค่าผิดปกติ
2. ระดับค่าผิดปกติ คือ ปานกลาง และรุนแรง
3. สัดส่วนการปลอมปน คือ 0.05, 0.10, 0.15 และ 0.20
4. ขนาดตัวอย่างที่ใช้ในการวิจัยคือ 20, 30, 40, 50, 60, 70, 80, 90 และ 100

การวิจัยครั้งนี้ผู้วิจัยศึกษาวิเคราะห์ด้วยเทคนิคมอนติคาร์โลโดยใช้โปรแกรมภาษาฟอร์แทน 77 บน PC Computer เพื่อสร้างข้อมูลตามสถานการณ์ที่กำหนดโดยกระทำซ้ำ 500 ครั้ง ในแต่ละสถานการณ์

สรุปผลการวิจัย

การวิจัยครั้งนี้ผู้วิจัยต้องการเปรียบเทียบตัวประมาณค่าพารามิเตอร์โดยพิจารณาจากค่า AMSE ของตัวประมาณทั้ง 3 ตัว ซึ่งจากผลการวิจัยพบว่า ระดับค่าผิดปกติ สัดส่วนการปลอมปน และขนาดตัวอย่าง ต่างก็มีผลต่อค่า AMSE ของตัวประมาณค่าพารามิเตอร์ทั้ง 3 ตัว โดยค่า AMSE มีแนวโน้มเพิ่มขึ้นเมื่อระดับค่าผิดปกติ หรือสัดส่วนการปลอมปนเพิ่มขึ้น แต่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น ผู้วิจัยได้สรุปผลการวิจัยออกเป็น 3 ส่วนดังนี้

<u>ส่วนที่ 1</u> ผลการเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่ไม่มีค่าผิดปกติในตัวแปรตามและในตัวแปรอิสระ

ผลการวิจัยในกรณีที่ไม่มีค่าผิดปกติในตัวแปรตามพบว่า ตัวประมาณ OLS ให้ค่า AMSE ต่ำที่สุด รองลงมาคือ ตัวประมาณ REWLS และตัวประมาณ AWLS ตามลำดับ เนื่องจากตัว

ประมาณ AWLS เป็นตัวประมาณที่ปรับค่าของข้อมูลที่ผิดปกติด้วยค่าถ่วงน้ำหนักเมื่อข้อมูลอยู่ต่ำกว่าเปอร์เซ็นไทล์ที่ 25 หรือสูงกว่าเปอร์เซ็นไทล์ที่ 75 ซึ่งในทางปฏิบัติมีความเป็นไปได้ว่าข้อมูลที่อยู่ต่ำกว่าเปอร์เซ็นไทล์ที่ 25 หรือสูงกว่าเปอร์เซ็นไทล์ที่ 75 อาจเป็นข้อมูลที่ปกติ ทำให้ค่าประมาณสัมประสิทธิ์การถดถอยที่ได้เบี่ยงเบนไปจากค่าจริง ส่งผลให้ค่า AMSE เพิ่มขึ้น

เมื่อตัวอย่างมีขนาดใหญ่ (n = 60, 70, 80, 90 และ 100) ตัวประมาณ OLS ตัวประมาณ AWLS และตัวประมาณ REWLS ให้ค่า AMSE ใกล้เคียงกัน และจะสังเกตได้ว่าค่า AMSE ของทุกวิธีมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น แต่ถึงอย่างไรก็ตามควรเลือกใช้ตัวประมาณ OLS ในการประมาณค่าพารามิเตอร์ในกรณีที่ไม่มีค่าผิดปกติ เนื่องจากตัวประมาณ OLS เป็นตัวประมาณที่มีกรรมวิธีคำนวณง่ายกว่า และได้ตัวประมาณไม่เอนเอียงเชิงเส้นที่ดีที่สุด

<u>ส่วนที่ 2</u> ผลการเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรตามในระดับไม่รุนแรง และไม่มีค่าผิดปกติในตัวแปรอิสระ

ผลการวิจัยในกรณีที่มีค่าผิดปกติในตัวแปรตามในระดับไม่รุนแรง พบว่าเมื่อสัดส่วนการปลอมปนน้อย ($p\in[0.05,0.10]$) และขนาดตัวอย่างมีขนาดเล็ก ($n\in[20,30]$) ตัวประมาณ REWLS ให้ประสิทธิภาพในการประมาณสูงที่สุด ในขณะที่สัดส่วนการปลอมปนน้อย ($p\in[0.05,0.10]$) และขนาดตัวอย่างเพิ่มขึ้น ($n\in(30,100]$) ตัวประมาณ AWLS ให้ประสิทธิภาพในการประมาณสูงที่สุด และสำหรับกรณีที่สัดส่วนการปลอมปนเพิ่มขึ้น ($p\in(0.10,0.20]$) ในทุกๆ ขนาดตัวอย่าง ($n\in[20,100]$) ตัวประมาณ AWLS ให้ประสิทธิภาพในการประมาณสูงที่สุด

นอกจากนี้ยังพบว่าเมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกๆ สัดส่วนการปลอมปน ค่า AMSE ของทุกตัวประมาณมีแนวโน้มลดลง ในขณะที่เมื่อสัดส่วนการปลอมปนเพิ่มขึ้นในทุกๆ ขนาดตัวอย่าง ค่า AMSE ของทุกตัวประมาณมีแนวโน้มเพิ่มขึ้น

เนื่องจากข้อมูลมีค่าผิดปกติในระดับไม่รุนแรง ทำให้ค่าประมาณสัมประสิทธิ์การถดถอยเบี่ยงเบนไปจากค่าจริงไม่มากนัก ดังนั้นการปรับข้อมูลด้วยค่าถ่วงน้ำหนักที่เหมาะสมโดยใช้ตัวประมาณ AWLS จึงทำให้ค่าประมาณสัมประสิทธิ์การถดถอยมีค่าเข้าใกล้ค่าจริงมากกว่าการตัดข้อมูลออกด้วยตัวประมาณ REWLS

<u>ส่วนที่ 3</u> ผลการเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรตามในระดับรุนแรง และไม่มีค่าผิดปกติในตัวแปรอิสระ

พบว่าในทุกสัดส่วนการปลอมปน (p =0.05, 0.10, 0.15 และ 0.20) และทุกขนาดตัวอย่าง (n= 20, 30, 40, 50, 60, 70, 80, 90 และ 100) ตัวประมาณ REWLS ให้ประสิทธิภาพในการ

ประมาณสูงที่สุด รองลงมาคือ ตัวประมาณ AWLS และตัวประมาณ OLS ตามลำดับ และพบว่า เมื่อสัดส่วนการปลอมปนเท่ากับ 0.05 ขนาดตัวอย่างตั้งแต่ 40 ขึ้นไป ตัวประมาณ REWLS และตัวประมาณ AWLS มีค่า AMSE ใกล้เคียงกัน

นอกจากนี้ยังพบว่าเมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกๆ สัดส่วนการปลอมปน ค่า AMSE ของทุกตัวประมาณมีแนวโน้มลดลง ในขณะที่เมื่อสัดส่วนการปลอมปนเพิ่มขึ้นในทุกๆ ขนาดตัวอย่าง ค่า AMSE ของทุกตัวประมาณมีแนวโน้มเพิ่มขึ้น

เนื่องจากข้อมูลมีค่าผิดปกติในระดับรุนแรง ส่งผลให้ค่าประมาณสัมประสิทธิ์การถดถอยเบี่ยงเบนไปจากค่าจริงมาก ดังนั้นการตัดข้อมูลที่จำเป็นจริงๆ ออกด้วยตัวประมาณ REWLS จึงทำให้ค่าประมาณสัมประสิทธิ์การถดถอยมีค่าเข้าใกล้ค่าจริงมากกว่าการปรับข้อมูลด้วยค่าถ่วงน้ำหนักโดยใช้ตัวประมาณ AWLS

ข้อเสนอแนะ

แนวทางในการเลือกใช้ตัวประมาณในการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นเมื่อมีค่าผิดปกติในตัวแปรตาม สามารถเสนอแนะแนวทางได้เป็น 2 ด้านดังนี้

<u>ด้านการนำไปใช้ประโยชน์</u>

ในการวิเคราะห์การถดถอยนั้นขั้นตอนแรกผู้วิเคราะห์ควรตรวจสอบในเบื้องต้นว่าข้อมูลมีค่าผิดปกติหรือไม่ โดยนำข้อมูลของตัวแปรอิสระและตัวแปรตามมาทำการพล็อตกราฟแบบ Box-plot เพื่อพิจารณาค่าของข้อมูล ซึ่งผู้วิเคราะห์สามารถทราบถึงชุดของข้อมูลที่เป็นค่าผิดปกติและทราบว่าข้อมูลชุดนั้นเกิดค่าผิดปกติในระดับใดได้อีกด้วย ถ้าผู้วิเคราะห์ทราบว่าสาเหตุของการเกิดค่าผิดปกติเกิดจากความผิดพลาดของการบันทึกข้อมูล ผู้วิเคราะห์อาจแก้ไขข้อมูลโดยการตัดค่าข้อมูลที่ผิดปกตินั้นออกไปถ้ามีข้อมูลจำนวนมาก แต่หากความผิดพลาดนั้นเกิดจากธรรมชาติของข้อมูลจะต้องนำค่าผิดปกตินั้นเข้าร่วมในการวิเคราะห์การถดถอยด้วย แต่ในกรณีที่ไม่ทราบถึงสาเหตุของการมีค่าผิดปกติ ผู้วิเคราะห์ไม่ควรที่จะตัดค่าผิดปกติออกทันที เนื่องจากค่าผิดปกติอาจให้ข้อมูลที่เป็นประโยชน์ในการหาตัวแบบการถดถอยที่เหมาะสม

ถ้าต้องนำข้อมูลที่ผิดปกตินั้นมาใช้ในการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยนั้น วิธีการแก้ไขสามารถทำได้โดยการใช้ตัวประมาณที่มีความแกร่งในการการลดอิทธิพลของค่าสังเกตที่มีค่าผิดปกติลง โดยทำการปรับค่าของข้อมูลที่ผิดปกติด้วยค่าถ่วงน้ำหนักที่เหมาะสม ซึ่งในงานวิจัยนี้ได้เสนอแนวทางในการเลือกตัวประมาณค่าพารามิเตอร์เมื่อไม่มีค่าผิดปกติและมีค่า

ผิดปกติในตัวแปรตามให้เหมาะสมในแต่ละสถานการณ์ โดยผู้วิจัยได้แสดงแผนผังการเลือกใช้ตัวประมาณค่าพารามิเตอร์ที่ทำให้ค่า AMSE ต่ำสุดในแต่ละสถานการณ์ตามปัจจัยต่างๆ ดังนี้

1. การมีค่าผิดปกติหรือไม่มีค่าผิดปกติในตัวแปรตาม
2. ระดับของค่าผิดปกติ
3. สัดส่วนการปลอมปนของค่าผิดปกติ
4. ขนาดตัวอย่าง

โดยกำหนดสัญลักษณ์ต่างๆ เพื่อแทนความหมายดังนี้

p หมายถึง สัดส่วนการปลอมปน

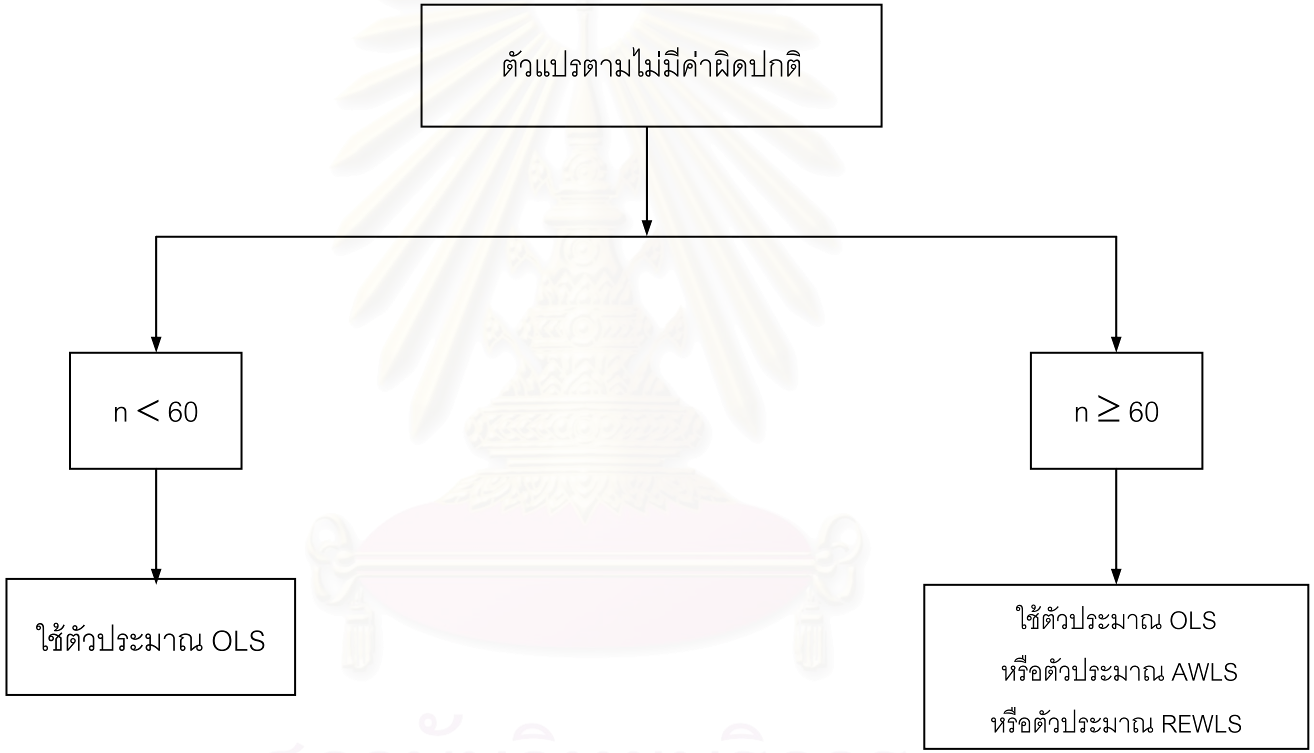
n หมายถึง ขนาดตัวอย่าง

OLS หมายถึง การประมาณค่าพารามิเตอร์ด้วยตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ

AWLS หมายถึง การประมาณค่าพารามิเตอร์ด้วยตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ได้รับการปรับ

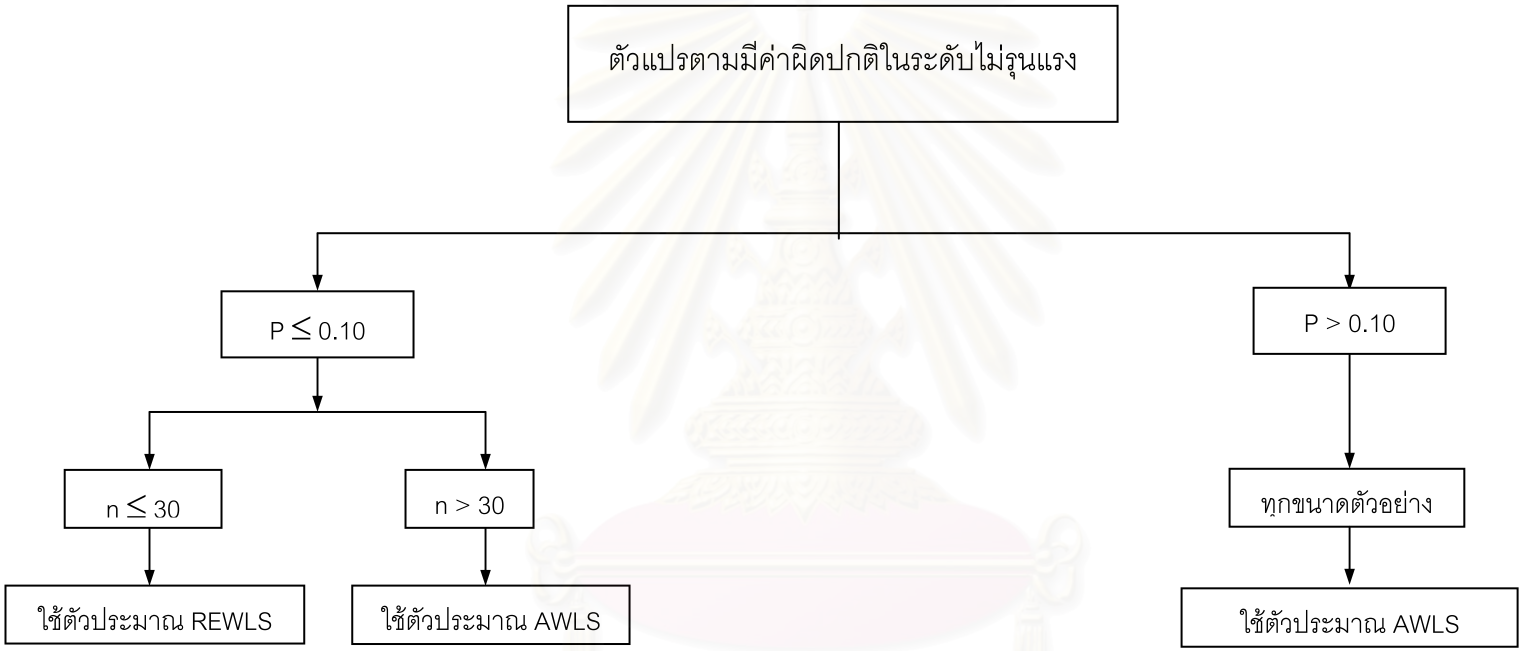
REWLS หมายถึง การประมาณค่าพารามิเตอร์ด้วยตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพ

แผนผังแสดงการเลือกใช้ตัวประมาณค่าพารามิเตอร์ในกรณีที่ไม่มีค่าผิดปกติในตัวแปรตามและในตัวแปรอิสระ

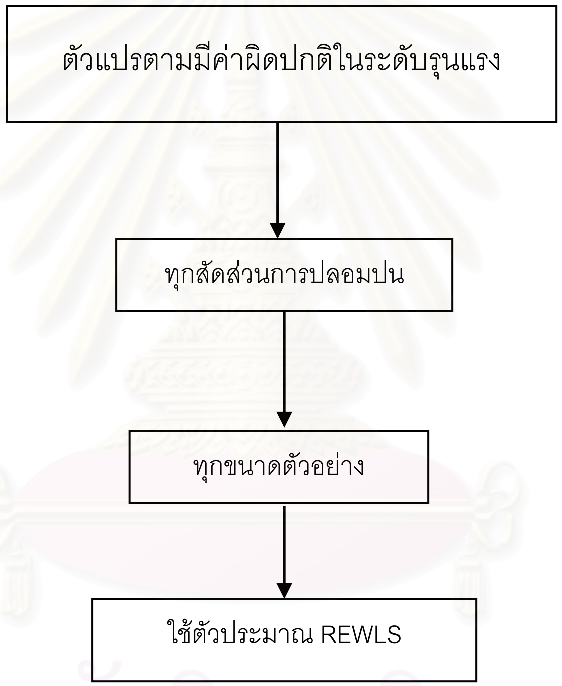


หมายเหตุ ในทุกขนาดตัวอย่างควรเลือกใช้ตัวประมาณ OLS เนื่องจากคำนวณได้ง่ายกว่า และได้ตัวประมาณไม่เอนเอียงเชิงเส้นที่ดีที่สุด

แผนผังแสดงการเลือกใช้ตัวประมาณค่าพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรตามในระดับไม่รุนแรง และไม่มีค่าผิดปกติในตัวแปรอิสระ



แผนผังแสดงการเลือกใช้ตัวประมาณค่าพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรตามในระดับรุนแรง และไม่มีค่าผิดปกติในตัวแปรอิสระ



<u>ด้านการศึกษาวิจัย</u>

เพื่อเป็นแนวทางให้ผู้ที่สนใจได้ศึกษาเพิ่มเติม เพื่อเป็นการขยายผลการวิจัยออกไปให้เกิดประโยชน์มากยิ่งขึ้น โดยทำการศึกษาในกรณีต่าง ๆ ดังต่อไปนี้

1. สามารถนำตัวประมาณที่ได้จากการวิจัยในครั้งนี้ไปทำการศึกษาเปรียบเทียบกับตัวประมาณอื่นๆ หรือนำไปศึกษาเพิ่มเติมในเชิงลึกเพื่อเพิ่มประสิทธิภาพของตัวประมาณในการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นเมื่อมีค่าผิดปกติในตัวแปรตาม

2. สามารถนำตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่มีความแกร่งและมีประสิทธิภาพที่ได้จากการวิจัยในครั้งนี้ไปทำการศึกษาในการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นเมื่อมีค่าผิดปกติในตัวแปรอิสระ

3. จากผลการวิจัยพบว่าค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) และค่าเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (S.D.) มีค่าสูง เนื่องจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\beta}_0$ มีค่าสูง

รายการอ้างอิง

ภาษาไทย

มานพ วราภักดิ์ . (2547) . การจำลองเบื้องต้น . กรุงเทพมหานคร : ศูนย์ผลิตตำราเรียนสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ .

ภาษาอังกฤษ

Rousseeuw, P.J. and Leroy, A. (1987) . Robust Regression and Outlier Detection . New York : Wiley .

Thomas W. O'Gorman . (2001) . Adaptive Estimation Using Weighted Least Squares . Australia New Zealand Journal of Statistics, 43(3) : 287-297.

บรรณานุกรม

ภาษาไทย.

จิตรวี วีระประดิษฐ . (2539) . การประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นพหุ เมื่อข้อมูลมีค่าผิดปกติ . วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย .

ทรงศิริ แต้สมบัติ . (2542) . การวิเคราะห์การถดถอย, กรุงเทพมหานคร : สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์ .

ปัทมวดี นันทนาเนตร์ . (2544) . การเปรียบเทียบตัวประมาณการถดถอยเมื่อมีพหุสัมพันธ์และ/หรือมีค่าผิดปกติ . วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย .

มานพ วราภักดิ์ . (2547) . การจำลองเบื้องต้น . กรุงเทพมหานคร : ศูนย์ผลิตตำราเรียนสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ .

ภาษาอังกฤษ

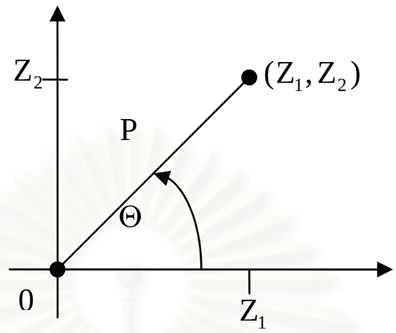
Daniel Gervini and Victor J. Yohai . (2002) . A Class of Robust and Fully Efficient Regression Estimator . The Annals of Statistics, 30(2) : 583-616.

ภาคผนวก

ภาคผนวก ก.

การจำลองตัวแปรสุ่มปกติด้วยวิธีบอกซ์-มูลเลอร์*

George E.P.Box และ Mervin E.Muller (1958) ได้คิดค้นวิธีการจำลองตัวแปรสุ่มปกติมาตรฐาน N(0,1) โดยใช้การแปลงตัวแปรสุ่มจากตัวแปรสุ่มมาตรฐาน Z_1 และ Z_2 ซึ่งอิสระกันได้จุดบนระนาบในระบบพิกัดคาร์ทีเซียน (Cartesian Coordinates) ดังรูป



ทำการแปลงตัวแปรสุ่มในระบบพิกัดคาร์ทีเซียนให้เป็นตัวแปรสุ่มในระบบพิกัดเชิงขั้ว (Polar Coordinates) เป็นจุด (P,Θ) โดยที่

$$Z_1 = P\cos\Theta$$

และ $$Z_2 = P\sin\Theta$$

เมื่อทราบการแจกแจงของ P และ Θ แล้ว เมื่อนำไปแทนค่าจะได้ Z_1 และ Z_2 ซึ่งวิธีการนี้เรียกว่า "วิธีบอกซ์-มูลเลอร์"

การแปลง $z_1 = \rho\cos\theta$ และ $z_2 = \rho\sin\theta$ เป็นการแปลงแบบหนึ่งต่อหนึ่ง (One-to-One Transformation) จากปริภูมิ $R_{Z_1,Z_2} = \{(z_1,z_2): -\infty < z_1 < \infty, -\infty < z_2 < \infty\}$ ของ (Z_1, Z_2) ไปยังปริภูมิ $R_{P,\Theta} = \{(\rho,\theta): 0 \le \rho < \infty, 0 \le \theta \le 2\pi\}$ โดยมีจาโคเบียนของการแปลงดังนี้

$$J = \left\{\begin{matrix} \dfrac{\partial z_1}{\partial \rho} & \dfrac{\partial z_1}{\partial \theta} \\ \dfrac{\partial z_2}{\partial \rho} & \dfrac{\partial z_2}{\partial \theta} \end{matrix}\right. = \begin{vmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{vmatrix}$$

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

เพราะฉะนั้นโดยเทคนิคของการแปลงในทฤษฎีความน่าจะเป็นจะได้ฟังก์ชันความหนาแน่นร่วมของ P และ Θ ดังนี้

$$f_{P,\Theta}(\rho,\theta) = f_{Z_1,Z_2}(\rho\cos\theta, \rho\sin\theta)\,|J|$$

เนื่องจาก Z_1 และ Z_2 มีฟังก์ชันความหนาแน่นร่วม

$$f_{Z_1,Z_2}(z_1,z_2) = f_{Z_1}(z_1) f_{Z_2}(z_2)$$

---

* ที่มา : มานพ วราภักดิ์, การจำลองเบื้องต้น (กรุงเทพฯ: ศูนย์ผลิตตำราเรียนสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2547), หน้า 142

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_1^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_2^2}$$

$$= \frac{1}{2} e^{-\frac{1}{2}(z_1^2+z_2^2)}$$

ดังนั้นโดยการแทนค่าได้ผลลัพธ์

$$f_{P,\Theta}(\rho,\theta) = \frac{1}{2\pi} e^{-\frac{1}{2}\rho^2} . \rho$$

$$= \frac{1}{2\pi} \rho e^{-\frac{1}{2}\rho^2} \quad , \; 0 \le \rho < \infty, 0 \le \theta \le 2\pi$$

$$= f_\Theta(\theta) f_P(\rho)$$

โดยที่ $f_\Theta(\theta) = \dfrac{1}{2\pi}$, $0 \le \theta \le 2\pi$ เป็นฟังก์ชันของ θ เท่านั้น ไม่ขึ้นกับ ρ

และ $f_P(\rho) = \rho e^{-\frac{1}{2}\rho^2}$, $\rho \ge 0$ เป็นฟังก์ชันของ ρ เท่านั้น ไม่ขึ้นกับ θ

เพราะฉะนั้นโดยคุณสมบัติของตัวแปรสุ่มอิสระจะได้ว่า P และ Θ เป็นอิสระกันในเชิงสถิติ นอกจากนี้ยังพบว่า f_Θ เป็นฟังก์ชันความหนาแน่นของการแจกแจงเอกรูป $U(0,2\pi)$ และ f_P เป็นฟังก์ชันความหนาแน่นของการแจกแจงเรย์ลี (Rayleigh Distribution) เพราะฉะนั้นในการจำลอง Z_1 และ Z_2 เราจะจำลอง P และ Θ อย่างอิสระกัน โดยจำลอง P จาก $f_P(\rho) = \rho e^{-\frac{1}{2}\rho^2}$ ซึ่งโดยวิธีการแปลงผกผันทำให้ได้ตัวแบบจำลอง $P = \sqrt{-2\ln R_1}$, $R_1 \sim U(0,1)$ และจำลอง Θ จากการแจกแจง $U(0,2\pi)$ ได้ $\Theta = 2\pi R_2$, $R_2 \sim U(0,1)$ ดังนั้นตัวแบบจำลองตัวแปรสุ่ม $Z_1 \sim N(0,1)$ และ $Z_2 \sim N(0,1)$ อิสระกัน ดังนี้

$$Z_1 = \sqrt{-2\ln R_1}\cos(2\pi R_2)$$
$$Z_2 = \sqrt{-2\ln R_1}\sin(2\pi R_2) \qquad \ldots\ldots\ldots\ldots(1)$$

โดยที่ $R_1, R_2 \sim U(0,1)$ และเป็นอิสระกัน

การจำลองตัวแปรสุ่มลาปลาซ*

ตัวแปรสุ่ม X มีการแจกแจงลาปลาซ (Laplace Distribution) หรือการแจกแจงแบบเลขชี้กำลังสองด้าน (Double-Exponential Distribution) ด้วยพารามิเตอร์ θ และ λ ซึ่งเขียนแทนด้วย $X \sim La(\theta,\lambda)$ และมีฟังก์ชันความหนาแน่นดังนี้

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x-\theta|} \quad , \; -\infty < x < \infty; -\infty < \theta < \infty \; , \; \lambda > 0$$

---

* ที่มา : มานพ วราภักดิ์, การจำลองเบื้องต้น (กรุงเทพฯ: ศูนย์ผลิตตำราเรียนสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2547), หน้า 154

โดยที่ $E(X)=\theta$ และ $Var(X)=\frac{2}{\lambda^2}$

การแจกแจงลาปลาซเป็นการแจกแจงที่นิยมใช้ในการจำลองข้อมูลให้มีค่าผิดปกติ และใช้เป็นการแจกแจงของความคลาดเคลื่อนสุ่ม

การจำลองตัวแปรสุ่มลาปลาซด้วยวิธีการแปลงผกผัน (Inverse Transformation Method) เริ่มด้วยการหาฟังก์ชันการแจกแจงสะสม F ของ X

สำหรับ $-\infty<x<\theta$ จะได้

$$F(x) = \int_{-\infty}^{x}\frac{\lambda}{2}e^{\lambda(y-\theta)}dy = \frac{1}{2}e^{\lambda(x-\theta)}$$

และสำหรับ $\theta<x<\infty$ จะได้

$$F(x) = F(\theta)+\int_{\theta}^{x}\frac{\lambda}{2}e^{-\lambda(y-\theta)}dy = 1-\frac{1}{2}e^{-\lambda(x-\theta)}$$

เพราะฉะนั้น

$$F(x)=\begin{cases}\frac{1}{2}e^{\lambda(x-\theta)} & ,-\infty<x<\theta \\ 1-\frac{1}{2}e^{-\lambda(x-\theta)} & ,\theta<x<\infty\end{cases}$$

จากนั้นแก้สมการหา x ในเทอมของ r จาก $F(x)=r$, $0\le r\le 1$ ได้ว่า

$$\frac{1}{2}e^{\lambda(x-\theta)}=r \text{ ก็ต่อเมื่อ } x=\theta+\frac{1}{\lambda}\ln(2r) \text{ สำหรับ } 0<r<\frac{1}{2}$$

และ

$$1-\frac{1}{2}e^{-\lambda(x-\theta)}=r \text{ ก็ต่อเมื่อ } x=\theta-\frac{1}{\lambda}\ln(2(1-r)) \text{ สำหรับ } \frac{1}{2}\le r<1$$

ดังนั้นตัวแบบการจำลองตัวแปรสุ่มสำหรับ $X\sim La(\theta,\lambda)$ เป็นดังนี้

$$X=\begin{cases}\theta+\frac{1}{\lambda}\ln(2R) & ,0<R<\frac{1}{2} \\ \theta-\frac{1}{\lambda}\ln(2(1-R)) & ,\frac{1}{2}\le R<1\end{cases}$$

โดยที่ $R_1,R_2\sim U(0,1)$ และเป็นอิสระกัน

การหาเปอร์เซ็นไทล์

การหาค่าเปอร์เซ็นไทล์ที่ P หาได้โดยคำนวณหาค่า $(n+1)\times\frac{P}{100}$

ถ้า $(n+1)\times\frac{P}{100}=$ r เป็นเลขจำนวนเต็ม ค่าเปอร์เซ็นไทล์ที่ P คือค่าของข้อมูลที่อยู่ในลำดับที่ r

ถ้า $(n+1)\times\frac{P}{100}$ = r + s โดยที่ r เป็นเลขหน้าจุดทศนิยม และ s เป็นเลขทศนิยม (0 < s < 1) แล้ว ค่าเปอร์เซ็นไทล์ที่ P คือ (1 – s)×(ค่าของข้อมูลในลำดับที่ r) + (s)×(ค่าของข้อมูลในลำดับที่ r+1)

ภาคผนวก ข.

```
!!!!! ********************** Program Main of OLS Estimator ********************** !!!!!
      Program Main
      Common/seed/IX,KK
      integer N,IX,KK,repeat
      real beta0,beta1,beta2,X(100,3),E(100),Y(100),lambda,X1_mean,X2_mean,mean,X1_var,X2_var,var
      double precision B_OLS(3),MS_OLS(3),BE_OLS(10000,3)
      Open(1,File='c:\ols.xls')
      print*,'Number of Example='
      read*,N
555   print*,'Distribution of Residual='
      read*,DR
      repeat=5000
      IX=11117
      KK=0
      c=3
      p=0.05
      beta0=5
      beta1=1
      beta2=1
      X1_mean=20
      X1_var=4
      X2_mean=30
      X2_var=9
      mean=0
      var=10
      theta=0
      lambda=0.125

      call gen_x(X1_mean,X1_var,X2_mean,X2_var,N,X)
      sum1=0
      sum2=0
      sum3=0
      sum4=0
      sum5=0
      sum6=0
      Do 17 z=1,repeat
      If (DR.EQ.1) then
            call cn_residual(mean ,var,c,p,N,E)
      else
            if (DR.EQ.2) then
                  call cl_residual(N,mean,var,p,theta,lambda,E)
            else
```

```
                    print*,'Mistake'
                    goto 555
          end if
      end if
      call gen_y(N,beta0,beta1,beta2,X,E,Y)
      call OLS(N,X,Y,B_OLS)
      sum4=sum4+B_OLS(1)
      sum5=sum5+B_OLS(2)
      sum6=sum6+B_OLS(3)
      BE_OLS(z,1)=B_OLS(1)
      BE_OLS(z,2)=B_OLS(2)
      BE_OLS(z,3)=B_OLS(3)
      call MSE(repeat,beta0,beta1,beta2,B_OLS,MS_OLS)
      sum1=sum1+MS_OLS(1)
      sum2=sum2+MS_OLS(2)
      sum3=sum3+MS_OLS(3)
17    continue
!!!   *** Find MSE of beta0, beta1 and beta2 *** !!!
      sum_m1=sum1              !**   sum_m1 is MSE of beta1     **!
      sum_m2=sum2              !**   sum_m2 is MSE of beta2     **!
      sum_m0=sum3              !**   sum_m0 is MSE of beta0     **!
      AMSE=(sum_m1+sum_m2+sum_m0)/3
!!!   *** Find standard deviation of MSE *** !!!
      D1=(sum_m1-AMSE)**2
      D2=(sum_m2-AMSE)**2
      D3=(sum_m0-AMSE)**2
      SD_MSE=((D1+D2+D3)/(3-1))**0.5
!!!   ***** Find mean of beta0, beta1 and beta2 ***** !!!
      sum_b1=sum4/repeat       !**   sum_b1 is mean of beta1    **!
      sum_b2=sum5/repeat       !**   sum_b2 is mean of beta2    **!
      sum_b0=sum6/repeat       !**   sum_b0 is mean of beta0    **!
!!!   ***** Find SD of beta0, beta1 and beta2 ***** !!!
      sum7=0
      sum8=0
      sum9=0
      Do 123 z=1,repeat
      D4=((BE_OLS(z,1)-sum_b1)**2)/repeat
      D5=((BE_OLS(z,2)-sum_b2)**2)/repeat
      D6=((BE_OLS(z,3)-sum_b0)**2)/repeat
      sum7=sum7+D4
      sum8=sum8+D5
```

```
      sum9=sum9+D6
123   continue
      SD_B1=sum7**0.5
      SD_B2=sum8**0.5
      SD_B0=sum9**0.5
      stop
      End
!!!!! ********************** Program Main of AWLS Estimator ********************** !!!!!
      Program Main
      Common/seed/IX,KK
      integer N,IX,KK,repeat
      real beta0,beta1,beta2,X(100,3),E(100),Y(100),lambda,X1_mean,X2_mean,mean,X1_var,X2_var,var
      real Y_hat(100),NE(100),W2(100)
      double precision B_OLS(3), B_AWLS(3), BE_AWLS(10000,3), MS_AWLS(3)
      Open(1,File='c:\awls.xls')
      print*,'Number of Example='
      read*,N
555   print*,'Distribution of Residual='
      read*,DR
      repeat=5000
      IX=11117
      KK=0
      c=3
      p=0.05
      beta0=5
      beta1=1
      beta2=1
      X1_mean=20
      X1_var=4
      X2_mean=30
      X2_var=9
      mean=0
      var=10
      theta=0
      lambda=0.125

      call gen_x(X1_mean,X1_var,X2_mean,X2_var,N,X)
      sum1=0
      sum2=0
      sum3=0
      sum4=0
      sum5=0
```

```
        sum6=0
        Do 17 z=1,repeat
        If (DR.EQ.1) then
                call cn_residual(mean,var,c,p,N,E)
        else
                if (DR.EQ.2) then
                        call cl_residual(mean,var,p,theta,lambda,N,E)
                else
                        print*,'Mistake'
                        goto 555
                end if
        end if
        call gen_y(N,beta0,beta1,beta2,X,E,Y)
        call OLS(N,X,Y,B_OLS)
        call est_y(N,B_OLS,X,Y_hat)
        call E_OLS(N,Y,Y_hat,NE)
        call weight2(N,NE,W2)
        call AWLS(N,X,Y,W2,B_AWLS)
        sum4=sum4+B_AWLS(1)
        sum5=sum5+B_AWLS(2)
        sum6=sum6+B_AWLS(3)
        BE_AWLS(z,1)=B_AWLS(1)
        BE_AWLS(z,2)=B_AWLS(2)
        BE_AWLS(z,3)=B_AWLS(3)
        call MSE(repeat,beta0,beta1,beta2,B_AWLS,MS_AWLS)
        sum1=sum1+MS_AWLS(1)
        sum2=sum2+MS_AWLS(2)
        sum3=sum3+MS_AWLS(3)
17      continue
!!!     *** Find MSE of beta0, beta1 and beta2 *** !!!
        sum_m1=sum1                     !**     sum_m1 is MSE of beta1 **!
        sum_m2=sum2                     !**     sum_m2 is MSE of beta2 **!
        sum_m0=sum3                     !**     sum_m0 is MSE of beta0 **!
        AMSE=(sum_m1+sum_m2+sum_m0)/3
!!!     *** Find standard deviation of MSE *** !!!
        D1=(sum_m1-AMSE)**2
        D2=(sum_m2-AMSE)**2
        D3=(sum_m0-AMSE)**2
        SD_MSE=((D1+D2+D3)/(3-1))**0.5
!!!      ***** Find mean of beta0, beta1 and beta2 ***** !!!
        sum_b1=sum4/repeat
```

```
        sum_b2=sum5/repeat
        sum_b0=sum6/repeat
!!!     ***** Find SD of beta0, beta1 and beta2 ***** !!!
        sum7=0
        sum8=0
        sum9=0
        Do 12 z=1,repeat
        D4=((BE_AWLS(z,1)-sum_b1)**2)/repeat
        D5=((BE_AWLS(z,2)-sum_b2)**2)/repeat
        D6=((BE_AWLS(z,3)-sum_b0)**2)/repeat
        sum7=sum7+D4
        sum8=sum8+D5
        sum9=sum9+D6
12      continue
        SD_B1=sum7**0.5
        SD_B2=sum8**0.5
        SD_B0=sum9**0.5
        stop
        End
        !!!!! ********************** Program Main of REWLS Estimator ********************** !!!!!
        Program Main
        Common/seed/IX,KK
        integer N,IX,KK,repeat
        real beta0,beta1,beta2,X(100,3),E(100),Y(100),lambda,X1_mean,X2_mean,mean,X1_var,X2_var,var
        real S_LMS, W3(100),SR(100)
        double precision B_P(500,3),B_LMS(3),B_REWLS(3), MS_REWLS(3),BE_REWLS(10000,3)
        Open(1,File='c:\rewls.xls')
        print*,'Number of Example='
        read*,N
555     print*,'Distribution of Residual='
        read*,DR
        repeat=5000
        IX=11117
        KK=0
        c=3
        p=0.05
        beta0=5
        beta1=1
        beta2=1
        X1_mean=20
        X1_var=4
        X2_mean=30
```

```
X2_var=9
mean=0
var=10
theta=0
lambda=0.125

call gen_x(X1_mean,X1_var,X2_mean,X2_var,N,X)
sum1=0
sum2=0
sum3=0
sum4=0
sum5=0
sum6=0
Do 17 z=1,repeat
If  (DR.EQ.1) then
        call cn_residual(mean ,var,c,p,N,E)
else
        if (DR.EQ.2) then
                call cl_residual(mean,var,p,theta,lambda,N,E)
        else
                print*,'Mistake'
                goto 555
        end if
end if
call gen_y(N,beta0,beta1,beta2,X,E,Y)
call beta_p(N,X,Y,B_P)
call LMS(N,X,Y,B_P,B_LMS)
call SN(N,X,Y,B_LMS,S_LMS)
call st_residual(N,X,Y,B_LMS,S_LMS,SR)
call weight3(N,SR,W3)
call REWLS(N,X,Y,W3,S_LMS,B_LMS,B_REWLS)
sum4=sum4+B_REWLS(1)
sum5=sum5+B_REWLS(2)
sum6=sum6+B_REWLS(3)
BE_REWLS(z,1)=B_REWLS(1)
BE_REWLS(z,2)=B_REWLS(2)
BE_REWLS(z,3)=B_REWLS(3)
call MSE(repeat,beta0,beta1,beta2,B_REWLS,MS_REWLS)
sum1=sum1+MS_REWLS(1)
sum2=sum2+MS_REWLS(2)
sum3=sum3+MS_REWLS(3)
```

```
17      continue
!!!     *** Find MSE of beta0, beta1 and beta2 *** !!!
        sum_m1=sum1                 !**     sum_m1 is MSE of beta1 **!
        sum_m2=sum2                 !**     sum_m2 is MSE of beta2 **!
        sum_m0=sum3                 !**     sum_m0 is MSE of beta0 **!
        AMSE=(sum_m1+sum_m2+sum_m0)/3
!!!     *** Find standard deviation of MSE *** !!!
        D1=(sum_m1-AMSE)**2
        D2=(sum_m2-AMSE)**2
        D3=(sum_m0-AMSE)**2
        SD_MSE=((D1+D2+D3)/(3-1))**0.5
!!!     ***** Find mean of beta0, beta1 and beta2 ***** !!!
        sum_b1=sum4/repeat
        sum_b2=sum5/repeat
        sum_b0=sum6/repeat
!!!     ***** Find SD of beta0, beta1 and beta2 ***** !!!
        sum7=0
        sum8=0
        sum9=0
        Do 123 repeat=1,1000
        D4=((BE_REWLS(z,1)-sum_b1)**2)/repeat
        D5=((BE_REWLS(z,2)-sum_b2)**2)/repeat
        D6=((BE_REWLS(z,3)-sum_b0)**2)/repeat
        sum7=sum7+D4
        sum8=sum8+D5
        sum9=sum9+D6
123     continue
        SD_B1=sum7**0.5
        SD_B2=sum8**0.5
        SD_B0=sum9**0.5
        stop
        end
!!!     OLS Estimator      !!!
        subroutine OLS(N,X,Y,B_OLS)
        integer N
        real X(100,3),Y(100),XT(3,100)
        double precision XTX(3,3),XTXI(3,3),XTY(3),B_OLS(3)
        Do 111 i=1,N
        Do 111 j=1,3
        XT(j,i)=X(i,j)
111     continue
```

```
      Do 112 i=1,3
      Do 113 j=1,3
      sum=0
      Do 114 k=1,N
      sum=sum+XT(i,k)*X(k,j)
114   continue
      XTX(i,j)=sum
113   continue
112   continue
      call inverse3(XTX,XTXI)
      Do 115 i=1,3
      sum=0
      Do 116 j=1,N
      sum=sum+XT(i,j)*Y(j)
116   continue
      XTY(i)=sum
115   continue
      Do 117 i=1,3
      sum=0
      Do 118 j=1,3
      sum=sum+XTXI(i,j)*XTY(j)
118   continue
      B_OLS(i)=sum
117   continue
      return
      end
!!!   Find Inverse Matrix for (3x3)  !!!
      subroutine inverse3(A,inv)
      double precision A(3,3),inv(3,3),E(4)
      D1=(A(1,1)*A(2,2)*A(3,3))+(A(1,2)*A(2,3)*A(3,1))+(A(1,3)*A(2,1)*A(3,2))
      D2=-(A(3,1)*A(2,2)*A(1,3))-(A(3,2)*A(2,3)*A(1,1))-(A(3,3)*A(2,1)*A(1,2))
      det=D1+D2
      Do 64 i=1,3
      Do 64 j=1,3
      m=0
      Do 65 k=1,3
      Do 66 l=1,3
      If ((k.EQ.i).OR.(l.EQ.j)) then
      goto 66
      else
            m=m+1
```

```
                 E(m)=A(k,l)
          end if
66        continue
65        continue
          sign=(-1)**(i+j)
          inv(j,i)=(((E(1)*E(4))-(E(2)*E(3)))*sign)/det
64        continue
          return
          end
!!!       AWLS Estimator     !!!
          subroutine AWLS(N,X,Y,W2,B_AWLS)
          integer N
          real X(100,3),Y(100),W2(100)
          double precision WX(100,3),WXT(3,100),WXTX(3,3),WXTXI(3,3)
          double precision WY(100),WXTY(3),B_AWLS(3)
          Do 119 i=1,N
          Do 121 j=1,3
          WX(i,j)=W2(i)*X(i,j)
          WXT(j,i)=WX(i,j)
121       continue
119       continue
          Do 122 i=1,3
          Do 123 j=1,3
          sum=0
          Do 124 k=1,N
          sum=sum+WXT(i,k)*WX(k,j)
124       continue
          WXTX(i,j)=sum
123       continue
122       continue
          call inverse3(WXTX,WXTXI)
          Do 129 i=1,N
          WY(i)= W2(i)*Y(i)
129       continue
          Do 125 i=1,3
          sum=0
          Do 126 j=1,N
          sum=sum+WXT(i,j)*WY(j)
126       continue
          WXTY(i)=sum
125       continue
```

```
        Do 127 i=1,3
        sum=0
        Do 128 j=1,3
        sum=sum+WXTXI(i,j)*WXTY(j)
128     continue
        B_AWLS(i)=sum
127     continue
        return
        end
!!! Find Weighted of AWLS Estimator !!!
        subroutine weight2(N,NE,W2)
        integer N
        real NE(100),DE(100),W2(100)
        Do 81 i=1,N
        DE(i)=NE(i)
81      continue
        call rank_error(N,DE)

        if  (N.LT.50)  then
                P2=DE(1)
        else
                H=(N+1)*0.02
                IH=H
                PL=IH+1
                H1=H-IH
                P2=(1-H1)*DE(IH)+H1*DE(PL)
        end if
        G=(N+1)*0.25
        IG=G
        PL=IG+1
        G1=G-IG
        P25=(1-G1)*DE(IG)+G1*DE(PL)
        J=(N+1)*0.75
        IJ=J
        PL=IJ+1
        J1=J-IJ
        P75=(1-J1)*DE(IJ)+J1*DE(PL)
        R=(N+1)*0.98
        IR=R
        PL=IR+1
        R1=R-IR
```

```
P98=(1-R1)*DE(IR)+R1*DE(PL)
T1=P25 - P2
T2=P98 - P75
T3=P75 - P25
TL=T1/T3
TR=T2/T3
if (N.EQ.20) then
        t_NL=0.736779
else
        if (N.EQ.30) then
                t_NL=0.870363
        else
                if (N.EQ.40) then
                        t_NL=0.960734
                else
                        t_NL=1.022
                end if
        end if
end if
WL=t_NL/TL
if (N.EQ.20) then
        t_NR=0.736779
else
        if (N.EQ.30) then
                t_NR=0.870363
        else
                if (N.EQ.40) then
                        t_NR=0.960734
                else
                        t_NR=1.022
                end if
        end if
end if
WR=t_NR/TR
SL1=(1-WL)/T1
SL2=(WR-1)/T2
Do 82 i=1,N
If ((NE(i).GE.P25).AND.(NE(i).LE.P75)) then
        W2(i)=1
else
        If (NE(i).EQ.P2) then
```

```
                W2(i)=WL
            else
                If (NE(i).EQ.P98) then
                    W2(i)=WR
                else
                    If ((NE(i).LT.P25).AND.(NE(i).GT.P2)) then
                        W2(i)=(SL1*(NE(i)-P2))+WL
                    else
                        If ((NE(i).GT.P75).AND.(NE(i).LT.P98)) then
                            W2(i)=(SL2*(NE(i)-P75))+1
                        else
                            W2(i)=0
                        end if
                    end if
                end if
            end if
        end if
82      continue
        return
        end
!!!     Rank Error !!!
        subroutine rank_error(N,DE)
        integer N
        real DE(100)
        N1=N-1
        Do 96 i=1,N1
        i1=i+1
        Do 97 k=i1,N
        If (DE(i).LE.DE(k)) goto 97
        T=DE(i)
        DE(i)=DE(k)
        DE(k)=T
97      continue
96      continue
        return
        end
!!!     **********Find Y_hat*********** !!!
        subroutine est_y(N,B_OLS,X,Y_hat)
        integer N
        real X(100,3),Y_hat(100)
        double precision B_OLS(3)
```

```
        Do 131 i=1,N
        Y_hat(i)=(B_OLS(3)*X(i,3))+(B_OLS(1)*X(i,1))+(B_OLS(2)*X(i,2))
131     continue
        return
        end
!!!     ****** Find Error from OLS Estimator ****** !!!
        subroutine E_OLS(N,Y,Y_hat,NE)
        integer N
        real Y(100),Y_hat(100), NE(100)
        Do 133 i=1,N
        NE(i)=Y(i)-Y_hat(i)
133     continue
        return
        end
!!!     *************** REWLS Estimator **************** !!!
        subroutine REWLS(N,X,Y,W3,S_LMS,B_LMS,B_REWLS)
        integer N
        real X(100,3),Y(100),XT(3,100)
        real W3(100),S_LMS
        double precision B_LMS(3),B_WLS(3),B_REWLS(3)
        double precision XTW(3,100),XTWX(3,3),XTWXI(3,3),XTWY(3)
        Do 71 i=1,N
        Do 72 j=1,3
        XT(j,i)=X(i,j)
72      continue
71      continue
        Do 73 i=1,3
        Do 74 j=1,N
        XTW(i,j)=XT(i,j)*W3(j)
74      continue
73      continue
        Do 75 i=1,3
        Do 76 j=1,3
        sum=0
        Do 78 k=1,N
        sum=sum+XTW(i,k)*X(k,j)
78      continue
        XTWX(i,j)=sum
76      continue
75      continue
        call inverse3(XTWX,XTWXI)
```

```
      Do 79 i=1,3
      sum=0
      Do 44 j=1,N
      sum=sum+XTW(i,j)*Y(j)
44    continue
      XTWY(i)=sum
79    continue
      Do 81 i=1,3
      sum=0
      Do 82 j=1,3
      sum=sum+XTWXI(i,j)*XTWY(j)
82    continue
      B_WLS(i)=sum
81    continue
      Do 83 i=1,3
      if (S_LMS.EQ.0) then
            B_REWLS(i)=B_LMS(i)
      else
            if (S_LMS.GT.0) then
                  B_REWLS(i)=B_WLS(i)
            else
                  print*,'Mistake'
            end if
      end if
83    continue
      return
      end
!!!   Find Weight of REWLS Estimator !!!
      subroutine weight3(N,SR,W3)
      integer N,OR1
      real DE(100),SR(100),cum(100)
      real DI(100),DII(100),W3(100),u(100)
      real cum4,DC,M
      Do 42 i=1,N
      DE(i)=abs(SR(i))
42    continue
      call rank_error(N,DE)
      a=0
      Do 43 i=1,N
      if (DE(i).LT.2.5) then
            a=a+1
```

```
        else
                a=a
        end if
43      continue
        OR1=a
        I2=N-OR1
        Do 45 i=1,I2
        j=i+OR1
        DC=DE(j)                !**     DE(j) is order statistics of SR(i>i0)     **!
        call cumulative(DC,cum4)
        cum(i)=cum4
45      continue
        b=OR1-1
        Do 333 i=1,I2
        j=i+OR1
        b=b+1
        DII(i)=cum(i)-b/N
        if (DII(i).GE.0) then
                DI(i)=DII(i)
        else
                DI(i)=0
        end if
333     continue
        Do 51 i=1,I2
        call rank_d(I2,DI)
51      continue
        DN=DI(I2)
        M=N*DN
        IM=M
        OR2=N-IM
        tn=DE(OR2)
        Do 46 i=1,N
        u(i)=(abs(SR(i)))/tn
        if (u(i).LT.1) then
                W3(i)=1
        else
                W3(i)=0
        end if
46      continue
        return
        end
```

```
!!!     Rank D(i) for Find DN !!!
        subroutine rank_d(I2,D)
        integer I2
        real D(100)
        N1=I2-1
        Do 61 i=1,N1
        i1=i+1
        Do 62 k=i1,I2
        If (D(i).LE.D(k)) goto 62
        T=D(i)
        D(i)=D(k)
        D(k)=T
62      continue
61      continue
        Do 63 i=1,I2
63      continue
        return
        end
!!!     Find Cumulative of Standard Normal Distribution !!!
        subroutine cumulative(DC,cum4)
        real DC,cum4,pdf(2000000),K5(2000000),K6(2000000)
        real K1,K2,K3,K4
        sum_k5=0
        sum_k6=0
        K=DC*100
        IK=K
        PL=IK
        K1=PL/100
        K2=(PL+1)/100
        K3=(K1+10)*100
        K4=(K2+10)*100
        Do 1 i=1,K3
        K5(i)=(-10.005)+(0.01*i)
        pdf(i)=(exp((-1)*((K5(i)**2)/2)))/2.506628275
        sum_k5=sum_k5+pdf(i)
1       continue
        cum1=sum_k5/100
        Do 2 i=1,K4
        K6(i)=(-10.005)+(0.01*i)
        pdf(i)=(exp((-1)*((K6(i)**2)/2)))/2.506628275
        sum_k6=sum_k6+pdf(i)
```

```
2       continue
        cum2=sum_k6/100
        if (DC.EQ.K3) then
                cum4=cum1
        else
                if (DC.EQ.K4) then
                        cum4=cum2
                else
                        cum3=((cum2-cum1)*100)*(DC-K1)
                        cum4=cum1+cum3
                end if
        end if
        return
        end
!!!     Find Standardized Residuals !!!
        subroutine st_residual(N,X,Y,B_LMS,S_LMS,SR)
        integer N
        real X(100,3),Y(100), S_LMS,SR(100),XB(100)
        double precision B_LMS(3)
        Do 41 i=1,N
        XB(i)=X(i,1)*B_LMS(1)+X(i,2)*B_LMS(2)+X(i,3)*B_LMS(3)
        SR(i)=(Y(i)-XB(i))/S_LMS
41      continue
        return
        end
!!!     ************ Scale Estimator of LMS Estimator ************ !!!
        subroutine SN(N,X,Y,B_LMS,S_LMS)
        integer N
        real X(100,3),Y(100),Y_est(100),DE(100),U(100),med,S_LMS
        double precision B_LMS(3)
        Do 91 i=1,N
        Y_est(i)=X(i,1)*B_LMS(1)+X(i,2)*B_LMS(2)+X(i,3)*B_LMS(3)
        U(i)=abs(Y(i)-Y_est(i))
91      continue
        Do 92 i=1,N
        DE(i)=U(i)
92      continue
        call rank_error(N,DE)
        a=N/2
        b=N/2+1
        med=(DE(a)+DE(b))/2
```

```
        S_LMS=med/0.67449
        return
        end
!!!     ***************** Generate LMS Estimator ***************** !!!
!!!     Find Beta of LMS Estimator !!!
        subroutine LMS(N,X,Y,B_P,B_LMS)
        integer N,pos(500)
        real X(100,3),Y(100),Y_P(100),E_SQR(100),med(500),DE(100)
        double precision B_P(500,3),Beta_p(3),B_LMS(3)
        if (N.LE.15) then
                H=((N)*(N-1)*(N-2))/6
        else
                H=400
        end if
        Do 33 i=1,H
        Do 23 j=1,3
        Beta_p(j)=B_P(i,j)
23      continue
        Do 34 k=1,N
        Y_P(k)=(X(k,1)*Beta_p(1))+(X(k,2)*Beta_p(2))+(X(k,3)*Beta_p(3))
        E_SQR(k)=(Y(k)-Y_P(k))**2
34      continue
        Do 24 k=1,N
        DE(k)=E_SQR(k)
24      continue
        call rank_error(N,DE)
        a=N/2
        b=N/2+1
        med(i)=(DE(a)+DE(b))/2
33      continue
        DO 31 j=1,H
        pos(j)=j
31      continue
        call rank_med(N,med,pos)
        Do 32 k=1,3
        B_LMS(k)=B_P(pos(1),k)
32      continue
        return
        end
!!!     Discrete Uniform  !!!
        function DU(N)
```

```
      common/seed/IX,KK
      integer N
      real T
      double precision IY
56    call random(IX,IY,RD)
      R=RD
      if ((R.EQ.0).OR.(R.EQ.1)) goto 56
      T=N*R
      IT=T
      if (IT.EQ.0) then
      L=IT+1
      goto 57
      end if
      if (T-IT.GE.0.5) then
            L=IT+1
      else
            L=IT
      end if
57    DU=L
      return
      end
!!!   Find 1 case of X and Y (p-observations) !!!
      subroutine sampling(N,S)
      integer N,S(3)
70    S(1)=DU(N)
80    S(2)=DU(N)
      if (S(2).EQ.S(1)) goto 80
90    S(3)=DU(N)
      if ((S(3).EQ.S(1)).OR.(S(3).EQ.S(2))) goto 90
      return
      end
!!!   Find H cases of X and Y (p-observations) !!!
      subroutine grouping(N,M)
      common/seed/IX,KK
      integer N,S(3),M(500,3)
      call sampling(N,S)
      Do 35 j=1,3
      M(1,j)=S(j)
35    continue
      if (N.LE.15) then
            H=((N)*(N-1)*(N-2))/6
```

```
        else
                H=400
        end if
        Do 110 i=2,H
200     call sampling(N,S)
        Do 130 j=1,3
        M(i,j)=S(j)
130     continue
        Do 180 l=1,I-1
        a=0
        Do 160 j=1,3
        Do 190 k=1,3
        if (M(i,j).NE.M(l,k)) then
                goto 190
        else
                a=a+1
        end if
190     continue
160     continue
        if(a.EQ.3) goto 200
180     continue
110     continue
        return
        end
!!!     Find Beta of p-observations !!!
        subroutine beta_p(N,X,Y,B_P)
        integer N,M(500,3)
        real X(100,3),Y(100)
        real X_1(100,3),Y_1(100)
        real XT(3,3)
        double precision B_P(500,3)
        double precision XTX(3,3),XTXI(3,3),XTY(3),B_OLSP(3)
        if (N.LE.15) then
                H=((N)*(N-1)*(N-2))/6
        else
                H=400
        end if
        call grouping(N,M)
        Do 11 i=1,H
        Do 12 p=1,3
        Y_1(p)=Y(M(i,p))
```

```
        X_1(p,1)=X(M(i,p),1)
        X_1(p,2)=X(M(i,p),2)
        X_1(p,3)=X(M(i,p),3)
12      continue
        Do 111 p=1,3
        Do 111 j=1,3
        XT(j,p)=X_1(p,j)
111     continue
        Do 112 p=1,3
        Do 113 j=1,3
        sum=0
        Do 114 k=1,3
        sum=sum+XT(p,k)*X_1(k,j)
114     continue
        XTX(p,j)=sum
113     continue
112     continue
        call inverse3(XTX,XTXI)
        Do 115 p=1,3
        sum=0
        Do 116 j=1,3
        sum=sum+XT(p,j)*Y_1(j)
116     continue
        XTY(p)=sum
115     continue
        Do 117 p=1,3
        sum=0
        Do 118 j=1,3
        sum=sum+XTXI(p,j)*XTY(j)
118     continue
        B_OLSP(p)=sum
        B_P(i,p)=B_OLSP(p)
117     continue
11      continue
        return
        end
!!!     Ranking of Median !!!
        subroutine rank_med(N,med,pos)
        real med(500)
        integer N,pos(500)
        if (N.LE.15) then
```

```
                H=N*(N-1)*(N-2)/6
        else
                H=400
        end if
        a=H-1
        Do 37 i=1,a
        I3=i+1
        Do 38 k=I3,H
        if(med(i).LE.med(K)) goto 38
        T3=med(i)
        med(i)=med(k)
        med(k)=T3
        H3=pos(i)
        pos(i)=pos(k)
        pos(k)=H3
38      continue
37      continue
        return
        end
!!!     ****************** Generate X, Y and Residual ***************** !!!
!!!     Random Uniform(0,1)!!!
        subroutine random(IX,IY,RD)
        real RD
        integer IX
        double precision IY
        IY=IX*16807
        If(IY.LT.0)IY=1+(IY+2147483647)
        RD=IY
        RD=RD/2147483647
        IX=IY
        return
        end
!!!     Normal Distribution !!!
        subroutine normal(mean,var,x_normal)
        Common/seed/IX,KK
        real pi,var,sd,mean
        integer IX,KK
        double precision IY
        sd=sqrt(var)
        pi=3.1415926
        If(KK.EQ.1)goto 10
```

```
      call random(IX,IY,RD)
      r1=RD
      call random(IX,IY,RD)
      r2=RD
      z1=sqrt(-2*alog(r1))*cos(2*pi*r2)
      z2=sqrt(-2*alog(r1))*sin(2*pi*r2)
      x_normal=mean+(z1*sd)
      KK=1
      goto 20
10    x_normal=mean+(z2*sd)
      KK=0
20    return
      end
!!!   Generated X1 and X2 Normal Distribution !!!
      subroutine gen_x(X1_mean,X1_var,X2_mean,X2_var,N,X)
      common/seed/IX,KK
      real X(100,3)
      integer IX,KK
      Do 40 i=1,N
      call normal(X1_mean,X1_var,x_normal)
      X(i,1)=x_normal
40    continue
      Do 47 j=1,N
      call normal(X2_mean,X2_var,x_normal)
      X(j,2)=x_normal
47    continue
      Do 48 k=1,N
      X(k,3)=1
48    continue
      return
      end
!!!   Contaminate Normal Distribution !!!
      subroutine cnormal(mean ,var,c,p,c_normal)
      common/seed/IX,KK
      integer IX,KK
      real var,c,p,c_var,mean
      double precision IY
      c_var=(c**2)*var
      call random(IX,IY,RD)
      if (RD.LE.p) then
            call normal(mean,c_var,x_normal)
```

```
                c_normal=x_normal
        else
                call normal(mean,var,x_normal)
                c_normal=x_normal
        end if
        return
        end
!!!     Generated Residual Distribution (Comtaminate Normal)!!!
        subroutine cn_residual(mean ,var,c,p,N,E)
        common/seed/IX,KK
        integer IX,KK
        real c,p,E(100),mean
        Do 30 i=1,N
        call cnormal(mean,var,c,p,c_normal)
        E(i)=c_normal
30      continue
        return
        end
!!!     Laplace Distribution!!!
        subroutine laplace(theta,lambda,dex)
        common/seed/IX,KK
        integer IX,KK
        real theta,lambda,dex
        double precision IY
        call random(IX,IY,RD)
        T=RD
        if(T.LT.0.5)then
        dex=alog(2*T)/lambda+theta
        else
        dex=-alog(2*(1-T))/lambda+theta
        end if
        return
        end
!!!     Contaminate Laplace-Normal Distribution!!!
        subroutine lnormal(mean,var,p,theta,lambda,l_normal)
        common/seed/IX,KK
        integer IX,KK
        real var,theta,lambda,mean,l_normal,dex
        double precision IY
        call random(IX,IY,RD)
        if (RD.LE.p) then
```

```
                call laplace(theta,lambda,dex)
                l_normal=dex
        else
                call normal(mean,var,x_normal)
                l_normal=x_normal
        end if
        return
        end
!!!     Generated Residual Distribution (Comtaminate Laplace)!!!
        subroutine cl_residual(N,mean,var,p,theta,lambda,E)
        common/seed/IX,KK
        integer IX,KK
        real p,E(100),lambda,mean,var,l_normal
        Do 480 i=1,N
        call lnormal(mean,var,p,theta,lambda,l_normal)
        E(i)=l_normal
480     continue
        return
        end
!!!     Generated Y !!!
        subroutine gen_y(N,beta0,beta1,beta2,X,E,Y)
        real beta0,beta1,beta2,X(100,3),E(100),Y(100)
        integer N
        Do 1 i=1,N
        Y(i)=(beta0*X(i,3))+(beta1*X(i,1))+(beta2*X(i,2))+E(i)
1       continue
        return
        end
!!!     ******************** MSE *******************  !!!
        subroutine MSE(repeat,beta0,beta1,beta2,be,MS)
        integer repeat
        real beta0,beta1,beta2
        double precision be(3),SSE(3),MS(3)
        SSE(1)=(beta1-be(1))**2
        SSE(2)=(beta2-be(2))**2
        SSE(3)=(beta0-be(3))**2
        Do 119 i=1,3
        MS(i)=SSE(i)/repeat
119     continue
        return
        end
```

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวกัญญารัตน์ โพธิสุทธิ์ เกิดเมื่อวันที่ 15 สิงหาคม พ.ศ. 2522 ที่จังหวัดพระนครศรีอยุธยา จบการศึกษาชั้นมัธยมศึกษาตอนปลายจากโรงเรียนพิบูลวิทยาลัย จังหวัดลพบุรี สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต สาขาวิชาคณิตศาสตร์ จากคณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2543 และเข้าศึกษาต่อในระดับปริญญาสถิติศาสตรมหาบัณฑิต สาขาวิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2545