การประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยโลจิสติคเมื่อมีค่าผิดปกติ

นางสาวอรนิต เกตุสุข

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2547

ISBN 974-17-6700-5

ESTIMATION OF PARAMETERS IN LOGISTIC REGRESSION MODEL HAVING OUTLIERS

Miss Orranit Ketsuk

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 974-17-6700-5

หัวข้อวิทยานิพนธ์	การประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยโลจิสติคเมื่อมีค่าผิดปกติ
โดย	นางสาวอรนิต เกตุสุข
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ร้อยเอก มานพ วราภักดิ์

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.... คณบดีคณะพาณิชยศาสตร์และการบัญชี

(ผู้ช่วยศาสตราจารย์ ดร.ดนุชา คุณพนิชกิจ)

คณะกรรมการสอบวิทยานิพนธ์

.... ประธานกรรมการ

(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

.... อาจารย์ที่ปรึกษา

(รองศาสตราจารย์ ร้อยเอก มานพ วราภักดิ์)

.... กรรมการ

(รองศาสตราจารย์ วัชราภรณ์ สุริยาภิวัฒน์)

อรนิต เกตุสุข : การประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยโลจิสติค เมื่อมีค่าผิดปกติ (ESTIMATION OF PARAMETERS IN LOGISTIC REGRESSION MODEL HAVING OUTLIERS) อ. ที่ปรึกษา : รศ.ร.อ.มานพ วราภักดิ์, 201 หน้า. ISBN 974-17-6700-5.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ เมื่อมีค่าผิดปกติในตัวแปรอิสระของตัวแบบการถดถอยโลจิสติค โดยทำการเปรียบเทียบวิธีความควรจะเป็นสูงสุด(ML) วิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และ Haesbroeck(WMLCH) และวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann(WMLRC) ซึ่งเกณฑ์การเปรียบเทียบคือ ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย(AMSE)ของพารามิเตอร์ ในการวิจัยครั้งนี้มีตัวแปรอิสระ x_1 และ x_2 โดยกำหนดตัวแปรอิสระ x_1 และ x_2 มีการแจกแจงแบบไม่มีค่าผิดปกติและแบบมีค่าผิดปกติ ซึ่งกำหนดระดับค่าผิดปกติเป็นระดับไม่รุนแรงและระดับรุนแรง แต่ละระดับจะกำหนดให้มีสัดส่วนการปลอมปนของขนาดตัวอย่างคือ 0.05, 0.10 และ 0.15 และขนาดตัวอย่างเท่ากับ 20, 30, 40, 50, 60, 70, 80, 90 และ 100 ข้อมูลที่ใช้ในการวิจัยได้จากการจำลองและใช้วิธีมอนติคาร์โลในการหาค่า AMSE ซึ่งกระทำซ้ำ 1,000 ครั้งในแต่ละสถานการณ์

ผลการวิจัยปรากฏว่าระดับค่าผิดปกติ สัดส่วนการปลอมปน และขนาดตัวอย่าง ต่างมีผลต่อการประมาณค่าพารามิเตอร์ของทั้งสามวิธี โดยค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของพารามิเตอร์จะเพิ่มขึ้นเมื่อระดับค่าผิดปกติ และสัดส่วนการปลอมปนเพิ่มขึ้น แต่จะมีค่าลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

กรณีที่ไม่มีค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2

ในทุกขนาดตัวอย่าง วิธี ML จะให้ค่า AMSE ต่ำที่สุด และเมื่อตัวอย่างใหญ่ขนาด 70 ขึ้นไป วิธี ML วิธี WMLCH และวิธี WMLRC จะมีค่า AMSE ใกล้เคียงกัน

กรณีที่มีค่าผิดปกติในตัวแปรอิสระหนึ่งตัว

ในทุกระดับค่าผิดปกติ สัดส่วนการปลอมปน และขนาดตัวอย่าง วิธี WMLRC จะให้ค่า AMSE ต่ำที่สุด และเมื่อตัวอย่างใหญ่ขนาด 60 ขึ้นไป วิธี WMLCH และวิธี WMLRC จะมีค่า AMSE ใกล้เคียงกัน

กรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2

ในทุกระดับค่าผิดปกติ สัดส่วนการปลอมปน และขนาดตัวอย่าง วิธี WMLRC จะให้ค่า AMSE ต่ำที่สุด รองลงมาคือวิธี WMLCH และวิธี ML ตามลำดับ ทั้งนี้วิธี WMLCH และวิธี WMLRC จะมีค่า AMSE ใกล้เคียงกัน

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

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

ปีการศึกษา………….. 2547 ……………..

4582436026 : MAJOR STATISTICS

KEY WORD: Logistic Regression / Outliers / Maximum Likelihood Method / Weighted Maximum Likelihood Method

ORRANIT KETSUK : ESTIMATION OF PARAMETERS IN LOGISTIC REGRESSION MODEL HAVING OUTLIERS. THESIS ADVISOR : PROF.CAPT.MANOP VARAPHAKDI, 201 pp.

ISBN 974-17-6700-5.

The objective of this research is to compare the estimation methods of parameters in the logistic regression model having outliers in independent variables. The estimation methods are Maximum Likelihood (ML) Method, Weighted Maximum Likelihood Method of Croux and Haesbroeck (WMLCH) and Weighted Maximum Likelihood Method of Rousseeuw and Christmann (WMLRC). The criterion of comparison is the average mean square error(AMSE) of parameters. There are two levels of outliers ,mild and extreme, and three proportions of contamination 0.05, 0.10 and 0.15. The sample sizes are 20, 30, 40, 50, 60, 70 80, 90 and 100. The study used data from simulation and used the Monte Carlo method to compute AMSE. The experiment was repeated 1,000 times under each situations.

The results of this research showed that the level of outliers, proportion of outliers contamination and sample sizes have effected on parameters estimates. The average values of mean square error of parameters increase when level of outliers and proportion of outliers contamination increase but they decrease when sample sizes increase.

In case of no outliers in independent variable x_1 and x_2

For all sample sizes , AMSE of ML method is smallest. Whereas sample sizes more than 70 , the AMSE of ML method , WMLCH method and WMLRC method are nearly the same.

In case of one independent variable has outliers

For all level of outliers, proportion of contamination and sample sizes, AMSE of WMLRC method is smallest. Whereas sample sizes more than 60 , the AMSE of WMLCH method and WMLRC method are nearly the same.

In case of independent variable x_1 and x_2 have outliers

For all level of outliers, proportion of contamination and sample sizes, The smallest AMSE is WMLRC method, WMLCH method and ML method respectively. Whereas the AMSE of WMLCH method and WMLRC method are nearly the same.

Department............. Statistics .... Student's signature..

Field of study........... Statistics .... Advisor's signature..

Academic year............ 2004 ....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของ รองศาสตราจารย์ ร้อยเอก มานพ วราภักดิ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำแนะนำ ข้อคิดเห็นต่าง ๆ ในการวิจัยตลอดจนช่วยเหลือแก้ไขข้อบกพร่องต่าง ๆ จนกระทั่งวิทยานิพนธ์เสร็จสมบูรณ์ ผู้วิจัยใคร่ขอกราบขอบพระคุณเป็นอย่างสูง

ผู้วิจัยขอกราบขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ซึ่งประกอบไปด้วย รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา ผู้เป็นประธานกรรมการ รองศาสตราจารย์ วัชราภรณ์ สุริยาภิวัฒน์ ผู้เป็นกรรมการ ที่ช่วยตรวจสอบและแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น และขอกราบขอบพระคุณคณาจารย์ประจำภาควิชาสถิติที่ประสิทธิประสาทความรู้ให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษา

ท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณ บิดา มารดา และพี่ชาย ซึ่งสนับสนุนและให้กำลังใจเสมอมาจนสำเร็จการศึกษา และขอขอบคุณเพื่อน ๆ ทุกคนที่คอยให้กำลังใจมาโดยตลอด

สารบัญ

หน้า

สารบัญตาราง

ตาราง หน้า

สารบัญภาพ

ภาพประกอบ หน้า

บทที่1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันวิธีการวิเคราะห์ทางด้านสถิติเป็นที่นิยมในงานหลาย ๆ ด้าน ทั้งนี้เนื่องจากวิธีการทางสถิติสามารถนำไปประยุกต์ในการวิเคราะห์ วิจัยด้านต่าง ๆ ได้อย่างมีประสิทธิภาพ ซึ่งในการวิเคราะห์ให้มีประสิทธิภาพนั้น จำเป็นต้องเลือกใช้วิธีการทางสถิติให้เหมาะสมกับข้อมูลและวัตถุประสงค์ของงานในด้านนั้น ๆ

วิธีการวิเคราะห์การถดถอยเชิงเส้นเป็นวิธีการหนึ่งที่นิยมนำไปใช้ในงานด้านต่าง ๆ ซึ่งการวิเคราะห์การถดถอยเชิงเส้นนี้เป็นการวิเคราะห์ข้อมูลเพื่ออธิบายถึงความสัมพันธ์ระหว่างตัวแปรตามที่เป็นเชิงปริมาณและตัวแปรอิสระที่เป็นตัวแปรเชิงปริมาณเพียงอย่างเดียว หรืออาจมีตัวแปรอิสระบางตัวที่เป็นตัวแปรเชิงคุณภาพหรือเชิงกลุ่ม แต่มีงานไม่น้อย เช่นงานทางด้านการแพทย์ทางด้านวิทยาศาสตร์ และทางด้านชีวสถิติ ที่พบว่าตัวแปรตามจะเป็นตัวแปรเชิงคุณภาพหรือเชิงกลุ่มที่มีค่าเป็นไปได้ 2 ค่า (Binary หรือ Dichotomous Dependent Variable) เช่น การแพร่กระจายของเชื้อโรคมะเร็ง (แพร่กระจาย ไม่แพร่กระจาย) การตรวจเชื้อ HIV ในร่างกาย (มีเชื้อ ไม่มีเชื้อ) สภาพของแมลงเมื่อได้รับยาพ่น (ตาย ไม่ตาย) และความพึงพอใจในสินค้า (พอใจ ไม่พอใจ) เป็นต้น ดังนั้นจึงเป็นการไม่เหมาะสมที่จะแสดงความสัมพันธ์ของตัวแปรตามและตัวแปรอิสระด้วยการวิเคราะห์การถดถอยเชิงเส้น เนื่องจากตัวแปรตามมีค่าเป็นไปได้ 2 ค่า ซึ่งไม่ใช่การแจกแจงปกติตรงตามข้อสมมติพื้นฐานทั่วไปของสมการถดถอยเชิงเส้นและทำให้ค่าคลาดเคลื่อนไม่มีการแจกแจงแบบปกติด้วย จึงจำเป็นต้องเลือกใช้วิธีการอื่นและวิธีการหนึ่งที่นิยมใช้กันมากคือ การวิเคราะห์การถดถอยโลจิสติค (Logistic regression analysis) มาวิเคราะห์ความสัมพันธ์ของข้อมูลประเภทนี้แทน ซึ่งวัตถุประสงค์และแนวคิดยังคงเหมือนกับการวิเคราะห์การถดถอยเชิงเส้น คือ เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ และนำสมการถดถอยที่ได้ไปประมาณหรือพยากรณ์ค่าตัวแปรตามเมื่อกำหนดค่าตัวแปรอิสระ

เนื่องจากตัวแบบที่แสดงความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระในการวิเคราะห์การถดถอยเชิงเส้นทั่วไป คือ

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{im} + \varepsilon_i \quad , \; i = 1 , 2 , \ldots , N$$

ค่าเฉลี่ยของ Y_i เมื่อกำหนด $\underset{\sim}{x} = (x_1, \ldots, x_m) = \underset{\sim i}{x} = (x_{i1}, \ldots, x_{im})$ คือ $E(Y_i \mid \underset{\sim i}{x})$ ซึ่งด้วยข้อตกลงเบื้องต้น $E(\varepsilon_i) = 0$ จะได้ว่า

$$E(Y_i \mid \underset{\sim i}{x}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im}$$

ซึ่งหมายความว่า $E(Y_i \mid \underset{\sim i}{x})$ อาจเป็นค่าใดๆ ณ ค่าของ $-\infty < \underset{\sim i}{x} < \infty$ ที่กำหนด แต่การวิเคราะห์การถดถอยโลจิสติค ตัวแปรตามเป็นแบบทวิภาค (binary) มีค่าได้เพียง 2 ค่า จึงทำให้ $0 \leq E(Y_i \mid \underset{\sim i}{x}) \leq 1$ ดังนั้นจึงต้องทำการแปลงค่า ดังนี้

$Y_i \mid \underset{\sim i}{x} \sim b(1, \pi(\underset{\sim i}{x}))$ โดยที่ $\pi(\underset{\sim i}{x}) = P(Y_i = 1) = E(Y_i \mid \underset{\sim i}{x}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im}$

ให้ $E(Y_i \mid \underset{\sim i}{x}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im} = \ln\left(\dfrac{\pi(\underset{\sim i}{x})}{1-\pi(\underset{\sim i}{x})}\right)$

เรียกฟังก์ชัน $\text{logit } \pi(\underset{\sim i}{x}) = \ln\left(\dfrac{\pi(\underset{\sim i}{x})}{1-\pi(\underset{\sim i}{x})}\right)$ ว่า “ ฟังก์ชันการถดถอยโลจิสติค “

$$\text{logit}\ (\pi(\underset{\sim i}{x})) = \ln\left(\frac{\pi(\underset{\sim i}{x})}{1-\pi(\underset{\sim i}{x})}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im}$$

เพราะฉะนั้น

$$\frac{\pi(\underset{\sim i}{x})}{1-\pi(\underset{\sim i}{x})} = e^{\beta_0 + \beta_1 x_{i1} + \ldots + \beta_m x_{im}}$$

$$\pi(\underset{\sim i}{x}) = (1 - \pi(\underset{\sim i}{x}))\, e^{\beta_0 + \beta_1 x_{i1} + \ldots + \beta_m x_{im}}$$

$$\pi(\underset{\sim}{x}_i) + \pi(\underset{\sim}{x}_i)\, e^{\beta_0+\beta_1 x_{i1}+\dots+\beta_m x_{im}} \quad = \quad e^{\beta_0+\beta_1 x_{i1}+\dots+\beta_m x_{im}}$$

$$\pi(\underset{\sim}{x}_i) \quad = \quad \frac{e^{\beta_0+\beta_1 x_{i1}+\dots+\beta_m x_{im}}}{1+e^{\beta_0+\beta_1 x_{i1}+\dots+\beta_m x_{im}}}$$

โดยเรียกตัวแบบที่ได้ว่า ตัวแบบการถดถอยโลจิสติค (Logistic regression model)

เมื่อได้ตัวแบบการถดถอยโลจิสติคแล้วจะประมาณค่าของพารามิเตอร์ (β_0 , β_1 , … , β_m) ซึ่งเป็นพารามิเตอร์ที่ไม่ทราบค่า โดยปกติการประมาณค่าพารามิเตอร์ของตัวแบบการถดถอยโลจิสติคนิยมใช้วิธีความควรจะเป็นสูงสุด (Maximum Likelihood) แต่ในทางปฏิบัติอาจพบว่ามีค่าสังเกต x_{ij} ที่ถูกเก็บรวบรวมเพื่อนำมาวิเคราะห์การถดถอยอาจมีค่าสูงกว่าหรือต่ำกว่าค่าสังเกตส่วนใหญ่มาก ซึ่งลักษณะของข้อมูลดังกล่าวจะเรียกว่า ค่าผิดปกติ (Outliers) ในกรณีนี้อาจส่งผลทำให้การประมาณค่าพารามิเตอร์โดยวิธีความควรจะเป็นสูงสุดไม่เหมาะสม

มีนักสถิติหลายท่านได้ทำการคิดวิธีการประมาณค่าพารามิเตอร์ที่มีความแกร่งโดยพยายามลดอิทธิพลของข้อมูลที่มีค่าผิดปกติลง เช่น

Bianco และ Yohai (1996) ได้เสนอวิธีการประมาณที่มีความแกร่งและคงเส้นคงวาด้วยการนำฟังก์ชันขอบเขต (bounded function) และเทอมปรับแก้ความเอนเอียง (a bias correction term) เข้ามาใช้ ต่อมา Croux และ Haesbroeck (2003) ได้ทำการปรับปรุงตัวประมาณของ Bianco และ Yohai ด้วยการเสนอฟังก์ชันขอบเขตใหม่ที่สามารถให้ค่าที่แน่นอนกว่าเดิมและขั้นตอนการคำนวณมีความสะดวกและรวดเร็วสามารถนำไปใช้ได้ง่ายขึ้น ทั้ง Croux และ Haesbroeck ยังได้เสนอวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนัก (Weighted Maximum Likelihood Method) ซึ่งได้ทำการถ่วงน้ำหนักฟังชันความควรจะเป็นในสมการถดถอยโลจิสติค โดยฟังก์ชันความควรจะเป็น มีรูปแบบ

$$\ell(\underset{\sim}{B}) = \prod_{i=1}^{N} (\pi(\underset{\sim}{x}_i))^{y_i} (1-\pi(\underset{\sim}{x}_i))^{1-y_i}$$

และเมื่อทำการถ่วงน้ำหนักจะได้ฟังก์ชันความควรจะเป็นแบบถ่วงน้ำหนัก ในรูปแบบ

$$\ell_{WMLCH}(\underset{\sim}{B}) = \prod_{i=1}^{N} (\pi(\underset{\sim}{x}_i))^{w_i y_i} (1-\pi(\underset{\sim}{x}_i))^{w_i(1-y_i)}$$

โดยที่ w_i เป็นค่าถ่วงน้ำหนักข้อมูลที่มาจากการเปรียบเทียบค่า Robust distance (RD^2 $(\underset{\sim}{x}_i')$) กับค่า $\chi^2_{p,0.975}$ ซึ่งจะให้น้ำหนักน้อยกับข้อมูลที่มีค่าผิดปกติ

Rousseeuw และ Christmann (2003) ได้เสนอวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนัก (Weighted Maximum Likelihood Method) เป็นการประยุกต์ฟังก์ชันความควรจะเป็นในสมการถดถอยโลจิสติคด้วยฟังก์ชันความควรจะเป็นแบบถ่วงน้ำหนัก อยู่ในรูปแบบ

$$\ell_{WMLRC}(\underset{\sim}{B}) \quad = \quad \prod_{i=1}^{N} (\boldsymbol{\pi}(\underset{\sim i}{x}))^{w_i \tilde{y}_i} (1-\boldsymbol{\pi}(\underset{\sim i}{x}))^{w_i (1-\tilde{y}_i)}$$

ซึ่ง $\tilde{y}_i$ เป็นค่าความน่าจะเป็นที่ y_i จะเกิดเหตุการณ์ที่สนใจอย่างมีเงื่อนไขบนการประมาณของสถานะจริงที่ไม่ได้เป็นค่าสังเกตและ w_i เป็นค่าถ่วงน้ำหนักข้อมูลที่คำนวณจากการเปรียบเทียบค่า $RD^2(\underset{\sim i}{x}')$ กับค่าเปอร์เซ็นไทล์ที่ 75 ของค่า $RD^2(\underset{\sim i}{x}')$ ทั้งหมด

จากการศึกษาวิธีการประมาณค่าพารามิเตอร์ข้างต้น ยังไม่มีผู้ศึกษาทำการเปรียบเทียบวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และ Haesbroeck และวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann จึงทำให้ผู้วิจัยมีความสนใจที่จะทำการศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ โดยประกอบไปด้วย

วิธีที่1 คือ วิธีความควรจะเป็นสูงสุด [Maximum Likelihood Method (ML)]

วิธีที่2 คือ วิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และ Haesbroeck [Weighted Maximum Likelihood Method of Croux and Haesbroeck (WMLCH)]

วิธีที่3 คือ วิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann [Weighted Maximum Likelihood Method of Rousseeuw and Christmann (WMLRC)]

เพื่อศึกษาว่าการประมาณค่าพารามิเตอร์วิธีใดมีประสิทธิภาพมากที่สุด เมื่อมีค่าผิดปกติในตัวแปรอิสระโดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยเป็นเกณฑ์ในการพิจารณา

1.2 วัตถุประสงค์การวิจัย

ในการวิจัยนี้มีวัตถุประสงค์ดังนี้

1. เพื่อศึกษาประสิทธิภาพของตัวประมาณที่เป็นความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่ได้จากวิธี ML วิธี WMLCH และวิธี WMLRC ในการประมาณค่าพารามิเตอร์ของตัวแบบการถดถอยโลจิสติค เมื่อตัวแปรอิสระมีค่าผิดปกติ ภายใต้กรณีต่าง ๆ

2. เพื่อหาข้อเสนอแนะวิธีการประมาณที่เหมาะสม ในการประมาณค่าพารามิเตอร์ของตัวแบบการถดถอยโลจิสติค เมื่อตัวแปรอิสระมีค่าผิดปกติ ภายใต้กรณีต่าง ๆ

1.3 สมมติฐานของการวิจัย

เมื่อตัวแปรอิสระมีค่าผิดปกติ วิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann จะให้ตัวประมาณที่มีประสิทธิภาพมากกว่า วิธีความควรจะเป็นสูงสุด และวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และ Haesbroeck

1.4 ข้อตกลงเบื้องต้น

การวิจัยครั้งนี้มีข้อตกลงเบื้องต้นสำหรับการดำเนินงานวิจัย ดังนี้

1. ตัวแบบการถดถอยโลจิสติค (Logistic regression model) อยู่ในรูปแบบ

$$\pi(\underset{\sim}{x}_i) = \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} \quad , \ i = 1, 2, \ldots, N$$

โดยที่ $\pi(\underset{\sim}{x}_i)$ เป็น ความน่าจะเป็นที่ y = 1 ณ ระดับ x_{i1} และ x_{i2}

x_{i1} , x_{i2} เป็น ตัวแปรอิสระตัวที่ 1 และ ตัวแปรอิสระตัวที่ 2

β_i เป็น พารามิเตอร์ที่ไม่ทราบค่า ; i = 0 , 1 , 2

N เป็น ขนาดตัวอย่าง

2. ข้อมูลที่มีค่าผิดปกติจะกำหนดให้เกิดในกรณีต่างๆดังนี้

2.1 กรณีที่มีค่าผิดปกติในตัวแปรอิสระเพียงหนึ่งตัวใด ๆ ในที่นี้ให้เป็น x_1 ตัวอย่างของข้อมูลกรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1 แสดงไว้ในตารางที่ 1.1 และในรูปที่ 1.1

2.2 กรณีที่มีค่าผิดปกติในตัวแปรอิสระทั้งหมดคือใน x_1 และ x_2 อย่างอิสระกัน ตัวอย่างของข้อมูลกรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 แสดงไว้ในตารางที่ 1.2 และในรูปที่ 1.2

3. กรณีที่มีค่าผิดปกติในตัวแปรอิสระ ในที่นี้จะกำหนดให้ตัวแปรอิสระที่มีค่าผิดปกติมี 2 ระดับ ตามเงื่อนไขของการตรวจสอบค่าผิดปกติโดยใช้กราฟแบบ Box และ Whisker คือ

1. ค่าผิดปกติระดับไม่รุนแรง (Mild Outliers) คือ ค่าของตัวแปรที่อยู่ในช่วง ($Q_1 - 3(IQR)$, $Q_1 - 1.5(IQR)$) หรือ ($Q_3 + 1.5(IQR)$, $Q_3 + 3(IQR)$)

เมื่อ Q_1 คือ ค่าควอไทล์ที่ 1 (The First Quartile)

Q_3 คือ ค่าควอไทล์ที่ 3 (The Third Quartile)

และ IQR คือ ระยะห่างระหว่างควอไทล์ (The Interquartile Range) ซึ่งเท่ากับ $Q_3 - Q_1$

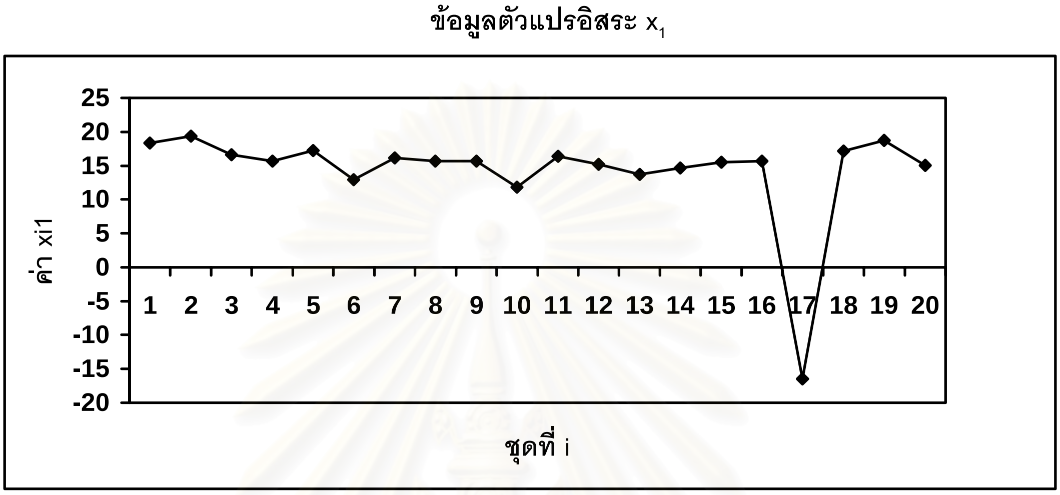
2. ค่าผิดปกติระดับรุนแรง (Extreme Outliers) คือ ค่าของตัวแปรที่อยู่ในช่วง $(-\infty, Q_1 - 3(IQR))$ หรือ $(Q_3 + 3(IQR), \infty)$

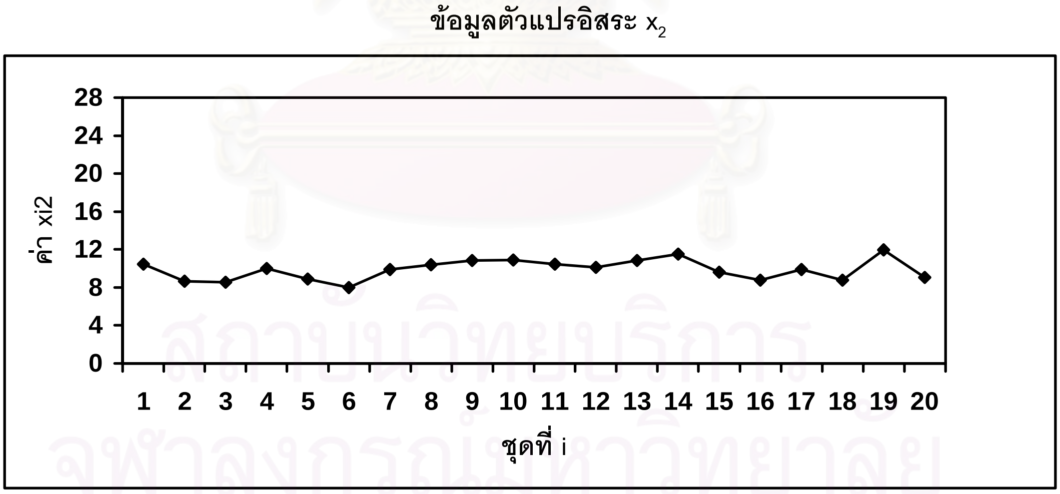
หมายเหตุ ผู้วิจัยได้แสดงรายละเอียดและเงื่อนไขของการตรวจสอบค่าผิดปกติโดยใช้กราฟแบบ Box Plot ในภาคผนวก ก

ตารางที่1.1 แสดงตัวเลขค่าของข้อมูลตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 กรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1 เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับรุนแรง และสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 เท่ากับ 0.05 โดยมีขนาดตัวอย่างเท่ากับ 20

สัดส่วนการปลอมปน PX1	**ชุดที่**	**ตัวแปรอิสระ x_1**	**ตัวแปรอิสระ x_2**
0.05	1	18.37056	10.46361
	2	19.33175	8.640879
	3	16.58217	8.542603
	4	15.6625	9.977243
	5	17.22291	8.889557
	6	12.94806	7.981412
	7	16.11109	9.901498
	8	15.65162	10.37625
	9	15.70206	10.86955
	10	11.82461	10.91223
	11	16.3951	10.4555
	12	15.16438	10.10795
	13	13.69994	10.86947
	14	14.62774	11.50791
	15	15.55188	9.605157
	16	15.64789	8.767909
	17	-16.50691	9.878396
	18	17.18581	8.790048
	19	18.75025	11.95257
	20	15.05794	9.078777

รูปที่1.1 แสดงตัวอย่างของข้อมูลกรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1 เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับรุนแรง และสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 เท่ากับ 0.05 โดยมีขนาดตัวอย่างเท่ากับ 20



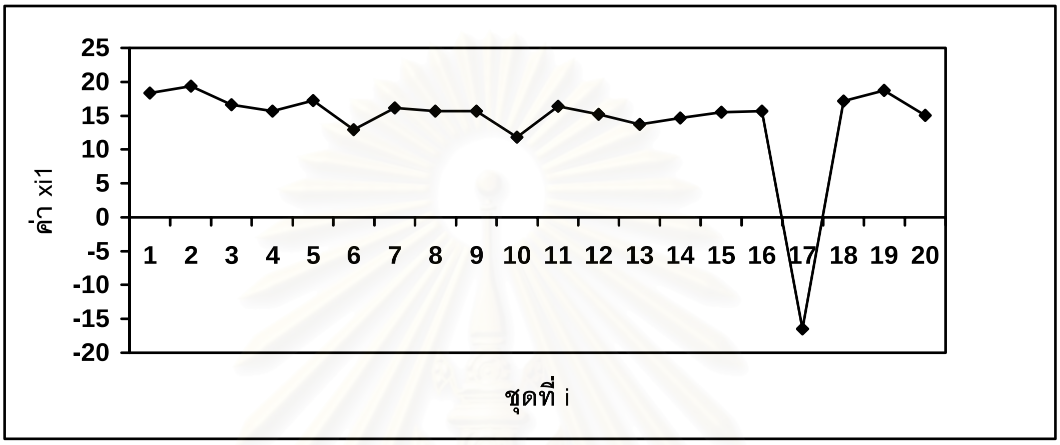


ตารางที่1.2 แสดงตัวเลขค่าของข้อมูลตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 กรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 เมื่อตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 มีค่าผิดปกติระดับรุนแรง และสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 เท่ากับ 0.05 โดยมีขนาดตัวอย่างเท่ากับ 20

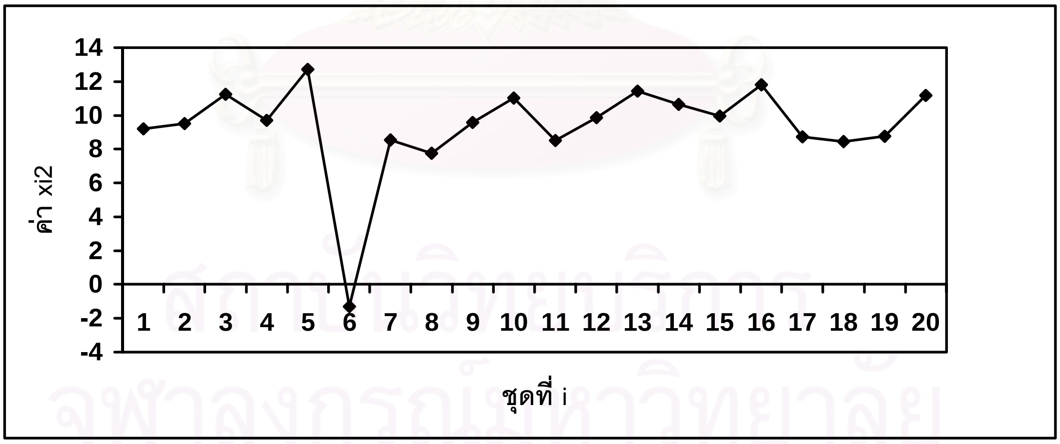
สัดส่วนการปลอมปน PX1,PX2	**ชุดที่**	**ตัวแปรอิสระ x_1**	**ตัวแปรอิสระ x_2**
0.05,0.05	1	18.37056	9.201059
	2	19.33175	9.507368
	3	16.58217	11.23151
	4	15.6625	9.707528
	5	17.22291	12.70403
	6	12.94806	-1.294142
	7	16.11109	8.53803
	8	15.65162	7.75778
	9	15.70206	9.578762
	10	11.82461	11.02919
	11	16.3951	8.511627
	12	15.16438	9.849738
	13	13.69994	11.42984
	14	14.62774	10.65632
	15	15.55188	9.954899
	16	15.64789	11.80784
	17	-16.50691	8.722265
	18	17.18581	8.438353
	19	18.75025	8.768954
	20	15.05794	11.18347

รูปที่1.2 แสดงตัวอย่างของข้อมูลกรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 เมื่อตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 มีค่าผิดปกติระดับรุนแรง และสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 เท่ากับ 0.05 โดยมีขนาดตัวอย่างเท่ากับ 20

ข้อมูลตัวแปรอิสระ x_1



ข้อมูลตัวแปรอิสระ x_2



1.5 ขอบเขตการวิจัย

ในการวิจัยครั้งนี้ ทำการศึกษาในกรณีที่ตัวแปรอิสระมีค่าผิดปกติ โดยทำการศึกษาในกรณีต่างๆ ดังนี้

1. ตัวแปรตาม (Y) เป็นข้อมูลเชิงคุณภาพที่มี 2 ค่า คือ 0 และ 1
2. ตัวแปรอิสระ (x) ในตัวแบบทำการศึกษา 2 ตัวแปร
3. ตัวแปรอิสระเป็นข้อมูลเชิงปริมาณ
4. ตัวแปรอิสระ x_1 และ x_2 ไม่มีความสัมพันธ์กัน
5. กำหนดค่าพารามิเตอร์ของสมการการถดถอยเป็นค่าใด ๆ ในที่นี้กำหนด $\beta_0 = 5$, $\beta_1 = -1$, $\beta_2 = 1$
6. กำหนดลักษณะตัวแปรอิสระตามการแจกแจงของข้อมูล ดังนี้

1. การแจกแจงแบบปกติ (Normal Distribution) สำหรับจำลองตัวแปรอิสระมีค่าปกติ

ฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad , \quad -\infty < x < \infty$$

เมื่อ $E(x) = \mu \quad ; \quad V(x) = \sigma^2$

2. การแจกแจงแบบปกติปลอมปน (Contaminated Normal Distribution) สำหรับจำลองตัวแปรอิสระมีค่าผิดปกติ

ฟังก์ชันความหนาแน่นอยู่ในรูปของ

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

เมื่อ P แทน สัดส่วนการปลอมปน (Proportion of Contamination)
และ C แทน ตัวประกอบสเกล (Scale Factor)

ในงานวิจัยครั้งนี้ จะศึกษาระดับค่าผิดปกติของตัวแปรอิสระ 2 ระดับ คือ

1. ระดับไม่รุนแรง
2. ระดับรุนแรง

โดยที่ เกณฑ์การกำหนดขนาดค่าผิดปกติด้วย Box Plot จะได้ว่า

เมื่อค่า C = 4 , 5 , 6 ตัวแปรอิสระจะมีค่าผิดปกติระดับไม่รุนแรง

และ เมื่อค่า C = 12 , 13 ตัวแปรอิสระจะมีค่าผิดปกติระดับรุนแรง

เลือกกำหนดค่า C = 5 เมื่อตัวแปรอิสระมีค่าผิดปกติระดับไม่รุนแรง และเลือกกำหนดค่า C = 12 เมื่อตัวแปรอิสระมีค่าผิดปกติระดับรุนแรง

แต่ละระดับจะกำหนดให้มีสัดส่วนการปลอมปนตามขนาดตัวอย่างเท่ากับ 0.05 , 0.10 และ 0.15

ในงานวิจัยครั้งนี้ จะศึกษาในกรณีที่ตัวแปรอิสระ x_1 มีการแจกแจงแบบปกติที่มีค่าพารามิเตอร์ μ = 15 และ σ^2 = 4 และตัวแปรอิสระ x_2 มีการแจกแจงแบบปกติที่มีค่าพารามิเตอร์ μ = 10 และ σ^2 = 2

กรณีมีค่าผิดปกติในตัวแปรอิสระ x_1 จะใช้การแจกแจงแบบปกติปลอมปน

กรณีมีค่าผิดปกติในตัวแปรอิสระ x_1 และ x_2 จะใช้การแจกแจงแบบปกติปลอมปนที่เหมือนกัน

7. ขนาดตัวอย่างที่ทำการศึกษาคือ N = 20 , 30 , 40 , 50 , 60 , 70 , 80 , 90 และ100

8. การวิจัยครั้งนี้ใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) เขียนด้วยโปรแกรมภาษาฟอร์แทน (Fortran) ทำการจำลองข้อมูลซ้ำ 1000 รอบ ในแต่ละสถานการณ์ที่ทำการศึกษา

1.6 เกณฑ์ที่ใช้การประเมินประสิทธิภาพของตัวประมาณ

เกณฑ์ที่ใช้ในการประเมินประสิทธิภาพของตัวประมาณว่าวิธีในการหาค่าประมาณของพารามิเตอร์ของสมการการถดถอยโลจิสติควิธีใดมีประสิทธิภาพมากที่สุดจะพิจารณาจากเกณฑ์ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย (Average Mean Square Error : AMSE) โดยวิธีที่ให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดจะเป็นวิธีที่มีประสิทธิภาพมากที่สุด โดยมีวิธีในการคำนวณดังนี้

$$MSE_j = \frac{1}{1000}\sum_{i=1}^{1000}\left[\hat{\beta}_{ij} - \beta_j\right]^2$$

$$\text{AMSE} = \frac{1}{k}\sum_{j=1}^{k} \text{MSE}_j$$

เมื่อ β_j แทน พารามิเตอร์ตัวที่ j ในสมการการถดถอยโลจิสติค

$\hat{\beta}_{ij}$ แทน ค่าประมาณของพารามิเตอร์ตัวที่ j (β_j) จากการประมาณครั้งที่ i

k แทน จำนวนพารามิเตอร์ ในที่นี้ k = 3

MSE_j แทน ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสำหรับพารามิเตอร์ β_j

AMSE แทน ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์จำนวน k ตัว

1.7 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่คาดว่าจะได้รับในการวิจัยครั้งนี้

1. เพื่อทราบประสิทธิภาพของตัวประมาณในสถานการณ์ต่าง ๆ ที่ศึกษา

2. เพื่อเป็นแนวทางในการตัดสินใจในการเลือกวิธีการประมาณค่าพารามิเตอร์ได้อย่างเหมาะสม เมื่อข้อมูลมีค่าผิดปกติ

3. เพื่อเป็นแนวทางในการศึกษาและเปรียบเทียบกับวิธีการประมาณค่าพารามิเตอร์เมื่อข้อมูลมีค่าผิดปกติในสถานการณ์อื่น ๆ ต่อไป

4. เพื่อเป็นแนวทางในการศึกษาวิเคราะห์ตัวประมาณอื่น ๆ ต่อไป

บทที่2

ทฤษฎีที่เกี่ยวข้องกับการวิจัย

ในบทนี้จะกล่าวถึงข้อสมมติของตัวแบบการถดถอยโลจิสติค การประมาณค่าพารามิเตอร์ของตัวแปรอิสระโดยใช้วิธีความควรจะเป็นสูงสุด วิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และ Haesbroeck และวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann

2.1 ข้อสมมติทั่วไปของตัวแบบการถดถอยโลจิสติค

ในการวิจัยครั้งนี้ข้อมูลที่นำมาพิจารณาอยู่ในรูป $(\underset{\sim}{x}_i, y_i)$; i = 1,2,...,N ซึ่งตัวแปรตาม y_i มีการแจกแจงแบบแบร์นูลลี (Bernoulli Distribution) [$y_i \sim b(1, \pi(\underset{\sim}{x}_i))$] ซึ่ง $\underset{\sim}{x}_i$ แทน $(x_{i1}, x_{i2}, ..., x_{im})$ และ $\underset{\sim}{x}_i$ เป็นตัวแปรอิสระร่วมกัน m ตัว (m-covariate explanatory variable) ซึ่งความน่าจะเป็นที่จะเกิดเหตุการณ์ที่สนใจคือ $\pi(\underset{\sim}{x}_i)$ นั่นคือ

$$Y_i = \begin{cases} 1 & \text{ด้วยความน่าจะเป็น} \quad \pi(\underset{\sim}{x}_i) \\ 0 & \text{ด้วยความน่าจะเป็น} \quad 1-\pi(\underset{\sim}{x}_i) \end{cases}$$

ให้ $\eta = \ln\left(\dfrac{\pi(\underset{\sim}{x}_i)}{1-\pi(\underset{\sim}{x}_i)}\right)$ เป็นฟังก์ชันเชื่อมโยงแบบโลจิสติค ซึ่งจะทำการแปลงค่า $\pi(\underset{\sim}{x}_i)$ จากช่วง (0,1) เป็นค่าที่อยู่ในช่วง $(-\infty,\infty)$ โดยที่

$$\eta = E(Y_i \mid \underset{\sim}{x}_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im}$$

เมื่อ η แทนฟังก์ชันเชื่อมโยง จะได้ว่า

$$\ln\left(\frac{\pi(\underset{\sim}{x}_i)}{1-\pi(\underset{\sim}{x}_i)}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im}$$

$$\frac{\pi(\underset{\sim}{x}_i)}{1-\pi(\underset{\sim}{x}_i)} = e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}}$$

$$\pi(\underset{\sim}{x}_i) = \frac{e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}}}{1+e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}}}$$

สามารถเขียนสมการถดถอยโลจิสติคเชิงเส้น ในรูปของเมทริกซ์ ดังนี้

$$\underset{\sim}{G}(X) = X\underset{\sim}{B}$$

เมื่อ

$$\underset{\sim}{G}(X) = \begin{bmatrix} \text{logit}(\pi(\underset{\sim}{x}_1)) \\ \text{logit}(\pi(\underset{\sim}{x}_2)) \\ \vdots \\ \text{logit}(\pi(\underset{\sim}{x}_N)) \end{bmatrix}_{N\times 1}$$

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{Nm} \end{bmatrix}_{N\times(m+1)}$$

และ

$$\underset{\sim}{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}_{(m+1)\times 1}$$

ลอการิทึมธรรมชาติของฟังก์ชันความควรจะเป็นในสมการถดถอยโลจิสติค (Log Likelihood Function)

ฟังก์ชันความควรจะเป็นของ $\underset{\sim}{y} = (y_1, y_2, \ldots, y_N)$ ให้แทนด้วย

$$\ell(\underset{\sim}{B}) = \prod_{i=1}^{N} (\pi(\underset{\sim}{x}_i))^{y_i} (1-\pi(\underset{\sim}{x}_i))^{1-y_i}$$

ลอการิทึมธรรมชาติของฟังก์ชันความควรจะเป็น คือ

$$\begin{aligned}
L(\underset{\sim}{B}) &= \ln(\ell(\underset{\sim}{B})) \\
&= \sum_{i=1}^{N}\left\{ y_i \ln(\pi(\underset{\sim}{x}_i)) + (1-y_i)\ln(1-\pi(\underset{\sim}{x}_i))\right\} \\
&= \sum_{i=1}^{N}\left\{ y_i \ln(\pi(\underset{\sim}{x}_i)) + \ln(1-\pi(\underset{\sim}{x}_i)) - y_i \ln(1-\pi(\underset{\sim}{x}_i))\right\} \\
&= \sum_{i=1}^{N}\left\{ y_i \left[\ln(\pi(\underset{\sim}{x}_i)) - \ln(1-\pi(\underset{\sim}{x}_i))\right] + \ln(1-\pi(\underset{\sim}{x}_i))\right\} \\
&= \sum_{i=1}^{N}\left\{ y_i \ln\left(\frac{\pi(\underset{\sim}{x}_i)}{1-\pi(\underset{\sim}{x}_i)}\right) + \ln(1-\pi(\underset{\sim}{x}_i))\right\} \\
&= \sum_{i=1}^{N}\left\{ y_i \ln(e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}}) + \ln\left[1-\frac{e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}}}{1+e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}}}\right]\right\} \\
&= \sum_{i=1}^{N}\left\{ y_i \ln(e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}}) + \ln\left(\frac{1}{1+e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}}}\right)\right\} \\
&= \sum_{i=1}^{N}\left\{ y_i \ln(e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}}) + \ln 1 - \ln(1+e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}})\right\} \\
&= \sum_{i=1}^{N}\left\{ y_i \ln(e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}}) - \ln(1+e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}})\right\}
\end{aligned}$$

$$= \sum_{i=1}^{N}\left\{ y_i(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_m x_{im}) - \ln(1 + e^{\beta_0 + \beta_1 x_{i1} + \ldots + \beta_m x_{im}})\right\} \quad (2.1)$$

2.2 การประมาณค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุด

หลักการของวิธีความควรจะเป็นสูงสุดมีจุดประสงค์เพื่อหาตัวประมาณพารามิเตอร์ที่ทำให้ $\ell(\underset{\sim}{B})$ มีค่ามากที่สุด โดยหาอนุพันธ์ (differentiate) เทียบกับ พารามิเตอร์ $\beta_0 , \beta_1 , \ldots , \beta_m$ และให้ผลลัพธ์มีค่าเท่ากับศูนย์ จะได้สมการที่ไม่เป็นเชิงเส้น m + 1 สมการ ซึ่งสามารถประมาณค่าพารามิเตอร์ $\beta_0 , \beta_1 , \ldots , \beta_m$ ได้ด้วยวิธี Newton – Raphson

วิธี Newton – Raphson จะหาอนุพันธ์ย่อย (partial differentiate) ของลอการิทึมธรรมชาติของฟังก์ชันความควรจะเป็น $L(\underset{\sim}{B})$ เทียบกับ $\beta_0 , \beta_1 , \ldots , \beta_m$ เรียกว่า efficient scores แล้วนำมาเป็นสมาชิกของเวกเตอร์ $\underset{\sim}{U}(\underset{\sim}{B})$ ที่มีอันดับ $(m+1) \times 1$

$$\underset{\sim}{U}(\underset{\sim}{B}) = \begin{bmatrix} \dfrac{\partial L(\underset{\sim}{B})}{\partial \beta_0} \\ \dfrac{\partial L(\underset{\sim}{B})}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial L(\underset{\sim}{B})}{\partial \beta_m} \end{bmatrix}_{(m+1)\times 1}$$

กำหนดเมทริกซ์ $H(\underset{\sim}{B})$ อันดับ $(m+1) \times (m+1)$ มีสมาชิกเป็นอนุพันธ์ย่อยอันดับที่สอง (second partial derivative) ของ $L(\underset{\sim}{B})$ ซึ่งเรียกเมทริกซ์ $H(\underset{\sim}{B})$ ว่า Hessian matrix

โดยที่สมาชิกตัวที่ (j , k) คือ $\dfrac{\partial^2 L(\underset{\sim}{B})}{\partial \beta_j \partial \beta_k}$, j , k = 0 , 1 , 2 , … , m

สามารถหาเวกเตอร์ $\underset{\sim}{U}(\hat{\underset{\sim}{B}})$ ซึ่งเป็นเวกเตอร์ที่ประมาณด้วยความควรจะเป็นสูงสุดของ $\underset{\sim}{U}(\underset{\sim}{B})$ โดยใช้ Taylor series กระจาย $\underset{\sim}{U}(\underset{\sim}{B})$ รอบ $\underset{\sim}{B}^*$ จะได้

$$\underset{\sim}{U}(\hat{\underset{\sim}{B}}) \approx \underset{\sim}{U}(\underset{\sim}{B}^*) + H(\underset{\sim}{B}^*)(\hat{\underset{\sim}{B}} - \underset{\sim}{B}^*)$$

โดยนิยามของความควรจะเป็นสูงสุดของ $\underset{\sim}{B}$ จะได้ว่า

$$\left.\frac{\partial L(\underset{\sim}{B})}{\partial \beta_j}\right|_{\underset{\sim}{\hat{B}}} = 0 \qquad \text{สำหรับ} \quad j = 0, 1, 2, \ldots, m$$

และ $\underset{\sim}{U}(\underset{\sim}{\hat{B}}) = 0$

ดังนั้น $\qquad \underset{\sim}{\hat{B}} = \underset{\sim}{B}^{*} - H^{-1}(\underset{\sim}{B}^{*})\,\underset{\sim}{U}(\underset{\sim}{B}^{*})$

ซึ่งชี้นำให้มีการประมาณ $\underset{\sim}{\hat{B}}$ โดยการคำนวณซ้ำๆ ซึ่งค่าประมาณ $\underset{\sim}{\hat{B}}$ ณ รอบที่ r+1 คือ

$$\underset{\sim}{\hat{B}}_{r+1} = \underset{\sim}{\hat{B}}_{r} - H^{-1}(\underset{\sim}{B}_r)\,\underset{\sim}{U}(\underset{\sim}{B}_r)$$

สำหรับ r = 0 , 1 , 2 , … ซึ่งเวกเตอร์ $\underset{\sim}{\hat{B}}_0$ เป็นเวกเตอร์ของตัวประมาณเริ่มต้น ในที่นี้กำหนดค่าตัวประมาณเริ่มต้นจากค่า $\underset{\sim}{\hat{B}}$ ของวิธีกำลังสองน้อยที่สุด(OLS) โดยวิธีการประมาณค่าพารามิเตอร์ทั้ง 3 วิธีใช้วิธี Newton-Raphson ในการแก้สมการเพื่อหาค่าประมาณพารามิเตอร์เหมือนกัน ดังนั้นจึงใช้ค่า $\underset{\sim}{\hat{B}}_0$ เดียวกันในการเป็นตัวประมาณเริ่มต้น

ถ้าผลต่างระหว่าง $\underset{\sim}{\hat{B}}$ ในรอบที่ r กับรอบที่ r+1 มีค่าน้อยมากจนถือว่าไม่แตกต่างกัน ซึ่งในที่นี้กำหนดเกณฑ์ว่า $\left|\underset{\sim}{\hat{B}}_{r+1} - \underset{\sim}{\hat{B}}_{r}\right| < 0.00001$ ค่า $\underset{\sim}{\hat{B}}_{r+1}$ นั้นจะเป็นค่าที่ยอมรับได้

วิธี Newton – Raphson สำหรับตัวแปรอิสระ 2 ตัว

ข้อมูลประกอบด้วยค่าสังเกตที่เป็นอิสระ N ค่า คือ (x_{i1}, x_{i2}, y_i) ; i = 1,2,…,N สมการถดถอยโลจิสติคเชิงเส้น คือ

$$\text{logit}\ (\pi(\underset{\sim}{x}_i)) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

เมื่อ $\qquad \underset{\sim}{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$

$\underset{\sim}{U}(\underset{\sim}{B})$ เป็นเวกเตอร์อันดับ 3×1 มีสมาชิกคือ $\dfrac{\partial L(\underset{\sim}{B})}{\partial \beta_j}$, j= 0,1,2

$$\underset{\sim}{U}(\underset{\sim}{B}) = \begin{bmatrix} \dfrac{\partial L(\underset{\sim}{B})}{\partial \beta_0} \\ \dfrac{\partial L(\underset{\sim}{B})}{\partial \beta_1} \\ \dfrac{\partial L(\underset{\sim}{B})}{\partial \beta_2} \end{bmatrix} \qquad (2.2)$$

และ Hessian matrix , $H(\underset{\sim}{B})$ เป็นเมทริกซ์อันดับ 3×3 มีสมาชิกตัวที่ (j,k) คือ $\dfrac{\partial^2 L(\underset{\sim}{B})}{\partial \beta_j \partial \beta_k}$

, j ,k = 0,1,2

$$H(\underset{\sim}{B}) = \begin{bmatrix} \dfrac{\partial^2 L(\underset{\sim}{B})}{\partial \beta_0^2} & \dfrac{\partial^2 L(\underset{\sim}{B})}{\partial \beta_0 \partial \beta_1} & \dfrac{\partial^2 L(\underset{\sim}{B})}{\partial \beta_0 \partial \beta_2} \\ \dfrac{\partial^2 L(\underset{\sim}{B})}{\partial \beta_1 \partial \beta_0} & \dfrac{\partial^2 L(\underset{\sim}{B})}{\partial \beta_1^2} & \dfrac{\partial^2 L(\underset{\sim}{B})}{\partial \beta_1 \partial \beta_2} \\ \dfrac{\partial^2 L(\underset{\sim}{B})}{\partial \beta_2 \partial \beta_0} & \dfrac{\partial^2 L(\underset{\sim}{B})}{\partial \beta_2 \partial \beta_1} & \dfrac{\partial^2 L(\underset{\sim}{B})}{\partial \beta_2^2} \end{bmatrix} \qquad (2.3)$$

สมาชิกในเวกเตอร์ $\underset{\sim}{U}(\underset{\sim}{B})$ และในเมทริกซ์ $H(\underset{\sim}{B})$ หาได้จาก $L(\underset{\sim}{B})$ ในสมการ (2.1) และสำหรับตัวแปรอิสระ 2 ตัว

$$L(\underset{\sim}{B}) = \sum_{i=1}^{N} \left\{ y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) - \ln(1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}}) \right\}$$

หาอนุพันธ์ของ $L(\underset{\sim}{B})$ เทียบกับ β_0 , β_1 , และ β_2 ดังนี้

$$\frac{\partial L(\underset{\sim}{B})}{\partial \beta_0} = \sum_{i=1}^{N} \left\{ y_i - \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}}} \right\}$$

$$= \sum_{i=1}^{N}\left\{y_i - \pi(\underset{\sim}{x}_i)\right\}$$

$$\frac{\partial L(\underset{\sim}{B})}{\partial \beta_1} = \sum_{i=1}^{N}\left\{y_i x_{i1} - x_{i1}\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right\}$$

$$= \sum_{i=1}^{N} x_{i1}\left\{y_i - \pi(\underset{\sim}{x}_i)\right\}$$

$$\frac{\partial L(\underset{\sim}{B})}{\partial \beta_2} = \sum_{i=1}^{N}\left\{y_i x_{i2} - x_{i2}\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right\}$$

$$= \sum_{i=1}^{N} x_{i2}\left\{y_i - \pi(\underset{\sim}{x}_i)\right\}$$

โดยที่ $\pi(\underset{\sim}{x}_i) = \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}$

และสามารถหาอนุพันธ์ย่อยอันดับที่ 2 ได้ดังนี้

$$\frac{\partial^2 L(\underset{\sim}{B})}{\partial \beta_0^2} = -\sum_{i=1}^{N}\left\{\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} - \left(\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N}\pi(\underset{\sim}{x}_i)(1-\pi(\underset{\sim}{x}_i))$$

$$\frac{\partial^2 L(\underset{\sim}{B})}{\partial \beta_1^2} = -\sum_{i=1}^{N}x_{i1}\left\{x_{i1}\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} - x_{i1}\left(\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N}x_{i1}^2\pi(\underset{\sim}{x}_i)(1-\pi(\underset{\sim}{x}_i))$$

$$\frac{\partial^2 L(\underset{\sim}{B})}{\partial \beta_2^2} = -\sum_{i=1}^{N}x_{i2}\left\{x_{i2}\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} - x_{i2}\left(\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N}x_{i2}^2\pi(\underset{\sim}{x}_i)(1-\pi(\underset{\sim}{x}_i))$$

$$\frac{\partial^2 L(\underset{\sim}{B})}{\partial\beta_0\partial\beta_1} = -\sum_{i=1}^{N} x_{i1}\left\{\frac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}-\left(\frac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N} x_{i1}\pi(\underset{\sim}{x}_i)(1-\pi(\underset{\sim}{x}_i))$$

ซึ่ง $\frac{\partial^2 L(\underset{\sim}{B})}{\partial\beta_1\partial\beta_0} = \frac{\partial^2 L(\underset{\sim}{B})}{\partial\beta_0\partial\beta_1}$

$$\frac{\partial^2 L(\underset{\sim}{B})}{\partial\beta_0\partial\beta_2} = -\sum_{i=1}^{N} x_{i2}\left\{\frac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}-\left(\frac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N} x_{i2}\pi(\underset{\sim}{x}_i)(1-\pi(\underset{\sim}{x}_i))$$

ซึ่ง $\frac{\partial^2 L(\underset{\sim}{B})}{\partial\beta_2\partial\beta_0} = \frac{\partial^2 L(\underset{\sim}{B})}{\partial\beta_0\partial\beta_2}$

และ

$$\frac{\partial^2 L(\underset{\sim}{B})}{\partial\beta_1\partial\beta_2} = -\sum_{i=1}^{N} x_{i2}\left\{x_{i1}\frac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}-x_{i1}\left(\frac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N} x_{i1}x_{i2}\pi(\underset{\sim}{x}_i)(1-\pi(\underset{\sim}{x}_i))$$

ซึ่ง $\frac{\partial^2 L(\underset{\sim}{B})}{\partial\beta_2\partial\beta_1} = \frac{\partial^2 L(\underset{\sim}{B})}{\partial\beta_1\partial\beta_2}$

โดยที่ $\pi(\underset{\sim}{x}_i)(1-\pi(\underset{\sim}{x}_i)) = \frac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}{(1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}})^2}$

นำสมาชิกของเวกเตอร์ $\underset{\sim}{U}(\underset{\sim}{B})$ และเมทริกซ์ $H(\underset{\sim}{B})$ ที่หามาได้ไปแทนในสมการ (2.2) และ (2.3) ดังนั้นคำนวณหาค่า $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ ได้จากสมการ

$$\underset{\sim}{\hat{B}}_{r+1} = \underset{\sim}{\hat{B}}_r - H^{-1}(\underset{\sim}{B}_r)\,\underset{\sim}{U}(\underset{\sim}{B}_r)$$

เมื่อ $\underset{\sim}{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$; $\underset{\sim}{U}(\underset{\sim}{B}) = \begin{bmatrix} \sum_{i=1}^{N}\left\{ y_i - \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} \right\} \\ \sum_{i=1}^{N} x_{i1}\left\{ y_i - \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} \right\} \\ \sum_{i=1}^{N} x_{i2}\left\{ y_i - \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} \right\} \end{bmatrix}$

และ

$$H(\underset{\sim}{B}) = \begin{bmatrix} -\sum_{i=1}^{N} \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} x_{i1} \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} x_{i2} \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} \\ -\sum_{i=1}^{N} x_{i1} \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} x_{i1}^2 \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} x_{i1}x_{i2} \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} \\ -\sum_{i=1}^{N} x_{i2} \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} x_{i1}x_{i2} \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} x_{i2}^2 \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} \end{bmatrix}$$

ถ้าผลต่างระหว่าง $\hat{\underset{\sim}{B}}$ ในรอบที่ r กับรอบที่ r+1 มีค่าน้อยมากจนถือว่าไม่แตกต่างกัน โดยมีเกณฑ์ว่า $\left|\hat{\underset{\sim}{B}}_{r+1} - \hat{\underset{\sim}{B}}_{r}\right| < 0.00001$ ค่า $\hat{\underset{\sim}{B}}_{r+1}$ นั้นจะเป็นค่าที่ยอมรับได้

2.3 การประมาณค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และ Haesbroeck

หลักการของวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และ Haesbroeck มีวัตถุประสงค์เหมือนกับวิธีความควรจะเป็นสูงสุด นั่นคือเพื่อหาตัวประมาณพารามิเตอร์ที่ทำให้ $\ell_{WMLCH}(\underset{\sim}{B})$ มีค่ามากที่สุด โดยหาอนุพันธ์เทียบกับพารามิเตอร์ β_0 , β_1 , … , β_m และให้ผลลัพธ์มีค่าเท่ากับศูนย์ จะได้สมการที่ไม่เป็นเชิงเส้น m + 1 สมการ ซึ่งสามารถประมาณค่าพารามิเตอร์ β_0 , β_1 , … , β_m ได้ด้วยวิธี Newton – Raphson

ฟังก์ชันความควรจะเป็นแบบถ่วงน้ำหนัก คือ

$$\ell_{WMLCH}(\underset{\sim}{B}) = \prod_{i=1}^{N} (\pi(\underset{\sim}{x}_i))^{w_i y_i} (1-\pi(\underset{\sim}{x}_i))^{w_i(1-y_i)}$$

ดังนั้นลอการิทึมธรรมชาติของฟังก์ชันความควรจะเป็นแบบถ่วงน้ำหนัก คือ

$$
\begin{aligned}
L_{WMLCH}(\underset{\sim}{B}) &= \ln(\ell_{WMLCH}(\underset{\sim}{B})) \\
&= \sum_{i=1}^{N}\left\{ w_i y_i \ln(\pi(\underset{\sim}{x}_i)) + w_i(1-y_i)\ln(1-\pi(\underset{\sim}{x}_i)) \right\} \\
&= \sum_{i=1}^{N}\left\{ w_i y_i \ln(\pi(\underset{\sim}{x}_i)) + w_i \ln(1-\pi(\underset{\sim}{x}_i)) - w_i y_i \ln(1-\pi(\underset{\sim}{x}_i)) \right\} \\
&= \sum_{i=1}^{N}\left\{ w_i y_i \left[\ln(\pi(\underset{\sim}{x}_i)) - \ln(1-\pi(\underset{\sim}{x}_i))\right] + w_i \ln(1-\pi(\underset{\sim}{x}_i)) \right\} \\
&= \sum_{i=1}^{N}\left\{ w_i y_i \ln\left(\frac{\pi(\underset{\sim}{x}_i)}{1-\pi(\underset{\sim}{x}_i)}\right) + w_i \ln(1-\pi(\underset{\sim}{x}_i)) \right\} \\
&= \sum_{i=1}^{N}\left\{ w_i y_i \ln(e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}}) - w_i \ln(1+e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}}) \right\} \\
&= \sum_{i=1}^{N}\left\{ w_i y_i (\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}) - w_i \ln(1+e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}}) \right\} \qquad (2.4)
\end{aligned}
$$

โดยที่ w_i เป็นการถ่วงน้ำหนักข้อมูล

ซึ่ง
$$
w_i = \begin{cases} 1 & \text{ถ้า } RD^2(\underset{\sim}{x}_i') \le \chi^2_{p,0.975} \\ 0 & \text{อื่น} \end{cases}
$$

โดยที่ $RD(\underset{\sim}{x}_i')$ เป็น Robust distance

ค่า $RD^2(\underset{\sim}{x}_i') = (\underset{\sim}{x}_i' - \bar{x})' S^{-1} (\underset{\sim}{x}_i' - \bar{x})$

ซึ่ง $\bar{x}$ เป็นค่าเฉลี่ยของชุดค่าสังเกต h ชุดจากทั้งหมด N ชุด

S เป็นเมทริกซ์ความแปรปรวนร่วมของชุดค่าสังเกต h ชุดจากทั้งหมด N ชุด

โดยใช้ตัวประมาณ MCD (Minimum Covariance Determinant) หาชุดค่าสังเกต h ชุด ที่ทำให้เมทริกซ์ความแปรปรวนร่วมมีค่าดีเทอร์มิแนนต์ (determinant) น้อยที่สุด

และ p เป็นจำนวนตัวแปรอิสระ

หมายเหตุ ผู้วิจัยได้แสดงตัวประมาณ MCD ไว้ในภาคผนวก ข

วิธี Newton – Raphson จะหาอนุพันธ์ย่อยของลอการิทึมธรรมชาติของฟังก์ชันความควรจะเป็นแบบถ่วงน้ำหนัก $L_{WMLCH}(\underset{\sim}{B})$ เทียบกับ β_0 , β_1 , … , β_m เรียกว่า efficient scores แล้วนำมาเป็นสมาชิกของเวกเตอร์ $\underset{\sim}{U}_{WMLCH}(\underset{\sim}{B})$ ที่มีอันดับ $(m+1) \times 1$

$$\underset{\sim}{U}_{WMLCH}(\underset{\sim}{B}) = \begin{bmatrix} \dfrac{\partial L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_0} \\ \dfrac{\partial L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_1} \\ \dfrac{\partial L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_2} \end{bmatrix}$$

กำหนดเมทริกซ์ $H_{WMLCH}(\underset{\sim}{B})$ อันดับ $(m+1) \times (m+1)$ มีสมาชิกเป็นอนุพันธ์ย่อยอันดับที่สอง (second partial derivative) ของ $L_{WMLCH}(\underset{\sim}{B})$ ซึ่งเรียกเมทริกซ์ $H_{WMLCH}(\underset{\sim}{B})$ ว่า Hessian matrix

โดยที่สมาชิกตัวที่ (j , k) คือ $\dfrac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_j \partial \beta_k}$, $j, k = 0, 1, 2, \ldots, m$

สามารถหาเวกเตอร์ $\underset{\sim}{U}_{WMLCH}(\hat{\underset{\sim}{B}})$ ซึ่งเป็นเวกเตอร์ที่ประมาณด้วยความควรจะเป็นสูงสุดของ $\underset{\sim}{U}_{WMLCH}(\underset{\sim}{B})$ โดยใช้ Taylor series กระจาย $\underset{\sim}{U}_{WMLCH}(\underset{\sim}{B})$ รอบ $\underset{\sim}{B}^*$ จะได้

$$\underset{\sim}{U}_{WMLCH}(\hat{\underset{\sim}{B}}) \approx \underset{\sim}{U}_{WMLCH}(\underset{\sim}{B}^*) + H_{WMLCH}(\underset{\sim}{B}^*)(\hat{\underset{\sim}{B}} - \underset{\sim}{B}^*)$$

โดยนิยามของความควรจะเป็นสูงสุดของ $\underset{\sim}{B}$ จะได้ว่า

$$\left.\frac{\partial L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_j}\right|_{\hat{\underset{\sim}{B}}} = 0 \qquad \text{สำหรับ } j = 0, 1, 2, \ldots, m$$

และ $\underset{\sim}{U}_{WMLCH}(\hat{\underset{\sim}{B}}) = 0$

ดังนั้น $$\hat{\underset{\sim}{B}} = \underset{\sim}{B}^{*} - H^{-1}_{WMLCH}(\underset{\sim}{B}^{*})\, \underset{\sim}{U}_{WMLCH}(\underset{\sim}{B}^{*})$$

ซึ่งชี้นำให้มีการประมาณ $\hat{\underset{\sim}{B}}$ โดยการคำนวณซ้ำๆ ซึ่งค่าประมาณ $\hat{\underset{\sim}{B}}$ ณ รอบที่ r+1 คือ

$$\hat{\underset{\sim}{B}}_{r+1} = \hat{\underset{\sim}{B}}_{r} - H^{-1}_{WMLCH}(\underset{\sim}{B}_{r})\, \underset{\sim}{U}_{WMLCH}(\underset{\sim}{B}_{r})$$

สำหรับ $r = 0, 1, 2, \ldots$ ซึ่งเวกเตอร์ $\hat{\underset{\sim}{B}}_0$ เป็นเวกเตอร์ของตัวประมาณเริ่มต้น ถ้าผลต่างระหว่าง $\hat{\underset{\sim}{B}}$ ในรอบที่ r กับรอบที่ r+1 มีค่าน้อยมากจนถือว่าไม่แตกต่างกัน ซึ่งในที่นี้กำหนดเกณฑ์ว่า $\left|\hat{\underset{\sim}{B}}_{r+1} - \hat{\underset{\sim}{B}}_{r}\right| < 0.00001$ ค่า $\hat{\underset{\sim}{B}}_{r+1}$ นั้นจะเป็นค่าที่ยอมรับได้

วิธี Newton – Raphson สำหรับตัวแปรอิสระ 2 ตัว

ข้อมูลประกอบด้วยค่าสังเกตที่เป็นอิสระ N ค่า คือ (x_{i1}, x_{i2}, y_i) ; $i = 1,2,\ldots,N$ สมการถดถอยโลจิสติคเชิงเส้น คือ

$$\text{logit}\,(\pi(\underset{\sim}{x}_i)) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

เมื่อ $$\underset{\sim}{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

$\underset{\sim}{U}_{WMLCH}(\underset{\sim}{B})$ เป็นเวกเตอร์อันดับ 3×1 มีสมาชิกคือ $\dfrac{\partial L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_j}$, j= 0,1,2

$$\underset{\sim}{U}_{WMLCH}(\underset{\sim}{B}) = \begin{bmatrix} \dfrac{\partial L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_0} \\ \dfrac{\partial L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_1} \\ \dfrac{\partial L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_2} \end{bmatrix} \qquad (2.5)$$

และ Hessian matrix , $H_{WMLCH}(\underset{\sim}{B})$ เป็นเมทริกซ์อันดับ 3×3 มีสมาชิกตัวที่ (j,k) คือ

$$\frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_j \partial\beta_k} \quad , \; j,k = 0,1,2$$

$$H_{WMLCH}(\underset{\sim}{B}) = \begin{bmatrix} \frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_0^2} & \frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_0 \partial\beta_1} & \frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_0 \partial\beta_2} \\ \frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_1 \partial\beta_0} & \frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_1^2} & \frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_1 \partial\beta_2} \\ \frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_2 \partial\beta_0} & \frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_2 \partial\beta_1} & \frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_2^2} \end{bmatrix} \tag{2.6}$$

สมาชิกในเวกเตอร์ $\underset{\sim}{U}_{WMLCH}(\underset{\sim}{B})$ และในเมทริกซ์ $H_{WMLCH}(\underset{\sim}{B})$ หาได้จาก $L_{WMLCH}(\underset{\sim}{B})$ ในสมการ (2.4) และสำหรับตัวแปรอิสระ 2 ตัว

$$L_{WMLCH}(\underset{\sim}{B}) = \sum_{i=1}^{N} \left\{ w_i y_i (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) - w_i \ln(1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}}) \right\}$$

หาอนุพันธ์ของ $L_{WMLCH}(\underset{\sim}{B})$ เทียบกับ β_0 , β_1 , และ β_2 ดังนี้

$$\frac{\partial L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_0} = \sum_{i=1}^{N} \left\{ w_i y_i - w_i \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}}} \right\}$$

$$= \sum_{i=1}^{N} w_i \left\{ y_i - \pi(\underset{\sim}{x}_i) \right\}$$

$$\frac{\partial L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_1} = \sum_{i=1}^{N} \left\{ w_i y_i x_{i1} - w_i x_{i1} \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}}} \right\}$$

$$= \sum_{i=1}^{N} w_i x_{i1} \left\{ y_i - \pi(\underset{\sim}{x}_i) \right\}$$

$$\frac{\partial L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_2} = \sum_{i=1}^{N}\left\{ w_i y_i x_{i2} - w_i x_{i2}\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right\}$$

$$= \sum_{i=1}^{N} w_i x_{i2}\left\{ y_i - \pi(\underset{\sim}{x}_i)\right\}$$

โดยที่ $\pi(\underset{\sim}{x}_i) = \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}$

และสามารถหาอนุพันธ์ย่อยอันดับที่ 2 ได้ดังนี้

$$\frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_0^2} = -\sum_{i=1}^{N} w_i\left\{\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} - \left(\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N} w_i \pi(\underset{\sim}{x}_i)(1-\pi(\underset{\sim}{x}_i))$$

$$\frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_1^2} = -\sum_{i=1}^{N} w_i x_{i1}\left\{x_{i1}\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} - x_{i1}\left(\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N} w_i x_{i1}^2 \pi(\underset{\sim}{x}_i)(1-\pi(\underset{\sim}{x}_i))$$

$$\frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_2^2} = -\sum_{i=1}^{N} w_i x_{i2}\left\{x_{i2}\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} - x_{i2}\left(\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N} w_i x_{i2}^2 \pi(\underset{\sim}{x}_i)(1-\pi(\underset{\sim}{x}_i))$$

$$\frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_0 \partial \beta_1} = -\sum_{i=1}^{N} w_i x_{i1}\left\{\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} - \left(\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N} w_i x_{i1} \pi(\underset{\sim}{x}_i)(1-\pi(\underset{\sim}{x}_i))$$

ซึ่ง $\dfrac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_1 \partial \beta_0} = \dfrac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial \beta_0 \partial \beta_1}$

$$\frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_0\partial\beta_2} = -\sum_{i=1}^{N} w_i x_{i2}\left\{\frac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}} - \left(\frac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N} w_i x_{i2}\boldsymbol{\pi}(\underset{\sim}{x}_i)(1-\boldsymbol{\pi}(\underset{\sim}{x}_i))$$

ซึ่ง $\dfrac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_2\partial\beta_0} = \dfrac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_0\partial\beta_2}$

และ

$$\frac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_1\partial\beta_2} = -\sum_{i=1}^{N} w_i x_{i2}\left\{x_{i1}\frac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}} - x_{i1}\left(\frac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N} w_i x_{i1} x_{i2}\boldsymbol{\pi}(\underset{\sim}{x}_i)(1-\boldsymbol{\pi}(\underset{\sim}{x}_i))$$

ซึ่ง $\dfrac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_2\partial\beta_1} = \dfrac{\partial^2 L_{WMLCH}(\underset{\sim}{B})}{\partial\beta_1\partial\beta_2}$

โดยที่ $\boldsymbol{\pi}(\underset{\sim}{x}_i)(1-\boldsymbol{\pi}(\underset{\sim}{x}_i)) = \dfrac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}{(1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}})^2}$

นำสมาชิกของเวกเตอร์ $\underset{\sim}{U}_{WMLCH}(\underset{\sim}{B})$ และเมทริกซ์ $H_{WMLCH}(\underset{\sim}{B})$ ที่หามาได้ไปแทนในสมการ (2.5) และ (2.6) ดังนั้นคำนวณหาค่า $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ ได้จากสมการ

$$\hat{\underset{\sim}{B}}_{r+1} = \hat{\underset{\sim}{B}}_r - H^{-1}_{WMLCH}(\underset{\sim}{B}_r)\,\underset{\sim}{U}_{WMLCH}(\underset{\sim}{B}_r)$$

เมื่อ $\underset{\sim}{B} = \begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix}$; $\underset{\sim}{U}_{WMLCH}(\underset{\sim}{B}) = \begin{bmatrix} \sum_{i=1}^{N} w_i\left\{y_i - \dfrac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}\right\} \\ \sum_{i=1}^{N} w_i x_{i1}\left\{y_i - \dfrac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}\right\} \\ \sum_{i=1}^{N} w_i x_{i2}\left\{y_i - \dfrac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}}}\right\} \end{bmatrix}$

และ

$$H_{WMLCH}(\underset{\sim}{B}) = \begin{bmatrix} -\sum_{i=1}^{N} w_i \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} w_i x_{i1} \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} w_i x_{i2} \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} \\ -\sum_{i=1}^{N} w_i x_{i1} \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} w_i x_{i1}^2 \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} w_i x_{i1} x_{i2} \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} \\ -\sum_{i=1}^{N} w_i x_{i2} \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} w_i x_{i1} x_{i2} \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} w_i x_{i2}^2 \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} \end{bmatrix}$$

ถ้าผลต่างระหว่าง $\hat{\underset{\sim}{B}}$ ในรอบที่ r กับรอบที่ r+1 มีค่าน้อยมากจนถือว่าไม่แตกต่างกัน โดยมีเกณฑ์ว่า $|\hat{\underset{\sim}{B}}_{r+1} - \hat{\underset{\sim}{B}}_{r}| < 0.00001$ ค่า $\hat{\underset{\sim}{B}}_{r+1}$ นั้นจะเป็นค่าที่ยอมรับได้

2.4 การประมาณค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann

หลักการของวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann มีวัตถุประสงค์เหมือนกับวิธีความควรจะเป็นสูงสุด นั่นคือ เพื่อหาตัวประมาณพารามิเตอร์ที่ทำให้ $\ell_{WMLRC}(\underset{\sim}{B})$ มีค่ามากที่สุด โดยหาอนุพันธ์เทียบกับพารามิเตอร์ β_0 , β_1 , … , β_m และให้ผลลัพธ์มีค่าเท่ากับศูนย์ จะได้สมการที่ไม่เป็นเชิงเส้น m + 1 สมการ ซึ่งสามารถประมาณค่าพารามิเตอร์ $\beta_0 , \beta_1 , \ldots , \beta_m$ ได้ด้วยวิธี Newton – Raphson

ฟังก์ชันความควรจะเป็นแบบถ่วงน้ำหนัก คือ

$$\ell_{WMLRC}(\underset{\sim}{B}) = \prod_{i=1}^{N} (\pi(\underset{\sim}{x}_i))^{w_i \tilde{y}_i} (1-\pi(\underset{\sim}{x}_i))^{w_i(1-\tilde{y}_i)}$$

ดังนั้นลอการิทึมธรรมชาติของฟังก์ชันความควรจะเป็นแบบถ่วงน้ำหนัก คือ

$$L_{WMLRC}(\underset{\sim}{B}) = \ln(\ell_{WMLRC}(\underset{\sim}{B}))$$

$$= \sum_{i=1}^{N} \left\{ w_i \tilde{y}_i \ln(\pi(\underset{\sim}{x}_i)) + w_i (1-\tilde{y}_i) \ln(1-\pi(\underset{\sim}{x}_i)) \right\}$$

$$= \sum_{i=1}^{N} \left\{ w_i \tilde{y}_i \ln(\pi(\underset{\sim}{x}_i)) + w_i \ln(1-\pi(\underset{\sim}{x}_i)) - w_i \tilde{y}_i \ln(1-\pi(\underset{\sim}{x}_i)) \right\}$$

$$= \sum_{i=1}^{N} \left\{ w_i \tilde{y}_i \left[\ln(\pi(\underset{\sim}{x}_i)) - \ln(1-\pi(\underset{\sim}{x}_i)) \right] + w_i \ln(1-\pi(\underset{\sim}{x}_i)) \right\}$$

$$= \sum_{i=1}^{N}\left\{ w_i \tilde{y}_i \ln\left(\frac{\pi(\underset{\sim}{x}_i)}{1-\pi(\underset{\sim}{x}_i)}\right) + w_i \ln(1-\pi(\underset{\sim}{x}_i))\right\}$$

$$= \sum_{i=1}^{N}\left\{ w_i\ \tilde{y}_i \ln(e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}}) - w_i \ln(1+e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}})\right\}$$

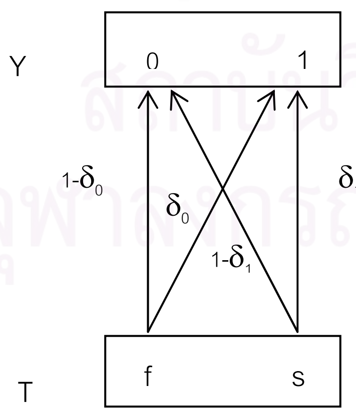
$$= \sum_{i=1}^{N}\left\{ w_i \tilde{y}_i(\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}) - w_i \ln(1+e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_m x_{im}})\right\} \qquad (2.7)$$

โดยเรียก $\tilde{y}_i$ ว่าค่าสังเกตเทียม (pseudo – observation) คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ที่สนใจอย่างมีเงื่อนไขบนการประมาณของสถานะจริง (true status)

ซึ่ง $\tilde{y}_i \quad = \quad (1-y_i)\,\delta_0 + y_i\,\delta_1$

โดย δ_0 และ δ_1 คือ ความน่าจะเป็นของตัวแปรตามเมื่อกำหนดสถานะจริง (true status) โดยสถานะจริงมีสถานการณ์ที่เป็นไปได้ 2 สถานการณ์ นั่นคือเกิดเหตุการณ์ที่สนใจ (s) และไม่เกิดเหตุการณ์ที่สนใจ (f)

กำหนดให้ T เป็นค่าตอบสนองจริง (true response) ที่ไม่เป็นค่าสังเกต (unobservable / latent) โดยที่ตัวแปรตาม (y) มีความสัมพันธ์อย่างมากกับ T ดังแสดงในรูป



จากรูปจะเห็นได้ว่า

ถ้า T = s แล้ว Y = 1 จะเกิดขึ้นด้วยความน่าจะเป็น $P(Y=1|T=s) = \delta_1$

และ ถ้า T = s แล้ว Y = 0 จะเกิดขึ้นด้วยความน่าจะเป็น $P(Y=0|T=s) = 1-\delta_1$

ในทำนองเดียวกัน

ถ้า T = f แล้ว Y = 1 จะเกิดขึ้นด้วยความน่าจะเป็น $P(Y=1|T=f) = \delta_0$

และ ถ้า T = f แล้ว Y = 0 จะเกิดขึ้นด้วยความน่าจะเป็น $P(Y=0|T=f) = 1-\delta_0$

ค่า δ_0 และ δ_1 หาได้จาก

$$\delta_0 = \frac{\hat{\Lambda}\delta}{1+\delta} \qquad \delta_1 = \frac{1+\hat{\Lambda}\delta}{1+\delta}$$

โดยที่ $\hat{\Lambda} = \max(\delta, \min(1-\delta, \overline{\Lambda}))$

และ $\overline{\Lambda} = \frac{1}{N}\sum_{i=1}^{N} y_i$ หรือ (จำนวนของ $y_i = 1$) / N

Rousseeuw และ Christmann เสนอแนะว่าการคำนวณ $\hat{\Lambda}$, δ_0 และ δ_1 ควรใช้ค่า $\delta = 0.01$ จึงจะให้ค่าการคำนวณที่เหมาะสม

ส่วน w_i เป็นการถ่วงน้ำหนักข้อมูล

ซึ่ง $$w_i = \frac{M}{\max\{RD^2(\underset{\sim}{x}_i'), M\}}$$

โดยที่ $RD(\underset{\sim}{x}_i')$ เป็น Robust distance

และ M เป็น เปอร์เซ็นไทล์ที่ 75 ของ ค่า $RD^2(\underset{\sim}{x}_i')$ ทั้งหมด

วิธี Newton – Raphson จะหาอนุพันธ์ย่อยของลอการิทึมธรรมชาติของฟังก์ชันความควรจะเป็นแบบถ่วงน้ำหนัก $L_{WMLRC}(\underset{\sim}{B})$ เทียบกับ β_0 , β_1 , … , β_m เรียกว่า efficient scores แล้วนำมาเป็นสมาชิกของเวกเตอร์ $\underset{\sim}{U}_{WMLRC}(\underset{\sim}{B})$ ที่มีอันดับ $(m+1) \times 1$

$$\underset{\sim}{U}_{WMLRC}(\underset{\sim}{B}) = \begin{bmatrix} \dfrac{\partial L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_0} \\ \dfrac{\partial L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_1} \\ \dfrac{\partial L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_2} \end{bmatrix}$$

กำหนดเมทริกซ์ $H_{WMLRC}(\underset{\sim}{B})$ อันดับ $(m+1) \times (m+1)$ มีสมาชิกเป็นอนุพันธ์ย่อยอันดับที่สอง (second partial derivative) ของ $L_{WMLRC}(\underset{\sim}{B})$ ซึ่งเรียกเมทริกซ์ $H_{WMLRC}(\underset{\sim}{B})$ ว่า Hessian matrix

โดยที่สมาชิกตัวที่ (j , k) คือ $\dfrac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_j \partial \beta_k}$, $j, k = 0, 1, 2, \ldots, m$

สามารถหาเวกเตอร์ $\underset{\sim}{U}_{WMLRC}(\hat{\underset{\sim}{B}})$ ซึ่งเป็นเวกเตอร์ที่ประมาณด้วยความควรจะเป็นสูงสุดของ $\underset{\sim}{U}_{WMLRC}(\underset{\sim}{B})$ โดยใช้ Taylor series กระจาย $\underset{\sim}{U}_{WMLRC}(\underset{\sim}{B})$ รอบ $\underset{\sim}{B}^{PP*}$ จะได้

$$\underset{\sim}{U}_{WMLRC}(\hat{\underset{\sim}{B}}) \approx \underset{\sim}{U}_{WMLRC}(\underset{\sim}{B}^*) + H_{WMLRC}(\underset{\sim}{B}^*)(\hat{\underset{\sim}{B}} - \underset{\sim}{B}^*)$$

โดยนิยามของความควรจะเป็นสูงสุดของ $\underset{\sim}{B}$ จะได้ว่า

$$\left.\frac{\partial L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_j}\right|_{\hat{\underset{\sim}{B}}} = 0 \qquad \text{สำหรับ } j = 0, 1, 2, \ldots, m$$

และ $\underset{\sim}{U}_{WMLRC}(\hat{\underset{\sim}{B}}) = 0$

ดังนั้น $$\hat{\underset{\sim}{B}} = \underset{\sim}{B}^{*} - H^{-1}_{WMLRC}(\underset{\sim}{B}^{*})\, \underset{\sim}{U}_{WMLRC}(\underset{\sim}{B}^{*})$$

ซึ่งชี้นำให้มีการประมาณ $\hat{\underset{\sim}{B}}$ โดยการคำนวณซ้ำๆ ซึ่งค่าประมาณ $\hat{\underset{\sim}{B}}$ ณ รอบที่ r+1 คือ

$$\hat{\underset{\sim}{B}}_{r+1} = \hat{\underset{\sim}{B}}_{r} - H^{-1}_{WMLRC}(\underset{\sim}{B}_{r})\, \underset{\sim}{U}_{WMLRC}(\underset{\sim}{B}_{r})$$

สำหรับ $r = 0, 1, 2, \ldots$ ซึ่งเวกเตอร์ $\hat{\underset{\sim}{B}}_0$ เป็นเวกเตอร์ของตัวประมาณเริ่มต้น ถ้าผลต่างระหว่าง $\hat{\underset{\sim}{B}}$ ในรอบที่ r กับรอบที่ r+1 มีค่าน้อยมากจนถือว่าไม่แตกต่างกัน ซึ่งในที่นี้กำหนดเกณฑ์ว่า $\left|\hat{\underset{\sim}{B}}_{r+1} - \hat{\underset{\sim}{B}}_{r}\right| < 0.00001$ ค่า $\hat{\underset{\sim}{B}}_{r+1}$ นั้นจะเป็นค่าที่ยอมรับได้

วิธี Newton – Raphson สำหรับตัวแปรอิสระ 2 ตัว

ข้อมูลประกอบด้วยค่าสังเกตที่เป็นอิสระ N ค่า คือ (x_{i1}, x_{i2}, y_i) ; $i = 1,2,\ldots,N$ สมการถดถอยโลจิสติคเชิงเส้น คือ

$$\text{logit}(\pi(\underset{\sim}{x}_i)) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

เมื่อ $$\underset{\sim}{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

$\underset{\sim}{U}_{WMLRC}(\underset{\sim}{B})$ เป็นเวกเตอร์อันดับ 3×1 มีสมาชิกคือ $\dfrac{\partial L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_j}$, j= 0,1,2

$$\underset{\sim}{U}_{WMLRC}(\underset{\sim}{B}) = \begin{bmatrix} \dfrac{\partial L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_0} \\ \dfrac{\partial L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_1} \\ \dfrac{\partial L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_2} \end{bmatrix} \qquad (2.8)$$

และ Hessian matrix , $H_{WMLRC}(\underset{\sim}{B})$ เป็นเมทริกซ์อันดับ 3×3 มีสมาชิกตัวที่ (j,k) คือ

$$\frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_j \partial\beta_k} \quad , \ j,k = 0,1,2$$

$$H_{WMLRC}(\underset{\sim}{B}) = \begin{bmatrix} \frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_0^2} & \frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_0\partial\beta_1} & \frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_0\partial\beta_2} \\ \frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_1\partial\beta_0} & \frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_1^2} & \frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_1\partial\beta_2} \\ \frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_2\partial\beta_0} & \frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_2\partial\beta_1} & \frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_2^2} \end{bmatrix} \qquad (2.9)$$

สมาชิกในเวกเตอร์ $\underset{\sim}{U}_{WMLRC}(\underset{\sim}{B})$ และในเมทริกซ์ $H_{WMLRC}(\underset{\sim}{B})$ หาได้จาก $L_{WMLRC}(\underset{\sim}{B})$ ในสมการ (2.7) และสำหรับตัวแปรอิสระ 2 ตัว

$$L_{WMLRC}(\underset{\sim}{B}) = \sum_{i=1}^{N}\left\{ w_i\tilde{y}_i(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}) - w_i\ln(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})\right\}$$

หาอนุพันธ์ของ $L_{WMLRC}(\underset{\sim}{B})$ เทียบกับ β_0 , β_1 , และ β_2 ดังนี้

$$\begin{aligned}\frac{\partial L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_0} &= \sum_{i=1}^{N}\left\{ w_i\tilde{y}_i - w_i\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right\} \\ &= \sum_{i=1}^{N} w_i\left\{\tilde{y}_i - \pi(\underset{\sim}{x}_i)\right\}\end{aligned}$$

$$\begin{aligned}\frac{\partial L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_1} &= \sum_{i=1}^{N}\left\{ w_i\tilde{y}_i x_{i1} - w_i x_{i1}\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right\} \\ &= \sum_{i=1}^{N} w_i x_{i1}\left\{\tilde{y}_i - \pi(\underset{\sim}{x}_i)\right\}\end{aligned}$$

$$\frac{\partial L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_2} = \sum_{i=1}^{N}\left\{ w_i\tilde{y}_i x_{i2} - w_i x_{i2}\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right\}$$

$$= \sum_{i=1}^{N} w_i x_{i2}\left\{\tilde{y}_i - \pi(\underset{\sim i}{x})\right\}$$

โดยที่ $\pi(\underset{\sim i}{x}) = \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}$

และสามารถหาอนุพันธ์ย่อยอันดับที่ 2 ได้ดังนี้

$$\frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_0^2} = -\sum_{i=1}^{N} w_i\left\{\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} - \left(\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N} w_i \pi(\underset{\sim i}{x})(1-\pi(\underset{\sim i}{x}))$$

$$\frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_1^2} = -\sum_{i=1}^{N} w_i x_{i1}\left\{x_{i1}\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} - x_{i1}\left(\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N} w_i x_{i1}^2 \pi(\underset{\sim i}{x})(1-\pi(\underset{\sim i}{x}))$$

$$\frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_2^2} = -\sum_{i=1}^{N} w_i x_{i2}\left\{x_{i2}\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} - x_{i2}\left(\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N} w_i x_{i2}^2 \pi(\underset{\sim i}{x})(1-\pi(\underset{\sim i}{x}))$$

$$\frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_0 \partial \beta_1} = -\sum_{i=1}^{N} w_i x_{i1}\left\{\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} - \left(\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N} w_i x_{i1} \pi(\underset{\sim i}{x})(1-\pi(\underset{\sim i}{x}))$$

ซึ่ง $\dfrac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_1 \partial \beta_0} = \dfrac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial \beta_0 \partial \beta_1}$

$$\frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_0\partial\beta_2} = -\sum_{i=1}^{N} w_i x_{i2}\left\{\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} - \left(\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N} w_i x_{i2}\pi(\underset{\sim}{x}_i)(1-\pi(\underset{\sim}{x}_i))$$

ซึ่ง $\frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_2\partial\beta_0} = \frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_0\partial\beta_2}$

และ

$$\frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_1\partial\beta_2} = -\sum_{i=1}^{N} w_i x_{i2}\left\{x_{i1}\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} - x_{i1}\left(\frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right)^2\right\}$$

$$= -\sum_{i=1}^{N} w_i x_{i1} x_{i2}\pi(\underset{\sim}{x}_i)(1-\pi(\underset{\sim}{x}_i))$$

ซึ่ง $\frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_2\partial\beta_1} = \frac{\partial^2 L_{WMLRC}(\underset{\sim}{B})}{\partial\beta_1\partial\beta_2}$

โดยที่ $\pi(\underset{\sim}{x}_i)(1-\pi(\underset{\sim}{x}_i)) = \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2}$

นำสมาชิกของเวกเตอร์ $\underset{\sim}{U}_{WMLRC}(\underset{\sim}{B})$ และเมทริกซ์ $H_{WMLRC}(\underset{\sim}{B})$ ที่หามาได้ไปแทนในสมการ (2.8) และ (2.9) ดังนั้นคำนวณหาค่า $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ ได้จากสมการ

$$\hat{\underset{\sim}{B}}_{r+1} = \hat{\underset{\sim}{B}}_r - H^{-1}_{WMLRC}(\underset{\sim}{B}_r)\,\underset{\sim}{U}_{WMLRC}(\underset{\sim}{B}_r)$$

เมื่อ $\underset{\sim}{B} = \begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix}$; $\underset{\sim}{U}_{WMLRC}(\underset{\sim}{B}) = \begin{bmatrix}\sum_{i=1}^{N} w_i\left\{\tilde{y}_i - \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right\}\\ \sum_{i=1}^{N} w_i x_{i1}\left\{\tilde{y}_i - \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right\}\\ \sum_{i=1}^{N} w_i x_{i2}\left\{\tilde{y}_i - \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}\right\}\end{bmatrix}$

และ

$$H_{WMLRC}(\underset{\sim}{B}) = \begin{bmatrix} -\sum_{i=1}^{N} w_i \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} w_i x_{i1} \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} w_i x_{i2} \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} \\ -\sum_{i=1}^{N} w_i x_{i1} \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} w_i x_{i1}^2 \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} w_i x_{i1} x_{i2} \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} \\ -\sum_{i=1}^{N} w_i x_{i2} \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} w_i x_{i1} x_{i2} \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} & -\sum_{i=1}^{N} w_i x_{i2}^2 \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{(1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}})^2} \end{bmatrix}$$

ถ้าผลต่างระหว่าง $\underset{\sim}{\hat{B}}$ ในรอบที่ r กับรอบที่ r+1 มีค่าน้อยมากจนถือว่าไม่แตกต่างกัน โดยมีเกณฑ์ว่า $\left|\underset{\sim}{\hat{B}}_{r+1} - \underset{\sim}{\hat{B}}_{r}\right| < 0.00001$ ค่า $\underset{\sim}{\hat{B}}_{r+1}$ นั้นจะเป็นค่าที่ยอมรับได้

บทที่3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาและเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของตัวแบบการถดถอยโลจิสติคเมื่อตัวแปรอิสระมีค่าผิดปกติ ด้วยวิธีการประมาณ 3 วิธี คือ

1.วิธีความควรจะเป็นสูงสุด

2.วิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และ Haesbroeck

3.วิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann

ในการเปรียบเทียบวิธีการประมาณจะพิจารณาจากค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย (Average Mean Square Error : AMSE)

สำหรับข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากการจำลองโดยใช้โปรแกรมภาษาฟอร์แทน 77 บน PC Computer ซึ่งมีขั้นตอนของการทดลองและโปรแกรมที่ใช้ในการศึกษาดังต่อไปนี้

3.1 ข้อกำหนดของการทดลอง

ในการวิจัยครั้งนี้กำหนดสถานการณ์ต่าง ๆ ที่ต้องการศึกษาดังนี้

1. ในกรณีตัวแปรอิสระมีค่าปกติ ผู้วิจัยสนใจศึกษาเมื่อตัวแปรอิสระx_1 มีการแจกแจงปกติ โดยมีพารามิเตอร์ μ เท่ากับ 15 และ σ^2 เท่ากับ 4 และตัวแปรอิสระx_2 มีการแจกแจงปกติ โดยมีพารามิเตอร์ μ เท่ากับ 10 และ σ^2 เท่ากับ 2

2. ขนาดตัวอย่างที่ทำการศึกษาเท่ากับ 20 , 30 ,40 , 50 , 60 , 70 , 80 , 90 และ100

3. ในกรณีตัวแปรอิสระมีค่าผิดปกติ ผู้วิจัยจึงกำหนดการมีค่าผิดปกติ ดังนี้

3.1 กรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1 ผู้วิจัยสนใจศึกษาเมื่อกำหนดให้ตัวแปรอิสระ x_1 มีการแจกแจงแบบปกติปลอมปน โดยมีระดับค่าผิดปกติ 2 ระดับ คือ ระดับไม่รุนแรงที่ตัวประกอบสเกลเท่ากับ 5 และระดับรุนแรงที่ตัวประกอบสเกลเท่ากับ 12 ซึ่งแต่ละระดับมีสัดส่วนการปลอมปนเท่ากับ 0.05 , 0.10 และ 0.15

3.2 กรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1 และ x_2 ผู้วิจัยสนใจศึกษาเมื่อกำหนดให้ตัวแปรอิสระ x_1 และ x_2 มีการแจกแจงแบบปกติปลอมปนที่เหมือนกัน โดยแต่ละตัวแปรอิสระมีระดับค่าผิดปกติ 2 ระดับ คือ ระดับไม่รุนแรงที่ตัวประกอบสเกลเท่ากับ 5 และระดับรุนแรงที่ตัวประกอบสเกลเท่ากับ 12 ซึ่งแต่ละระดับค่าผิดปกติของแต่ละตัวแปรอิสระมีสัดส่วนการปลอมปนเท่ากับ 0.05 , 0.10 และ 0.15

3.2 ขั้นตอนในการวิจัย

ขั้นตอนในการวิจัยมีดังนี้

1. สร้างข้อมูลของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 ตามลักษณะที่กำหนดในขอบเขตการวิจัย

2. สร้างข้อมูลของความคลาดเคลื่อน (ε) ให้มีการแจกแจงเอกรูปแบบต่อเนื่องในช่วง [-1,1]

3. สร้างข้อมูลของตัวแปรตาม (y) จากรูปแบบความสัมพันธ์

$$Y_i = \begin{cases} 0 & \text{ถ้า} \quad \pi(\underset{\sim}{x}_i) + \varepsilon_i < 0.5 \\ 1 & \text{ถ้า} \quad \pi(\underset{\sim}{x}_i) + \varepsilon_i \geq 0.5 \end{cases}$$

โดยที่ $$\pi(\underset{\sim}{x}_i) = \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}} \quad , \; i = 1, 2, \ldots, N$$

4. ประมาณค่าพารามิเตอร์ของตัวแบบการถดถอยโลจิสติคด้วยวิธีที่สนใจศึกษา 3 วิธี คือ วิธีความควรจะเป็นสูงสุด วิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และ Haesbroeck และวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann

5. หาค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์พร้อมทั้งเปรียบเทียบค่าและสรุปผลที่ได้

3.2.1 การสร้างตัวแปรอิสระให้มีการแจกแจงตามที่กำหนด

การสร้างตัวแปรอิสระให้มีลักษณะการแจกแจงตามที่ต้องการศึกษานั้นใช้โปรแกรมภาษาฟอร์แทน 77 กับ PC Computer โดยการสร้างลักษณะการแจกแจงแบบปกติปลอมปนที่มีค่าเฉลี่ยและความแปรปรวนตามที่กำหนดโดยพิจารณาการแจกแจง ซึ่งแปลงมาจากการแจกแจงแบบปกติที่มีฟังก์ชันการแจกแจงอยู่ในรูป

$$f(x) = (1\text{-}P)\, N(\mu, \sigma^2) + P\, N(\mu, C^2\sigma^2)$$

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

μ และ σ^2 เป็นค่าที่กำหนดของค่าเฉลี่ยและความแปรปรวนของการแจกแจงปกติ

P และ C เป็นค่าที่กำหนดสัดส่วนการปลอมปนและตัวประกอบสเกล

ในการวิจัยครั้งนี้ กำหนดให้ ตัวแปรอิสระ x_1 มีการแจกแจงปกติ โดยมีพารามิเตอร์ μ เท่ากับ 15 และ σ^2 เท่ากับ 4 และตัวแปรอิสระ x_2 มีการแจกแจงปกติ โดยมีพารามิเตอร์ μ เท่ากับ 10 และ σ^2 เท่ากับ 2 และกำหนดระดับค่าผิดปกติ 2 ระดับ คือ ระดับไม่รุนแรง(C=5) และระดับรุนแรง(C=12) ตามเกณฑ์การกำหนดขนาดค่าผิดปกติด้วย Box Plot โดยแต่ละระดับจะกำหนดให้มีสัดส่วนการปลอมปน(P) เท่ากับ 0.05 , 0.10 และ 0.15 โดยได้กำหนดจำนวนของค่าผิดปกติตามขนาดตัวอย่างที่กำหนดโดยจำนวนค่าผิดปกติของแต่ละขนาดตัวอย่างที่ศึกษามีดังนี้

ขนาดตัวอย่าง	สัดส่วนการปลอมปน		
	0.05	0.10	0.15
20	1	2	3
30	2	3	5
40	2	4	6
50	3	5	8
60	3	6	9
70	4	7	11
80	4	8	12
90	5	9	14
100	5	10	15

การสร้างตัวแปรอิสระให้มีการแจกแจงแบบปกติปลอมปนจะต้องอาศัยการสร้างลักษณะการแจกแจงแบบปกติ ซึ่งการแจกแจงแบบปกติจะใช้เลขสุ่ม (random number) ซึ่งมีการแจกแจงเอกรูป (uniform distribution) ในช่วง [0,1] เป็นองค์ประกอบหลัก

การสร้างเลขสุ่มที่เป็นอิสระกันและมีการแจกแจงเอกรูปในช่วง [0,1][1]

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

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

โดยที่ค่า c , a และ m เป็นค่าคงที่จำนวนเต็มค่าไม่เป็นลบ และความหมายของตัวแบบคือ 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 $(c + aX_{i-1})/m$) ดังนั้น ค่าเป็นไปได้ของ X_i คือ 0 , 1 , … , m-1 และก่อนที่จะได้ค่าของ $X_1, X_2, \ldots$ ต้องกำหนดค่าของ c , a , m และ X_0 เราเรียก X_0 ว่า ซีด(seed) หรือ ค่าเริ่มต้น (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 ซึ่งได้ผ่านการตรวจสอบคุณสมบัติแล้วอย่างมาก คือ กำหนด c = 0 , $m = 2^{31}-1 = 2147483647$, $a = 7^5 = 16807$ และ X_0 เป็นจำนวนเต็มบวกที่เป็นเลขคี่ไม่เกิน m ฟังก์ชันการจำลองเลขสุ่มที่เป็นอิสระกันและมีการแจกแจงเอกรูปในช่วง [0,1] คือ SUBROUTINE RANDOM

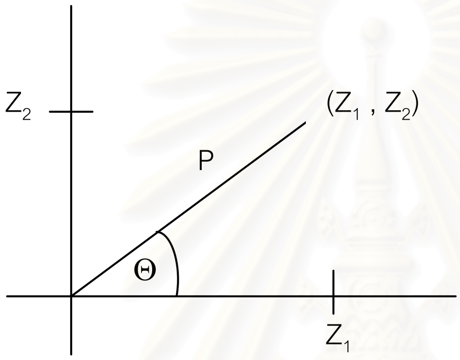
[1]มานพ วราภักดิ์, <u>การจำลองเบื้องต้น</u> (กรุงเทพฯ: ศูนย์ผลิตตำราเรียนสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2547), หน้า 43

การสร้างการแจกแจงแบบปกติ[1]

การผลิตเลขสุ่มที่มีการแจกแจงแบบปกติใช้วิธีของ George E.P Box และ Mervin E.Muller (1958) โดยใช้การแปลงตัวแปรสุ่ม คือ จากตัวแปรสุ่มมาตรฐานอิสระกัน Z_1 และ Z_2 ได้จุดบนระนาบในระบบพิกัดคาร์ทีเซียน (Cartesian coordinates) ดังรูปที่ 3.2.1 แปลงตัวแปรสุ่มในระบบพิกัดคาร์ทีเซียนเป็นตัวแปรสุ่มในระบบพิกัดเชิงขั้ว (polar coordinates) เป็นจุด (P,Θ) โดยที่

$$Z_1 = P\cos\Theta$$

$$Z_2 = P\sin\Theta$$



รูปที่ 3.2.1

การแปลง $z_1 = \rho\cos\theta$ และ $z_2 = \rho\sin\theta$ เป็นการแปลงแบบหนึ่งต่อหนึ่งจากปริภูมิ $R_{z_1,z_2} = \{(z_1,z_2) : -\infty < z_1 < \infty , -\infty < z_2 < \infty\}$ ของ (Z_1 ,Z_2) ไปยังปริภูมิ $R_{P,\Theta} = \{(\rho,\theta) : 0 \le \rho < \infty , 0 \le \theta \le 2\pi\}$ โดยมีจาโคเบียน (Jacobian) ของการแปลง J ดังนี้

[1]มานพ วราภักดิ์, <u>การจำลองเบื้องต้น</u> (กรุงเทพฯ: ศูนย์ผลิตตำราเรียนสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2547), หน้า 142

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

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

จากเทคนิคการแปลงในทฤษฎีความน่าจะเป็น ได้ว่า P และ Θ มีฟังก์ชันความหนาแน่นร่วม (joint density function) คือ

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

เนื่องจาก Z_1 และ Z_2 มีฟังก์ชันความหนาแน่นร่วม คือ

$$\begin{aligned} f_{Z_1,Z_2}(z_1,z_2) &= f_{Z_1}(z_1) f_{Z_2}(z_2) \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_1^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_2^2} \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}(z_1^2+z_2^2)} \end{aligned}$$

ดังนั้นโดยการแทนค่าจะได้ผลลัพธ์

$$\begin{aligned} f_{P,\Theta}(\rho,\theta) &= \frac{1}{2\pi} e^{-\frac{1}{2}\rho^2} \rho \\ &= \frac{1}{2\pi} \rho e^{-\frac{1}{2}\rho^2}, 0 \le \theta \le 2\pi, 0 \le \rho \le \infty \\ &= f_{\Theta}(\theta) f_P(\rho) \end{aligned}$$

โดยที่ $f_{\Theta}(\theta) = \dfrac{1}{2\pi}$, $0 \le \theta \le 2\pi$ เป็นฟังก์ชันของ θ เท่านั้น ไม่ขึ้นอยู่กับ ρ และ $f_P(\rho) = \rho e^{-\frac{1}{2}\rho^2}$, $\rho \ge 0$ เป็นฟังก์ชันของ ρ เท่านั้น ไม่ขึ้นอยู่กับ θ เพราะฉะนั้นจากคุณสมบัติของตัวแปรสุ่มอิสระ ได้ว่า P และ Θ เป็นอิสระกัน(เชิงสถิติ)

ในการจำลอง Z_1 และ Z_2 เราจะจำลอง P และ Θ อย่างอิสระกัน โดยจำลอง P จาก $f_P(\rho) = \rho e^{-\frac{1}{2}\rho^2}$ ซึ่งด้วยวิธีการแปลงผกผันได้ตัวแบบจำลอง $P = \sqrt{-2\ln R_1}$, $R_1 \sim U(0,1)$ และจำลอง Θ จากการแจกแจง $U(0,2\pi)$ ได้ $\Theta = 2\pi R_2$, $R_2 \sim U(0,1)$ ดังนั้นเราจะได้ตัวแบบจำลองตัวแปรสุ่ม $Z_1 \sim N(0,1)$ และ $Z_2 \sim N(0,1)$ ซึ่งเป็นอิสระกัน คือ

$$Z_1 = (-2\ln R_1)^{\frac{1}{2}} \cos(2\pi R_2)$$

$$Z_2 = (-2\ln R_1)^{\frac{1}{2}} \sin(2\pi R_2)$$

ซึ่ง R_1 และ R_2 เป็นเลขสุ่มที่สร้างจาก SUBROUTINE RANDOM เมื่อเราได้เลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานแล้ว จากนั้นจึงแปลงค่าเลขสุ่มดังกล่าวโดยใช้สมการ

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

หรือ

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

ทำให้ได้ $NORMAL_i$ ที่มีการแจกแจงแบบปกติมีค่าเฉลี่ยเท่ากับ μ และค่าความแปรปรวนเท่ากับ σ^2 ($NORMAL_i \sim N(\mu , \sigma^2)$) และเป็นอิสระกัน ฟังก์ชันของการจำลองแบบประชากรที่มีการแจกแจงเป็นแบบปกติที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 คือ SUBROUTINE NORMAL

3.2.2 การสร้างตัวแปรตามให้มีค่าตามที่กำหนด

การสร้างตัวแปรตามจะสร้างจากรูปแบบความสัมพันธ์

$$Y_i = \begin{cases} 0 & \text{ถ้า} \quad \pi(\underset{\sim}{x}_i) + \varepsilon_i < 0.5 \\ 1 & \text{ถ้า} \quad \pi(\underset{\sim}{x}_i) + \varepsilon_i \geq 0.5 \end{cases}$$

โดยที่ $$\pi(\underset{\sim}{x}_i) = \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}}} \quad , \quad i = 1, 2, \ldots, N$$

และ ε_i เป็นความคลาดเคลื่อนที่มีการแจกแจงเอกรูปแบบต่อเนื่องในช่วง[-1,1]

การสร้างความคลาดเคลื่อนที่มีการแจกแจงเอกรูปแบบต่อเนื่อง

การผลิตเลขสุ่มที่มีการแจกแจงเอกรูปแบบต่อเนื่องในช่วง [a,b] ใช้วิธีการแปลงผกผันโดยใช้เลขสุ่ม (random number) ซึ่งมีการแจกแจงแบบเอกรูป (uniform distribution) ในช่วง [0,1] เป็นองค์ประกอบหลัก ได้ตัวแบบจำลองสำหรับตัวแปรสุ่ม X ~U(a,b) คือ

$$X = a + (b - a)R \quad , R \sim U(0,1)$$

ซึ่ง R เป็นเลขสุ่มที่สร้างจาก SUBROUTINE RANDOM และเราสามารถจำลองเลขสุ่มที่มีการแจกแจงเอกรูปแบบต่อเนื่องในช่วง [a,b] ได้จาก SUBROUTINE CONU

3.2.3 การประมาณค่าพารามิเตอร์ของตัวแบบการถดถอยโลจิสติคด้วยวิธีความควรจะเป็นสูงสุด วิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และHaesbroeck และวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann

จากหัวข้อ (3.2.1) และ (3.2.2) เราสามารถสร้างตัวแปรอิสระให้มีค่าผิดปกติตรงตามที่กำหนด และสร้างตัวแปรตามได้ จากนั้นจึงทำการประมาณค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุด วิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และHaesbroeckและวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann ดังที่ได้แสดงไว้ใน SUBROUTINE ML SUBROUTINE WMLCH และ SUBROUTINE WMLRC ตามลำดับ

3.2.4 การหาค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์ มีวิธีการคำนวณดังนี้

1. คำนวณค่าความคลาดเคลื่อนกำลังสองของพารามิเตอร์แต่ละตัวเมื่อกระทำซ้ำ 1000 ครั้ง

$$MSE_j = \frac{1}{1000}\sum_{i=1}^{1000}\left[\hat{\beta}_{ij} - \beta_j\right]^2$$

เมื่อ β_j แทน พารามิเตอร์ตัวที่ j ในสมการการถดถอยโลจิสติค

$\hat{\beta}_{ij}$ แทน ค่าประมาณของพารามิเตอร์ตัวที่ j (β_j) จากการประมาณครั้งที่ i

MSE_j แทน ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสำหรับพารามิเตอร์ β_j

2. คำนวณหาค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์

$$AMSE(ML) = \frac{1}{k}\sum_{j=1}^{k} MSE(ML)_j$$

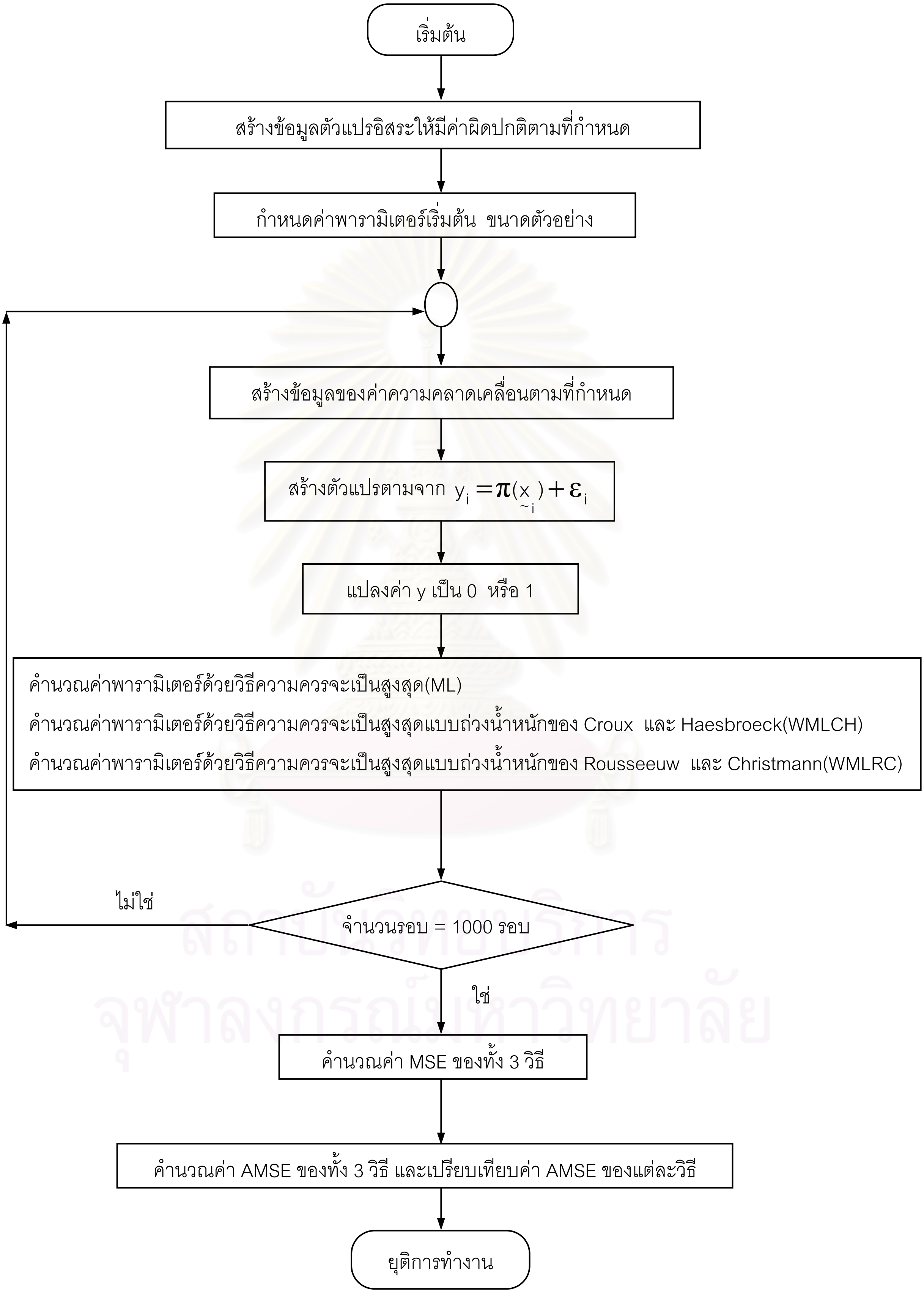
$$AMSE(WMLCH) = \frac{1}{k}\sum_{j=1}^{k} MSE(WMLCH)_j$$

$$AMSE(WMLRC) = \frac{1}{k}\sum_{j=1}^{k} MSE(WMLRC)_j$$

เมื่อ k แทน จำนวนพารามิเตอร์ในตัวแบบการถดถอยโลจิสติค ในที่นี้ $k = 3$

โดยผู้วิจัยได้แสดงตารางขั้นตอนการจำลองที่ใช้ในการวิจัยรวมทั้งตารางโปรแกรมดังต่อไปนี้

แผนผังขั้นตอนวิธีการจำลอง



ตารางที่3.1 โปรแกรมทั้งหมดที่ใช้ในการวิจัย

อันดับที่	ชื่อโปรแกรม	การทำงานของโปรแกรม	ชื่อโปรแกรมย่อยที่เรียกใช้
โปรแกรมหลัก	MAIN	-อ่านค่าพารามิเตอร์ที่กำหนด	
		-สร้างตัวแปรอิสระกรณี x_1 มีค่าผิดปกติ	GENX1CON
		-สร้างตัวแปรอิสระกรณี x_1 , x_2 มีค่าผิดปกติ	GENX1X2CON
		-สร้างตัวแปรอิสระกรณี x_1 , x_2 มีค่าปกติ	GENX1X2NOR
		-คำนวณหาค่า Robust distance	MCD
		-คำนวณหาค่าถ่วงน้ำหนักของวิธี WMLCH	W_MLCH
		-คำนวณหาค่าถ่วงน้ำหนักของวิธี WMLRC	W_MLRC
		-สร้างตัวแปรตาม	GENY
		-คำนวณหาค่าพารามิเตอร์ของวิธี OLS	OLS
		-คำนวณหาค่าพารามิเตอร์ของวิธีML	ML
		-คำนวณหาค่าพารามิเตอร์ของวิธี WMLCH	WMLCH
		-คำนวณหาค่าพารามิเตอร์ของวิธี WMLRC	WMLRC

ตารางที่3.1 (ต่อ)

อันดับที่	ชื่อโปรแกรม	การทำงานของโปรแกรม	ชื่อโปรแกรมย่อยที่เรียกใช้
โปรแกรมย่อย			
1	GENX1CON	-สร้างตัวแปรอิสระ x_1 ให้มีค่าผิดปกติ	GENX1,NORMAL XRANK,QUATILE
2	GENX1X2CON	-สร้างตัวแปรอิสระ x_1 , x_2 ให้มีค่าผิดปกติ	GENX1,GENX2
3	GENX1X2NOR	-สร้างตัวแปรอิสระ x_1 , x_2 ให้มีค่าปกติ	NORMAL, XRANK,QUATILE
4	MCD	-คำนวณหาค่า Robust distance	COMBINATION, DISTANCE, CSTEP,RANKD, RANKDET, RANKD10 CSTEP10, RANKDET10
5	W_MLCH	-คำนวณหาค่าถ่วงน้ำหนักของวิธี WMLCH	
6	W_MLRC	-คำนวณหาค่าถ่วงน้ำหนักของวิธี WMLRC	RANKW
7	GENY	-สร้างตัวแปรตาม	CONU
8	OLS	-คำนวณหาค่าพารามิเตอร์ของวิธีกำลังสองน้อยที่สุด	
9	ML	-คำนวณหาค่าพารามิเตอร์ของวิธีML และค่า SMSE_ML	MSE_ML,INV, MULT
10	WMLCH	-คำนวณหาค่าพารามิเตอร์ของวิธี WMLCHและค่าSMSE_WMLCH	MSE_WMLCH INV,MULT

ตารางที่3.1 (ต่อ)

อันดับที่	ชื่อโปรแกรม	การทำงานของโปรแกรม	ชื่อโปรแกรมย่อยที่เรียกใช้
โปรแกรมย่อย			
11	WMLRC	-คำนวณหาค่าพารามิเตอร์ของวิธี WMLRCและค่าSMSE_WMLRC	MSE_WMLRC INV,MULT
12	GENX1	-สร้างตัวแปรอิสระ x_1 ให้มีค่าผิดปกติ	CONTANOR, XRANK,QUATILE
13	GENX2	-สร้างตัวแปรอิสระ x_2 ให้มีค่าผิดปกติ	CONTANOR, XRANK,QUATILE
14	NORMAL	-สร้างการแจกแจงแบบปกติ	RANDOM
15	XRANK	-เรียงลำดับค่า x และตำแหน่ง	
14	QUATILE	-หาค่า quatile ที่ 1 และ 3	
15	COMBINATION	-สุ่มเลือกชุดค่าสังเกต	STAIL
16	DISTANCE	-หาระยะทางของค่าสังเกต	
17	RANKD	-เรียงลำดับค่าระยะทางและตำแหน่งของค่าสังเกต 500 ชุด	
18	CSTEP	-หาชุดค่าสังเกต 500 ชุดที่ให้ค่า det น้อยที่สุด	
19	RANKDET	-เรียงลำดับค่า det ของค่าสังเกต 10 ชุด	
20	RANKD10	-เรียงลำดับค่าระยะทางและตำแหน่งของค่าสังเกต 10 ชุด	
21	CSTEP10	-หาชุดค่าสังเกต 10 ชุดที่ให้ค่า det น้อยที่สุด	
22	RANKDET10	-เรียงลำดับค่า det ของค่าสังเกต 10 ชุด	
23	RANKW	-เรียงลำดับค่า Robust Distance	
24	CONU	-สร้างการแจกแจงเอกรูปแบบต่อเนื่อง	RANDOM

ตารางที่3.1 (ต่อ)

อันดับที่	ชื่อโปรแกรม	การทำงานของโปรแกรม	ชื่อโปรแกรมย่อยที่เรียกใช้
โปรแกรมย่อย			
25	MSE_ML	-หาผลรวมค่าคลาดเคลื่อนกำลังสองของพารามิเตอร์ของ ML	
26	MSE_WMLCH	-หาผลรวมค่าคลาดเคลื่อนกำลังสองของพารามิเตอร์ของ WMLCH	
27	MSE_WMLRC	-หาผลรวมค่าคลาดเคลื่อนกำลังสองของพารามิเตอร์ของ WMLRC	
28	INV	-หา inverse มิติ 3*3	
29	MULT	-หาผลคูณเมทริกซ์	
30	CONTANOR	-สร้างตัวแปรอิสระให้มีการแจกแจงแบบปกติปลอมปน	
31	STAIL	-หาชุดค่าสังเกตเริ่มต้น	DET_3
32	RANDOM	-สร้างเลขสุ่มที่เป็นอิสระกัน	
33	DET_3	-หา det มิติ 3*3	

บทที่4

ผลการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของตัวแบบการถดถอยโลจิสติค เมื่อตัวแปรอิสระมีค่าผิดปกติ โดยผู้วิจัยได้ทำการเปรียบเทียบวิธีการทั้ง 3 วิธี ซึ่งได้แก่ วิธีความควรจะเป็นสูงสุด วิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และ Haesbroeck และวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann โดยใช้เกณฑ์การเปรียบเทียบคือ ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์ ซึ่งมีวิธีการคำนวณเกณฑ์ที่ใช้ในการเปรียบเทียบดังนี้

$$MSE_j = \frac{1}{1000}\sum_{i=1}^{1000}\left[\hat{\beta}_{ij} - \beta_j\right]^2$$

$$AMSE = \frac{1}{k}\sum_{j=1}^{k} MSE_j$$

เมื่อ β_j แทน พารามิเตอร์ตัวที่ j ในสมการถดถอยโลจิสติค

$\hat{\beta}_{ij}$ แทน ค่าประมาณของพารามิเตอร์ตัวที่ j (β_j) จากการประมาณครั้งที่ i

k แทน จำนวนพารามิเตอร์ ในที่นี้ k = 3

MSE_j แทน ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสำหรับพารามิเตอร์ β_j

AMSE แทน ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์จำนวน k ตัว

ผู้วิจัยได้เสนอผลการวิจัยโดยแบ่งออกเป็น 3 ส่วน คือ

ส่วนที่ 1 ผลการเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่ไม่มีค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2

ส่วนที่ 2 ผลการเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1

ส่วนที่ 3 ผลการเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2

สำหรับการนำเสนอผลการวิจัยจะนำเสนอในรูปแบบตารางและรูปกราฟ เพื่อความสะดวกในการอธิบาย จึงใช้สัญลักษณ์ต่อไปนี้ เพื่อแทนความหมายต่างๆ

PX1 หมายถึง สัดส่วนการปลอมปนของตัวแปรอิสระ x_1

PX2 หมายถึง สัดส่วนการปลอมปนของตัวแปรอิสระ x_2

N หมายถึง ขนาดตัวอย่าง

ML หมายถึง การประมาณค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุด

WMLCH หมายถึง การประมาณค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และ Haesbroeck

WMLRC หมายถึง การประมาณค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann

s.d. หมายถึง ส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

AMSE หมายถึง ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์

4.1 การเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่ไม่มีค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2

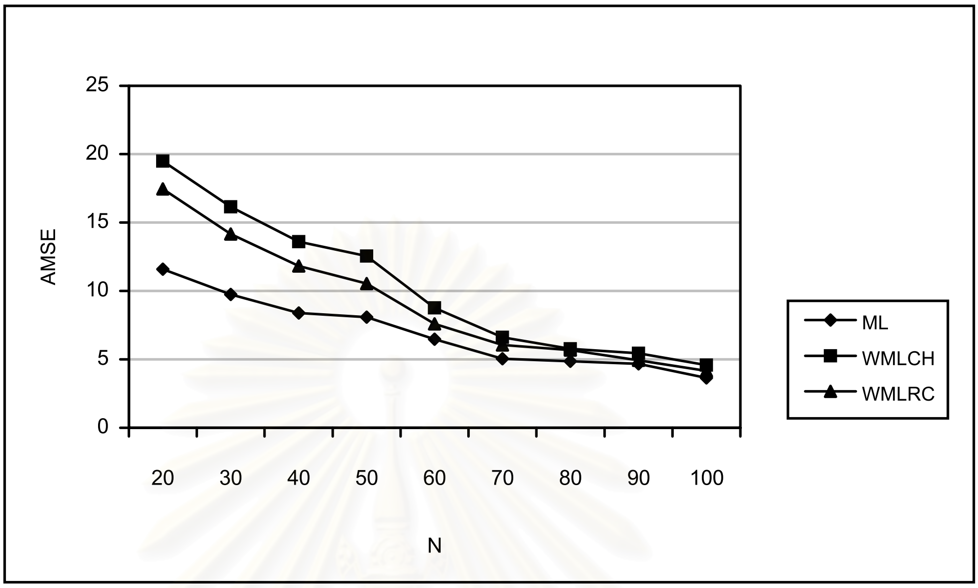
การวิจัยครั้งนี้ ผู้วิจัยทำการศึกษาค่าปกติของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 4.1 และกราฟรูปที่ 4.1

ตารางที่4.1 การเปรียบเทียบค่าประมาณของพารามิเตอร์จากวิธี ML วิธี WMLCH และวิธี WMLRC ด้วยค่า AMSE เมื่อตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 ไม่มีค่าผิดปกติ

ขนาดตัวอย่าง N	AMSE		
	ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
20	11.596263* (19.192911)	19.491837 (30.927593)	17.455846 (28.867454)
30	9.743810* (16.505950)	16.140560 (27.902350)	14.167040 (24.449880)
40	8.391153* (13.646440)	13.587850 (23.639340)	11.821427 (19.517430)
50	8.092108* (13.285330)	12.535960 (21.389460)	10.529560 (17.315840)
60	6.464125* (10.154650)	8.772758 (14.342720)	7.595036 (12.287640)
70	5.044994* (8.165687)	6.609529 (10.580660)	6.060069 (9.613885)
80	4.866234* (7.796138)	5.765262 (9.251051)	5.679097 (9.029369)
90	4.671854* (7.434920)	5.439378 (8.785590)	4.936002 (7.991730)
100	3.635285* (5.882727)	4.586470 (7.233841)	4.165995 (6.660778)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

รูปที่4.1 แสดงการเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 ไม่มีค่าผิดปกติ



จากตารางที่4.1 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 ไม่มีค่าผิดปกติ ซึ่งสามารถสรุปผลได้ดังนี้

เมื่อตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 ไม่มีค่าผิดปกติ ในทุก ๆ ขนาดตัวอย่าง คือ N = 20,30,40,50,60,70,80,90 และ100 วิธี ML ให้ค่า AMSE น้อยที่สุด รองลงมาคือ วิธี WMLRC และวิธี WMLCH ตามลำดับ และจะพบว่า เมื่อตัวอย่างมีขนาดใหญ่(N = 70,80,90 และ 100) วิธี ML วิธี WMLCH และวิธี WMLRC ให้ค่า AMSE ใกล้เคียงกัน

ข้อสรุป

จากตารางที่4.1 ทุกขนาดตัวอย่างวิธี ML ให้ค่า AMSE ต่ำกว่าวิธีอื่น ๆ โดยที่วิธี WMLCH จะเป็นวิธีที่ให้ค่า AMSE สูงที่สุด และจะพบว่าเมื่อตัวอย่างมีขนาดใหญ่ (N = 70,80,90 และ 100) วิธี ML วิธี WMLCH และวิธี WMLRC ให้ค่า AMSE ที่ใกล้เคียงกัน และจะสังเกตได้ว่าค่า AMSE ของทุกวิธีมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น เพราะเมื่อขนาดตัวอย่างเพิ่มขึ้นจะทำให้ความแปรปรวนลดลงและค่าประมาณจะเข้าใกล้ค่าจริงมากขึ้นหรือมีความเอนเอียงน้อยลงจึงส่งผลให้ค่า AMSE ลดลง

4.2 การเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1

การวิจัยครั้งนี้ ผู้วิจัยทำการศึกษาระดับค่าผิดปกติของตัวแปรอิสระ x_1 2 ระดับ คือ

4.2.1 ตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับไม่รุนแรง(C=5) ทำการศึกษาที่สัดส่วนการปลอมปนเป็น 0.05 , 0.10 และ 0.15 ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 4.2.1 และกราฟรูปที่ 4.2.1

ตารางที่4.2.1 การเปรียบเทียบค่าประมาณของพารามิเตอร์จากวิธี ML วิธี WMLCH และวิธี WMLRC ด้วยค่า AMSE เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ ไม่รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1

สัดส่วนการปลอมปน PX1	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05	20	15.815250 (21.440121)	14.218100 (20.739447)	12.865340* (18.991332)
	30	10.012390 (16.282900)	9.145673 (15.025680)	7.839638* (12.722680)
	40	8.874522 (14.334340)	8.263473 (13.441072)	7.018407* (11.032070)
	50	8.343843 (13.610445)	7.760020 (12.534311)	6.834788* (11.142360)
	60	6.882869 (11.204220)	6.269540 (10.582670)	6.163590* (9.822516)
	70	6.012746 (9.608660)	5.400862 (8.526520)	5.183291* (8.239480)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.2.1 (ต่อ)

สัดส่วนการปลอมปน PX1	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05	80	5.524580 (8.332720)	5.093114 (7.997530)	4.875261* (7.514330)
	90	5.092584 (7.983750)	4.652280 (7.242560)	4.466670* (7.057190)
	100	4.481638 (7.084289)	4.034474 (6.807610)	3.993544* (6.736465)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.2.1 (ต่อ)

สัดส่วนการปลอมปน PX1	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.10	20	17.929150 (24.419090)	15.914430 (21.583650)	14.659410* (20.845880)
	30	10.133940 (16.573530)	9.992354 (15.458730)	8.249493* (13.891083)
	40	9.350096 (15.204670)	8.548240 (13.995530)	7.437162* (12.212360)
	50	8.828650 (14.568990)	8.002160 (13.038090)	7.177440* (11.128390)
	60	7.169284 (11.340810)	6.448200 (10.290140)	6.224416* (10.102141)
	70	6.543552 (10.402870)	5.973520 (9.584476)	5.842308* (9.481385)
	80	6.380802 (10.138370)	5.765019 (8.985060)	5.637029* (8.714687)
	90	5.730252 (8.953670)	5.137548 (8.144398)	5.014943* (8.012962)
	100	5.044331 (8.053160)	4.365096 (7.050290)	4.249579* (6.986760)

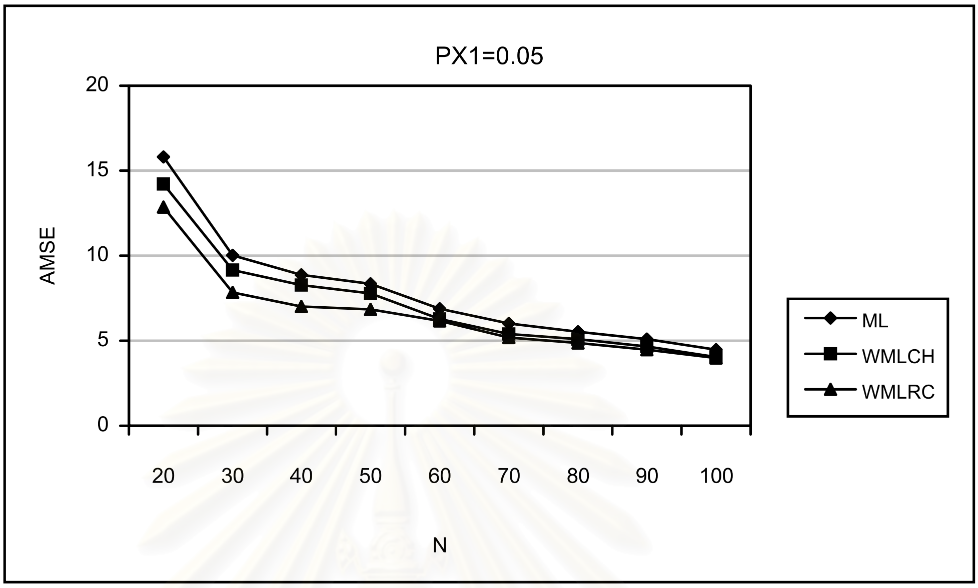
* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

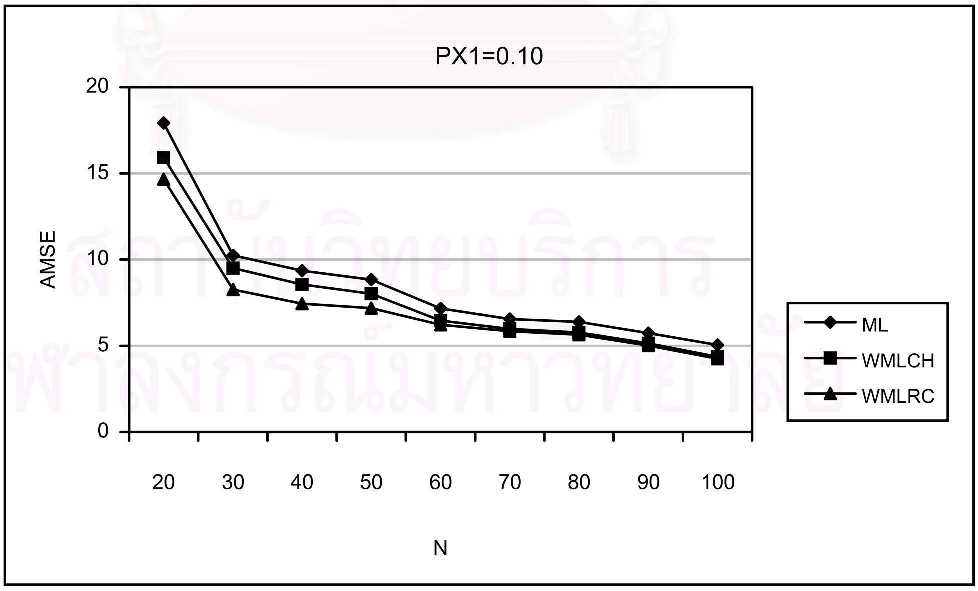
ตารางที่4.2.1 (ต่อ)

สัดส่วนการปลอมปน PX1	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.15	20	20.023536 (30.133217)	16.592942 (23.025141)	15.718240* (21.145781)
	30	15.049653 (20.508394)	13.267942 (19.593247)	12.054381* (18.286200)
	40	11.661150 (17.975390)	10.525540 (16.994470)	9.410365* (15.357880)
	50	10.125320 (16.446860)	9.037810 (15.043370)	8.250312* (13.459460)
	60	8.356366 (13.859030)	7.339150 (11.658240)	7.128163* (11.450940)
	70	7.329405 (11.629890)	6.480495 (10.443160)	6.332237* (9.832030)
	80	7.182118 (11.447840)	6.101369 (9.718800)	6.035255* (9.695070)
	90	6.390467 (9.902920)	5.524567 (8.672440)	5.459386* (8.597620)
	100	5.646428 (8.737870)	4.814347 (7.833810)	4.782903* (7.794240)

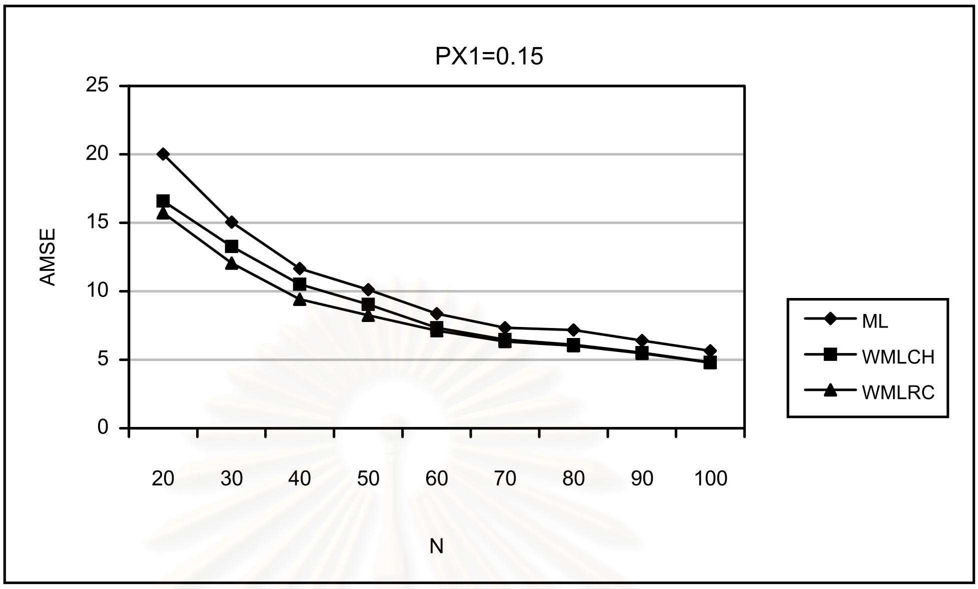
* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

รูปที่4.2.1 แสดงการเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ ไม่รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1





รูปที่4.2.1 (ต่อ)


PX1=0.15
AMSE
25
20
15
10
5
0
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC

จากตารางที่4.2.1 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับไม่รุนแรง ซึ่งมีสัดส่วนการปลอมปนเท่ากับ 0.05 , 0.10 และ 0.15 สามารถสรุปผลได้ดังนี้

เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับไม่รุนแรง ในทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (P = 0.05,0.10 และ 0.15) โดยในทุก ๆ ขนาดตัวอย่าง คือ N = 20,30,40,50,60,70,80, 90 และ 100 วิธี WMLRCให้ค่า AMSE น้อยที่สุด รองลงมาคือ วิธี WMLCH และวิธี ML ตามลำดับ และจะพบว่าเมื่อตัวอย่างมีขนาดใหญ่(N = 60,70,80,90 และ 100) วิธี WMLCH และวิธี WMLRC ให้ค่า AMSE ใกล้เคียงกัน

ข้อสรุป

จากตารางที่4.2.1 ทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และทุกขนาดตัวอย่างวิธี WMLRC ให้ค่า AMSE ต่ำกว่าวิธีอื่น ๆ โดยที่วิธี ML จะเป็นวิธีที่ให้ค่า AMSE สูงที่สุด และจะพบว่าเมื่อตัวอย่างมีขนาดใหญ่ (N = 60,70,80,90 และ 100) วิธี WMLCH และวิธี WMLRC ให้ค่า AMSE ที่ใกล้เคียงกัน

ค่า AMSE ของทุกวิธีมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น ขณะที่สัดส่วนการปลอมปนคงที่

ค่า AMSE ของทุกวิธีมีแนวโน้มเพิ่มขึ้นเมื่อสัดส่วนการปลอมปนเพิ่มขึ้น ขณะที่ขนาดตัวอย่างคงที่

4.2.2 ตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับรุนแรง(C=12) ทำการศึกษาที่สัดส่วนการปลอมปนเป็น 0.05 , 0.10 และ 0.15 ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 4.2.2 และกราฟรูปที่ 4.2.2

ตารางที่4.2.2 การเปรียบเทียบค่าประมาณของพารามิเตอร์จากวิธี ML วิธี WMLCH และวิธี WMLRC ด้วยค่า AMSE เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1

สัดส่วนการปลอมปน PX1	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05	20	17.740790 (29.578700)	16.224900 (27.350900)	14.966270* (24.877060)
	30	12.266830 (18.581790)	10.783531 (17.625680)	9.587802* (15.646380)
	40	11.058040 (16.758190)	9.794290 (16.053490)	8.601069* (13.983880)
	50	10.495380 (17.340960)	9.131880 (15.043320)	8.386696* (13.721180)
	60	9.465770 (15.374870)	8.297283 (13.281340)	8.115396* (13.219730)
	70	9.291430 (15.257470)	8.049920 (13.146100)	7.931949* (12.823090)
	80	7.884260 (12.577250)	6.730745 (10.868690)	6.624678* (10.601520)
	90	7.584740 (12.941120)	6.485651 (10.376040)	6.389822* (10.214960)
	100	7.449580 (11.843820)	6.331259 (10.114230)	6.265774* (9.973194)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.2.2 (ต่อ)

สัดส่วนการปลอมปน PX1	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.10	20	20.301532 (34.143325)	18.245539 (31.762622)	17.022592* (27.824078)
	30	14.792300 (24.455000)	12.636130 (20.752220)	11.803120* (19.438280)
	40	13.659220 (22.424250)	11.649530 (18.993590)	10.951440* (17.735630)
	50	12.382330 (18.590720)	10.426024 (17.310360)	9.921807* (16.250110)
	60	10.317480 (17.210590)	8.490380 (14.744500)	8.313716* (13.593460)
	70	9.932380 (16.386700)	8.222170 (13.502240)	8.139442* (13.254650)
	80	9.006820 (14.834470)	7.729259 (12.593260)	7.634104* (12.407300)
	90	8.815130 (14.501880)	7.198894 (11.490360)	7.123164* (11.424480)
	100	8.380470 (13.702270)	6.754774 (10.977650)	6.686470* (10.755610)

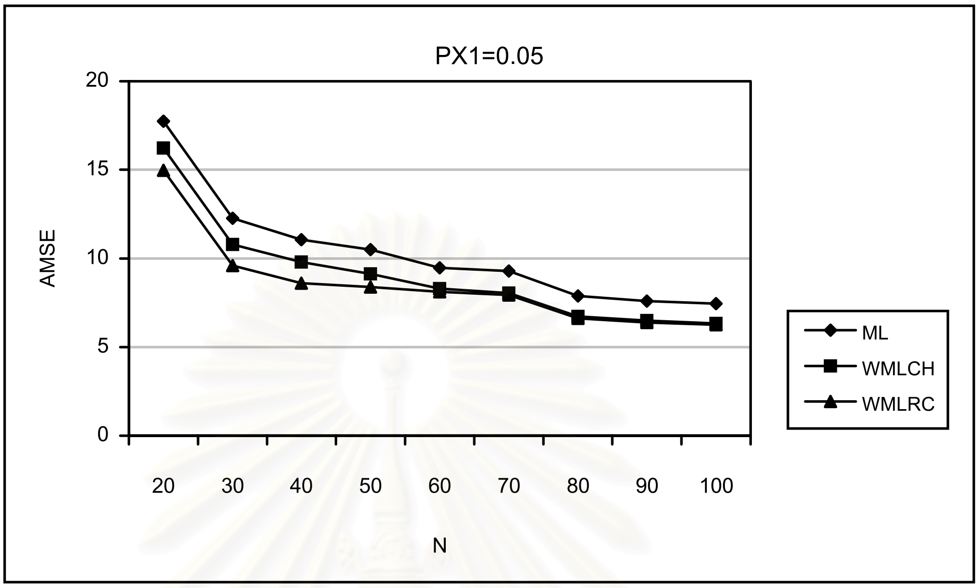
* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

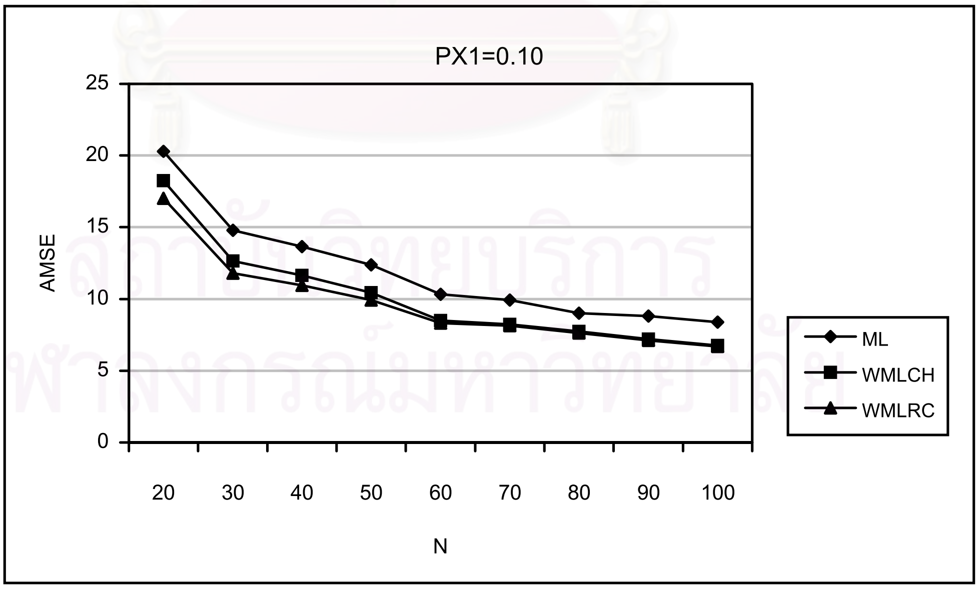
ตารางที่4.2.2 (**ต่อ**)

สัดส่วนการปลอมปน PX1	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.15	20	22.627483 (36.240783)	19.729156 (32.876798)	18.577258* (31.930281)
	30	18.440700 (30.649150)	15.518560 (25.011130)	14.765820* (24.611840)
	40	16.647200 (27.751300)	13.746410 (22.599110)	12.996908* (21.586540)
	50	13.314770 (22.038830)	11.215490 (18.592310)	10.606943* (17.677040)
	60	11.634410 (18.983550)	9.539993 (15.558170)	9.385119* (15.340830)
	70	11.134090 (18.320850)	9.030902 (14.767990)	8.947356* (14.608320)
	80	10.746500 (17.643850)	8.846206 (14.682300)	8.775593* (14.338660)
	90	9.814040 (16.078230)	7.919766 (12.560133)	7.856375* (12.504840)
	100	9.634760 (15.963370)	7.731048 (12.411510)	7.674270* (12.378990)

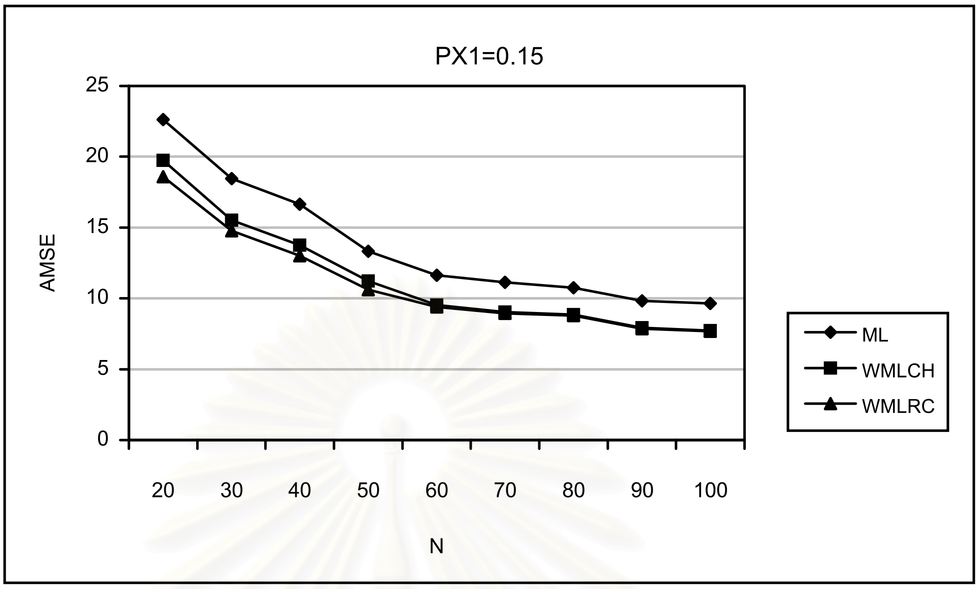
* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

รูปที่4.2.2 แสดงการเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1





รูปที่4.2.2 (ต่อ)


PX1=0.15
AMSE
25
20
15
10
5
0
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC

จากตารางที่4.2.2 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ รุนแรง ซึ่งมีสัดส่วนการปลอมปนเท่ากับ 0.05 , 0.10 และ 0.15 สามารถสรุปผลได้ดังนี้

เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับรุนแรง ในทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1(P = 0.05,0.10 และ 0.15) โดยในทุก ๆ ขนาดตัวอย่าง คือ N = 20,30,40,50,60,70,80,90 และ 100 วิธี WMLRC ให้ค่า AMSE น้อยที่สุด รองลงมาคือ วิธี WMLCH และวิธี ML ตามลำดับ และจะพบว่าเมื่อตัวอย่างมีขนาดใหญ่(N = 60,70,80,90 และ 100) วิธี WMLCH และวิธี WMLRC ให้ค่า AMSE ใกล้เคียงกัน

ข้อสรุป

จากตารางที่4.2.2 ทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และทุกขนาดตัวอย่างวิธี WMLRC ให้ค่า AMSE ต่ำกว่าวิธีอื่น ๆ โดยที่วิธี ML จะเป็นวิธีที่ให้ค่า AMSE สูงที่สุด และจะพบว่าเมื่อตัวอย่างมีขนาดใหญ่ (N = 60,70,80,90 และ 100) วิธี WMLCH และวิธี WMLRC ให้ค่า AMSE ที่ใกล้เคียงกัน

ค่า AMSE ของทุกวิธีมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น ขณะที่สัดส่วนการปลอมปนคงที่

ค่า AMSE ของทุกวิธีมีแนวโน้มเพิ่มขึ้นเมื่อสัดส่วนการปลอมปนเพิ่มขึ้น ขณะที่ขนาดตัวอย่างคงที่

4.3 การเปรียบเทียบค่าประมาณของพารามิเตอร์ในกรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2

การวิจัยครั้งนี้ ผู้วิจัยทำการศึกษาระดับค่าผิดปกติของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 ในระดับค่าผิดปกติ 2 ระดับ คือ ระดับ ไม่รุนแรง และระดับ รุนแรง โดยได้ทำการศึกษาในกรณีต่าง ๆ ดังนี้

4.3.1 ตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับไม่รุนแรง(C=5) และตัวแปรอิสระ x_2 มีระดับค่าผิดปกติระดับไม่รุนแรง(C=5) โดยแต่ละระดับทำการศึกษาที่สัดส่วนการปลอมปนเป็น 0.05 , 0.10 และ 0.15 ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 4.3.1 และกราฟรูปที่ 4.3.1

ตารางที่4.3.1 การเปรียบเทียบค่าประมาณของพารามิเตอร์จากวิธี ML วิธี WMLCH และวิธี WMLRC ด้วยค่า AMSE เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ ไม่รุนแรงและตัวแปรอิสระ x_2 มีค่าผิดปกติระดับ ไม่รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.05	20	19.670681 (32.746457)	14.925488 (24.856976)	14.138422* (23.972248)
	30	17.588772 (29.224723)	13.371553 (22.176187)	12.587319* (20.931959)
	40	14.212450 (23.856490)	10.338700 (17.178630)	9.916656* (16.326790)
	50	11.762925 (19.587680)	8.009106 (13.097950)	7.703659* (12.515320)
	60	10.959692 (17.929350)	7.479154 (11.908140)	7.338941* (11.815520)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.1 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.05	70	10.149651 (17.117440)	6.620574 (10.679160)	6.543541* (10.566810)
	80	9.387999 (15.431020)	6.099840 (9.711670)	6.013754* (9.655310)
	90	9.222574 (15.040020)	5.958650 (9.497680)	5.909167* (9.421103)
	100	8.283166 (13.508735)	5.152247 (8.892260)	5.109167* (8.821103)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.1 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.10	20	21.363910 (35.255784)	16.475893 (27.470839)	15.791564* (25.348703)
	30	18.568622 (30.710853)	13.817339 (22.615021)	13.194967* (21.936054)
	40	16.439380 (27.444240)	12.086140 (19.981820)	11.674562* (18.851380)
	50	13.595946 (22.384310)	9.432907 (15.923320)	9.394064* (15.879860)
	60	12.268847 (20.277910)	8.344110 (13.772960)	8.251512* (13.441780)
	70	11.002513 (18.303076)	7.169833 (11.537070)	7.091746* (11.424660)
	80	10.627665 (17.519390)	6.914503 (11.284790)	6.888373* (11.084980)
	90	10.527581 (17.492290)	6.851207 (11.019220)	6.823487* (10.972000)
	100	10.107932 (16.904706)	6.649934 (10.753460)	6.613621* (10.600040)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.1 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.15	20	21.635641 (35.745511)	16.781668 (27.826456)	16.127483* (27.087894)
	30	20.091565 (33.914124)	15.210191 (25.145858)	14.738309* (24.608024)
	40	17.271629 (29.750090)	12.465350 (20.839120)	12.028459* (20.012140)
	50	15.392360 (25.746830)	10.661120 (17.432450)	10.574111* (17.393140)
	60	13.742936 (22.241630)	9.069810 (14.801090)	8.994323* (14.780510)
	70	12.895270 (21.323134)	8.343146 (13.461320)	8.274878* (13.352370)
	80	12.438942 (20.750250)	7.912280 (12.823550)	7.899779* (12.760560)
	90	11.706825 (19.114294)	7.395332 (11.712520)	7.377878* (11.679530)
	100	11.385551 (18.879004)	6.955795 (11.362060)	6.932679* (11.120078)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.1 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.10,0.05	20	21.165182 (34.973041)	16.266824 (27.143622)	15.501938* (25.057465)
	30	18.722926 (30.924424)	14.063353 (22.896162)	13.386805* (22.138656)
	40	15.328220 (24.731790)	10.919340 (17.713800)	10.533090* (17.318120)
	50	13.061274 (21.202170)	8.563670 (13.903350)	8.495672* (13.818570)
	60	12.242231 (20.254030)	8.312350 (13.556700)	8.228745* (13.360390)
	70	11.775965 (18.919680)	7.875487 (12.949210)	7.805591* (12.626790)
	80	11.438057 (18.649500)	7.705654 (12.548980)	7.672676* (12.431180)
	90	10.719071 (17.800270)	7.017800 (11.233500)	6.985486* (11.211420)
	100	10.086759 (16.667090)	6.554494 (10.547220)	6.525798* (10.440450)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.1 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.10,0.10	20	21.801057 (35.791421)	16.760795 (27.824368)	16.111765* (27.112598)
	30	19.357209 (31.915159)	14.481402 (22.970274)	13.844573* (22.687953)
	40	17.537473 (28.591140)	12.803620 (20.960630)	12.596219* (20.702630)
	50	15.022131 (24.226110)	10.652534 (17.421650)	10.566364* (17.396940)
	60	13.380060 (22.163410)	9.237690 (15.149230)	9.147582* (15.101980)
	70	12.659344 (20.640030)	8.561679 (13.987140)	8.503163* (13.881590)
	80	12.073461 (19.921680)	8.168070 (13.200870)	8.133299* (13.144400)
	90	11.689031 (18.903960)	7.794493 (12.651320)	7.768805* (12.590390)
	100	10.713987 (17.601010)	7.060261 (11.528510)	7.047015* (11.304590)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.1 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.10,0.15	20	22.532293 (37.234050)	17.454883 (28.434157)	16.952036* (28.054322)
	30	21.240562 (35.195378)	16.329145 (27.314342)	15.816406* (25.914875)
	40	18.792880 (30.967470)	13.938690 (22.971250)	13.546044* (22.466020)
	50	15.860079 (25.576380)	11.166550 (18.954830)	11.078842* (18.872430)
	60	14.270710 (23.892430)	9.509190 (15.418130)	9.433870* (15.332500)
	70	13.506595 (22.419518)	9.191149 (14.816660)	9.135905* (14.761432)
	80	12.768478 (20.764130)	8.673190 (14.077380)	8.640095* (14.045150)
	90	12.234674 (20.280430)	8.183551 (13.350290)	8.160604* (13.338640)
	100	11.704337 (19.348260)	7.589977 (12.293600)	7.562749* (12.217235)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.1 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.15,0.05	20	21.709561 (35.989426)	16.813751 (27.933898)	16.202658* (27.282821)
	30	19.865973 (33.379202)	15.076682 (24.809038)	14.548037* (24.257341)
	40	19.005480 (32.273840)	14.201590 (23.957270)	13.892630* (22.544800)
	50	17.541590 (30.440570)	12.810150 (21.484860)	12.728170* (21.144740)
	60	14.611626 (24.290570)	9.917290 (16.524530)	9.847464* (16.215800)
	70	12.873541 (21.313500)	8.350777 (13.582850)	8.288600* (13.435290)
	80	12.628502 (20.974080)	8.208866 (13.397730)	8.180225* (13.311970)
	90	12.172030 (20.384320)	7.755680 (12.444890)	7.733069* (12.408610)
	100	11.694948 (19.019960)	7.273691 (11.719590)	7.254109* (11.671260)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.1 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.15,0.10	20	22.804431 (37.519068)	17.762052 (28.801117)	17.203695* (28.173343)
	30	20.817046 (34.879497)	15.989539 (26.075287)	15.495422* (25.370964)
	40	19.270041 (31.794215)	14.457578 (24.134853)	14.062884* (23.792167)
	50	18.395733 (30.109320)	13.688760 (22.564440)	13.595146* (22.423910)
	60	15.835209 (25.983970)	10.175910 (16.517590)	10.100377* (16.153280)
	70	13.548091 (22.829770)	9.177036 (15.372420)	9.114584* (15.286320)
	80	13.086700 (22.080610)	8.907940 (14.640960)	8.875796* (14.588560)
	90	12.661150 (20.546720)	8.517720 (13.720220)	8.494471* (13.616950)
	100	11.939767 (19.395390)	7.357602 (11.793890)	7.330849* (11.783250)

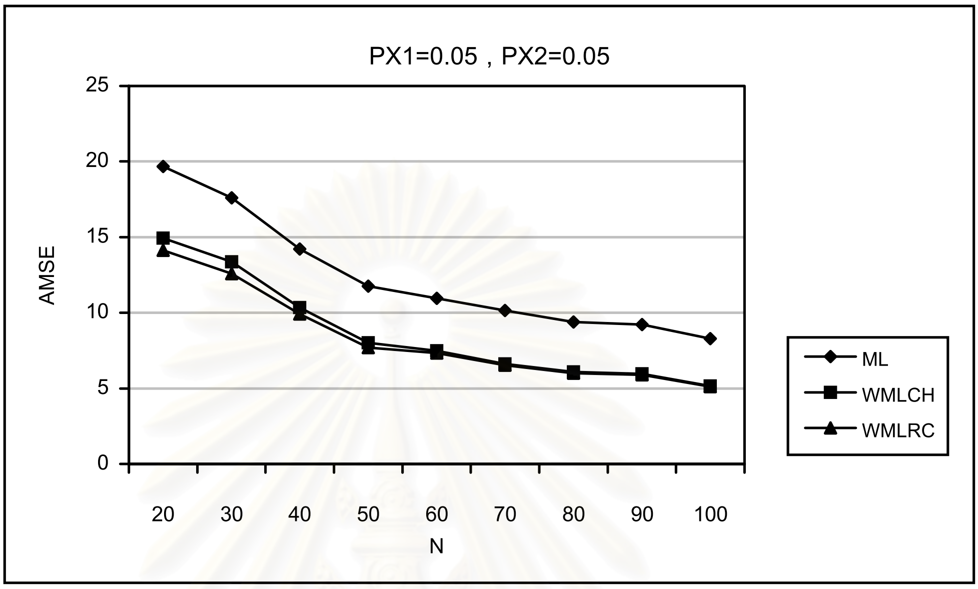
* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

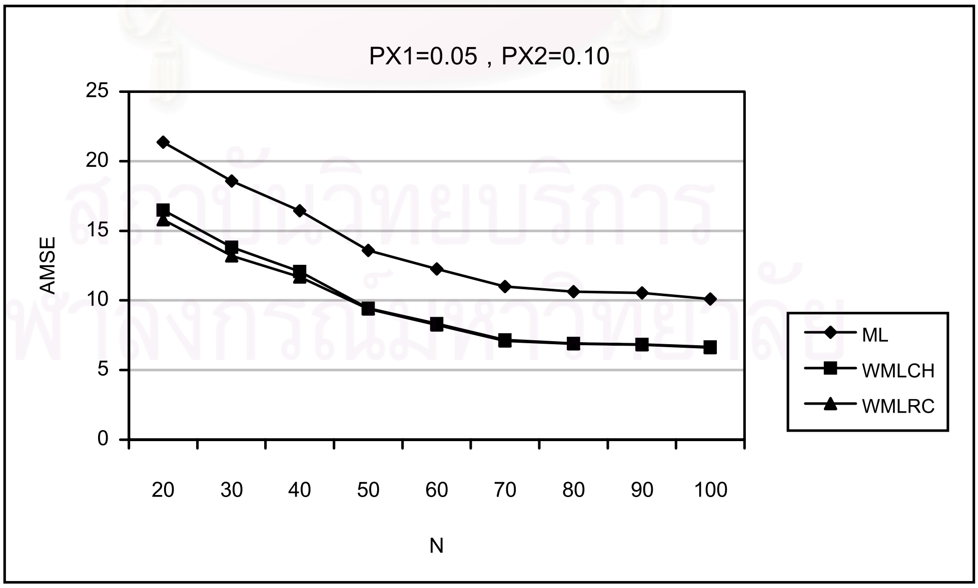
ตารางที่4.3.1 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.15,0.15	20	23.655406 (39.754176)	18.357418 (30.347575)	17.970384* (28.971853)
	30	21.676365 (35.990636)	16.582422 (27.610780)	16.173356* (26.825770)
	40	19.759318 (32.394100)	14.835550 (24.690280)	14.506811* (24.438930)
	50	18.811304 (30.943500)	14.027760 (23.247590)	13.944812* (22.803200)
	60	16.331150 (26.322893)	11.623930 (19.182890)	11.558852* (19.028630)
	70	15.277720 (25.101870)	10.720814 (17.734610)	10.674597* (17.610510)
	80	13.416053 (22.226930)	9.107920 (14.806530)	9.083552* (14.751970)
	90	13.026892 (21.988150)	8.700572 (13.611420)	8.680792* (13.579120)
	100	12.394633 (20.282630)	8.209026 (13.559400)	8.188543* (13.494430)

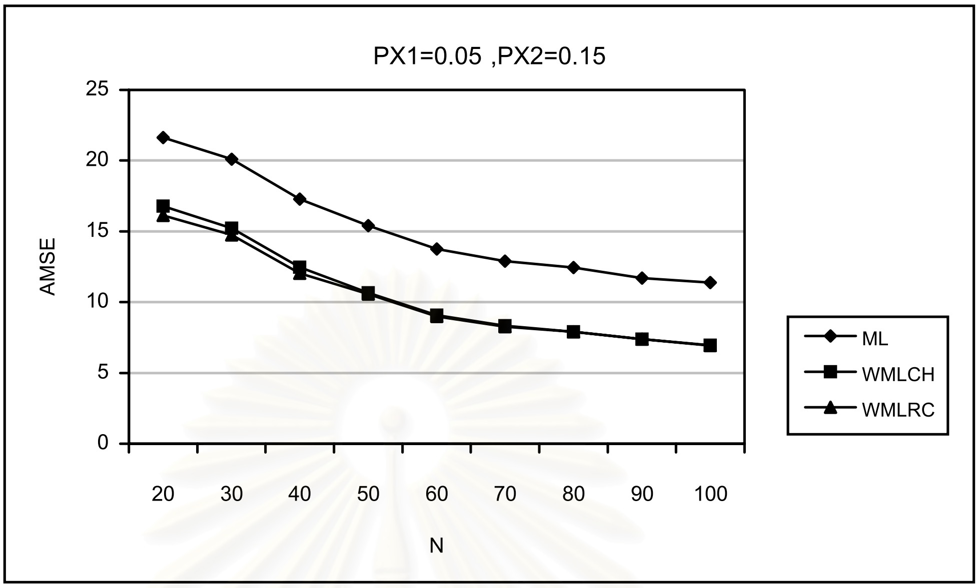
* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

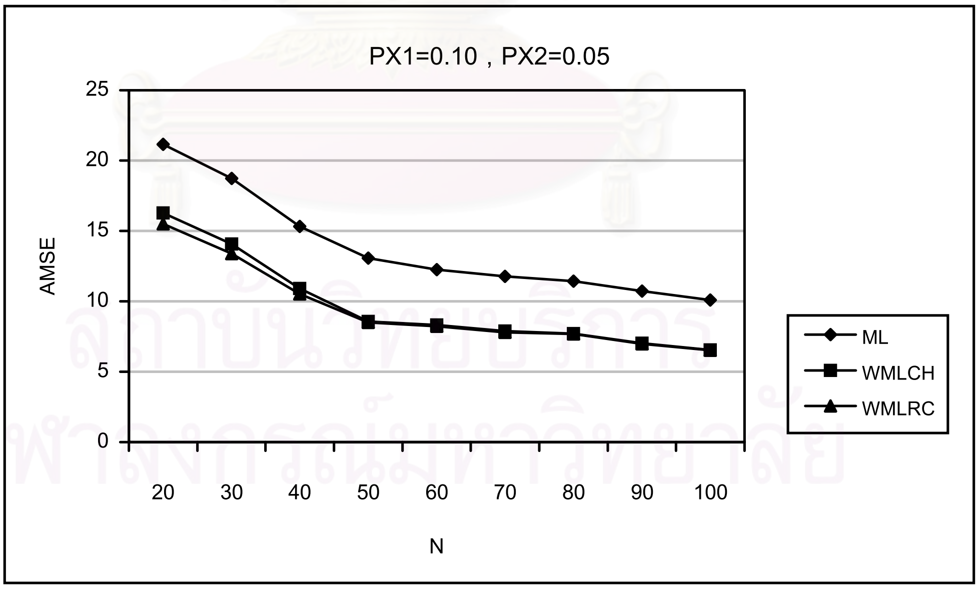
รูปที่4.3.1 แสดงการเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ ไม่รุนแรงและตัวแปรอิสระ x_2 มีค่าผิดปกติระดับ ไม่รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2



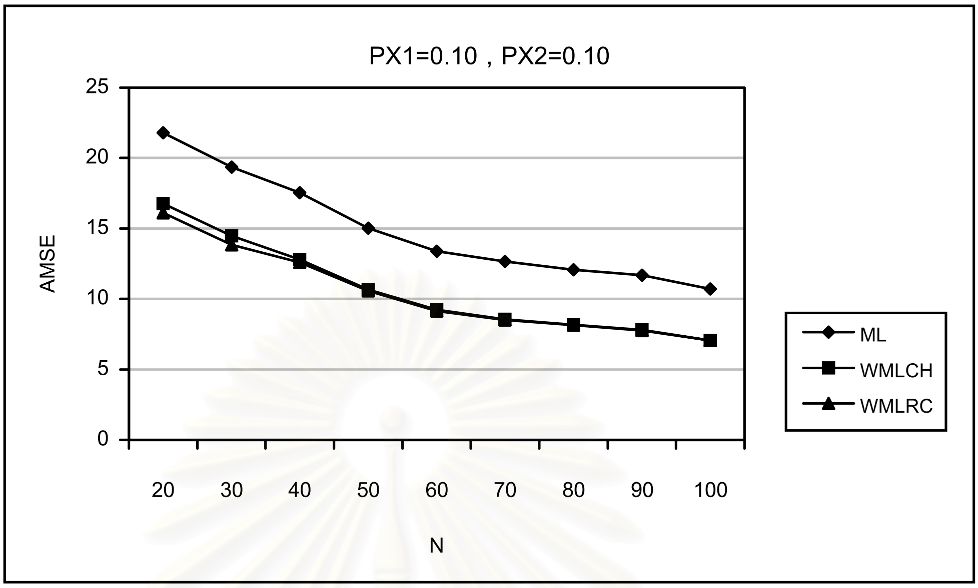


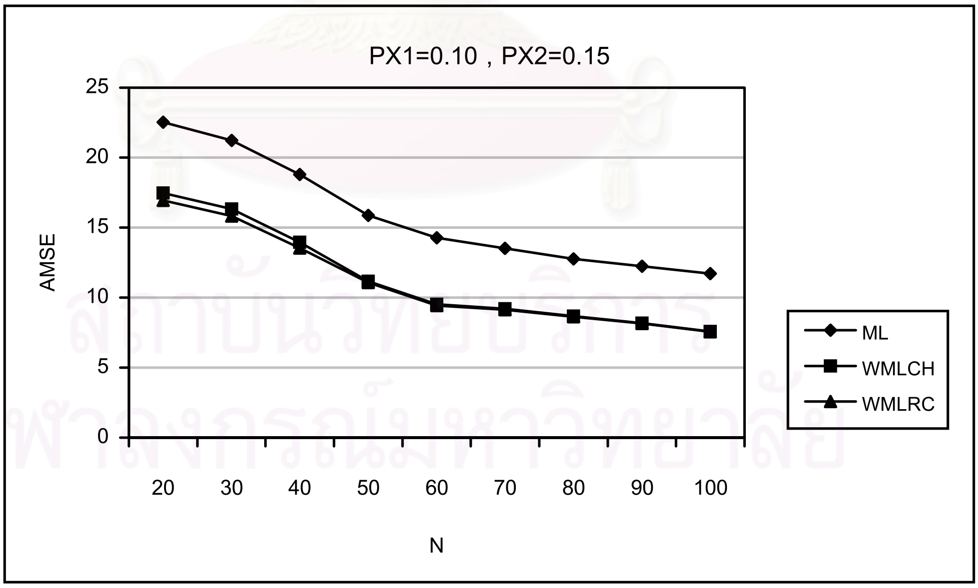
รูปที่4.3.1 (ต่อ)


PX1=0.05 ,PX2=0.15
AMSE
N
ML
WMLCH
WMLRC

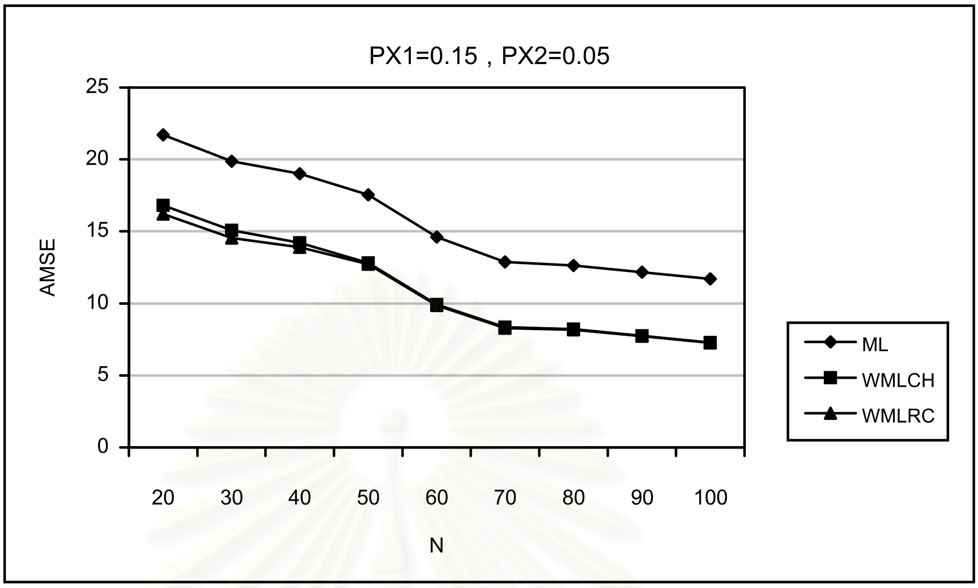

PX1=0.10 , PX2=0.05
AMSE
N
ML
WMLCH
WMLRC

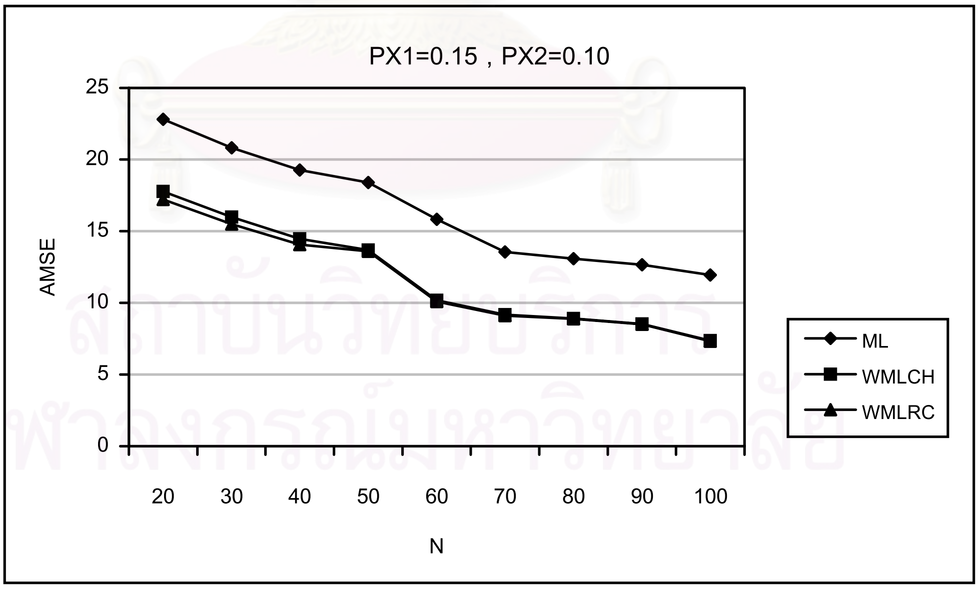
รูปที่4.3.1 (ต่อ)


PX1=0.10 , PX2=0.10
25
20
15
10
5
0
AMSE
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC

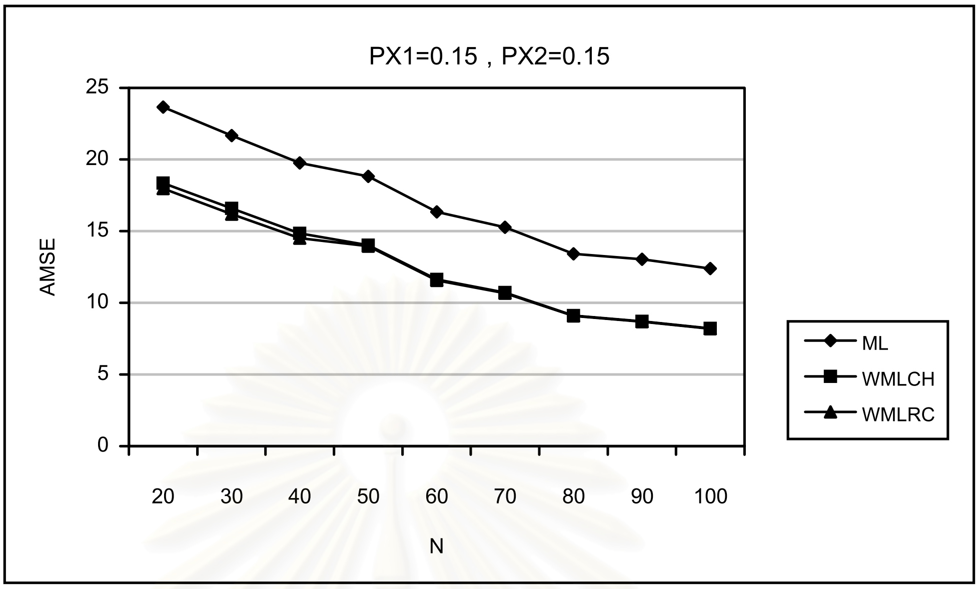

PX1=0.10 , PX2=0.15
25
20
15
10
5
0
AMSE
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC

รูปที่4.3.1 (ต่อ)


PX1=0.15 , PX2=0.05
AMSE
N
25
20
15
10
5
0
20
30
40
50
60
70
80
90
100
ML
WMLCH
WMLRC


PX1=0.15 , PX2=0.10
AMSE
N
25
20
15
10
5
0
20
30
40
50
60
70
80
90
100
ML
WMLCH
WMLRC

รูปที่4.3.1 (ต่อ)


PX1=0.15 , PX2=0.15
25
20
15
10
5
0
AMSE
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC

จากตารางที่4.3.1 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับไม่รุนแรง ซึ่งมีสัดส่วนการปลอมปนเท่ากับ 0.05 , 0.10 และ 0.15 และตัวแปรอิสระ x_2 มีค่าผิดปกติระดับไม่รุนแรง ซึ่งมีสัดส่วนการปลอมปน เท่ากับ 0.05 , 0.10 และ 0.15 สามารถสรุปผลได้ดังนี้

เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับไม่รุนแรงและตัวแปรอิสระ x_2 มีค่าผิดปกติระดับไม่รุนแรง ในทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1 = 0.05 , 0.10 และ 0.15) และทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_2 (PX2 = 0.05 , 0.10 และ 0.15) โดยในทุก ๆ ขนาดตัวอย่าง คือ N = 20,30,40,50,60,70,80,90 และ 100 วิธี WMLRC ให้ค่า AMSE น้อยที่สุด รองลงมาคือ วิธี WMLCH และวิธี ML ตามลำดับ และจะพบว่าวิธี WMLCH และวิธี WMLRC ให้ค่า AMSE ใกล้เคียงกันโดยเฉพาะอย่างยิ่งเมื่อตัวอย่างมีขนาดใหญ่ (N = 50,60,70,80,90 และ 100)

ข้อสรุป

จากตารางที่4.3.1 ทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 และทุกขนาดตัวอย่างวิธี WMLRC ให้ค่า AMSE ต่ำกว่าวิธีอื่น ๆ โดยที่วิธี ML จะเป็นวิธีที่ให้ค่า AMSE สูงที่สุด และจะพบว่าวิธี WMLCH และ วิธี WMLRC ให้ค่า AMSE ที่ใกล้เคียงกันโดยเฉพาะอย่างยิ่งเมื่อตัวอย่างมีขนาดใหญ่ (N = 50,60,70,80,90 และ 100)

ค่า AMSE ของทุกวิธีมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น ขณะที่สัดส่วนการปลอมปนคงที่

ค่า AMSE ของทุกวิธีมีแนวโน้มเพิ่มขึ้นเมื่อสัดส่วนการปลอมปนเพิ่มขึ้น ขณะที่ขนาดตัวอย่างคงที่

4.3.2 ตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับไม่รุนแรง(C=5) และตัวแปรอิสระ x_2 มีระดับค่าผิดปกติระดับรุนแรง(C=12) โดยแต่ละระดับทำการศึกษาที่สัดส่วนการปลอมปนเป็น 0.05 , 0.10 และ 0.15 ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 4.3.2 และกราฟรูปที่ 4.3.2

ตารางที่4.3.2 การเปรียบเทียบค่าประมาณของพารามิเตอร์จากวิธี ML วิธี WMLCH และวิธี WMLRC ด้วยค่า AMSE เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ ไม่รุนแรงและตัวแปรอิสระ x_2 มีค่าผิดปกติระดับ รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.05	20	33.698288 (54.876531)	17.330393 (27.252735)	16.931546* (26.731813)
	30	29.238420 (47.967510)	14.171960 (24.350290)	14.001220* (24.050240)
	40	27.071790 (44.692450)	13.112650 (21.962200)	13.065040* (21.655920)
	50	24.542166 (41.099030)	11.421820 (18.885710)	11.330820* (18.740190)
	60	22.874193 (37.876280)	10.661290 (17.190800)	10.576966* (16.947110)
	70	20.543040 (34.734880)	8.874884 (14.711200)	8.795013* (14.514000)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.2 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.05	80	19.378906 (32.258745)	8.088790 (13.864670)	8.015581* (13.079720)
	90	18.077716 (30.963163)	7.574930 (12.448120)	7.522664* (12.309490)
	100	17.351440 (27.782202)	7.117500 (11.528110)	7.069777* (11.352650)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.2 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.10	20	35.261149 (57.167299)	18.549053 (31.854229)	17.893778* (28.583302)
	30	31.647150 (50.147380)	15.369360 (25.394110)	15.262500* (25.148570)
	40	29.786557 (48.683450)	14.649830 (24.707860)	14.552077* (24.673770)
	50	26.821685 (43.785084)	12.336480 (20.794340)	12.248757* (20.567280)
	60	24.938977 (41.648473)	11.253450 (17.965030)	11.185849* (17.770320)
	70	22.441700 (37.178160)	10.004057 (16.077650)	9.931753* (15.995550)
	80	21.680660 (35.665216)	9.501014 (15.663330)	9.441853* (15.618350)
	90	20.887032 (34.903179)	9.013907 (15.069520)	8.968044* (14.903040)
	100	19.255033 (32.167468)	8.153715 (14.037890)	8.111430* (13.939296)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.2 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.15	20	37.593851 (59.418624)	20.231612 (34.016549)	19.053482* (31.913694)
	30	34.536548 (56.864869)	17.279847 (27.461041)	17.146126* (27.391231)
	40	31.879686 (50.951223)	15.181460 (25.102920)	15.096192* (25.069130)
	50	30.240729 (48.797814)	14.377380 (24.588070)	14.298051* (24.502750)
	60	27.594217 (45.707854)	12.779200 (21.016500)	12.708227* (20.972570)
	70	25.526341 (42.702066)	12.001964 (20.092070)	11.947479* (19.955190)
	80	24.023154 (40.057238)	11.112469 (17.737490)	11.056170* (17.634820)
	90	22.303745 (36.959377)	10.273947 (16.407760)	10.233135* (16.352071)
	100	21.407075 (35.318502)	9.695772 (15.777850)	9.654213* (15.704630)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.2 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.10,0.05	20	34.492366 (55.918646)	17.546193 (27.876533)	17.116148* (26.891386)
	30	30.497524 (48.667321)	14.671878 (24.746115)	14.568432* (24.692167)
	40	27.755193 (44.806110)	13.259330 (22.053220)	13.169860* (21.987030)
	50	25.463230 (42.325840)	11.762731 (18.905910)	11.677690* (18.839520)
	60	24.153683 (39.865736)	11.035019 (17.898090)	10.949667* (17.727070)
	70	21.519452 (35.707470)	9.473498 (15.661070)	9.394789* (15.610950)
	80	20.892576 (35.169128)	8.988135 (14.911550)	8.917597* (14.711311)
	90	19.552474 (32.786531)	8.208218 (13.185420)	8.154448* (13.168800)
	100	18.465237 (31.539727)	7.845428 (12.956630)	7.796806* (12.914570)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.2 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.10,0.10	20	36.717599 (58.312771)	19.026658 (31.907141)	18.795686* (31.815867)
	30	32.810835 (53.676320)	15.981790 (26.082610)	15.869587* (25.957940)
	40	30.797682 (50.727190)	14.972450 (25.229210)	14.886746* (25.119680)
	50	28.302680 (45.369614)	13.432656 (22.415660)	13.342041* (22.353960)
	60	26.954338 (43.429489)	12.835380 (20.810210)	12.776719* (20.791820)
	70	24.318921 (40.582220)	11.038239 (18.292680)	10.979635* (18.056590)
	80	22.846631 (37.965970)	10.212426 (17.799090)	10.164362* (17.477500)
	90	22.130940 (36.785911)	10.005640 (16.934270)	9.964394* (16.905640)
	100	20.637382 (34.274332)	8.931495 (14.583370)	8.896269* (14.451878)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.2 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.10,0.15	20	38.938836 (61.429582)	21.048375 (35.125718)	19.836719* (33.903131)
	30	35.296665 (57.323674)	17.792456 (28.465483)	17.625923* (28.247161)
	40	32.738231 (53.256121)	15.638746 (26.922458)	15.543751* (26.721609)
	50	31.412459 (50.096275)	15.056230 (24.935930)	14.979405* (24.306630)
	60	28.589289 (45.723200)	13.270930 (22.041700)	13.216690* (21.841981)
	70	26.881754 (43.215810)	12.759061 (20.717310)	12.696708* (20.656890)
	80	25.776429 (42.810745)	12.115988 (20.144350)	12.068466* (20.022289)
	90	24.715040 (40.854487)	11.813551 (19.950290)	11.770646* (19.858750)
	100	23.024985 (38.953814)	10.965689 (18.143080)	10.933794* (18.088728)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.2 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.15,0.05	20	36.193146 (58.120723)	18.804851 (31.919768)	18.550257* (31.415602)
	30	32.436867 (53.282972)	15.928699 (25.989368)	15.856561* (25.951147)
	40	29.403440 (47.950390)	14.005940 (23.375420)	13.915820* (23.179250)
	50	27.520547 (44.804130)	12.854200 (21.557590)	12.778140* (21.345570)
	60	25.878532 (42.797220)	12.091100 (20.190700)	12.014440* (20.155680)
	70	23.771353 (38.799678)	10.870661 (17.500840)	10.840975* (17.304060)
	80	21.555562 (35.675338)	9.493821 (15.845680)	9.436042* (15.610610)
	90	20.553079 (34.887053)	8.771839 (14.537090)	8.730637* (14.317820)
	100	19.335482 (32.657313)	8.229220 (13.229040)	8.183140* (13.127090)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.2 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.15,0.10	20	37.213581 (59.153367)	19.374416 (32.214419)	19.163012* (32.044861)
	30	33.575283 (55.216717)	16.472695 (26.981835)	16.389467* (26.824156)
	40	31.425476 (51.950830)	15.184830 (26.136780)	15.101060* (26.035770)
	50	29.719867 (48.337290)	14.816390 (25.296970)	14.739079* (25.083980)
	60	27.470036 (44.584927)	13.693668 (23.188120)	13.628117* (23.130820)
	70	24.792009 (41.371630)	11.263021 (19.243790)	11.204843* (19.022160)
	80	23.574959 (38.506807)	10.620637 (17.444860)	10.573905* (17.427470)
	90	22.567240 (36.841670)	10.150605 (16.912460)	10.111933* (16.873660)
	100	21.149598 (35.129773)	9.228179 (15.166415)	9.193074* (15.017122)

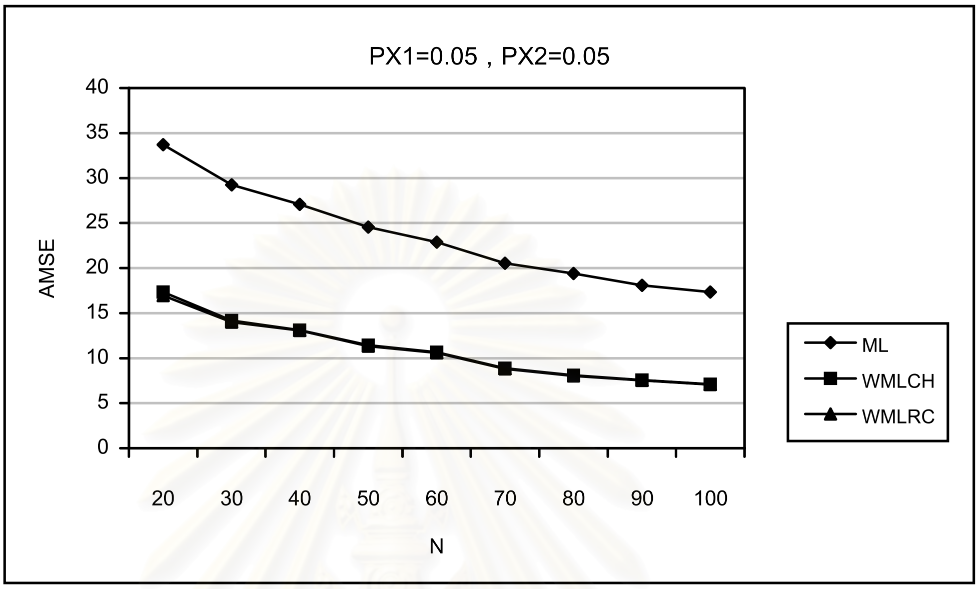
* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

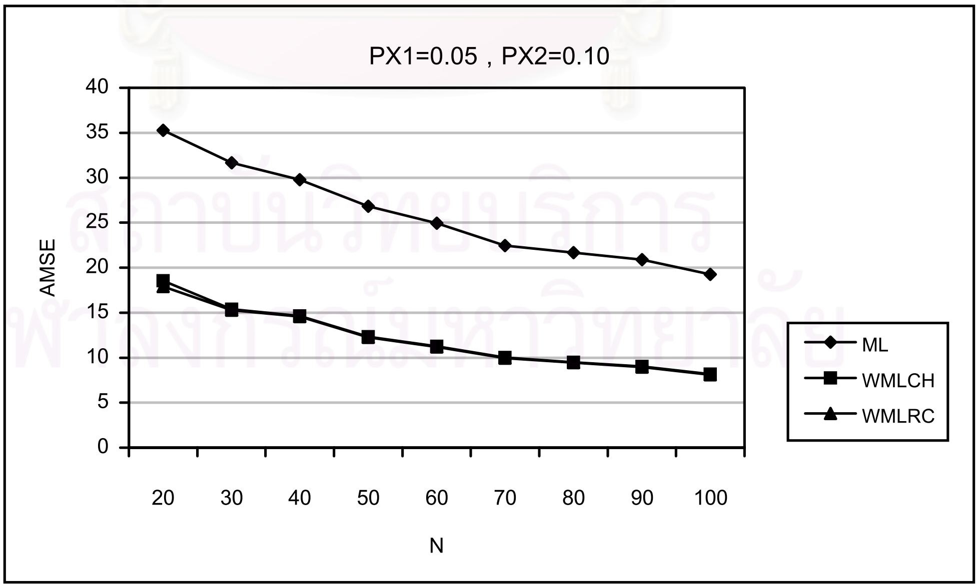
ตารางที่4.3.2 **(ต่อ)**

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.15,0.15	20	39.712926 (63.319569)	21.342461 (35.483157)	20.279856* (35.119725)
	30	37.042731 (58.983851)	18.945687 (32.494215)	18.856388* (32.352728)
	40	34.469193 (56.712768)	16.565625 (27.174961)	16.482869* (27.086797)
	50	32.398655 (53.016463)	15.270340 (26.150180)	15.183550* (26.016450)
	60	30.412638 (50.836590)	14.191790 (24.973030)	14.139732* (24.798420)
	70	29.221550 (47.616881)	13.409062 (22.717310)	13.349729* (22.605660)
	80	27.617046 (44.963912)	12.761630 (21.466090)	12.724002* (21.321070)
	90	25.749314 (42.516816)	12.154030 (20.628690)	12.114766* (20.519410)
	100	24.336848 (41.471609)	11.542077 (19.640280)	11.516719* (19.538990)

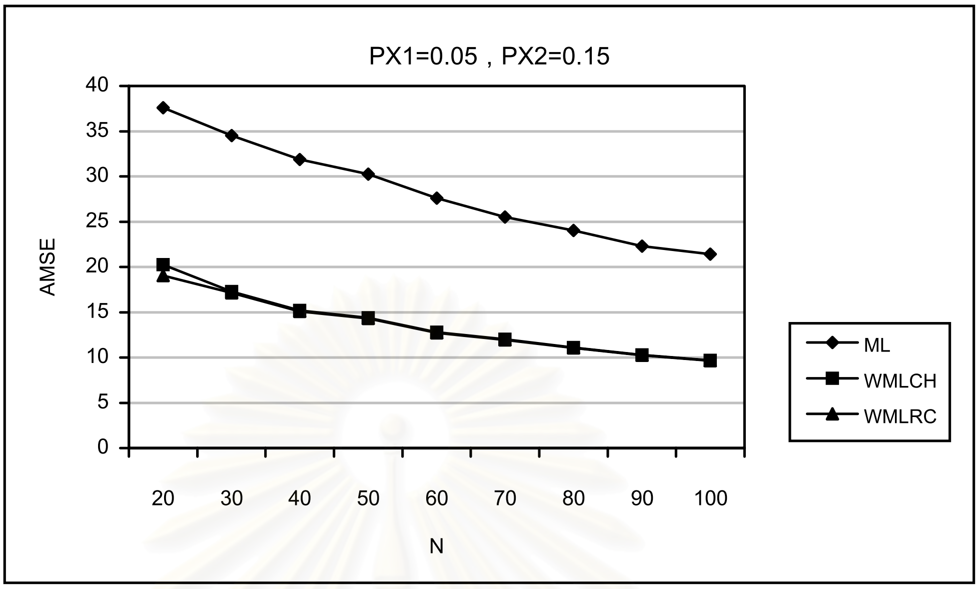
* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

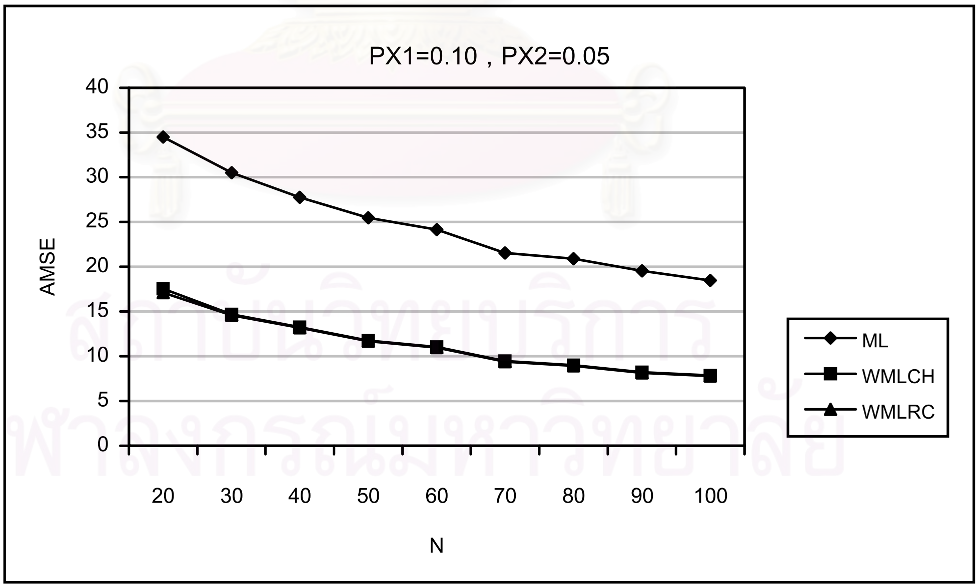
รูปที่4.3.2 แสดงการเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ ไม่รุนแรงและตัวแปรอิสระ x_2 มีค่าผิดปกติระดับ รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2



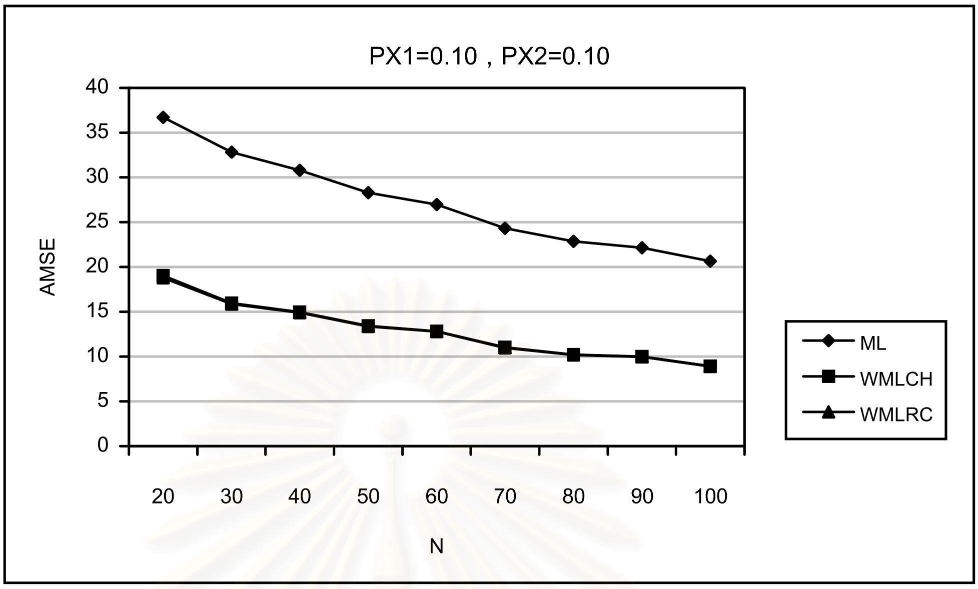


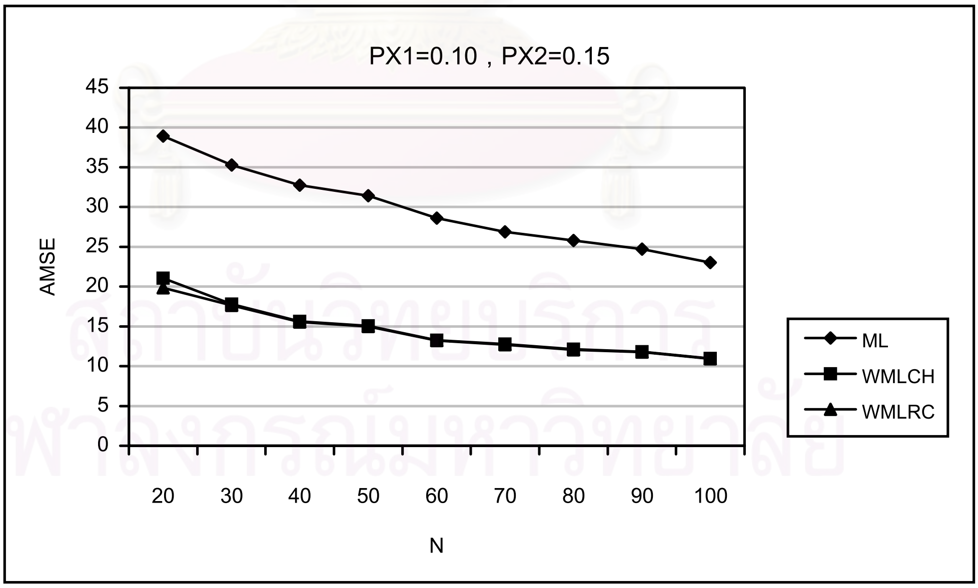
รูปที่4.3.2 (**ต่อ**)


PX1=0.05 , PX2=0.15
AMSE
N
0
5
10
15
20
25
30
35
40
20
30
40
50
60
70
80
90
100
ML
WMLCH
WMLRC

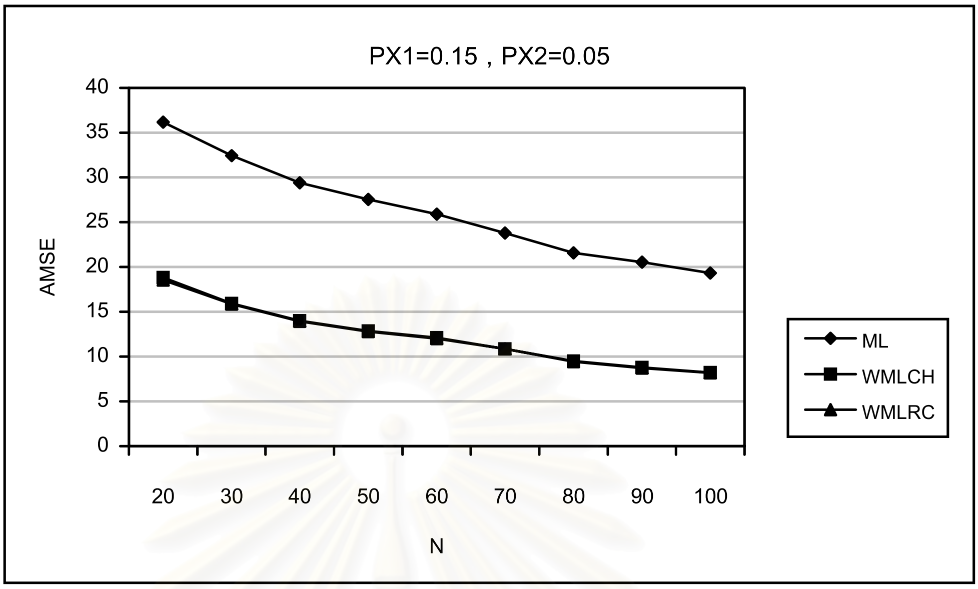

PX1=0.10 , PX2=0.05
AMSE
N
0
5
10
15
20
25
30
35
40
20
30
40
50
60
70
80
90
100
ML
WMLCH
WMLRC

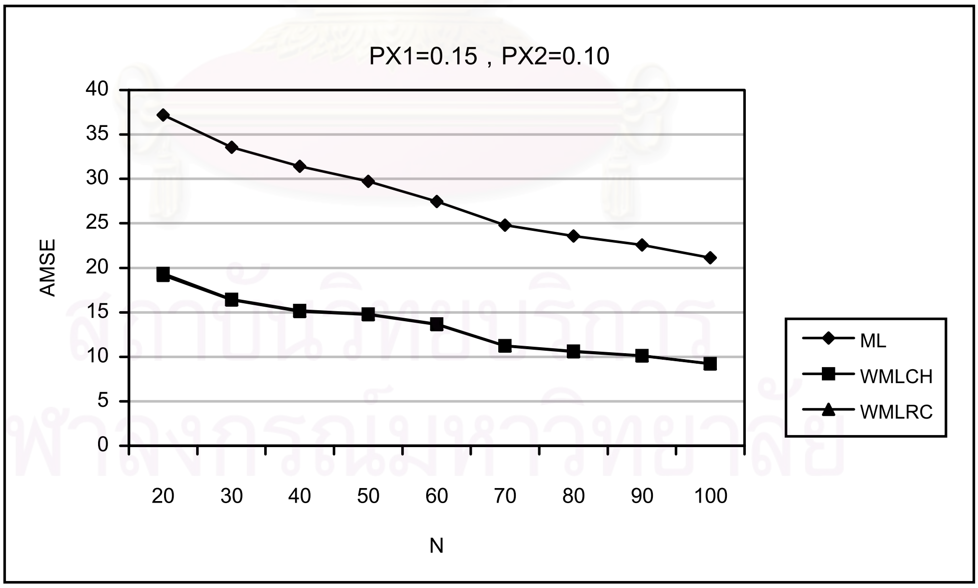
รูปที่4.3.2 (ต่อ)


PX1=0.10 , PX2=0.10
AMSE
N
20
30
40
50
60
70
80
90
100
ML
WMLCH
WMLRC

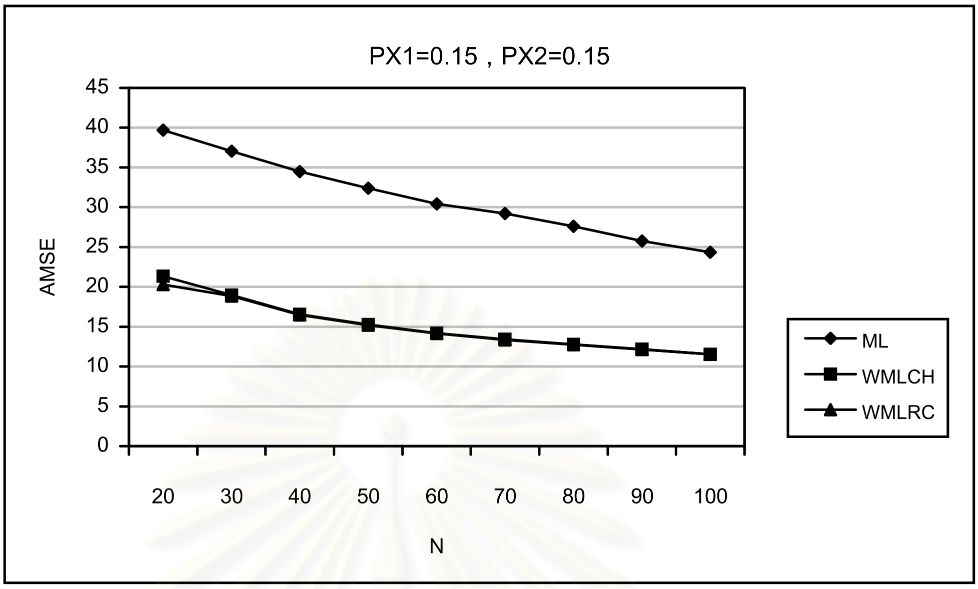

PX1=0.10 , PX2=0.15
AMSE
N
20
30
40
50
60
70
80
90
100
ML
WMLCH
WMLRC

รูปที่4.3.2 (ต่อ)


PX1=0.15 , PX2=0.05
AMSE
40
35
30
25
20
15
10
5
0
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC


PX1=0.15 , PX2=0.10
AMSE
40
35
30
25
20
15
10
5
0
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC

รูปที่4.3.2 (**ต่อ**)


PX1=0.15 , PX2=0.15
AMSE
N
ML
WMLCH
WMLRC
0
5
10
15
20
25
30
35
40
45
20
30
40
50
60
70
80
90
100

จากตารางที่4.3.2 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับไม่รุนแรง ซึ่งมีสัดส่วนการปลอมปนเท่ากับ 0.05 , 0.10 และ 0.15 และตัวแปรอิสระ x_2 มีค่าผิดปกติระดับรุนแรง ซึ่งมีสัดส่วนการปลอมปนเท่ากับ 0.05 , 0.10 และ 0.15 สามารถสรุปผลได้ดังนี้

เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับไม่รุนแรงและตัวแปรอิสระ x_2 มีค่าผิดปกติระดับรุนแรง ในทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1 = 0.05 , 0.10 และ 0.15) และทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_2 (PX2 = 0.05 , 0.10 และ 0.15) โดยในทุก ๆ ขนาดตัวอย่างคือ N = 20,30,40,50,60,70,80,90 และ 100 วิธี WMLRC ให้ค่า AMSE น้อยที่สุด รองลงมาคือ วิธี WMLCH และวิธี ML ตามลำดับ และจะพบว่า วิธี WMLCH และวิธี WMLRC ให้ค่า AMSE ใกล้เคียงกันมากในทุกขนาดตัวอย่าง โดยวิธี ML ให้ค่า AMSEสูงในทุกขนาดตัวอย่าง

ข้อสรุป

จากตารางที่4.3.2 ทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 และทุกขนาดตัวอย่างวิธี WMLRC ให้ค่า AMSE ต่ำกว่าวิธีอื่น ๆ โดยที่วิธี ML จะเป็นวิธีที่ให้ค่า AMSE สูงที่สุด และจะพบว่าในทุก ๆ ขนาดตัวอย่าง วิธี WMLCH และวิธี WMLRC ให้ค่า AMSE ใกล้เคียงกันมาก โดยวิธี ML ให้ค่า AMSE สูงในทุกขนาดตัวอย่าง

ค่า AMSE ของทุกวิธีมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น ขณะที่สัดส่วนการปลอมปนคงที่

ค่า AMSE ของทุกวิธีมีแนวโน้มเพิ่มขึ้นเมื่อสัดส่วนการปลอมปนเพิ่มขึ้น ขณะที่ขนาดตัวอย่างคงที่

4.3.3 ตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับรุนแรง(C=12) และตัวแปรอิสระ x_2 มีระดับค่าผิดปกติระดับไม่รุนแรง(C=5) โดยแต่ละระดับทำการศึกษาที่สัดส่วนการปลอมปนเป็น 0.05 , 0.10 และ 0.15 ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 4.3.3 และกราฟรูปที่ 4.3.3

ตารางที่4.3.3 การเปรียบเทียบค่าประมาณของพารามิเตอร์จากวิธี ML วิธี WMLCH และวิธี WMLRC ด้วยค่า AMSE เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ รุนแรง และตัวแปรอิสระ x_2 มีค่าผิดปกติระดับ ไม่รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.05	20	33.951176 (55.203616)	16.898509 (27.063391)	16.389928* (26.862519)
	30	29.718050 (48.421910)	13.436000 (22.452380)	13.377190* (22.244980)
	40	27.899120 (44.904140)	12.181520 (20.175820)	12.091259* (20.084330)
	50	25.573810 (42.116270)	11.094850 (18.231310)	11.007117* (18.092030)
	60	23.627260 (38.273050)	10.399650 (17.807560)	10.328580* (17.724860)
	70	21.150830 (35.397720)	8.969489 (14.793100)	8.890897* (14.776970)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.3 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.05	80	19.179690 (32.044650)	7.827116 (12.877460)	7.773952* (12.666270)
	90	18.122570 (31.048690)	7.322342 (11.845280)	7.273575* (11.758662)
	100	16.986340 (27.485760)	6.800396 (10.938455)	6.757505* (10.814958)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.3 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.10	20	36.188958 (58.178215)	17.963373 (27.116993)	17.575732* (27.012751)
	30	32.752166 (53.937262)	14.996382 (24.876401)	14.892759* (24.816797)
	40	30.396117 (49.413585)	14.097758 (22.933891)	14.016525* (22.859125)
	50	27.830550 (44.801780)	12.973530 (21.445790)	12.896610* (21.284240)
	60	26.420170 (43.106430)	12.303060 (20.249960)	12.232920* (20.186910)
	70	24.256760 (39.515180)	10.515014 (17.407050)	10.448519* (17.236020)
	80	21.704030 (35.805400)	9.179101 (14.934240)	9.123462* (14.837660)
	90	19.832910 (32.626740)	8.434541 (13.952050)	8.387762* (13.832320)
	100	18.461670 (31.592020)	7.758979 (12.706220)	7.706951* (12.595440)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.3 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.15	20	38.097785 (61.614503)	19.153176 (31.915738)	18.749367* (31.258431)
	30	34.485590 (56.298440)	16.362230 (26.815680)	16.282680* (26.627050)
	40	33.393010 (54.856680)	15.591030 (25.999170)	15.553240* (25.865780)
	50	29.934960 (48.284040)	13.688810 (22.970960)	13.641379* (22.906060)
	60	28.490650 (45.377180)	12.921159 (21.542380)	12.844219* (21.496180)
	70	25.646972 (42.415870)	11.097610 (18.330750)	11.028563* (18.289760)
	80	23.379950 (38.187790)	10.114220 (16.996580)	10.053457* (16.810430)
	90	21.038470 (35.047200)	9.189035 (15.137060)	9.131823* (15.068852)
	100	19.807960 (32.561800)	8.714109 (14.232450)	8.670077* (14.131530)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.3 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.10,0.05	20	36.191989 (58.308115)	18.453936 (27.836228)	18.128642* (27.239326)
	30	32.101320 (53.465430)	15.228720 (25.144010)	15.169260* (25.023270)
	40	30.462530 (49.242630)	14.518900 (24.298090)	14.428700* (24.229800)
	50	28.195860 (45.226800)	13.597260 (23.385270)	13.583290* (23.177310)
	60	26.468670 (43.277210)	12.761983 (20.887740)	12.692349* (20.838290)
	70	23.244570 (38.479230)	10.814805 (17.994840)	10.751369* (17.892260)
	80	21.216660 (35.617080)	9.475404 (15.688580)	9.411906* (15.507250)
	90	19.992420 (32.933370)	8.691350 (14.068140)	8.641492* (14.000310)
	100	18.820430 (31.894510)	8.041220 (13.353910)	8.000001* (13.268510)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.3 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.10,0.10	20	37.656251 (60.608499)	19.595857 (32.596761)	19.217643* (32.174911)
	30	34.013941 (56.783517)	16.548612 (26.975218)	16.462861* (26.927385)
	40	31.637560 (50.419480)	15.384010 (25.459790)	15.293880* (25.297660)
	50	30.252480 (49.116040)	14.854850 (24.786350)	14.768489* (24.686950)
	60	28.763360 (45.509240)	13.908040 (23.207940)	13.847915* (23.159570)
	70	26.273110 (42.911300)	12.042510 (20.170560)	11.982255* (20.117280)
	80	23.602450 (38.702780)	10.749629 (17.729830)	10.690463* (17.664890)
	90	22.091910 (36.606940)	9.970036 (16.366010)	9.930764* (16.351780)
	100	20.471360 (34.083020)	8.759811 (14.177330)	8.712969* (14.108670

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.3 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.10,0.15	20	38.498032 (61.286227)	19.952587 (33.369527)	19.729656* (32.912085)
	30	35.319430 (58.515570)	17.548960 (29.232340)	17.476420* (28.940940)
	40	33.573840 (55.906720)	16.514820 (26.956200)	16.421360* (26.892280)
	50	31.494070 (50.120930)	15.587290 (25.709340)	15.505570* (25.588160)
	60	29.455490 (47.831970)	14.242710 (24.263640)	14.172301* (24.183380)
	70	26.902160 (43.695350)	12.272620 (20.857680)	12.214493* (20.765730)
	80	24.927040 (41.742380)	11.625794 (19.702560)	11.576034* (19.564790)
	90	23.602130 (38.688010)	10.818154 (17.991440)	10.779481* (17.781930)
	100	22.161310 (36.961870)	10.059208 (16.835030)	10.021130* (16.615260)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.3 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.15,0.05	20	38.436651 (62.277231)	20.091411 (32.946376)	19.897781* (32.745531)
	30	35.960960 (57.897600)	18.737250 (31.456800)	18.645040* (31.303410)
	40	33.535880 (54.253870)	16.985870 (27.109010)	16.892440* (27.078750)
	50	30.900220 (49.096610)	15.049908 (24.799660)	14.966682* (24.650300)
	60	28.310920 (45.247840)	13.286713 (22.697420)	13.213967* (22.402750)
	70	25.586830 (42.276290)	11.378674 (18.878820)	11.305830* (18.748870)
	80	24.256060 (40.682930)	10.829587 (17.251360)	10.775068* (17.166720)
	90	22.215980 (36.712600)	9.920142 (15.935510)	9.887481* (15.891810)
	100	21.053380 (35.177460)	9.267895 (15.437850)	9.232297* (15.359300)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.3 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.15,0.10	20	39.418822 (63.233423)	20.352257 (34.401381)	20.137892* (34.114888)
	30	36.468562 (58.324765)	18.950447 (31.950318)	18.865567* (31.782862)
	40	34.532790 (57.875090)	17.615860 (29.686850)	17.536500* (29.597850)
	50	31.919270 (52.161210)	15.798937 (25.948580)	15.727804* (25.874810)
	60	30.716350 (48.591080)	14.941613 (24.661170)	14.887726* (24.567970)
	70	28.683640 (45.783400)	13.380683 (21.841060)	13.328018* (21.736110)
	80	26.645120 (43.175750)	12.593234 (21.156820)	12.547846* (21.155550)
	90	25.301500 (42.055980)	11.924903 (19.153190)	11.880341* (19.050980)
	100	23.569010 (39.379680)	11.318500 (18.713730)	11.280632* (18.637140)

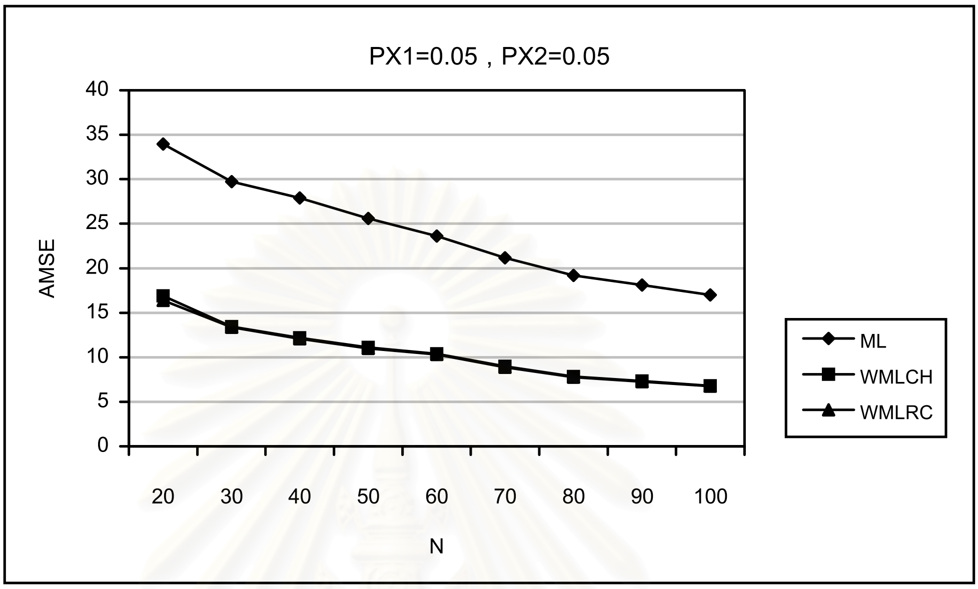
* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

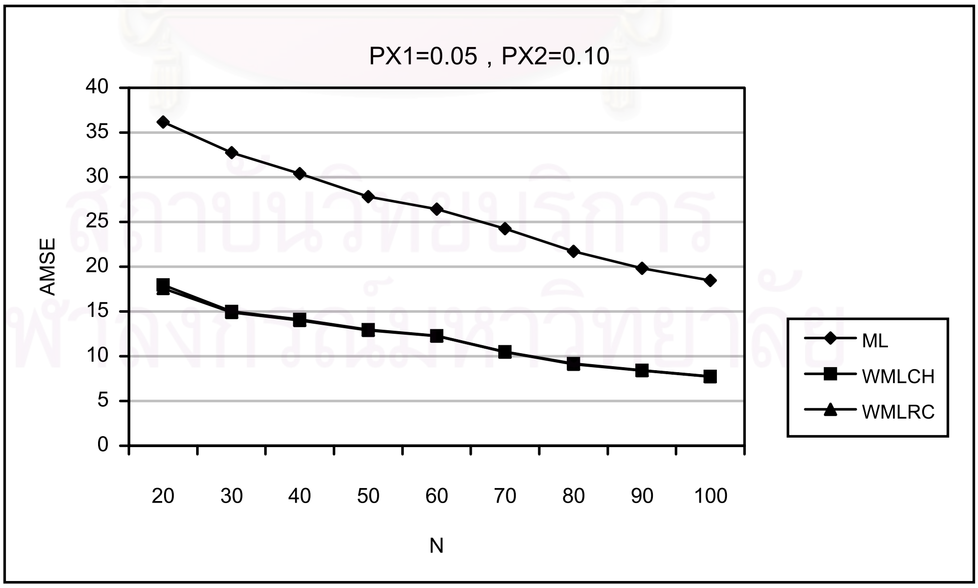
ตารางที่4.3.3 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.15,0.15	20	40.175238 (65.796601)	20.679085 (35.387953)	20.592245* (35.277085)
	30	37.517986 (59.368817)	19.339694 (32.986854)	19.241621* (32.834747)
	40	35.273460 (58.419680)	17.962300 (30.096390)	17.880676* (29.935700)
	50	33.526990 (55.926370)	16.677840 (27.176330)	16.599771* (27.029200)
	60	31.586560 (51.547220)	15.471491 (25.508090)	15.400038* (25.413280)
	70	30.884750 (48.858950)	15.092181 (25.043430)	15.037670* (24.988450)
	80	28.513320 (45.323690)	13.668134 (23.149340)	13.625289* (23.119900)
	90	26.745120 (43.381680)	12.997667 (21.704395)	12.955791* (21.634820)
	100	25.215510 (41.884360)	12.187458 (20.728600)	12.154246* (20.668130)

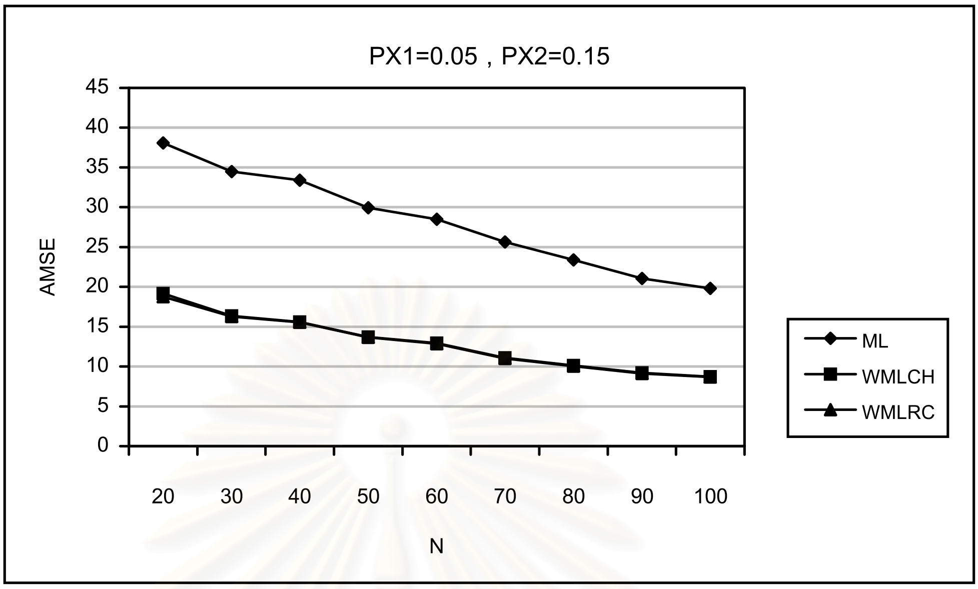
* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

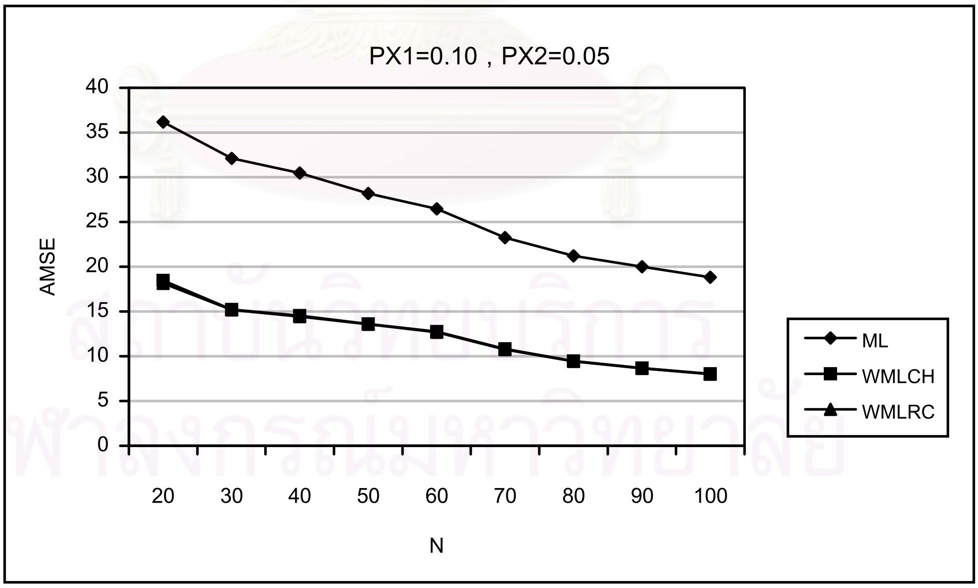
รูปที่4.3.3 แสดงการเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ รุนแรงและตัวแปรอิสระ x_2 มีค่าผิดปกติระดับ ไม่รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2



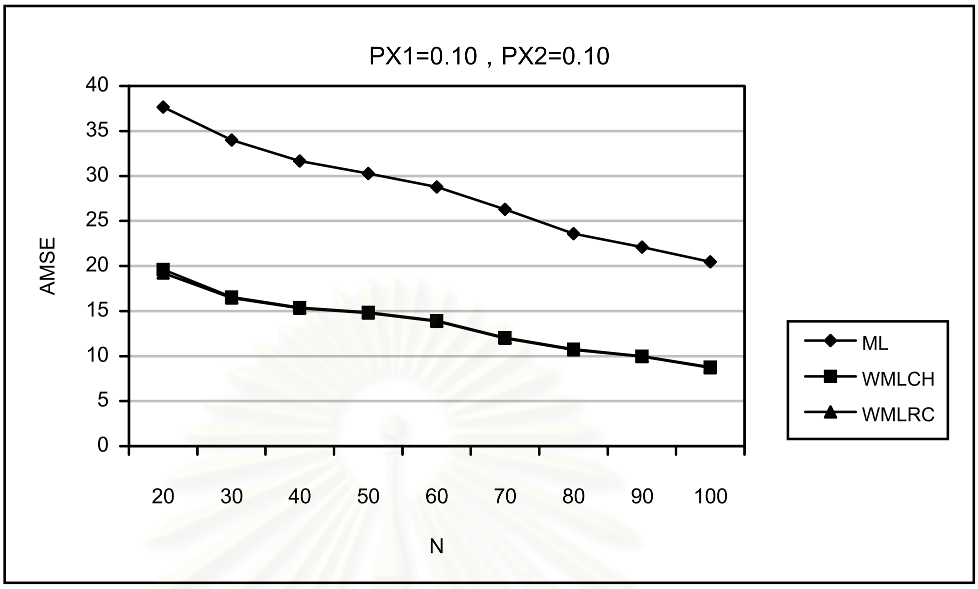


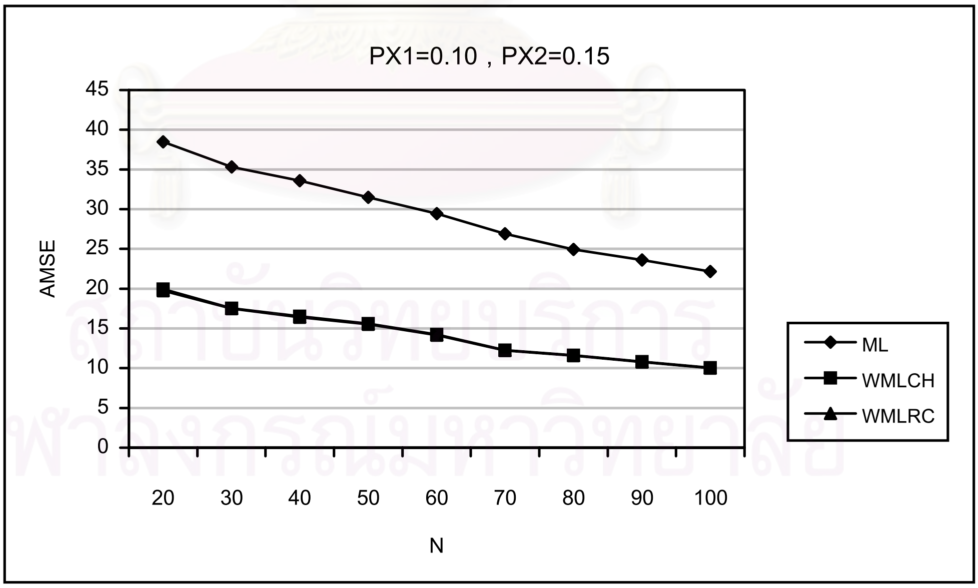
รูปที่4.3.3 (ต่อ)


PX1=0.05 , PX2=0.15
45
40
35
30
25
20
15
10
5
0
AMSE
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC

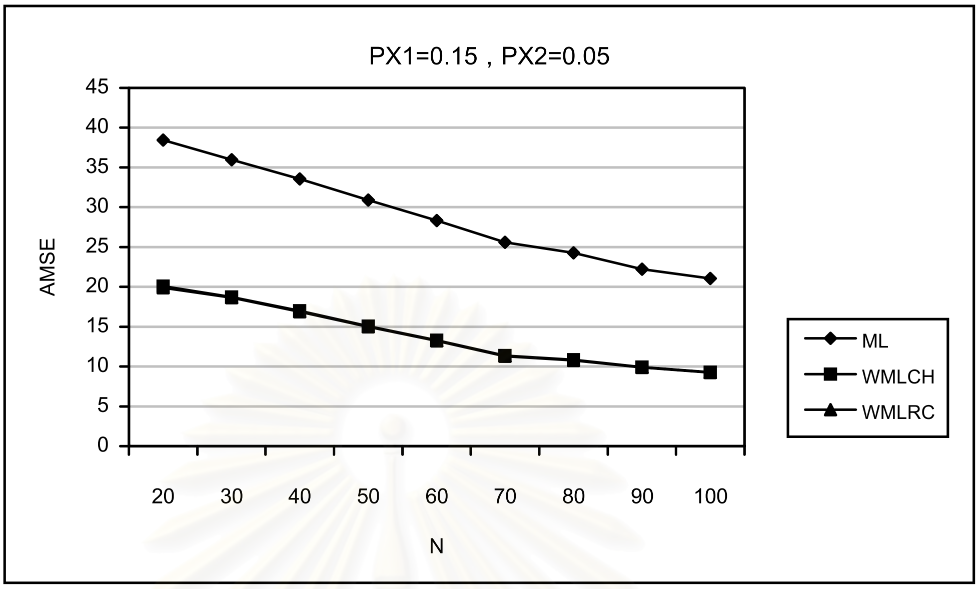

PX1=0.10 , PX2=0.05
40
35
30
25
20
15
10
5
0
AMSE
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC

รูปที่4.3.3 (ต่อ)

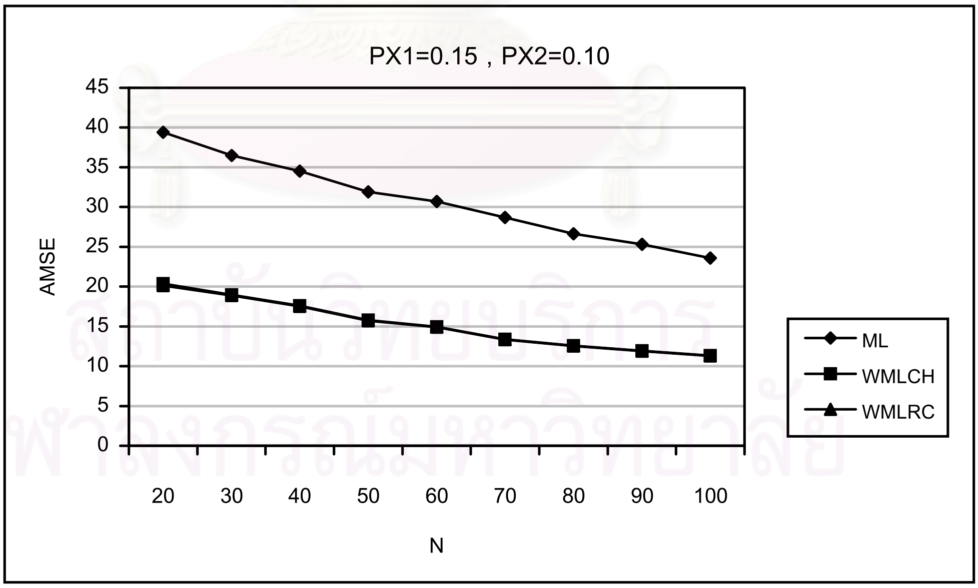

PX1=0.10 , PX2=0.10
AMSE
N
40
35
30
25
20
15
10
5
0
20
30
40
50
60
70
80
90
100
ML
WMLCH
WMLRC


PX1=0.10 , PX2=0.15
AMSE
N
45
40
35
30
25
20
15
10
5
0
20
30
40
50
60
70
80
90
100
ML
WMLCH
WMLRC

รูปที่4.3.3 (**ต่อ**)

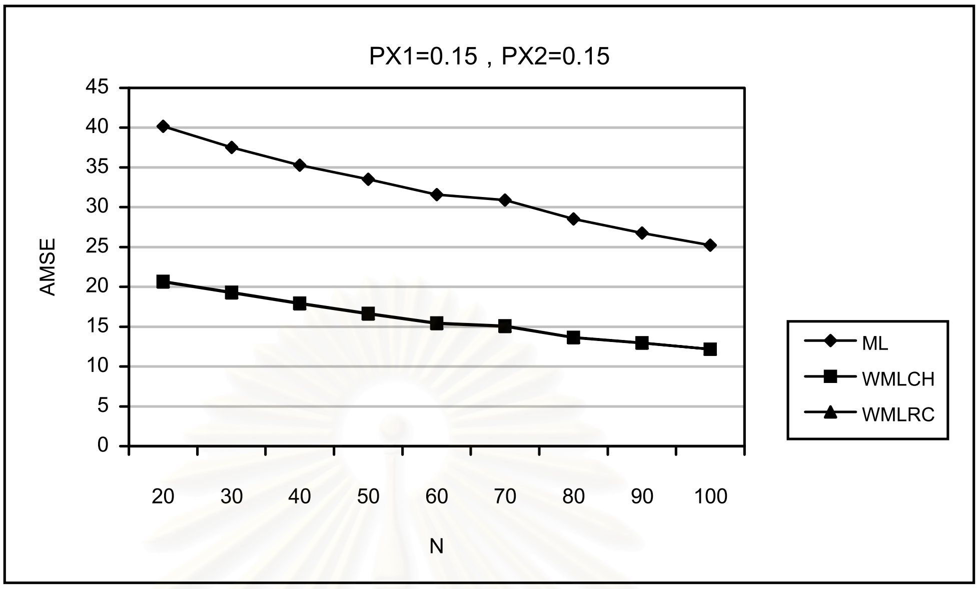


PX1=0.15 , PX2=0.05
AMSE
45
40
35
30
25
20
15
10
5
0
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC



PX1=0.15 , PX2=0.10
AMSE
45
40
35
30
25
20
15
10
5
0
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC

รูปที่4.3.3 (ต่อ)


PX1=0.15 , PX2=0.15
45
40
35
30
25
20
15
10
5
0
AMSE
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC

จากตารางที่4.3.3 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับรุนแรง ซึ่งมีสัดส่วนการปลอมปน เท่ากับ 0.05 , 0.10 และ 0.15 และตัวแปรอิสระ x_2 มีค่าผิดปกติระดับไม่รุนแรง ซึ่งมีสัดส่วนการปลอมปน เท่ากับ 0.05 , 0.10 และ 0.15 สามารถสรุปผลได้ดังนี้

เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับรุนแรงและตัวแปรอิสระ x_2 มีค่าผิดปกติระดับไม่รุนแรง ในทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1 = 0.05 , 0.10 และ 0.15) และทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_2 (PX2 = 0.05 , 0.10 และ 0.15) โดยในทุก ๆ ขนาดตัวอย่างคือ N = 20,30,40,50,60,70,80,90 และ 100 วิธี WMLRC ให้ค่า AMSE น้อยที่สุด รองลงมาคือ วิธี WMLCH และวิธี ML ตามลำดับ และจะพบว่า วิธี WMLCH และวิธี WMLRC ให้ค่า AMSE ใกล้เคียงกันมากในทุกขนาดตัวอย่าง โดยวิธี ML ให้ค่า AMSEสูงในทุกขนาดตัวอย่าง

ข้อสรุป

จากตารางที่4.3.3 ทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 และทุกขนาดตัวอย่างวิธี WMLRC ให้ค่า AMSE ต่ำกว่าวิธีอื่น ๆ โดยที่วิธี ML จะเป็นวิธีที่ให้ค่า AMSE สูงที่สุด และจะพบว่าในทุก ๆ ขนาดตัวอย่าง วิธี WMLCH และวิธี WMLRC ให้ค่า AMSE ใกล้เคียงกันมาก โดยวิธี ML ให้ค่า AMSE สูงในทุกขนาดตัวอย่าง

ค่า AMSE ของทุกวิธีมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น ขณะที่สัดส่วนการปลอมปนคงที่

ค่า AMSE ของทุกวิธีมีแนวโน้มเพิ่มขึ้นเมื่อสัดส่วนการปลอมปนเพิ่มขึ้น ขณะที่ขนาดตัวอย่างคงที่

4.3.4 ตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับรุนแรง(C=12) และตัวแปรอิสระ x_2 มีระดับค่าผิดปกติระดับรุนแรง(C=12) โดยแต่ละระดับทำการศึกษาที่สัดส่วนการปลอมปนเป็น 0.05 , 0.10 และ 0.15 ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 4.3.4 และกราฟรูปที่ 4.3.4

ตารางที่4.3.4 การเปรียบเทียบค่าประมาณของพารามิเตอร์จากวิธี ML วิธี WMLCH และวิธี WMLRC ด้วยค่า AMSE เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ รุนแรง และตัวแปรอิสระ x_2 มีค่าผิดปกติระดับ รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.05	20	75.151839 (119.145812)	17.867299 (30.187082)	17.525671* (29.639225)
	30	64.416991 (108.263125)	15.976828 (25.874792)	15.825765* (25.689637)
	40	55.912260 (98.265100)	14.790600 (24.501380)	14.706463* (24.440580)
	50	47.691615 (77.971020)	12.969910 (21.414600)	12.886259* (21.224110)
	60	37.038040 (59.485760)	12.493899 (20.400750)	12.443882* (20.329430)
	70	32.573186 (53.750890)	11.706535 (19.301900)	11.669538* (19.258850)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.4 **(ต่อ)**

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.05	80	30.409190 (49.630550)	11.565139 (18.905940)	11.527858* (18.847310)
	90	28.099841 (45.193980)	9.973230 (16.375580)	9.941752* (16.277080)
	100	27.141284 (44.493690)	9.447862 (15.650230)	9.413485* (15.541140)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.4 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.10	20	77.572310 (121.482711)	18.886851 (31.125087)	18.246996* (30.656586)
	30	66.339830 (112.444780)	16.709620 (27.924330)	16.615930* (27.839410)
	40	58.763831 (100.289140)	16.021880 (26.003490)	15.938416* (25.948300)
	50	51.721985 (81.027830)	15.405980 (25.225630)	15.333728* (25.143570)
	60	39.385701 (63.657570)	14.184805 (23.880650)	14.133743* (23.821760)
	70	34.789041 (56.624210)	13.273802 (22.981260)	13.234574* (22.851234)
	80	31.798454 (51.044090)	11.996054 (19.654100)	11.961464* (19.595404)
	90	30.700502 (49.970450)	11.553099 (18.872090)	11.520010* (18.842527)
	100	29.371626 (47.855710)	10.790913 (17.929780)	10.760518* (17.913608)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.4 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.05,0.15	20	80.873216 (124.219713)	20.754951 (32.785471)	20.121268* (32.548689)
	30	69.718836 (115.629295)	18.933187 (31.294966)	18.878651* (31.115173)
	40	60.286929 (102.867640)	17.001600 (29.180660)	16.926613* (29.046320)
	50	54.505583 (92.517570)	16.374160 (26.738160)	16.301014* (26.679470)
	60	40.597523 (64.131190)	15.796313 (25.820060)	15.745548* (25.765880)
	70	36.507582 (59.104490)	14.493518 (24.101020)	14.456444* (24.089910)
	80	34.609086 (57.068870)	13.235862 (22.865780)	13.201778* (22.806600)
	90	32.610142 (54.557960)	12.782133 (21.050940)	12.765844* (21.025092)
	100	31.285374 (50.317874)	12.076220 (20.116990)	12.050798* (20.085194)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.4 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.10,0.05	20	78.532991 (122.924714)	19.157233 (31.456425)	18.942021* (31.185624)
	30	67.142020 (114.685620)	17.423050 (29.514510)	17.383680* (29.480130)
	40	59.059860 (101.192310)	16.121580 (26.205610)	16.039340* (26.121670)
	50	50.960998 (80.132960)	15.452169 (25.380740)	15.392550* (25.279180)
	60	40.175240 (63.957200)	14.349910 (24.082030)	14.309506* (24.026940)
	70	35.526170 (57.201450)	13.807029 (23.417830)	13.761591* (23.388710)
	80	32.291350 (51.967910)	12.150399 (20.294930)	12.123690* (20.271800)
	90	30.601590 (49.832210)	11.471655 (18.711920)	11.449066* (18.668280)
	100	29.142626 (47.759710)	10.351870 (16.995930)	10.328123* (16.989180)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.4 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.10,0.10	20	80.014664 (123.659248)	19.979135 (32.237988)	19.865328* (32.156391)
	30	69.542622 (115.259599)	18.242589 (30.655892)	18.152743* (30.601874)
	40	60.415121 (103.484790)	17.017100 (29.354750)	16.934768* (29.048740)
	50	53.992073 (91.219870)	16.286247 (26.213670)	16.223910* (26.103560)
	60	41.603236 (65.980080)	15.409929 (25.296330)	15.352713* (25.215270)
	70	36.923437 (59.988470)	14.504396 (24.128300)	14.450207* (24.004980)
	80	34.123124 (56.783330)	13.304990 (22.944300)	13.268850* (22.892620)
	90	32.840240 (54.907890)	12.998113 (21.427900)	12.969278* (21.312618)
	100	31.826202 (50.561880)	12.551561 (20.965971)	12.538253* (20.941995)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.4 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.10,0.15	20	82.178488 (126.101108)	21.479857 (32.635176)	21.323122* (32.523237)
	30	71.241751 (117.725539)	19.147278 (31.968543)	19.051655* (31.909562)
	40	60.940436 (104.749220)	18.175880 (30.461900)	18.118581* (30.410230)
	50	55.012310 (92.695890)	16.724710 (27.391640)	16.668448* (27.308380)
	60	42.563684 (67.917840)	15.957321 (25.701560)	15.909941* (25.532160)
	70	37.821317 (60.477280)	15.306160 (25.156250)	15.264636* (25.116948)
	80	36.197822 (58.868890)	14.501316 (24.124690)	14.473902* (24.021270)
	90	34.109809 (56.235220)	13.901325 (23.611360)	13.885011* (23.572330)
	100	32.741671 (54.752790)	13.213323 (22.871950)	13.194370* (22.797370)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.4 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	**ขนาดตัวอย่าง** N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.15,0.05	20	80.927986 (124.795571)	19.327969 (31.719158)	19.312645* (31.697546)
	30	70.524432 (116.213565)	18.011899 (30.047322)	17.937638* (29.998512)
	40	60.362730 (102.924570)	17.185910 (29.334170)	17.104000* (29.269960)
	50	55.417880 (93.298990)	16.794612 (27.657240)	16.729142* (27.589850)
	60	41.955020 (67.572830)	16.032393 (26.337250)	15.982495* (26.271380)
	70	37.940600 (60.514160)	15.631191 (25.628360)	15.599602* (25.594150)
	80	35.001350 (57.926870)	13.978411 (23.559220)	13.947904* (23.517850)
	90	33.744730 (55.498030)	13.322671 (22.933600)	13.303612* (22.908090)
	100	32.217988 (51.664520)	12.740505 (20.985510)	12.728718* (20.975860)

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่4.3.4 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.15,0.10	20	82.974715 (127.024665)	20.631164 (32.986132)	20.529639* (32.877543)
	30	72.636203 (118.531460)	19.539670 (32.032010)	19.440650* (32.011090)
	40	61.966060 (105.597670)	17.791440 (30.150440)	17.738980* (30.118000)
	50	56.813260 (94.292390)	17.352610 (29.860130)	17.295820* (29.839940)
	60	43.470737 (68.592420)	16.708163 (27.361230)	16.662767* (27.306580)
	70	38.981551 (61.103220)	16.197230 (26.211220)	16.158677* (26.171690)
	80	36.492888 (59.224760)	14.576072 (24.183260)	14.554529* (24.176700)
	90	34.502894 (56.435630)	13.962001 (23.720690)	13.947277* (23.656740)
	100	32.833351 (54.997630)	13.018727 (22.409880)	13.001815* (22.312960)

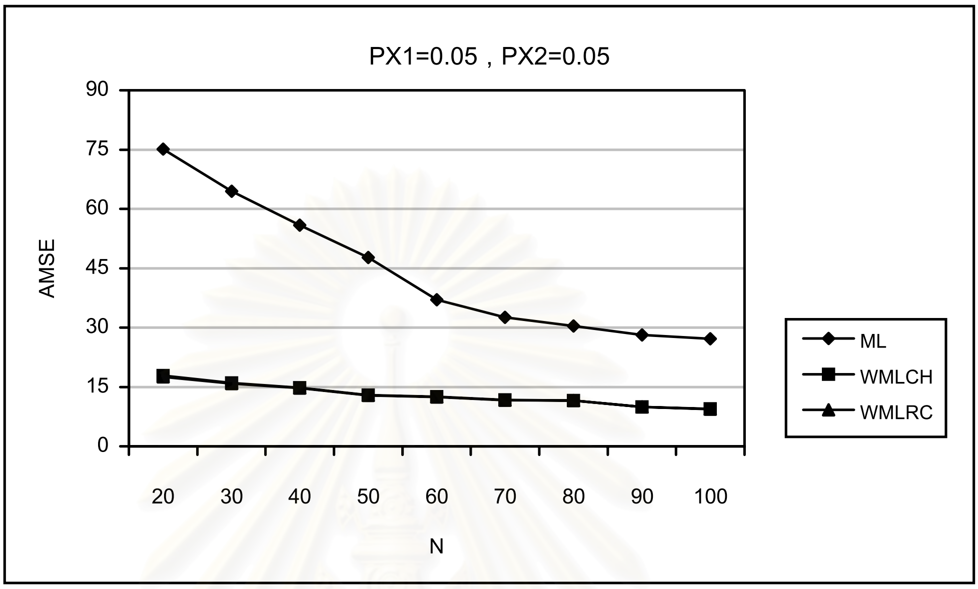
* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

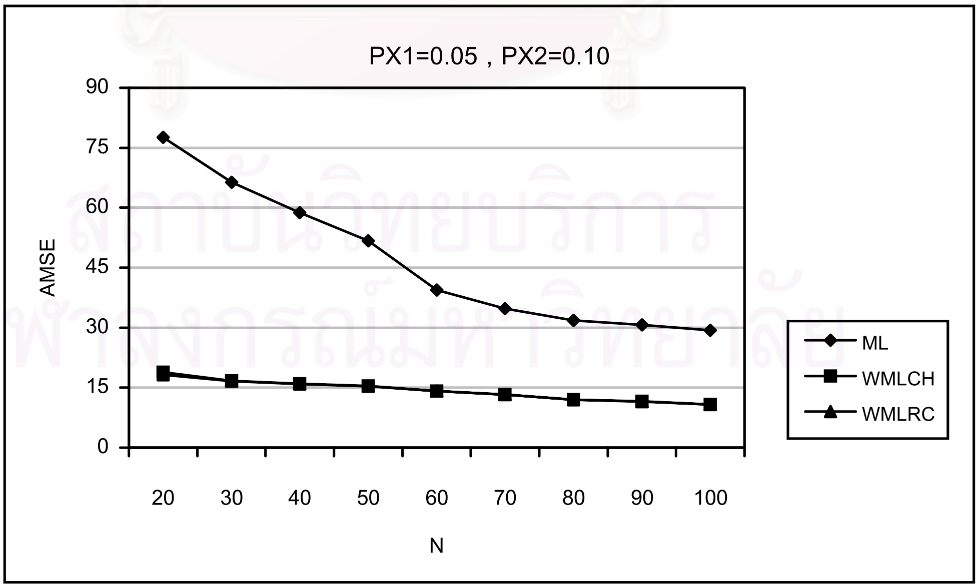
ตารางที่4.3.4 (ต่อ)

สัดส่วนการปลอมปน PX1,PX2	ขนาดตัวอย่าง N	AMSE		
		ML (s.d.)	WMLCH (s.d.)	WMLRC (s.d.)
0.15,0.15	20	85.126735 (129.175978)	22.225961 (34.817845)	22.131863* (34.751632)
	30	73.952461 (120.771452)	20.457978 (32.617941)	20.374675* (32.576636)
	40	62.825691 (106.930680)	18.236390 (31.126380)	18.171044* (31.033570)
	50	57.341280 (95.803170)	17.695376 (29.977700)	17.632025* (29.930100)
	60	44.221500 (71.244870)	17.139705 (29.297640)	17.081049* (29.213250)
	70	39.427640 (62.097980)	16.558706 (27.133570)	16.522559* (27.100362)
	80	37.699504 (60.344530)	15.479913 (26.481730)	15.442180* (26.422940)
	90	35.372912 (57.362690)	14.528511 (24.096400)	14.509410* (24.028650)
	100	33.951192 (56.088110)	13.811262 (23.551036)	13.805672* (23.519638)

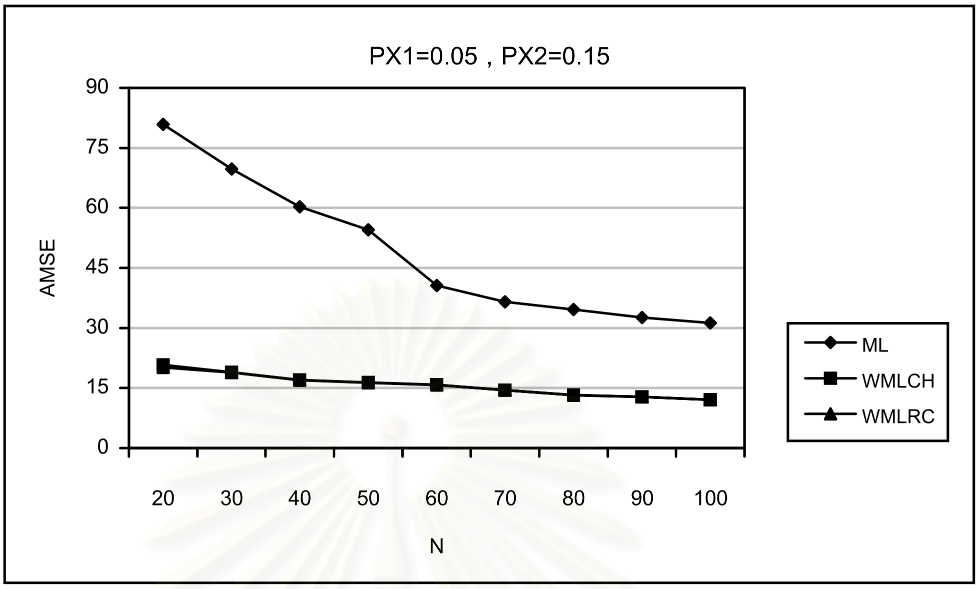
* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

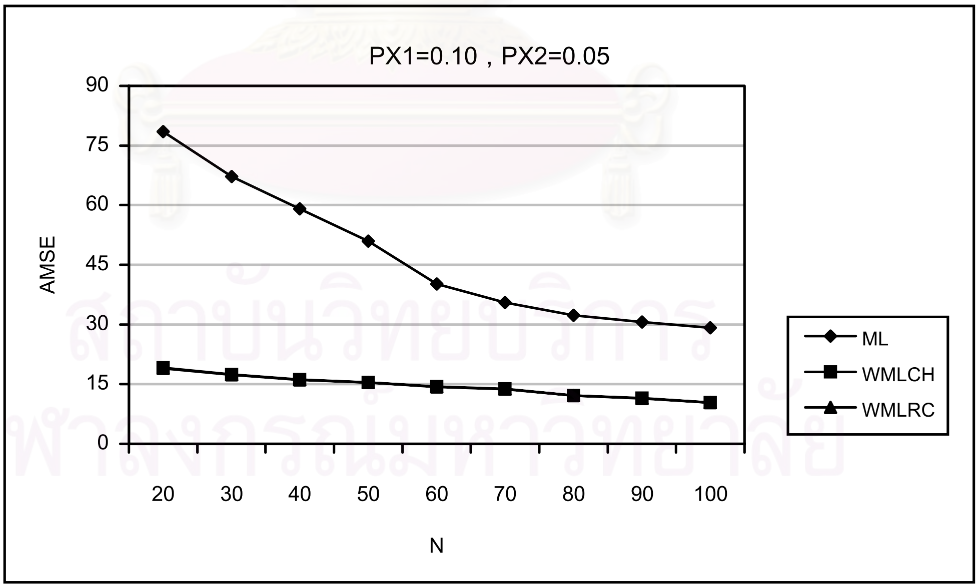
รูปที่4.3.4 แสดงการเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ รุนแรงและตัวแปรอิสระ x_2 มีค่าผิดปกติระดับ รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2



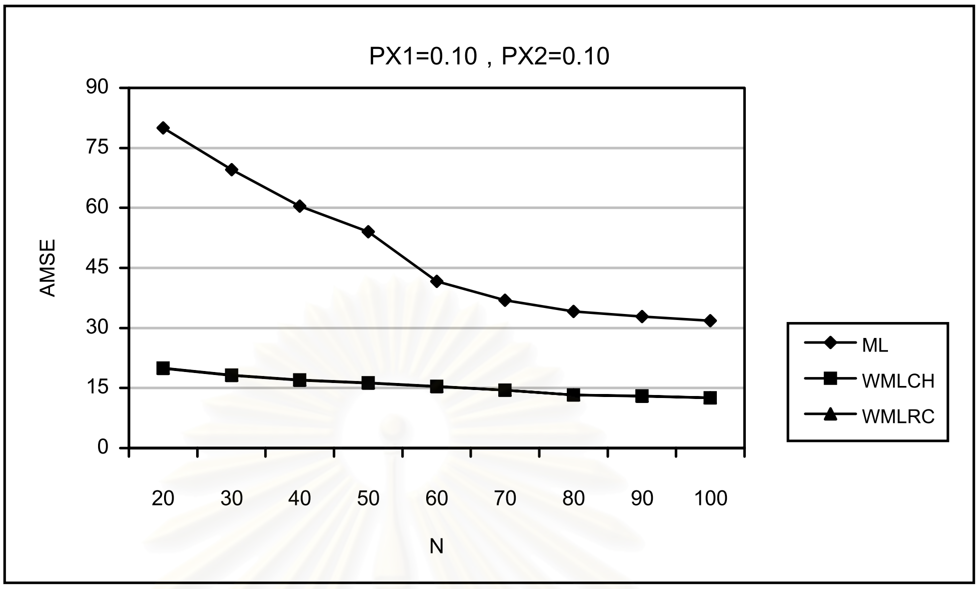


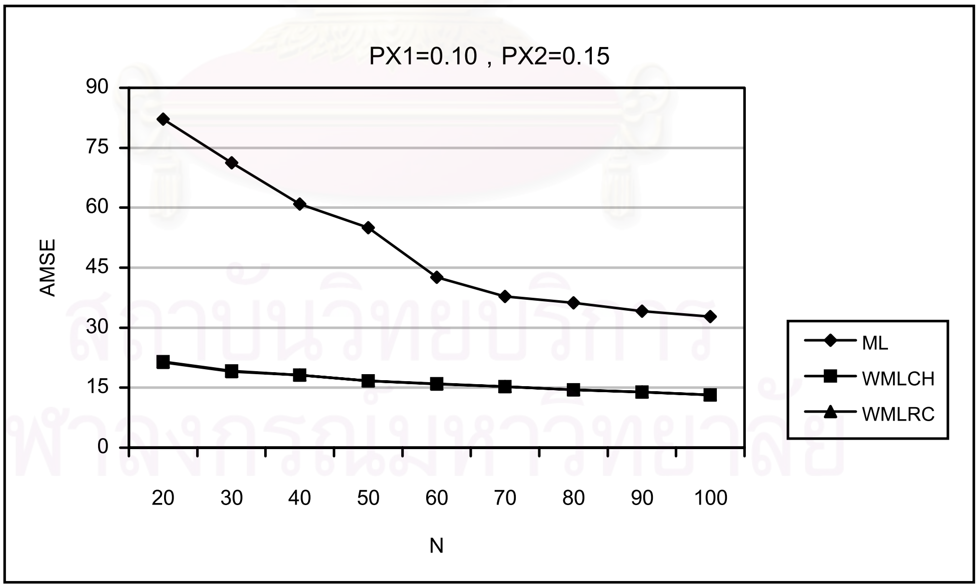
รูปที่4.3.4 (ต่อ)


PX1=0.05 , PX2=0.15
90
75
60
45
30
15
0
AMSE
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC

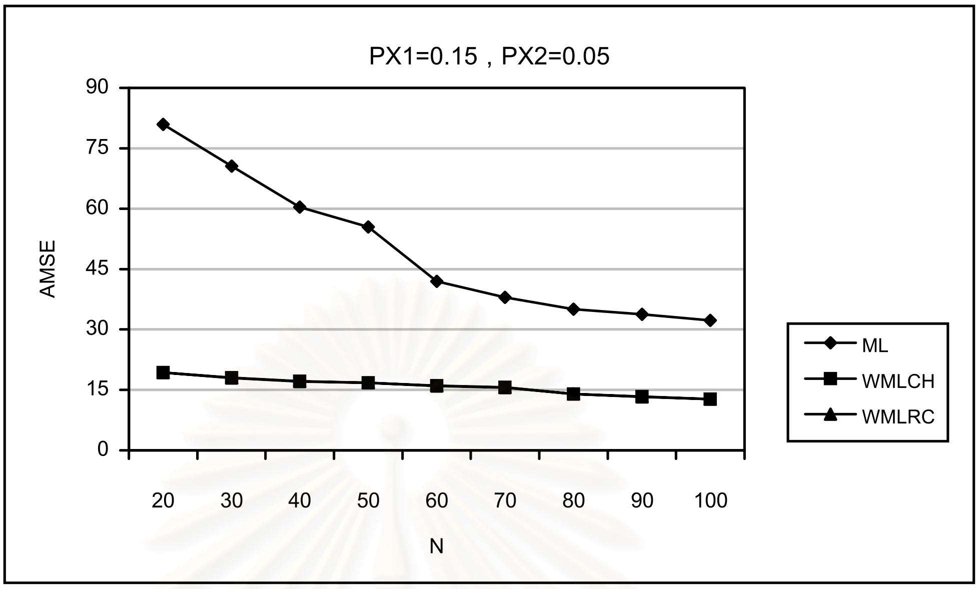

PX1=0.10 , PX2=0.05
90
75
60
45
30
15
0
AMSE
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC

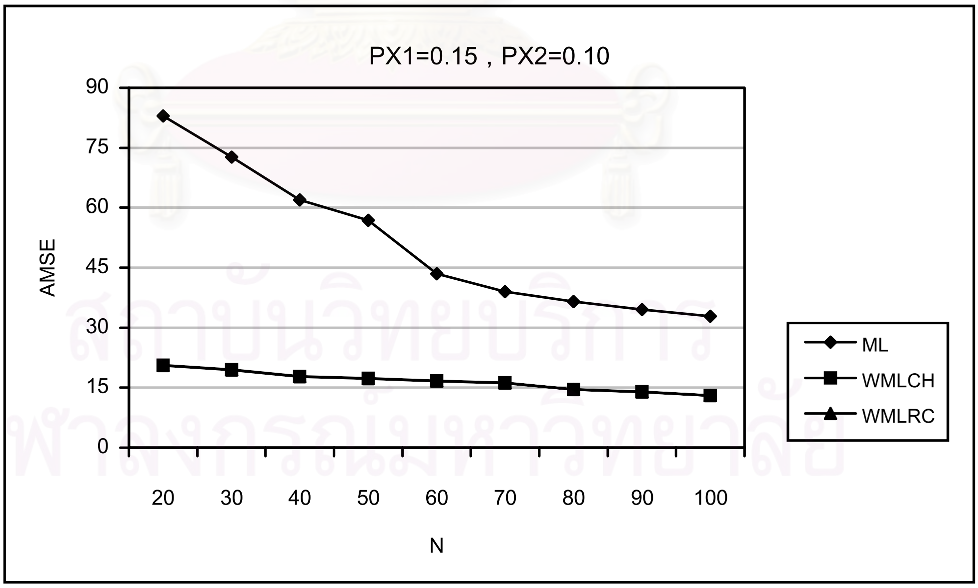
รูปที่4.3.4 (ต่อ)


PX1=0.10 , PX2=0.10
AMSE
90
75
60
45
30
15
0
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC

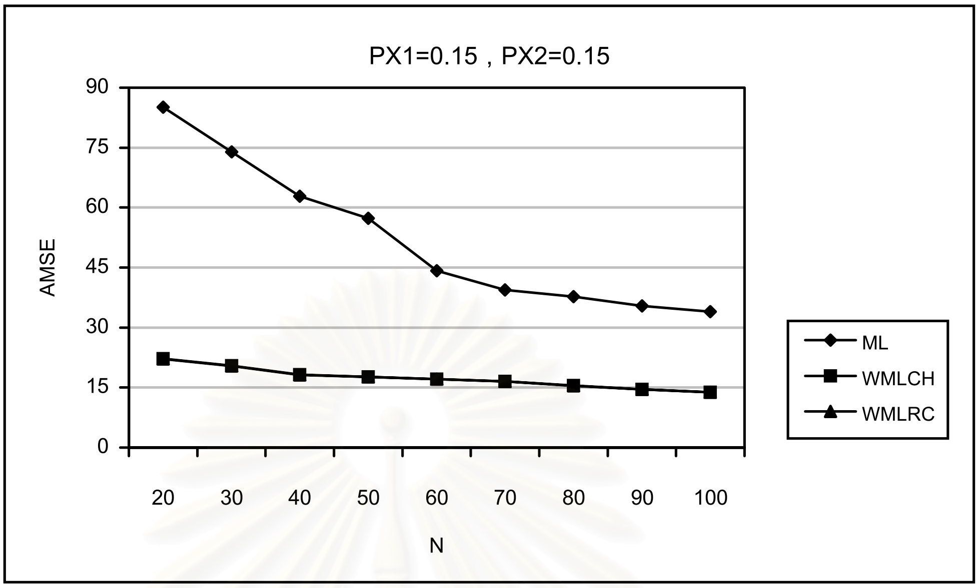

PX1=0.10 , PX2=0.15
AMSE
90
75
60
45
30
15
0
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC

รูปที่4.3.4 (ต่อ)


PX1=0.15 , PX2=0.05
AMSE
N
90
75
60
45
30
15
0
20
30
40
50
60
70
80
90
100
ML
WMLCH
WMLRC


PX1=0.15 , PX2=0.10
AMSE
N
90
75
60
45
30
15
0
20
30
40
50
60
70
80
90
100
ML
WMLCH
WMLRC

รูปที่4.3.4 (ต่อ)


PX1=0.15 , PX2=0.15
90
75
60
45
30
15
0
AMSE
20
30
40
50
60
70
80
90
100
N
ML
WMLCH
WMLRC

จากตารางที่4.3.4 เราสามารถสรุปผลการเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับรุนแรง ซึ่งมีสัดส่วนการปลอมปนเท่ากับ 0.05 , 0.10 และ 0.15 และตัวแปรอิสระ x_2 มีค่าผิดปกติระดับรุนแรง ซึ่งมีสัดส่วนการปลอมปนเท่ากับ 0.05 , 0.10 และ 0.15 สามารถสรุปผลได้ดังนี้

เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับรุนแรงและตัวแปรอิสระ x_2 มีค่าผิดปกติระดับรุนแรง ในทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1 = 0.05 , 0.10 และ 0.15) และทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_2 (PX2 = 0.05 , 0.10 และ 0.15) โดยในทุก ๆ ขนาดตัวอย่างคือ N = 20,30,40,50,60,70,80,90 และ 100 วิธี WMLRC ให้ค่า AMSE น้อยที่สุด รองลงมาคือ วิธี WMLCH และวิธี ML ตามลำดับ และจะพบว่า วิธี WMLCH และวิธี WMLRC ให้ค่า AMSE ใกล้เคียงกันมากในทุกขนาดตัวอย่าง โดยวิธี ML ให้ค่า AMSE สูงมากในทุกขนาดโดยเฉพาะตัวอย่างขนาดเล็ก(N = 20,30,40 และ 50)

ข้อสรุป

จากตารางที่4.3.4 ทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 และทุกขนาดตัวอย่างวิธี WMLRC ให้ค่า AMSE ต่ำกว่าวิธีอื่น ๆ โดยที่วิธี ML จะเป็นวิธีที่ให้ค่า AMSE สูงที่สุด และจะพบว่าในทุก ๆ ขนาดตัวอย่าง วิธี WMLCH และวิธี WMLRC ให้ค่า AMSE ใกล้เคียงกันมาก โดยวิธี ML ให้ค่า AMSE สูงมากในทุกขนาดตัวอย่าง โดยเฉพาะตัวอย่างขนาดเล็ก(N = 20,30,40 และ 50)

ค่า AMSE ของทุกวิธีมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น ขณะที่สัดส่วนการปลอมปนคงที่

ค่า AMSE ของทุกวิธีมีแนวโน้มเพิ่มขึ้นเมื่อสัดส่วนการปลอมปนเพิ่มขึ้น ขณะที่ขนาดตัวอย่างคงที่

บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลองเพื่อการศึกษาและเปรียบเทียบวิธีประมาณค่าพารามิเตอร์เมื่อมีค่าผิดปกติในตัวแปรอิสระ ซึ่งในที่นี้มี 3 วิธีคือ วิธีความควรจะเป็นสูงสุด(ML) วิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และ Haesbroeck(WMLCH) และวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann(WMLRC) โดยสถานการณ์ที่ศึกษามีดังนี้

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

การวิจัยครั้งนี้ผู้วิจัยศึกษาวิเคราะห์ด้วยเทคนิคมอนติคาร์โลโดยใช้โปรแกรมภาษาฟอร์แทน 77 บน PC Computer เพื่อสร้างข้อมูลตามสถานการณ์ที่กำหนดโดยกระทำซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์

5.1 สรุปผลการวิจัย

การวิจัยครั้งนี้ผู้วิจัยต้องการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์โดยพิจารณาจากค่า AMSE ของแต่ละวิธี ซึ่งจากผลการวิจัยพบว่า ระดับค่าผิดปกติ สัดส่วนการปลอมปน และขนาดตัวอย่าง ต่างก็มีผลต่อค่า AMSE ของวิธีประมาณค่าพารามิเตอร์ของวิธีทั้ง 3 โดยค่า AMSE มีแนวโน้มเพิ่มขึ้นเมื่อระดับค่าผิดปกติ และสัดส่วนการปลอมปนเพิ่มขึ้น แต่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น ผู้วิจัยได้สรุปผลการวิจัยตามกรณีต่างๆดังนี้

5.1.1 กรณีที่ไม่มีค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2

พบว่าในทุกขนาดตัวอย่าง วิธี ML จะเป็นวิธีที่ให้ค่า AMSE น้อยที่สุดเมื่อเทียบกับวิธีอื่น ๆ รองลงมาคือ วิธี WMLRC และวิธี WMLCH ตามลำดับ

เมื่อขนาดตัวอย่าง(N) ตั้งแต่ 70 ขึ้นไป วิธี ML วิธี WMLCH และวิธี WMLRC จะให้ค่า AMSE ใกล้เคียงกัน

5.1.2 กรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1

5.1.2.1 ตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับไม่รุนแรง

พบว่าในทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1 = 0.05 , 0.10 และ 0.15) และทุกขนาดตัวอย่าง วิธี WMLRC จะเป็นวิธีที่ให้ค่า AMSE น้อยที่สุดเมื่อเทียบกับวิธีอื่น ๆ รองลงมาคือ วิธี WMLCH และวิธี ML ตามลำดับ

ในทุกสัดส่วนการปลอมปนที่กำหนดเมื่อขนาดตัวอย่าง(N) ตั้งแต่ 60 ขึ้นไป วิธี WMLCH และวิธี WMLRC จะให้ค่า AMSE ใกล้เคียงกัน

5.1.2.2 ตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับรุนแรง

พบว่าในทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1 = 0.05 , 0.10 และ 0.15) และทุกขนาดตัวอย่าง วิธี WMLRC จะเป็นวิธีที่ให้ค่า AMSE น้อยที่สุดเมื่อเทียบกับวิธีอื่น ๆ รองลงมาคือ วิธี WMLCH และวิธี ML ตามลำดับ

ในทุกสัดส่วนการปลอมปนที่กำหนดเมื่อขนาดตัวอย่าง(N) ตั้งแต่ 60 ขึ้นไป วิธี WMLCH และวิธี WMLRC จะให้ค่า AMSE ใกล้เคียงกัน นอกจากนี้พบว่าเมื่อตัวแปรอิสระ x_1 มีระดับค่าผิดปกติมากขึ้นจะมีผลทำให้ค่า AMSE มากขึ้นด้วย เมื่อสัดส่วนการปลอมปนและขนาดตัวอย่างมีค่าคงที่

5.1.3 กรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2

5.1.3.1 ตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับไม่รุนแรงและตัวแปรอิสระ x_2 มีระดับค่าผิดปกติระดับไม่รุนแรง

พบว่าในทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1 = 0.05 , 0.10 และ 0.15) และทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_2 (PX2 = 0.05 , 0.10 และ 0.15) โดยทุกขนาดตัวอย่าง วิธี WMLRC จะเป็นวิธีที่ให้ค่า AMSE น้อยที่สุดเมื่อเทียบกับวิธีอื่น ๆ รองลงมาคือ วิธี WMLCH และวิธี ML ตามลำดับ

ในทุกสัดส่วนการปลอมปนที่กำหนดเมื่อขนาดตัวอย่าง(N) มีขนาดตั้งแต่ 50 ขึ้นไป วิธี WMLCH และวิธีWMLRC จะให้ค่า AMSE ใกล้เคียงกันมาก

5.1.3.2 ตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับไม่รุนแรงและตัวแปรอิสระ x_2 มีระดับค่าผิดปกติระดับรุนแรง

พบว่าในทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1 = 0.05 , 0.10 และ 0.15) และทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_2 (PX2 = 0.05 , 0.10 และ 0.15) โดยทุกขนาดตัวอย่าง วิธี WMLRC จะเป็นวิธีที่ให้ค่า AMSE น้อยที่สุดเมื่อเทียบกับวิธีอื่น ๆ รองลงมาคือ วิธี WMLCH และวิธี ML ตามลำดับ

ในทุกสัดส่วนการปลอมปนที่กำหนดวิธี WMLCH และวิธีWMLRC ให้ค่า AMSE ใกล้เคียงกันมากในทุกขนาดตัวอย่าง โดยวิธี ML ให้ค่า AMSE สูงในทุกขนาดตัวอย่าง

5.1.3.3 ตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับรุนแรงและตัวแปรอิสระ x_2 มีระดับค่าผิดปกติระดับไม่รุนแรง

พบว่าในทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1 = 0.05 , 0.10 และ 0.15) และทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_2 (PX2 = 0.05 , 0.10 และ 0.15) โดยทุกขนาดตัวอย่าง วิธี WMLRC จะเป็นวิธีที่ให้ค่า AMSE น้อยที่สุดเมื่อเทียบกับวิธีอื่น ๆ รองลงมาคือ วิธี WMLCH และวิธี ML ตามลำดับ

ในทุกสัดส่วนการปลอมปนที่กำหนดวิธี WMLCH และวิธีWMLRC ให้ค่า AMSE ใกล้เคียงกันมากในทุกขนาดตัวอย่าง โดยวิธี ML ให้ค่า AMSE สูงในทุกขนาดตัวอย่าง

5.1.3.4 ตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับรุนแรงและตัวแปรอิสระ x_2 มีระดับค่าผิดปกติระดับรุนแรง

พบว่าในทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1 = 0.05 , 0.10 และ 0.15) และทุกสัดส่วนการปลอมปนของตัวแปรอิสระ x_2 (PX2 = 0.05 , 0.10 และ 0.15) โดยทุกขนาดตัวอย่าง วิธี WMLRC จะเป็นวิธีที่ให้ค่า AMSE น้อยที่สุดเมื่อเทียบกับวิธีอื่น ๆ รองลงมาคือ วิธี WMLCH และวิธี ML ตามลำดับ

ในทุกสัดส่วนการปลอมปนที่กำหนดวิธี WMLCH และวิธีWMLRC ให้ค่า AMSE ใกล้เคียงกันมากในทุกขนาดตัวอย่าง โดยวิธี ML ให้ค่า AMSE สูงมากในทุกขนาดตัวอย่าง โดยเฉพาะขนาดตัวอย่าง(N) ที่น้อยกว่า 60 นอกจากนี้พบว่าเมื่อตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 มีระดับค่าผิดปกติมากขึ้นจะมีผลทำให้ค่า AMSE มากขึ้นด้วย เมื่อสัดส่วนการปลอมปนและขนาดตัวอย่างมีค่าคงที่

5.2 ข้อเสนอแนะ

ผลการวิจัยครั้งนี้จะเสนอแนะเป็น 2 ด้าน คือ

5.2.1 ด้านการนำไปใช้ประโยชน์

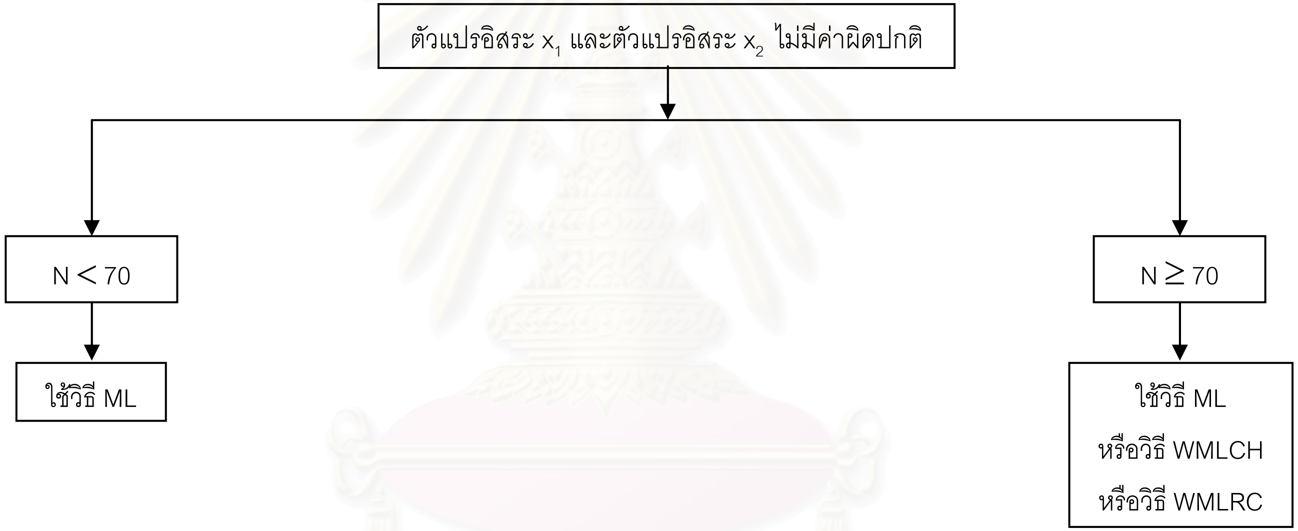
ในทางปฏิบัติของการวิเคราะห์กาถดถอยโลจิสติค เมื่อผู้วิเคราะห์ได้รับข้อมูลของตัวแปรตามและตัวแปรอิสระ ขั้นแรกควรทำการตรวจสอบว่า ข้อมูลของตัวแปรอิสระแต่ละตัวมีค่าผิดปกติหรือไม่อาจใช้วิธีการเขียนกราฟแบบ Box Plot โดยใช้โปรแกรมสำเร็จรูป เช่นโปรแกรมสำเร็จรูป SPSS ซึ่งผู้วิเคราะห์ยังสามารถทราบถึงชุดของข้อมูลที่เป็นค่าผิดปกติและทราบว่าข้อมูลชุดนั้นเกิดค่าเป็นปกติอยู่ในระดับใดได้อีกด้วย โดยได้แสดงตัวอย่างการตรวจสอบไว้ในภาคผนวก ค

ถ้าผลของการตรวจสอบพบว่า มีค่าผิดปกติและผู้วิเคราะห์ทราบว่ามีสาเหตุมาจากความผิดพลาดของการวัด เช่น การบันทึกข้อมูลผิดพลาด ก็ควรจะทำการแก้ไขโดยอาจใช้วิธีตัดค่าผิดปกติออกไปเมื่อข้อมูลมีจำนวนมาก แต่ถ้าข้อมูลมีจำนวนไม่มากอาจใช้วิธีการปรับแก้ค่านั้นตามความเหมาะสมก่อนที่จะนำข้อมูลไปใช้ในการวิเคราะห์ต่อไป แต่ถ้าไม่ทราบถึงสาเหตุของการมีค่าผิดปกติ ผู้วิเคราะห์ไม่ควรที่จะตัดค่าผิดปกติออกทันที เนื่องจากค่าผิดปกติอาจให้ข้อมูลที่เป็นประโยชน์ในการหาตัวแบบการถดถอยที่เหมาะสม

ถ้าผู้วิเคราะห์ต้องการนำข้อมูลเหล่านี้มาใช้ในการวิเคราะห์การถดถอยโลจิสติค วิธีแก้ไข คือ อาจใช้วิธีการประมาณค่าพารามิเตอร์ที่มีความแกร่งซึ่งวิธีการเหล่านี้จะใช้หลักการลดอิทธิพลของค่าสังเกตที่มีค่าผิดปกติลง โดยอาจทำการถ่วงน้ำหนักข้อมูลที่มีค่าผิดปกติ และในที่นี้ได้เสนอแนวทางในการเลือกวิธีการประมาณค่าพารามิเตอร์ เมื่อตัวแปรอิสระไม่มีค่าผิดปกติและมีค่าผิดปกติ ให้เหมาะสมในแต่ละสถานการณ์ ผู้วิจัยได้แสดงแผนผังการเลือกวิธีการประมาณค่าพารามิเตอร์ที่ทำให้ได้ค่า AMSE ต่ำสุดในสถานการณ์ต่าง ๆ เพื่อที่ไปใช้ ซึ่งในการเลือกใช้จะต้องทราบปัจจัยต่าง ๆ คือ กรณีไม่มีค่าผิดปกติและกรณีมีค่าผิดปกติในตัวแปรอิสระ ระดับค่าผิดปกติ สัดส่วนการปลอมปน และขนาดตัวอย่าง โดยใช้สัญลักษณ์ต่อไปนี้แทนความหมายต่าง ๆ กล่าวคือ

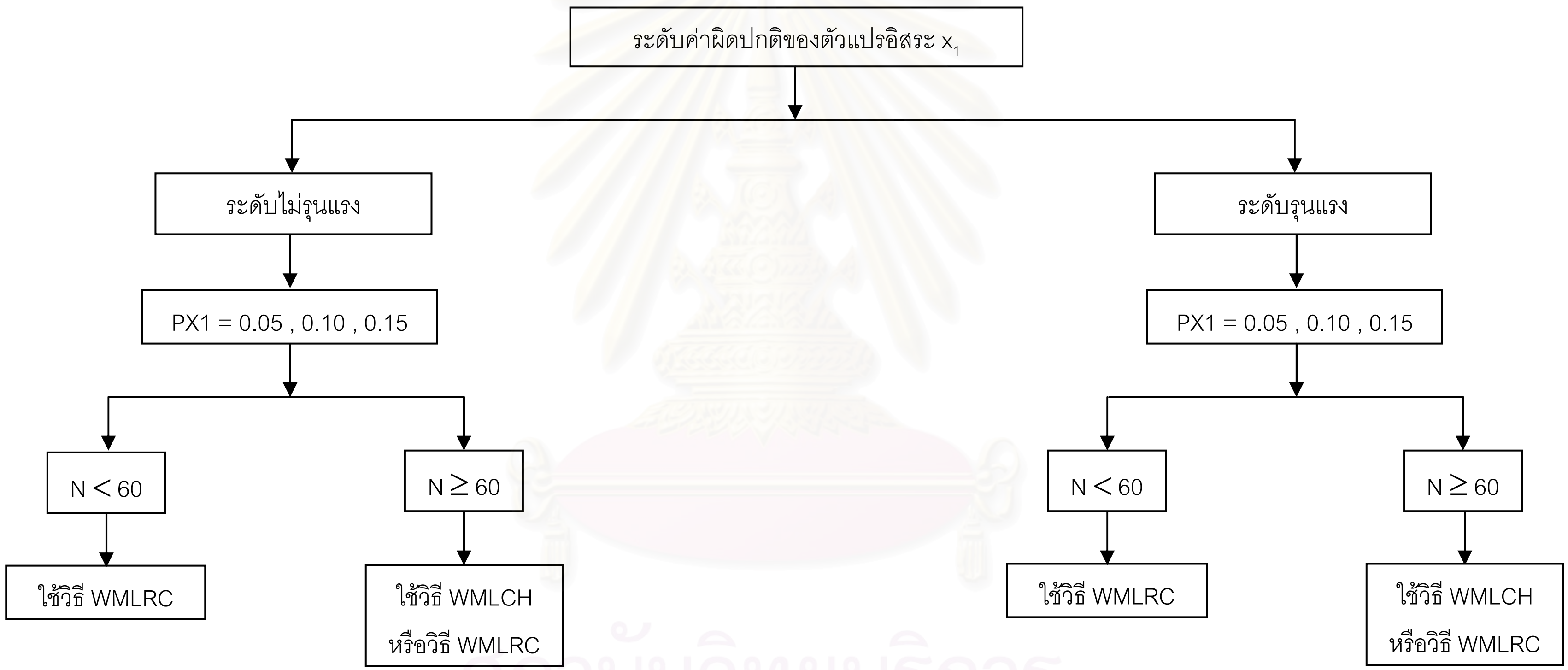
ML	หมายถึง	วิธีความควรจะเป็นสูงสุด
WMLCH	หมายถึง	วิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และ Haesbroeck
WMLRC	หมายถึง	วิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann
PX1	หมายถึง	สัดส่วนการปลอมปนของตัวแปรอิสระ x_1
PX2	หมายถึง	สัดส่วนการปลอมปนของตัวแปรอิสระ x_2
N	หมายถึง	ขนาดตัวอย่าง

แผนผังแสดงการเลือกใช้ค่าประมาณของพารามิเตอร์ กรณีที่ไม่มีค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2



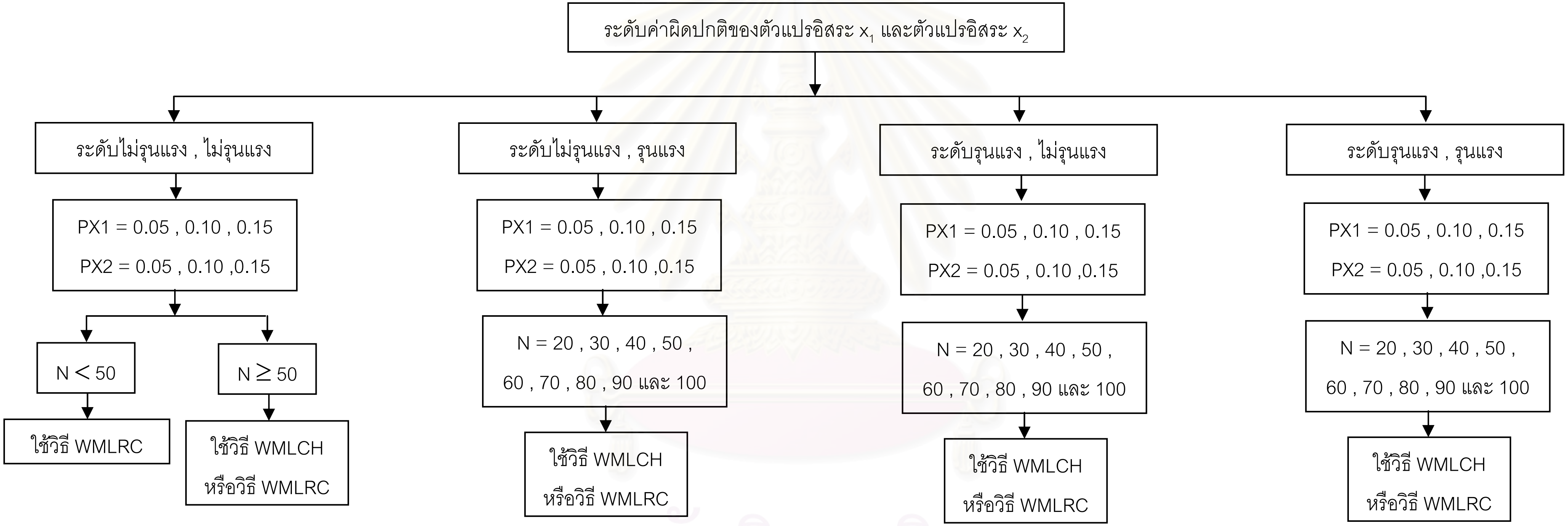
หมายเหตุ ในทุกขนาดตัวอย่างควรเลือกใช้วิธี ML เนื่องจากวิธี ML ให้ค่า AMSE น้อยกว่าวิธี WMLCH และวิธี WMLRC ซึ่งวิธี ML ไม่มีความยุ่งยากในการประมาณค่าพารามิเตอร์และสามารถนำไปใช้งานได้ง่ายกว่า

แผนผังแสดงการเลือกใช้ค่าประมาณของพารามิเตอร์ กรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1



หมายเหตุ ในทุกระดับค่าผิดปกติ สัดส่วนการปลอมปน และขนาดตัวอย่าง ควรเลือกใช้วิธี WMLRC เนื่องจากวิธี WMLRC ให้ค่า AMSE น้อยที่สุด

แผนผังแสดงการเลือกใช้ค่าประมาณของพารามิเตอร์ กรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2



หมายเหตุ ในทุกระดับค่าผิดปกติ สัดส่วนการปลอมปน และขนาดตัวอย่าง ควรเลือกใช้วิธี WMLRC เนื่องจากวิธี WMLRC ให้ค่า AMSE น้อยที่สุด

5.2.2 ด้านการศึกษาวิจัย

เพื่อเป็นแนวทางให้ผู้ที่สนใจได้ศึกษาเพิ่มเติม เพื่อเป็นการขยายผลการวิจัยออกไปให้เกิดประโยชน์มากยิ่งขึ้น โดยทำการศึกษาในกรณีต่าง ๆ ดังต่อไปนี้

1. ควรศึกษาเมื่อตัวแปรอิสระมีมากกว่า 2 ตัวแปร และจำนวนตัวแปรอิสระที่มีค่าผิดปกติมากกว่า 2 ตัว

2. ควรศึกษาและเปรียบเทียบวิธีการทางสถิติวิธีอื่นสำหรับการประมาณค่าพารามิเตอร์ของตัวแบบการถดถอยโลจิสติคเมื่อมีค่าผิดปกติในตัวแปรอิสระ เช่น วิธีของ Bianco และ Yohai (Bianco and Yohai Method (1996))

รายการอ้างอิง

ภาษาไทย

กัลยา วานิชย์บัญชา. (2546). <u>การใช้ SPSS for Windows ในการวิเคราะห์ข้อมูล</u>.

มานพ วราภักดิ์. (2547). <u>การจำลองเบื้องต้น</u>. กรุงเทพมหานคร : ศูนย์ผลิตตำราเรียนสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

ภาษาอังกฤษ

Bianco, A. M., and Yohai, V. J. (1996). <u>Robust Estimation in the Logistic Regression Model</u>, in Robust Statistics, Data Analysis, and Computer Intensive Methods, 17-34; Lecture Notes in Statistics 109, New York: Springer Verlag, Ed. H. Rieder.

Croux, C., and Haesbroeck, G. (2003). <u>Implementing the Bianco and Yohai estimator for logistic regression</u>, Computational Statistics & Data Analysis, 44: 273-295.

Rousseeuw, R. J., and Christmann, A. (2003). <u>Robustness against separation and outlier in logistic regression</u>, Computational Statistics & Data Analysis, 43: 315-332.

บรรณานุกรม

ภาษาไทย

กาญจนา พานิชการ. (2539). การประมาณค่าพารามิเตอร์ในสมการถดถอยโลจิสติคด้วยภาวะน่าจะเป็นสูงสุด และฟังก์ชันจำแนกประเภท. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

จิตรวี วีระประดิษฐ. (2538). การประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นพหุ เมื่อข้อมูลมีค่าผิดปกติ. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

วีรานันท์ พงศาภักดี. (ม.ป.ป.). การวิเคราะห์ข้อมูลเชิงกลุ่ม : ทฤษฎีและการประยุกต์(กับGLIM และ SPSS/FW). โรงพิมพ์มหาวิทยาลัยศิลปากร.

ภาษาอังกฤษ

Carroll, R. J., and Pederson, S. (1993). On Robustness in Logistic Regression Model, J. Roy. Statist. Soc. B, 55: 693-706.

Jobson, J. D. (1992). Applied Multivatiate Data Analysis, Volume II : Categorical and Multivariate Method . New York: Springer – Verlag.

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

Rousseeuw, R. J., and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator, Technometrics 41: 212-223.

Rousseeuw, R. J., and Van Zomeren, B. C. (1990).Unmasking Multivariate Outliers and Leverage Points, Journal of the American Statistical Association 85: 633-639.

ภาคผนวก

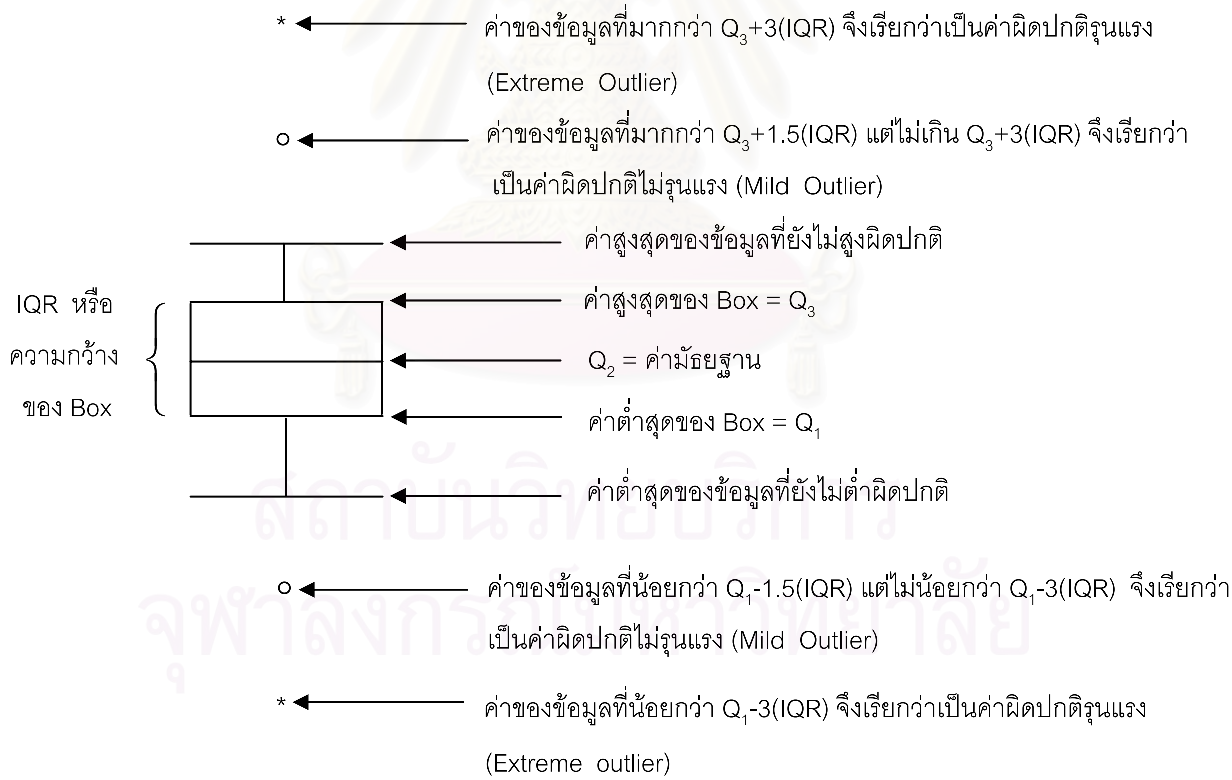
ภาคผนวก ก

Box Plot[1]

Box Plot เป็นเทคนิคที่ให้รายละเอียดเกี่ยวกับการแจกแจงของข้อมูล และแสดงรายละเอียดของค่าสถิติเพื่อตรวจสอบสำหรับการแจกแจง นั่นคือ จะแสดงค่ามัธยฐาน เปอร์เซ็นไทล์ที่ 25 , 75 และให้ค่าข้อมูลที่มีค่าผิดปกติ นั่นคือ ค่าที่สูงมากหรือต่ำมาก (Outliers) จากค่ากลาง ดังนั้น การสร้าง Box Plot จะใช้ค่าสถิติด้วยกัน 5 ค่า คือ

1. ค่าต่ำสุดของข้อมูลที่ยังไม่ต่ำผิดปกติ
2. ควอไทล์ที่ 1 (Q_1)
3. ค่ามัธยฐาน หรือควอไทล์ที่ 2 (Q_2)
4. ควอไทล์ที่ 3 (Q_3)
5. ค่าสูงสุดของข้อมูลที่ยังไม่สูงผิดปกติ

โดยลักษณะของ Box Plot ได้มีการแสดงดังในรูปนี้



---

[1]กัลยา วานิชย์บัญชา, <u>การใช้ SPSS for Windows ในการวิเคราะห์ข้อมูล</u> (ฉบับปรับปรุงใหม่, 2546), หน้า 216

ค่าต่ำสุดของ Box = เปอร์เซ็นไทล์ที่ 25 หรือควอไทล์ที่ 1 (Q_1) ของข้อมูล

ค่าสูงสุดของ Box = เปอร์เซ็นไทล์ที่ 75 หรือควอไทล์ที่ 3 (Q_3) ของข้อมูล

ค่ากลางของ Box = ค่ามัธยฐาน หรือเปอร์เซ็นไทล์ที่ 50 หรือควอไทล์ที่ 2 (Q_2) ของข้อมูล

ความกว้างของ Box = Q_3 - Q_1 = IQR (Interquatile Range) หรือกล่าวว่า 50% ของข้อมูลอยู่ใน Box

ค่าสูงสุดของข้อมูลที่ยังไม่สูงผิดปกติ คือ ค่าสูงสุดของข้อมูลชุดนั้น ๆ ที่มีค่าไม่เกิน Q_3+1.5(IQR)

ค่าต่ำสุดของข้อมูลที่ยังไม่ต่ำผิดปกติ คือ ค่าต่ำสุดของข้อมูลชุดนั้น ๆ ที่มีค่าไม่เกิน Q_1-1.5(IQR)

นอกจากนี้ Box Plot ยังแสดงค่าผิดปกติ 2 ลักษณะ คือ

1. ข้อมูลที่มีค่าระหว่าง 1.5 ถึง 3 เท่าของความกว้างของ Box นั่นคือ ข้อมูลที่มีค่ามากกว่า Q_3+1.5(IQR) แต่ไม่เกิน Q_3+3(IQR) หรือข้อมูลที่มีค่าน้อยกว่า Q_1-1.5(IQR) แต่ไม่น้อยกว่า Q_1-3(IQR) และจะเรียกค่าเหล่านี้ว่า Mild Outliers (ค่าผิดปกติในระดับไม่รุนแรง แสดงด้วยเครื่องหมาย วงกลม (o))

2. ข้อมูลที่มีค่ามากกว่า 3 เท่าของความกว้างของ Box นั่นคือ ข้อมูลที่มีค่ามากกว่า Q_3+3(IQR) หรือข้อมูลที่มีค่าน้อยกว่า Q_1-3(IQR) และจะเรียกค่าเหล่านี้ว่า Extreme Outliers (ค่าผิดปกติในระดับรุนแรง แสดงด้วยเครื่องหมาย ดอกจัน (*))

ภาคผนวก ข

Minimum Covariance Determinant Estimater

วิธี Minimum Covariance Determinant(MCD) เป็นการหาชุดค่าสังเกต h ชุดที่ทำให้เมทริกซ์ความแปรปรวนร่วมมีค่าดีเทอร์มิแนนต์(determinant) น้อยที่สุด โดยมีขั้นตอนดังนี้

ขั้นที่ 1 ทำการสุ่มชุดค่าสังเกต p+1 ชุด (p เท่ากับ จำนวนตัวแปรอิสระ) แล้วคำนวณค่า distance แต่ละ i (สำหรับ i = 1,2,…,N) จากสูตร

$$d(i) = \sqrt{(\underset{\sim}{x}_i' - \bar{x})' S^{-1} (\underset{\sim}{x}_i' - \bar{x})}$$

โดย $\underset{\sim}{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})$

$\bar{x} = \frac{1}{p+1}\sum_{j=1}^{p+1} x_j$ ของชุดค่าสังเกตที่สุ่มได้

$S = \frac{1}{N-1}\sum_{i=1}^{N}(\underset{\sim}{x}_i' - \bar{x})(\underset{\sim}{x}_i' - \bar{x})$ ซึ่งเป็นเมทริกซ์ความแปรปรวนร่วม

จากนั้นทำการเรียงลำดับค่า distance จากน้อยไปมาก แล้วเลือกลำดับชุดค่าสังเกตที่ 1 ถึง h มาทำต่อในขั้นที่ 2 โดย h = จำนวนเต็มของ 0.75N

ขั้นที่ 2 - นำชุดค่าสังเกต h ชุด จากขั้นที่ 1 มาคำนวณหาค่า distance แต่ละ i จากสูตรข้างต้น โดย $\bar{x} = \frac{1}{h}\sum_{j=1}^{h} x_j$ ของชุดค่าสังเกต h ชุด

- เรียงลำดับค่า distance จากน้อยไปมาก แล้วเลือกลำดับชุดค่าสังเกตที่ 1 ถึง h มาทำการคำนวณหาค่า distance แต่ละ i

- เรียงลำดับค่า distance จากน้อยไปมาก แล้วเลือกลำดับชุดค่าสังเกตที่ 1 ถึง h มาทำการคำนวณหาค่า determinant ของเมทริกซ์ความแปรปรวนร่วม แล้วทำการเก็บค่าไว้ โดยขั้นที่ 2 นี้เรียกว่า C-step

* ทำซ้ำขั้นที่ 1 และ 2 เป็นจำนวน 500 รอบ โดยชุดค่าสังเกต 500 ชุดต้องไม่ซ้ำกัน

ขั้นที่ 3 - เลือก 10 ชุดของค่า determinant ที่น้อยที่สุดจาก 500 ชุด

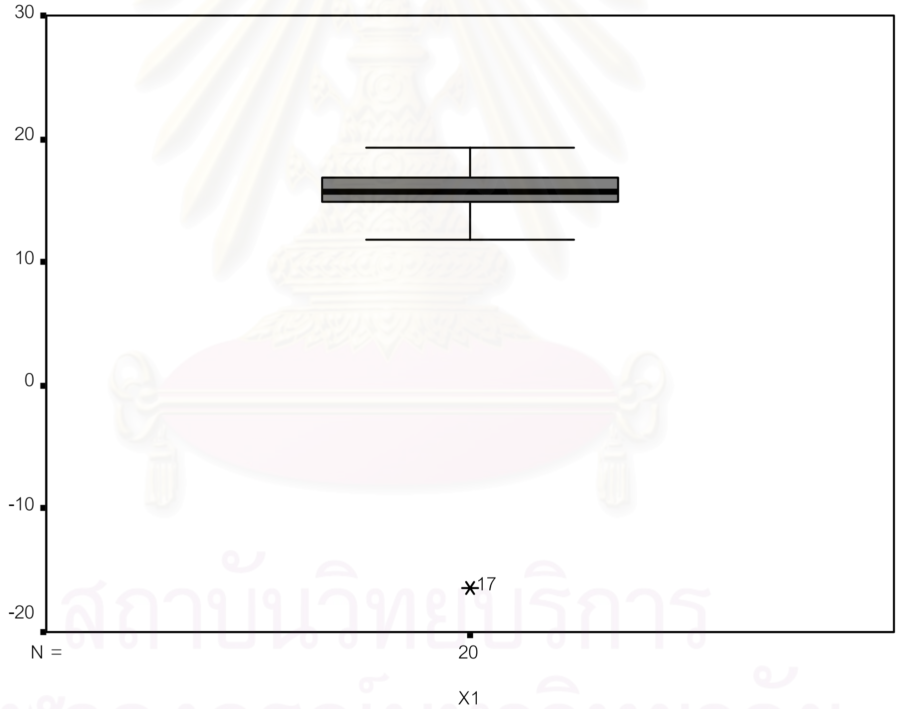
- คำนวณแต่ละชุดทั้ง 10 ชุดด้วย C-step จนกว่าค่า determinant ณ รอบใด ๆ เท่ากับ 0 หรือ ผลต่างระหว่างค่า determinant รอบปัจจุบันกับค่า determinant รอบก่อนมีค่าน้อยกว่า 0.000001

- เลือกชุดที่ให้ค่า determinant น้อยที่สุด เป็นชุดค่าสังเกต h ชุด เพื่อนำไปคำนวณหาค่า Robust distance

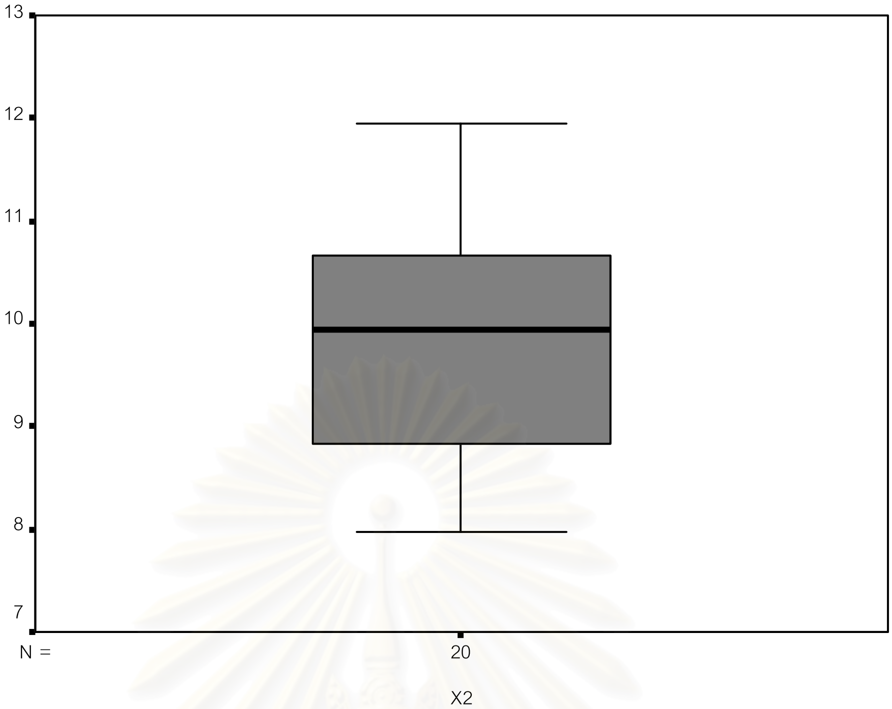
ภาคผนวก ค

การตรวจสอบด้วยวิธี Box Plot

จากข้อมูลในตารางที่ 1.1 ซึ่งแสดงตัวเลขค่าของข้อมูลตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 กรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1 เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับรุนแรง และสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 เท่ากับ 0.05 มีขนาดตัวอย่างเท่ากับ 20 โดยผู้วิจัยจะนำข้อมูลชุดนี้มาแสดงการตรวจสอบวิธี Box Plot ในโปรแกรมสำเร็จรูป SPSS โดยใช้คำสั่ง Explore ซึ่งอยู่ในคำสั่ง Analyze และ Descriptive Statistics โดยจะได้รูป Box Plot แสดงการตรวจสอบข้อมูลดังนี้



รูปที่ 1 แสดงการตรวจสอบตัวแปรอิสระ x_1



รูปที่ 2 แสดงการตรวจสอบตัวแปรอิสระ x_2

จากรูปที่ 1 แสดงการตรวจสอบตัวแปรอิสระ x_1 มีค่าผิดปกติในข้อมูลชุดที่ 17 และมีค่าผิดปกติอยู่ในระดับรุนแรง เนื่องจากมีเครื่องหมายดอกจัน (*) ออกนอก Box และจากรูปที่ 2 จะเห็นได้ว่าตัวแปรอิสระ x_2 มีค่าปกติ เนื่องจากไม่มีเครื่องหมายใดออกนอก Box

ภาคผนวก ง

ตารางแสดงการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ ของวิธีความควรจะเป็นสูงสุด (ML) วิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Croux และ Haesbroeck (WMLCH) และวิธีความควรจะเป็นสูงสุดแบบถ่วงน้ำหนักของ Rousseeuw และ Christmann (WMLRC) ด้วยค่า AMSE เมื่อขนาดตัวอย่างคงที่ ในขณะที่สัดส่วนการปลอมปนเปลี่ยนแปลงไป ผู้วิจัยได้ทำการกรณีต่าง ๆ ดังนี้

1. กรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1

โดยผู้วิจัยได้แสดงผลการวิจัยออกเป็น

1.1 กรณีตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับไม่รุนแรง(C=5) ทำการศึกษาที่สัดส่วนการปลอมปนเป็น 0.05 , 0.10 และ 0.15 ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 1.1

1.2 กรณีตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับรุนแรง(C=12) ทำการศึกษาที่สัดส่วนการปลอมปนเป็น 0.05 , 0.10 และ 0.15 ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 1.2

ตารางที่1.1 การเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ ไม่รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1)

ขนาดตัวอย่าง N	PX1	AMSE		
		ML	WMLCH	WMLRC
20	0.05	15.815250	14.218100	12.865340*
	0.10	17.929150	15.914430	14.659410*
	0.15	20.023536	16.592942	15.718240*
30	0.05	10.012390	9.145673	7.839638*
	0.10	10.133940	9.992354	8.249493*
	0.15	15.049653	13.267942	12.054381*
40	0.05	8.874522	8.263473	7.018407*
	0.10	9.350096	8.548240	7.437162*
	0.15	11.661150	10.525540	9.410365*
50	0.05	8.343843	7.760020	6.834788*
	0.10	8.828650	8.002160	7.177440*
	0.15	10.125320	9.037810	8.250312*
60	0.05	6.882869	6.269540	6.163590*
	0.10	7.169284	6.448200	6.224416*
	0.15	8.356366	7.339150	7.128163*
70	0.05	6.012746	5.400862	5.183291*
	0.10	6.543552	5.973520	5.842308*
	0.15	7.329405	6.480495	6.332237*
80	0.05	5.524580	5.093114	4.875261*
	0.10	6.380802	5.765019	5.637029*
	0.15	7.182118	6.101369	6.035255*
90	0.05	5.092584	4.652280	4.466670*
	0.10	5.730252	5.137548	5.014943*
	0.15	6.390467	5.524567	5.459386*
100	0.05	4.481638	4.034474	3.993544*
	0.10	5.044331	4.365096	4.249579*
	0.15	5.646428	4.814347	4.782903*

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่1.2 การเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1)

ขนาดตัวอย่าง N	PX1	AMSE		
		ML	WMLCH	WMLRC
20	0.05	17.740790	16.224900	14.966270*
	0.10	20.301532	18.245539	17.022592*
	0.15	22.627483	19.729156	18.577258*
30	0.05	12.266830	10.783531	9.587802*
	0.10	14.792300	12.636130	11.803120*
	0.15	18.440700	15.518560	14.765820*
40	0.05	11.058040	9.794290	8.601069*
	0.10	13.659220	11.649530	10.951440*
	0.15	16.647200	13.746410	12.996908*
50	0.05	10.495380	9.131880	8.386696*
	0.10	12.382330	10.426024	9.921807*
	0.15	13.314770	11.215490	10.606943*
60	0.05	9.465770	8.297283	8.115396*
	0.10	10.317480	8.490380	8.313716*
	0.15	11.634410	9.539993	9.385119*
70	0.05	9.291430	8.049920	7.931949*
	0.10	9.932380	8.222170	8.139442*
	0.15	11.134090	9.030902	8.947356*
80	0.05	7.884260	6.730745	6.624678*
	0.10	9.006820	7.729259	7.634104*
	0.15	10.746500	8.846206	8.775593*
90	0.05	7.584740	6.485651	6.389822*
	0.10	8.815130	7.198894	7.123164*
	0.15	9.814040	7.919766	7.856375*
100	0.05	7.449580	6.331259	6.265774*
	0.10	8.380470	6.754774	6.686470*
	0.15	9.634760	7.731048	7.674270*

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

2 กรณีที่มีค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2

โดยผู้วิจัยได้แสดงผลการวิจัยออกเป็น

2.1 กรณีตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับไม่รุนแรง(C=5) และตัวแปรอิสระ x_2 มีระดับค่าผิดปกติระดับไม่รุนแรง(C=5) โดยแต่ละระดับทำการศึกษาที่สัดส่วนการปลอมปนเป็น 0.05 , 0.10 และ 0.15 ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 2.1

2.2 กรณีตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับไม่รุนแรง(C=5) และตัวแปรอิสระ x_2 มีระดับค่าผิดปกติระดับรุนแรง(C=12) โดยแต่ละระดับทำการศึกษาที่สัดส่วนการปลอมปนเป็น 0.05 , 0.10 และ 0.15 ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 2.2

2.3 กรณีตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับรุนแรง(C=12) และตัวแปรอิสระ x_2 มีระดับค่าผิดปกติระดับไม่รุนแรง(C=5) โดยแต่ละระดับทำการศึกษาที่สัดส่วนการปลอมปนเป็น 0.05 , 0.10 และ 0.15 ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 2.3

2.4 กรณีตัวแปรอิสระ x_1 มีระดับค่าผิดปกติระดับรุนแรง(C=12) และตัวแปรอิสระ x_2 มีระดับค่าผิดปกติระดับรุนแรง(C=12) โดยแต่ละระดับทำการศึกษาที่สัดส่วนการปลอมปนเป็น 0.05 , 0.10 และ 0.15 ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 2.4

ตารางที่2.1 การเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ ไม่รุนแรง และตัวแปรอิสระ x_2 มีค่าผิดปกติระดับ ไม่รุนแรงโดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1) และตัวแปรอิสระ x_2 (PX2)

ขนาดตัวอย่าง N	PX1	PX2	วิธีการ		
			ML	WMLCH	WMLRC
20	0.05	0.05	19.670681	14.925488	14.138422*
		0.10	21.363910	16.475893	15.791564*
		0.15	21.635641	16.781668	16.127483*
	0.10	0.05	21.165182	16.266824	15.501938*
		0.10	21.801057	16.760795	16.111765*
		0.15	22.532293	17.454883	16.952036*
	0.15	0.05	21.709561	16.813751	16.202658*
		0.10	22.804431	17.762052	17.203695*
		0.15	23.655406	18.357418	17.970384*
30	0.05	0.05	17.588772	13.371553	12.587319*
		0.10	18.568622	13.817339	13.194967*
		0.15	20.091565	15.210191	14.738309*
	0.10	0.05	18.722926	14.063353	13.386805*
		0.10	19.357209	14.481402	13.844573*
		0.15	21.240562	16.329145	15.816406*
	0.15	0.05	19.865973	15.076682	14.548037*
		0.10	20.817046	15.989539	15.495422*
		0.15	21.676365	16.582422	16.173356*
40	0.05	0.05	14.212450	10.338700	9.916656*
		0.10	16.439380	12.086140	11.674562*
		0.15	17.271629	12.465350	12.028459*
	0.10	0.05	15.328220	10.919340	10.533090*
		0.10	17.537473	12.803620	12.596219*
		0.15	18.792880	13.938690	13.546044*
	0.15	0.05	19.005480	14.201590	13.892630*
		0.10	19.270041	14.457578	14.062884*
		0.15	19.759318	14.835550	14.506811*

***** หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่2.1 (ต่อ)

ขนาดตัวอย่าง N	PX1	PX2	วิธีการ		
			ML	WMLCH	WMLRC
50	0.05	0.05	11.762925	8.009106	7.703659*
		0.10	13.595946	9.432907	9.394064*
		0.15	15.392360	10.661120	10.574111*
	0.10	0.05	13.061274	8.563670	8.495672*
		0.10	15.022131	10.652534	10.566364*
		0.15	15.860079	11.166550	11.078842*
	0.15	0.05	17.541590	12.810150	12.728170*
		0.10	18.395733	13.688760	13.595146*
		0.15	18.811304	14.027760	13.944812*
60	0.05	0.05	10.959692	7.479154	7.338941*
		0.10	12.268847	8.344110	8.251512*
		0.15	13.742936	9.069810	8.994323*
	0.10	0.05	12.242231	8.312350	8.228745*
		0.10	13.380060	9.237690	9.147582*
		0.15	14.270710	9.509190	9.433870*
	0.15	0.05	14.611626	9.917290	9.847464*
		0.10	15.835209	10.175910	10.100377*
		0.15	16.331150	11.623930	11.558852*
70	0.05	0.05	10.149651	6.620574	6.543541*
		0.10	11.002513	7.169833	7.091746*
		0.15	12.895270	8.343146	8.274878*
	0.10	0.05	11.775965	7.875487	7.805591*
		0.10	12.659344	8.561679	8.503163*
		0.15	13.506595	9.191149	9.135905*
	0.15	0.05	12.873541	8.350777	8.288600*
		0.10	13.548091	9.177036	9.114584*
		0.15	15.277720	10.720814	10.674597*

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่2.1 (ต่อ)

ขนาดตัวอย่าง N	PX1	PX2	วิธีการ		
			ML	WMLCH	WMLRC
80	0.05	0.05	9.387999	6.099840	6.013754*
		0.10	10.627665	6.914503	6.888373*
		0.15	12.438942	7.912280	7.899779*
	0.10	0.05	11.438057	7.705654	7.672676*
		0.10	12.073461	8.168070	8.133299*
		0.15	12.768478	8.673190	8.640095*
	0.15	0.05	12.628502	8.208866	8.180225*
		0.10	13.086700	8.907940	8.875796*
		0.15	13.416053	9.107920	9.083552*
90	0.05	0.05	9.222574	5.958650	5.909167*
		0.10	10.527581	6.851207	6.823487*
		0.15	11.706825	7.395332	7.377878*
	0.10	0.05	10.719071	7.017800	6.985486*
		0.10	11.689031	7.794493	7.768805*
		0.15	12.234674	8.183551	8.160604*
	0.15	0.05	12.172030	7.755680	7.733069*
		0.10	12.661150	8.517720	8.494471*
		0.15	13.026892	8.700572	8.680792*
100	0.05	0.05	8.283166	5.152247	5.109167*
		0.10	10.107932	6.649934	6.613621*
		0.15	11.385551	6.955795	6.932679*
	0.10	0.05	10.086759	6.554494	6.525798*
		0.10	10.713987	7.060261	7.047015*
		0.15	11.704337	7.589977	7.562749*
	0.15	0.05	11.694948	7.273691	7.254109*
		0.10	11.939767	7.357602	7.330849*
		0.15	12.394633	8.209026	8.188543*

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่2.2 การเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ ไม่รุนแรง และตัวแปรอิสระ x_2 มีค่าผิดปกติระดับ รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1) และตัวแปรอิสระ x_2 (PX2)

ขนาดตัวอย่าง N	PX1	PX2	วิธีการ		
			ML	WMLCH	WMLRC
20	0.05	0.05	33.698288	17.330393	16.931546*
		0.10	35.261149	18.549053	17.893778*
		0.15	37.593851	20.231612	19.053482*
	0.10	0.05	34.492366	17.546193	17.116148*
		0.10	36.717599	19.026658	18.795686*
		0.15	38.938836	21.048375	19.836719*
	0.15	0.05	36.193146	18.804851	18.550257*
		0.10	37.213581	19.374416	19.163012*
		0.15	39.712926	21.342461	20.279856*
30	0.05	0.05	29.238420	14.171960	14.001220*
		0.10	31.647150	15.369360	15.262500*
		0.15	34.536548	17.279847	17.146126*
	0.10	0.05	30.497524	14.671878	14.568432*
		0.10	32.810835	15.981790	15.869587*
		0.15	35.296665	17.792456	17.625923*
	0.15	0.05	32.436867	15.928699	15.856561*
		0.10	33.575283	16.472695	16.389467*
		0.15	37.042731	18.945687	18.856388*
40	0.05	0.05	27.071790	13.112650	13.065040*
		0.10	29.786557	14.649830	14.552077*
		0.15	31.879686	15.181460	15.096192*
	0.10	0.05	27.755193	13.259330	13.169860*
		0.10	30.797682	14.972450	14.886746*
		0.15	32.738231	15.638746	15.543751*
	0.15	0.05	29.403440	14.005940	13.915820*
		0.10	31.425476	15.184830	15.101060*
		0.15	34.469193	16.565625	16.482869*

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่2.2 (ต่อ)

ขนาดตัวอย่าง N	PX1	PX2	วิธีการ ML	วิธีการ WMLCH	วิธีการ WMLRC
50	0.05	0.05	24.542166	11.421820	11.330820*
		0.10	26.821685	12.336480	12.248757*
		0.15	30.240729	14.377380	14.298051*
	0.10	0.05	25.463230	11.762731	11.677690*
		0.10	28.302680	13.432656	13.342041*
		0.15	31.412459	15.056230	14.979405*
	0.15	0.05	27.520547	12.854200	12.778140*
		0.10	29.719867	14.816390	14.739079*
		0.15	32.398655	15.270340	15.183550*
60	0.05	0.05	22.874193	10.661290	10.576966*
		0.10	24.938977	11.253450	11.185849*
		0.15	27.594217	12.779200	12.708227*
	0.10	0.05	24.153683	11.035019	10.949667*
		0.10	26.954338	12.835380	12.776719*
		0.15	28.589289	13.270930	13.216690*
	0.15	0.05	25.878532	12.091100	12.014440*
		0.10	27.470036	13.693668	13.628117*
		0.15	30.412638	14.191790	14.139732*
70	0.05	0.05	20.543040	8.874884	8.795013*
		0.10	22.441700	10.004057	9.931753*
		0.15	25.526341	12.001964	11.947479*
	0.10	0.05	21.519452	9.473498	9.394789*
		0.10	24.318921	11.038239	10.979635*
		0.15	26.881754	12.759061	12.696708*
	0.15	0.05	23.771353	10.870661	10.840975*
		0.10	24.792009	11.263021	11.204843*
		0.15	29.221550	13.409062	13.349729*

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่2.2 (ต่อ)

ขนาดตัวอย่าง N	PX1	PX2	วิธีการ		
			ML	WMLCH	WMLRC
80	0.05	0.05	19.378906	8.088790	8.015581*
		0.10	21.680660	9.501014	9.441853*
		0.15	24.023154	11.112469	11.056170*
	0.10	0.05	20.892576	8.988135	8.917597*
		0.10	22.846631	10.212426	10.164362*
		0.15	25.776429	12.115988	12.068466*
	0.15	0.05	21.555562	9.493821	9.436042*
		0.10	23.574959	10.620637	10.573905*
		0.15	27.617046	12.761630	12.724002*
90	0.05	0.05	18.077716	7.574930	7.522664*
		0.10	20.887032	9.013907	8.968044*
		0.15	22.303745	10.273947	10.233135*
	0.10	0.05	19.552474	8.208218	8.154448*
		0.10	22.130940	10.005640	9.964394*
		0.15	24.715040	11.813551	11.770646*
	0.15	0.05	20.553079	8.771839	8.730637*
		0.10	22.567240	10.150605	10.111933*
		0.15	25.749314	12.154030	12.114766*
100	0.05	0.05	17.351440	7.117500	7.069777*
		0.10	19.255033	8.153715	8.111430*
		0.15	21.407075	9.695772	9.654213*
	0.10	0.05	18.465237	7.845428	7.796806*
		0.10	20.637382	8.931495	8.896269*
		0.15	23.024985	10.965689	10.933794*
	0.15	0.05	19.335482	8.229220	8.183140*
		0.10	21.149598	9.228179	9.193074*
		0.15	24.336848	11.542077	11.516719*

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่2.3 การเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ รุนแรง และตัวแปรอิสระ x_2 มีค่าผิดปกติระดับ ไม่รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1) และตัวแปรอิสระ x_2 (PX2)

ขนาดตัวอย่าง N	PX1	PX2	วิธีการ		
			ML	WMLCH	WMLRC
20	0.05	0.05	33.951176	16.898509	16.389928*
		0.10	36.188958	17.963373	17.575732*
		0.15	38.097785	19.153176	18.749367*
	0.10	0.05	36.191989	18.453936	18.128642*
		0.10	37.656251	19.595857	19.217643*
		0.15	38.498032	19.952587	19.729656*
	0.15	0.05	38.436651	20.091411	19.897781*
		0.10	39.418822	20.352257	20.137892*
		0.15	40.175238	20.679085	20.592245*
30	0.05	0.05	29.718050	13.436000	13.377190*
		0.10	32.752166	14.996382	14.892759*
		0.15	34.485590	16.362230	16.282680*
	0.10	0.05	32.101320	15.228720	15.169260*
		0.10	34.013941	16.548612	16.462861*
		0.15	35.319430	17.548960	17.476420*
	0.15	0.05	35.960960	18.737250	18.645040*
		0.10	36.468562	18.950447	18.865567*
		0.15	37.517986	19.339694	19.241621*
40	0.05	0.05	27.899120	12.181520	12.091259*
		0.10	30.396117	14.097758	14.016525*
		0.15	33.393010	15.591030	15.553240*
	0.10	0.05	30.462530	14.518900	14.428700*
		0.10	31.637560	15.384010	15.293880*
		0.15	33.573840	16.514820	16.421360*
	0.15	0.05	33.535880	16.985870	16.892440*
		0.10	34.532790	17.615860	17.536500*
		0.15	35.273460	17.962300	17.880676*

***** หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่2.3 (ต่อ)

ขนาดตัวอย่าง N	PX1	PX2	วิธีการ		
			ML	WMLCH	WMLRC
50	0.05	0.05	25.573810	11.094850	11.007117*
		0.10	27.830550	12.973530	12.896610*
		0.15	29.934960	13.688810	13.641379*
	0.10	0.05	28.195860	13.597260	13.583290*
		0.10	30.252480	14.854850	14.768489*
		0.15	31.494070	15.587290	15.505570*
	0.15	0.05	30.900220	15.049908	14.966682*
		0.10	31.919270	15.798937	15.727804*
		0.15	33.526990	16.677840	16.599771*
60	0.05	0.05	23.627260	10.399650	10.328580*
		0.10	26.420170	12.303060	12.232920*
		0.15	28.490650	12.921159	12.844219*
	0.10	0.05	26.468670	12.761983	12.692349*
		0.10	28.763360	13.908040	13.847915*
		0.15	29.455490	14.242710	14.172301*
	0.15	0.05	28.310920	13.286713	13.213967*
		0.10	30.716350	14.941613	14.887726*
		0.15	31.586560	15.471491	15.400038*
70	0.05	0.05	21.150830	8.969489	8.890897*
		0.10	24.256760	10.515014	10.448519*
		0.15	25.646972	11.097610	11.028563*
	0.10	0.05	23.244570	10.814805	10.751369*
		0.10	26.273110	12.042510	11.982255*
		0.15	26.902160	12.272620	12.214493*
	0.15	0.05	25.586830	11.378674	11.305830*
		0.10	28.683640	13.380683	13.328018*
		0.15	30.884750	15.092181	15.037670*

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่2.3 (ต่อ)

ขนาดตัวอย่าง N	PX1	PX2	วิธีการ		
			ML	WMLCH	WMLRC
80	0.05	0.05	19.179690	7.827116	7.773952*
		0.10	21.704030	9.179101	9.123462*
		0.15	23.379950	10.114220	10.053457*
	0.10	0.05	21.216660	9.475404	9.411906*
		0.10	23.602450	10.749629	10.690463*
		0.15	24.927040	11.625794	11.576034*
	0.15	0.05	24.256060	10.829587	10.775068*
		0.10	26.645120	12.593234	12.547846*
		0.15	28.513320	13.668134	13.625289*
90	0.05	0.05	18.122570	7.322342	7.273575*
		0.10	19.832910	8.434541	8.387762*
		0.15	21.038470	9.189035	9.131823*
	0.10	0.05	19.992420	8.691350	8.641492*
		0.10	22.091910	9.970036	9.930764*
		0.15	23.602130	10.818154	10.779481*
	0.15	0.05	22.215980	9.920142	9.887481*
		0.10	25.301500	11.924903	11.880341*
		0.15	26.745120	12.997667	12.955791*
100	0.05	0.05	16.986340	6.800396	6.757505*
		0.10	18.461670	7.758979	7.706951*
		0.15	19.807960	8.714109	8.670077*
	0.10	0.05	18.820430	8.041220	8.000001*
		0.10	20.471360	8.759811	8.712969*
		0.15	22.161310	10.059208	10.021130*
	0.15	0.05	21.053380	9.267895	9.232297*
		0.10	23.569010	11.318500	11.280632*
		0.15	25.215510	12.187458	12.154246*

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่2.4 การเปรียบเทียบค่า AMSE ของวิธี ML วิธี WMLCH และวิธี WMLRC เมื่อตัวแปรอิสระ x_1 มีค่าผิดปกติระดับ รุนแรง และตัวแปรอิสระ x_2 มีค่าผิดปกติระดับ รุนแรง โดยจำแนกตามสัดส่วนการปลอมปนของตัวแปรอิสระ x_1 (PX1) และตัวแปรอิสระ x_2 (PX2)

ขนาดตัวอย่าง N	PX1	PX2	วิธีการ		
			ML	WMLCH	WMLRC
20	0.05	0.05	75.151839	17.867299	17.525671*
		0.10	77.572310	18.886851	18.246996*
		0.15	80.873216	20.754951	20.121268*
	0.10	0.05	78.532991	19.157233	18.942021*
		0.10	80.014664	19.979135	19.865328*
		0.15	82.178488	21.479857	21.323122*
	0.15	0.05	80.927986	19.327969	19.312645*
		0.10	82.974715	20.631164	20.529639*
		0.15	85.126735	22.225961	22.131863*
30	0.05	0.05	64.416991	15.976828	15.825765*
		0.10	66.339830	16.709620	16.615930*
		0.15	69.718836	18.933187	18.878651*
	0.10	0.05	67.142020	17.423050	17.383680*
		0.10	69.542622	18.242589	18.152743*
		0.15	71.241751	19.147278	19.051655*
	0.15	0.05	70.524432	18.011899	17.937638*
		0.10	72.636203	19.539670	19.440650*
		0.15	73.952461	20.457978	20.374675*
40	0.05	0.05	55.912260	14.790600	14.706463*
		0.10	58.763831	16.021880	15.938416*
		0.15	60.286929	17.001600	16.926613*
	0.10	0.05	59.059860	16.121580	16.039340*
		0.10	60.415121	17.017100	16.934768*
		0.15	60.940436	18.175880	18.118581*
	0.15	0.05	60.362730	17.185910	17.104000*
		0.10	61.966060	17.791440	17.738980*
		0.15	62.825691	18.236390	18.171044*

***** หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่2.4 (ต่อ)

ขนาดตัวอย่าง N	PX1	PX2	วิธีการ		
			ML	WMLCH	WMLRC
50	0.05	0.05	47.691615	12.969910	12.886259*
		0.10	51.721985	15.405980	15.333728*
		0.15	54.505583	16.374160	16.301014*
	0.10	0.05	50.960998	15.452169	15.392550*
		0.10	53.992073	16.286247	16.223910*
		0.15	55.012310	16.724710	16.668448*
	0.15	0.05	55.417880	16.794612	16.729142*
		0.10	56.813260	17.352610	17.295820*
		0.15	57.341280	17.695376	17.632025*
60	0.05	0.05	37.038040	12.493899	12.443882*
		0.10	39.385701	14.184805	14.133743*
		0.15	40.597523	15.796313	15.745548*
	0.10	0.05	40.175240	14.349910	14.309506*
		0.10	41.603236	15.409929	15.352713*
		0.15	42.563684	15.957321	15.909941*
	0.15	0.05	41.955020	16.032393	15.982495*
		0.10	43.470737	16.708163	16.662767*
		0.15	44.221500	17.139705	17.081049*
70	0.05	0.05	32.573186	11.706535	11.669538*
		0.10	34.789041	13.273802	13.234574*
		0.15	36.507582	14.493518	14.456444*
	0.10	0.05	35.526170	13.807029	13.761591*
		0.10	36.923437	14.504396	14.450207*
		0.15	37.821317	15.306160	15.264636*
	0.15	0.05	37.940600	15.631191	15.599602*
		0.10	38.981551	16.197230	16.158677*
		0.15	39.427640	16.558706	16.522559*

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ตารางที่2.4 (ต่อ)

ขนาดตัวอย่าง N	PX1	PX2	วิธีการ		
			ML	WMLCH	WMLRC
80	0.05	0.05	30.409190	11.565139	11.527858*
		0.10	31.798454	11.996054	11.961464*
		0.15	34.609086	13.235862	13.201778*
	0.10	0.05	32.291350	12.150399	12.123690*
		0.10	34.123124	13.304990	13.268850*
		0.15	36.197822	14.501316	14.473902*
	0.15	0.05	35.001350	13.978411	13.947904*
		0.10	36.492888	14.576072	14.554529*
		0.15	37.699504	15.479913	15.442180*
90	0.05	0.05	28.099841	9.973230	9.941752*
		0.10	30.700502	11.553099	11.520010*
		0.15	32.610142	12.782133	12.765844*
	0.10	0.05	30.601590	11.471655	11.449066*
		0.10	32.840240	12.998113	12.969278*
		0.15	34.109809	13.901325	13.885011*
	0.15	0.05	33.744730	13.322671	13.303612*
		0.10	34.502894	13.962001	13.947277*
		0.15	35.372912	14.528511	14.509410*
100	0.05	0.05	27.141284	9.447862	9.413485*
		0.10	29.371626	10.790913	10.760518*
		0.15	31.285374	12.076220	12.050798*
	0.10	0.05	29.142626	10.351870	10.328123*
		0.10	31.826202	12.551561	12.538253*
		0.15	32.741671	13.213323	13.194370*
	0.15	0.05	32.217988	12.740505	12.728718*
		0.10	32.833351	13.018727	13.001815*
		0.15	33.951192	13.811262	13.805672*

* หมายถึง วิธีที่ทำการประมาณค่าพารามิเตอร์ได้ค่า AMSE ต่ำที่สุด

ภาคผนวก จ

โปรแกรมแสดงการคำนวณค่าพารามิเตอร์ในตัวแบบการถดถอยโลจิสติคของวิธี ML วิธี WMLCH และวิธี WMLRC

```
      PROGRAM MAIN
      COMMON/SEED/IX,KK
      REAL VARX1,VARX2,X(100,3),BETA(3),W(100),RDW2(100)
      REAL BE_ML(3),SMSE_ML,BE_WMLCH(3),SMSE_WMLCH,BE_WMLRC(3),SMSE_WMLRC
      REAL AMSE_ML, AMSE_WMLCH, AMSE_WMLRC
      INTEGER N,MEANX1,MEANX2,IX,KK,RDW1(100),Y(100)
      DOUBLE PRECISION BOLS(3),PIX(100),EAD,ADDER
      OPEN(1,FILE='C:\AMSE.XLS')
      PRINT*,'NUMBER OF EXAMPLE='
      READ*,N
99    PRINT*,'CASE OF CONTAMINATION='
      READ*,T
      IX=65539
      KK=0
      BETA(1)=5.0
      BETA(2)=-1.0
      BETA(3)=1.0
      MEANX1=15
      MEANX2=10
      VARX1=4
      VARX2=2
      IF(T.EQ.1)THEN
      CALL GENX1CON(MEANX1,MEANX2,VARX1,VARX2,N,X)
      ELSE
            IF(T.EQ.2)THEN
            CALL GENX1X2CON(MEANX1,MEANX2,VARX1,VARX2,N,X)
            ELSE
                  IF(T.EQ.3)THEN
                  CALL GENX1X2NOR(MEANX1,MEANX2,VARX1,VARX2,N,X)
                  ELSE
                  PRINT*,'MISTAKE OF CASE OF CONTAMINATION'
                  GOTO 99
                  END IF
```

```
            END IF
      END IF
      DO 91 I=1,N
      ADDER=BETA(1)+(BETA(2)*X(I,1))+(BETA(3)*X(I,2))
      EAD=DEXP(ADDER)
      PIX(I)=EAD/(1+EAD)
91    CONTINUE
      CALL MCD(X,N,W)
      CALL W_MLCH(N,W,RDW1)
      CALL W_MLRC(N,W,RDW2)
      DO 999 REPEAT=1,1000
      CALL GENY(N,PIX,Y)
      CALL OLS(N,X,Y,BOLS)
      CALL ML(BETA,BOLS,N,Y,X,BE_ML,SMSE_ML)
      CALL WMLCH(BETA,BOLS,RDW1,N,Y,X,BE_WMLCH,SMSE_WMLCH)
      CALL WMLRC(BETA,BOLS,RDW2,N,Y,X,BE_WMLRC,SMSE_WMLRC)
999   CONTINUE
      AMSE_ML=SMSE_ML/3
      AMSE_WMLCH=SMSE_WMLCH/3
      AMSE_WMLRC=SMSE_WMLRC/3
      WRITE(1,*)'AMSE',AMSE_ML,AMSE_WMLCH,AMSE_WMLRC
      STOP
      END
```

FIND PARAMETERS IN WEIGHT MAXIMUM LIKELIHOOD METHOD OF RC

```
      SUBROUTINE WMLRC(BETA,BOLS,RDW2,N,Y,X,BE_WMLRC,SMSE_WMLRC)
      INTEGER Y(100)
      REAL BETA(3),YTEL(100),X(100,3),RDW2(100),BE_WMLRC (3),SMSE_WMLRC
      DOUBLE PRECISION PIXI,EADD,AD,UWE(3,1),HWE(3,3),HWEI(3,3),HIU_WE(3,1)
      DOUBLE PRECISION BET0,BET1,BET2,B0,B1,B2,ER0,ER1,ER2,BOLS(3)
      B0=BOLS(1)
      B1=BOLS(2)
      B2=BOLS(3)
      BET0=0.0
      BET1=0.0
      BET2=0.0
      PIXI=0.0
      BE_WMLRC(1)=0.0
```

```
      BE_WMLRC(2)=0.0
      BE_WMLRC(3)=0.0
      ICOUNT=0
      SUM=0
      DO 80 I=1,N
      IF(Y(I).EQ.1) THEN
      SUM=SUM+1
      END IF
80    CONTINUE
      LAM_BAR=SUM/N
      IF(LAM_BAR.LT.0.99) THEN
      MIN1=LAM_BAR
      ELSE
            MIN1=0.99
      END IF
      IF(MIN1.GT.0.01) THEN
      LAM_HAT=MIN1
      ELSE
            LAM_HAT=0.01
      END IF
      DELTA0=(LAM_HAT*0.01)/(1+0.01)
      DELTA1=(1+(LAM_HAT*0.01))/(1+0.01)
      DO 81 I=1,N
      YTEL(I)=((1-Y(I))*DELTA0)+(Y(I)*DELTA1)
81    CONTINUE
      DO 82 I=1,3
      UWE(I,1)=0.0
      DO 82 J=1,3
      HWE(I,J)=0.0
82    CONTINUE
86    DO 83 L=1,N
      AD=B0+(B1*X(L,1))+(B2*X(L,2))
      EADD=DEXP(AD)
      PIXI=EADD/(1+EADD)
      DO 84 I=1,3
      IF(I.EQ.1) THEN
      UWE(I,1)=UWE(I,1)+(RDW2(L)*(YTEL(L)-PIXI))
      ELSE
```

```
            IF(I.EQ.2) THEN
            UWE(I,1)=UWE(I,1)+(RDW2(L)*X(L,1)*(YTEL(L)-PIXI))
            ELSE
                  UWE(I,1)=UWE(I,1)+(RDW2(L)*X(L,2)*(YTEL(L)-PIXI))
            END IF
      END IF
      DO 84 J=1,3
      IF ((I.EQ.1).AND.(J.EQ.1)) THEN
      HWE(I,J)=HWE(I,J)+((-1)*RDW2(L)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.1).AND.(J.EQ.2)) THEN
      HWE(I,J)=HWE(I,J)+((-1)*RDW2(L)*X(L,1)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.1).AND.(J.EQ.3)) THEN
      HWE(I,J)=HWE(I,J)+((-1)*RDW2(L)*X(L,2)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.2).AND.(J.EQ.1)) THEN
      HWE(I,J)=HWE(I,J)+((-1)*RDW2(L)*X(L,1)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.2).AND.(J.EQ.2)) THEN
      HWE(I,J)=HWE(I,J)+((-1)*RDW2(L)*(X(L,1)**2)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.2).AND.(J.EQ.3)) THEN
      HWE(I,J)=HWE(I,J)+((-1)*RDW2(L)*X(L,1)*X(L,2)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.3).AND.(J.EQ.1)) THEN
      HWE(I,J)=HWE(I,J)+((-1)*RDW2(L)*X(L,2)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.3).AND.(J.EQ.2)) THEN
      HWE(I,J)=HWE(I,J)+((-1)*RDW2(L)*X(L,1)*X(L,2)*PIXI*(1-PIXI))
      ELSE
            HWE(I,J)=HWE(I,J)+((-1)*RDW2(L)*(X(L,2)**2)*PIXI*(1-PIXI))
      END IF
84    CONTINUE
83    CONTINUE
      CALL INV3(HWE,HWEI)
      CALL MULT(HWEI,UWE,HIU_WE)
      BET0=(B0-HIU_WE(1,1))
      BET1=(B1-HIU_WE(2,1))
      BET2=(B2-HIU_WE(3,1))
      ER0=ABS(B0-BET0)
      ER1=ABS(B1-BET1)
      ER2=ABS(B2-BET2)
      IF((ER0.LT.0.00001).AND.(ER1.LT.0.00001).AND.(ER2.LT.0.00001)) THEN
      BE_WMLRC(1)=BET0
```

```
      BE_WMLRC(2)=BET1
      BE_WMLRC(3)=BET2
      DO 89 I=1,3
89    CONTINUE
      ELSE
            DO 85 I=1,3
            UWE(I,1)=0.0
            DO 85 J=1,3
            HWE(I,J)=0.0
85          CONTINUE
            B0=BET0
            B1=BET1
            B2=BET2
            GOTO 86
      END IF
      CALL MSE_WMLRC(BETA,BE_WMLRC,SMSE_WMLRC)
      RETURN
      END

      FIND MEAN SQUARE ERROR OF WMLRC
      SUBROUTINE MSE_WMLRC(BETA,BE_WMLRC,SMSE_WMLRC)
      REAL BE_WMLRC (3),SMSE_WMLRC,DIF_WMLRC(3),TMSE_WMLRC(3),BETA(3)
      DO 87 I=1,3
      DIF_WMLRC(I)=(BE_WMLRC(I)-(BETA(I)))**2
      TMSE_WMLRC(I)=0.001*DIF_WMLRC(I)
87    CONTINUE
      DO 88 I=1,3
      SMSE_WMLRC=SMSE_WMLRC+TMSE_WMLRC(I)
88    CONTINUE
      RETURN
      END

      FIND PARAMETERS IN WEIGHT MAXIMUM LIKELIHOOD METHOD OF CH
      SUBROUTINE WMLCH(BETA,BOLS,RDW1,N,Y,X,BE_WMLCH,SMSE_WMLCH)
      INTEGER Y(100),RDW1(100)
      REAL BETA(3),X(100,3),BE_WMLCH (3),SMSE_WMLCH
      DOUBLE PRECISION PIXI,EADD,AD,UW(3,1),HW(3,3),HWI(3,3),HIU_W(3,1)
      DOUBLE PRECISION BET0,BET1,BET2,B0,B1,B2,ER0,ER1,ER2,BOLS(3)
```

```
      B0=BOLS(1)
      B1=BOLS(2)
      B2=BOLS(3)
      BET0=0.0
      BET1=0.0
      BET2=0.0
      PIXI=0.0
      BE_WMLCH(1)=0.0
      BE_WMLCH(2)=0.0
      BE_WMLCH(3)=0.0
      ICOUNT=0
      DO 72 I=1,3
      UW(I,1)=0.0
      DO 72 J=1,3
      HW(I,J)=0.0
72    CONTINUE
76    DO 73 L=1,N
      AD=B0+(B1*X(L,1))+(B2*X(L,2))
      EADD=DEXP(AD)
      PIXI=EADD/(1+EADD)
      DO 74 I=1,3
      IF(I.EQ.1) THEN
      UW(I,1)=UW(I,1)+(RDW1(L)*(Y(L)-PIXI))
      ELSE
            IF(I.EQ.2) THEN
            UW(I,1)=UW(I,1)+(RDW1(L)*X(L,1)*(Y(L)-PIXI))
            ELSE
                  UW(I,1)=UW(I,1)+(RDW1(L)*X(L,2)*(Y(L)-PIXI))
            END IF
      END IF
      DO 74 J=1,3
      IF ((I.EQ.1).AND.(J.EQ.1)) THEN
      HW(I,J)=HW(I,J)+((-1)*RDW1(L)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.1).AND.(J.EQ.2)) THEN
      HW(I,J)=HW(I,J)+((-1)*RDW1(L)*X(L,1)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.1).AND.(J.EQ.3)) THEN
      HW(I,J)=HW(I,J)+((-1)*RDW1(L)*X(L,2)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.2).AND.(J.EQ.1)) THEN
```

```
      HW(I,J)=HW(I,J)+((-1)*RDW1(L)*X(L,1)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.2).AND.(J.EQ.2)) THEN
      HW(I,J)=HW(I,J)+((-1)*RDW1(L)*(X(L,1)**2)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.2).AND.(J.EQ.3)) THEN
      HW(I,J)=HW(I,J)+((-1)*RDW1(L)*X(L,1)*X(L,2)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.3).AND.(J.EQ.1)) THEN
      HW(I,J)=HW(I,J)+((-1)*RDW1(L)*X(L,2)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.3).AND.(J.EQ.2)) THEN
      HW(I,J)=HW(I,J)+((-1)*RDW1(L)*X(L,1)*X(L,2)*PIXI*(1-PIXI))
      ELSE
            HW(I,J)=HW(I,J)+((-1)*RDW1(L)*(X(L,2)**2)*PIXI*(1-PIXI))
      END IF
74    CONTINUE
73    CONTINUE
      CALL INV3(HW,HWI)
      CALL MULT(HWI,UW,HIU_W)
      BET0=(B0-HIU_W(1,1))
      BET1=(B1-HIU_W(2,1))
      BET2=(B2-HIU_W(3,1))
      ER0=ABS(B0-BET0)
      ER1=ABS(B1-BET1)
      ER2=ABS(B2-BET2)
      IF((ER0.LT.0.00001).AND.(ER1.LT.0.00001).AND.(ER2.LT.0.00001)) THEN
      BE_WMLCH(1)=BET0
      BE_WMLCH(2)=BET1
      BE_WMLCH(3)=BET2
      DO 79 I=1,3
79    CONTINUE
      ELSE
            DO 75 I=1,3
            UW(I,1)=0.0
            DO 75 J=1,3
            HW(I,J)=0.0
75          CONTINUE
            B0=BET0
            B1=BET1
            B2=BET2
            GOTO 76
```

```
      END IF
      CALL MSE_WMLCH(BETA,BE_WMLCH,SMSE_WMLCH)
      RETURN
      END

      FIND MEAN SQUARE ERROR OF WMLCH
      SUBROUTINE MSE_WMLCH(BETA,BE_WMLCH,SMSE_WMLCH)
      REAL BE_WMLCH(3),SMSE_WMLCH,DIF_WMLCH(3),TMSE_WMLCH(3),BETA(3)
      DO 77 I=1,3
      DIF_WMLCH(I)=(BE_WMLCH(I)-(BETA(I)))**2
      TMSE_WMLCH(I)=0.001*DIF_WMLCH(I)
77    CONTINUE
      DO 78 I=1,3
      SMSE_WMLCH=SMSE_WMLCH+TMSE_WMLCH(I)
78    CONTINUE
      RETURN
      END

      FIND PARAMETERS IN MAXIMUM LIKELIHOOD METHOD
      SUBROUTINE ML(BETA,BOLS,N,Y,X,BE_ML,SMSE_ML)
      INTEGER Y(100)
      REAL BETA(3),X(100,3),BE_ML(3),SMSE_ML
      DOUBLE PRECISION PIXI,EADD,AD,U(3,1),H(3,3),HI(3,3),HIU(3,1)
      DOUBLE PRECISION BET0,BET1,BET2,B0,B1,B2,ER0,ER1,ER2,BOLS(3)
      B0=BOLS(1)
      B1=BOLS(2)
      B2=BOLS(3)
      BET0=0.0
      BET1=0.0
      BET2=0.0
      PIXI=0.0
      BE_ML(1)=0.0
      BE_ML(2)=0.0
      BE_ML(3)=0.0
      ICOUNT=0
      DO 59 I=1,3
      U(I,1)=0.0
      DO 59 J=1,3
```

```
      H(I,J)=0.0
59    CONTINUE
63    DO 60 L=1,N
      AD=B0+(B1*X(L,1))+(B2*X(L,2))
      EADD=DEXP(AD)
      PIXI=EADD/(1+EADD)
      DO 61 I=1,3
      IF(I.EQ.1) THEN
      U(I,1)=U(I,1)+(Y(L)-PIXI)
      ELSE
            IF(I.EQ.2) THEN
            U(I,1)=U(I,1)+(X(L,1)*(Y(L)-PIXI))
            ELSE
                  U(I,1)=U(I,1)+(X(L,2)*(Y(L)-PIXI))
            END IF
      END IF
      DO 61 J=1,3
      IF ((I.EQ.1).AND.(J.EQ.1)) THEN
      H(I,J)=H(I,J)+((-1)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.1).AND.(J.EQ.2)) THEN
      H(I,J)=H(I,J)+((-1)*X(L,1)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.1).AND.(J.EQ.3)) THEN
      H(I,J)=H(I,J)+((-1)*X(L,2)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.2).AND.(J.EQ.1)) THEN
      H(I,J)=H(I,J)+((-1)*X(L,1)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.2).AND.(J.EQ.2)) THEN
      H(I,J)=H(I,J)+((-1)*(X(L,1)**2)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.2).AND.(J.EQ.3)) THEN
      H(I,J)=H(I,J)+((-1)*X(L,1)*X(L,2)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.3).AND.(J.EQ.1)) THEN
      H(I,J)=H(I,J)+((-1)*X(L,2)*PIXI*(1-PIXI))
      ELSE IF ((I.EQ.3).AND.(J.EQ.2)) THEN
      H(I,J)=H(I,J)+((-1)*X(L,1)*X(L,2)*PIXI*(1-PIXI))
      ELSE
            H(I,J)=H(I,J)+((-1)*(X(L,2)**2)*PIXI*(1-PIXI))
      END IF
61    CONTINUE
60    CONTINUE
```

```
          CALL INV3(H,HI)
          CALL MULT(HI,U,HIU)
          BET0=(B0-HIU(1,1))
          BET1=(B1-HIU(2,1))
          BET2=(B2-HIU(3,1))
          ER0=ABS(B0-BET0)
          ER1=ABS(B1-BET1)
          ER2=ABS(B2-BET2)
          ICOUNT=ICOUNT+1
          IF (((ER0.LT.0.00001).AND.(ER1.LT.0.00001)).AND.(ER2.LT.0.00001)) THEN
          BE_ML(1)=BET0
          BE_ML(2)=BET1
          BE_ML(3)=BET2
          ELSE
                    DO 62 I=1,3
                    U(I,1)=0.0
                    DO 62 J=1,3
                    H(I,J)=0.0
62                  CONTINUE
                    B0=BET0
                    B1=BET1
                    B2=BET2
                    GOTO 63
          END IF
          CALL MSE_ML(BETA,BE_ML,SMSE_ML)
          RETURN
          END

          FIND MEAN SQUARE ERROR OF ML
          SUBROUTINE MSE_ML(BETA,BE_ML,SMSE_ML)
          REAL BE_ML(3),SMSE_ML,DIF_ML(3),TMSE_ML(3),BETA(3)
          DO 69 I=1,3
          DIF_ML(I)=(BE_ML(I)-(BETA(I)))**2
          TMSE_ML(I)=0.001*DIF_ML(I)
69        CONTINUE
          DO 70 I=1,3
          SMSE_ML=SMSE_ML+TMSE_ML(I)
70        CONTINUE
```

```
      RETURN
      END

      FIND PARAMETERS IN OLS METHOD
      SUBROUTINE OLS(N,X,Y,BOLS)
      INTEGER N,Y(100)
      REAL X(100,3),X_C(100,3),XT_C(3,100),XTY_C(3)
      DOUBLE PRECISION BOLS(3),XTX_C(3,3),XTXI_C(3,3)
      DO 105 I=1,N
      DO 106 J=1,3
      IF(J.EQ.1) THEN
      X_C(I,J)=1.0
      ELSE IF(J.GE.2) THEN
            X_C(I,J)=X(I,J-1)
      END IF
106   CONTINUE
105   CONTINUE
      DO 107 I=1,N
      DO 107 J=1,3
      XT_C(J,I)=X_C(I,J)
107   CONTINUE
      DO 108 I=1,3
      DO 108 J=1,3
      XTX_C(I,J)=0.0
      DO 108 K=1,N
      XTX_C(I,J)=XTX_C(I,J)+XT_C(I,K)*X_C(K,J)
108   CONTINUE
      DO 109 I=1,3
      XTY_C(I)=0.0
      DO 109 K=1,N
      XTY_C(I)=XTY_C(I)+XT_C(I,K)*Y(K)
109   CONTINUE
      CALL INV3(XTX_C,XTXI_C)
      DO 112 I=1,3
      BOLS(I)=0.0
      DO 112 K=1,3
      BOLS(I)=BOLS(I)+XTXI_C(I,K)*XTY_C(K)
112   CONTINUE
```

```
      RETURN
      END

      FIND INVERSE MATRIX FOR 3*3
      SUBROUTINE INV3(A,D)
      DOUBLE PRECISION A(3,3),D(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
      L=0
      DO 65 II=1,3
      DO 66 JJ=1,3
      IF((II.EQ.I).OR.(JJ.EQ.J)) THEN
      GOTO 66
      ELSE
            L=L+1
            E(L)=A(II,JJ)
      END IF
66    CONTINUE
65    CONTINUE
      SIGN=(-1)**(I+J)
      D(J,I)=(((E(1)*E(4))-(E(2)*E(3)))*SIGN)/DET
64    CONTINUE
      RETURN
      END

      FIND MULTIPLICATION
      SUBROUTINE MULT(A,B,C)
      DOUBLE PRECISION A(3,3),B(3,1),C(3,1)
      DO 67 I=1,3
      C(I,1)=0.0
67    CONTINUE
      DO 68 I=1,3
      DO 68 K=1,3
      C(I,1)=C(I,1)+(A(I,K)*B(K,1))
68    CONTINUE
```

```
      RETURN
      END

      GENERRATED Y
      SUBROUTINE GENY(N,PIX,Y)
      COMMON/SEED/IX,KK
      INTEGER IX,KK,N,Y(100)
      REAL ERR(100),YINTER(100),COU
      DOUBLE PRECISION PIX(100)
      DO 114 I=1,N
      CALL CONU(-1,1,COU)
      ERR(I)=COU
      YINTER(I)=PIX(I)+ERR(I)
114   CONTINUE
      DO 103 I=1,N
      IF (YINTER(I).LT.0.5) THEN
      Y(I)=0
      ELSE
            Y(I)=1
      END IF
103   CONTINUE
      RETURN
      END

      CONTINUE UNIFORM
      SUBROUTINE CONU(A,B,COU)
      COMMON/SEED/IX,KK
      INTEGER A,B
      REAL COU
      DOUBLE PRECISION IY
      CALL RANDOM(IX,IY,RD)
      AA=(B-A)*RD
      COU=A+AA
      RETURN
      END

      FIND RD OF WMLRC METHOD
      SUBROUTINE W_MLRC(N,W,RDW2)
      INTEGER N,POW(100)
```

```
      REAL W(100),RDW2(100),VW(100)
      DO 54 I=1,N
      VW(I)=W(I)
      POW(I)=I
54    CONTINUE
      CALL RANKW(N,VW,POW)
      B=(N+1)*0.75
      IB=B
      CC=VW(IB+1)-VW(IB)
      IF(B-IB.EQ.0.25) THEN
        C3=CC*0.25
        PL=VW(IB)+C3
      ELSE
              IF(B-IB.EQ.0.75) THEN
                C3=CC*0.75
          PL=VW(IB)+C3
              END IF
      END IF
      PER = PL
      DO 55 I=1,N
      IF(W(I)-PER.GT.0) THEN
      MDIF=W(I)
      DI_R=PER/MDIF
      RDW2(I)=DI_R
      ELSE
              DI_I=1
              RDW2(I)=DI_I
      END IF
55    CONTINUE
      RETURN
      END
```

RANK WEIGHT

```
      SUBROUTINE RANKW(N,VW,POW)
      INTEGER N,POW(100)
      REAL VW(100)
      N1=N-1
      DO 56 I=1,N1
```

```
      I1=I+1
      DO 57 K=I1,N
      IF(VW(I).LE.VW(K)) GOTO 57
      T=VW(I)
      VW(I)=VW(K)
      VW(K)=T
      HOLD=POW(I)
      POW(I)=POW(K)
      POW(K)=HOLD
57    CONTINUE
56    CONTINUE
      RETURN
      END

      FIND RD OF WMLCH METHOD
      SUBROUTINE W_MLCH(N,W,RDW1)
      INTEGER N,RDW1(100)
      REAL W(100)
      DO 53 I=1,N
      IF(W(I).LT.7.38) THEN
      A=1
      ELSE
      A=0
      END IF
      RDW1(I)=A
53    CONTINUE
      RETURN
      END

      FIND MCD OF X
      SUBROUTINE MCD(X,N,W)
      COMMON/SEED/IX,KK
      INTEGER N,Q(3),PO(500,100),H,PODET(500),PO10(10,100),POD(10),M(500,3)
      REAL X(100,3),W(100),DIS(500,100),DET(500),A(100),DET1(500),DIS10(10,100)
      REAL DETBE(10),DETCOM(10),DISCOM(10,100),DET10(10)
      CALL COMBINATION(X,N,M)
      H=3
      DO 14 I=1,500
```

```
      DO 15 J=1,H
      Q(J)=M(I,J)
15    CONTINUE
      CALL DISTANCE(X,N,H,Q,DETS,A)
      DO 16 K=1,N
      DIS(I,K)=A(K)
      PO(I,K)=K
16    CONTINUE
      DET(I)=DETS
14    CONTINUE
      CALL RANKD(N,DIS,PO)
      B=N*0.75
      IB=B
      IF(B-IB.GE.0.5)THEN
      H=IB+1
      ELSE
            H=IB
      END IF
      CALL CSTEP(X,N,H,DET,DIS,PO)
      CALL RANKD(N,DIS,PO)
      CALL CSTEP(X,N,H,DET,DIS,PO)
      DO 31 J=1,500
      DET1(J)=DET(J)
      PODET(J)=J
31    CONTINUE
      CALL RANKDET(DET1,PODET)
      DO 35 I=1,10
      DO 36 K=1,N
      DIS10(I,K)=DIS(PODET(I),K)
      PO10(I,K)=K
36    CONTINUE
      DET10(I)=DET1(I)
35    CONTINUE
      CALL RANKD10(N,DIS10,PO10)
      DO 41 I=1,10
      DETBE(I)=DET10(I)
43    CALL CSTEP10(X,N,H,I,DET10,DIS10,PO10)
      IF((DET10(I).EQ.0).OR.((DETBE(I)-DET10(I)).LT.0.000001)) THEN
```

```
      DETCOM(I)=DET10(I)
      DO 42 J=1,N
      DISCOM(I,J)=DIS10(I,J)
42    CONTINUE
      ELSE
          DETBE(I)=DET10(I)
          CALL RANKONE(N,I,DIS10,PO10)
          GOTO 43
      END IF
41    CONTINUE
      DO 48 J=1,10
      POD(J)=J
48    CONTINUE
      CALL RANKDET10(DETCOM,POD)
      DO 49 I=1,N
      W(I)=DISCOM(POD(1),I)
49    CONTINUE
      RETURN
      END

      RANK DETERMINANT 10 SET
      SUBROUTINE RANKDET10(DETCOM,POD)
      REAL DETCOM(10)
      INTEGER POD(10)
      DO 50 I=1,9
      I3=I+1
      DO 51 K=I3,10
      IF(DETCOM(I).LE.DETCOM(K)) GOTO 51
      T3=DETCOM(I)
      DETCOM(I)=DETCOM(K)
      DETCOM(K)=T3
      H3=POD(I)
      POD(I)=POD(K)
      POD(K)=H3
51    CONTINUE
50    CONTINUE
      RETURN
      END
```

RANK DISTANCE AND POSITION OF ONE SET

```
      SUBROUTINE RANKONE(N,I,DIS10,PO10)
      REAL DIS10(10,100)
      INTEGER PO10(10,100),N,I
      N1=N-1
      DO 46 J=1,N1
      J1=J+1
      DO 47 K=J1,N
      IF(DIS10(I,J).LE.DIS10(I,K)) GOTO 47
      T=DIS10(I,J)
      DIS10(I,J)=DIS10(I,K)
      DIS10(I,K)=T
      HOLD=PO10(I,J)
      PO10(I,J)=PO10(I,K)
      PO10(I,K)=HOLD
47    CONTINUE
46    CONTINUE
      RETURN
      END
```

CALCULATE DISTANCE H SUBSET 10 SET

```
      SUBROUTINE CSTEP10(X,N,H,I,DET10,DIS10,PO10)
      REAL DIS10(10,100),DETS,A(100),X(100,3),DET10(10)
      INTEGER PO10(10,100),N,H,I,POSI(50)
      DO 44 J=1,H
      POSI(J)=PO10(I,J)
44    CONTINUE
      CALL DISTANCE(X,N,H,POSI,DETS,A)
      DO 45 K=1,N
      DIS10(I,K)=A(K)
      PO10(I,K)=K
45    CONTINUE
      DET10(I)=DETS
      RETURN
      END
```

RANK DISTANCE AND POSITION

```
      SUBROUTINE  RANKD10(N,DIS10,PO10)
      REAL DIS10(10,100)
```

```
      INTEGER PO10(10,100),N
      DO 37 I=1,10
      N1=N-1
      DO 38 J=1,N1
      J1=J+1
      DO 39 K=J1,N
      IF(DIS10(I,J).LE.DIS10(I,K)) GOTO 39
      T=DIS10(I,J)
      DIS10(I,J)=DIS10(I,K)
      DIS10(I,K)=T
      HOLD=PO10(I,J)
      PO10(I,J)=PO10(I,K)
      PO10(I,K)=HOLD
39    CONTINUE
38    CONTINUE
37    CONTINUE
      RETURN
      END
```

RANK DETERMINANT

```
      SUBROUTINE RANKDET(DET1,PODET)
      REAL DET1(500)
      INTEGER PODET(500)
      DO 32 I=1,499
      I2=I+1
      DO 33 K=I2,500
      IF(DET1(I).LE.DET1(K)) GOTO 33
      T=DET1(I)
      DET1(I)=DET1(K)
      DET1(K)=T
      HOLD=PODET(I)
      PODET(I)=PODET(K)
      PODET(K)=HOLD
33    CONTINUE
32    CONTINUE
      RETURN
      END
```

```
      CALCULATE DISTANCE H SUBSET 500 SET
      SUBROUTINE CSTEP(X,N,H,DET,DIS,PO)
      INTEGER PO(500,100),PO1(50),H,N
      REAL DIS(500,100),DET(500),X(100,3),A(100),DETS
      DO 27 I=1,500
      DO 28 J=1,H
      PO1(J)=PO(I,J)
28    CONTINUE
      CALL DISTANCE(X,N,H,PO1,DETS,A)
      DO 29 K=1,N
      DIS(I,K)=A(K)
      PO(I,K)=K
29    CONTINUE
      DET(I)=DETS
27    CONTINUE
      RETURN
      END

      RANK DISTANCE AND POSITION
      SUBROUTINE RANKD(N,DIS,PO)
      REAL DIS(500,100)
      INTEGER N,PO(500,100)
      DO 23 I=1,500
      N1=N-1
      DO 24 J=1,N1
      J1=J+1
      DO 25 K=J1,N
      IF(DIS(I,J).LE.DIS(I,K)) GOTO 25
      T=DIS(I,J)
      DIS(I,J)=DIS(I,K)
      DIS(I,K)=T
      HOLD=PO(I,J)
      PO(I,J)=PO(I,K)
      PO(I,K)=HOLD
25    CONTINUE
24    CONTINUE
23    CONTINUE
      RETURN
```

```
      END

      FIND DISTANCE
      SUBROUTINE DISTANCE(X,N,H,R,DETS,A)
      INTEGER R(50),N,H
      REAL B(2,2),INVS(2,2),DIF(100,2),Z(2),A(100),X(100,3),DETS
      SUM_X1=0
      SUM_X2=0
      DO 17 J=1,H
      SUM_X1=SUM_X1+X(R(J),1)
      SUM_X2=SUM_X2+X(R(J),2)
17    CONTINUE
      X1_BAR=SUM_X1/H
      X2_BAR=SUM_X2/H
      DO 18 K=1,N
      DIF(K,1)=X(K,1)-X1_BAR
      DIF(K,2)=X(K,2)-X2_BAR
18    CONTINUE
      SUMX1_DIF=0
      SUMX1X2_DIF=0
      SUMX2_DIF=0
      DO 19 J=1,H
      DIFX1_SQ=(X(R(J),1)-X1_BAR)**2
      DIFX1X2_SQ=(X(R(J),1)-X1_BAR)*(X(R(J),2)-X2_BAR)
      DIFX2_SQ=(X(R(J),2)-X2_BAR)**2
      SUMX1_DIF=SUMX1_DIF+DIFX1_SQ
      SUMX1X2_DIF=SUMX1X2_DIF+DIFX1X2_SQ
      SUMX2_DIF=SUMX2_DIF+DIFX2_SQ
19    CONTINUE
      AVGX1_DIF=SUMX1_DIF/(H-1)
      AVGX1X2_DIF=SUMX1X2_DIF/(H-1)
      AVGX2_DIF=SUMX2_DIF/(H-1)
      B(1,1)=AVGX1_DIF
      B(1,2)=AVGX1X2_DIF
      B(2,1)=AVGX1X2_DIF
      B(2,2)=AVGX2_DIF
      DETS=(B(1,1)*B(2,2))-(B(2,1)*B(1,2))
      B(1,2)=-B(1,2)
```

```
      B(2,1)=-B(2,1)
      T=B(1,1)
      B(1,1)=B(2,2)
      B(2,2)=T
      DO 21 K=1,2
      DO 21 J=1,2
      INVS(K,J)=B(K,J)/DETS
21    CONTINUE
      DO 22 K=1,N
      Z(1)=(DIF(K,1)*INVS(1,1))+(DIF(K,2)*INVS(2,1))
      Z(2)=(DIF(K,1)*INVS(1,2))+(DIF(K,2)*INVS(2,2))
      A(K)=(Z(1)*DIF(K,1))+(Z(2)*DIF(K,2))
22    CONTINUE
      RETURN
      END
```

COMBINATION

```
      SUBROUTINE COMBINATION(X,N,M)
      COMMON/SEED/IX,KK
      REAL X(100,3)
      INTEGER N,S(3),M(500,3)
      CALL STAIL(X,N,S)
      DO 120 J=1,3
      M(1,J)=S(J)
120   CONTINUE
      DO 110 I=2,500
200   CALL STAIL(X,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
```

STAIL COMBINATION

```
      SUBROUTINE STAIL(X,N,S)
      INTEGER N,S(3)
      REAL X(100,3)
11    S(1)=DU(N)
100   S(2)=DU(N)
      IF(S(2).EQ.S(1)) GOTO 100
150   S(3)=DU(N)
      IF((S(3).EQ.S(1)).OR.(S(3).EQ.S(2))) GOTO 150
      CALL DET_3(X,S,DETC)
      IF(DETC.EQ.0) GOTO 11
      RETURN
      END
```

FIND DET OF 3 SUBSET

```
      SUBROUTINE DET_3(X,S,DETC)
      REAL X(100,3),C(2,2)
      INTEGER S(3)
      SUMX1=0
      SUMX2=0
      DO 8 J=1,3
      SUMX1=SUMX1+X(S(J),1)
      SUMX2=SUMX2+X(S(J),2)
8     CONTINUE
      X1BAR=SUMX1/3
      X2BAR=SUMX2/3
      X1_D=0
      X1X2_D=0
      X2_D=0
```

```
      DO 10 J=1,3
      DX1_SQ=(X(S(J),1)-X1BAR)**2
      DX1X2_SQ=(X(S(J),1)-X1BAR)*(X(S(J),2)-X2BAR)
      DX2_SQ=(X(S(J),2)-X2BAR)**2
      X1_D=X1_D+DX1_SQ
      X1X2_D=X1X2_D+DX1X2_SQ
      X2_D=X2_D+DX2_SQ
10    CONTINUE
      DX1_A=X1_D/2
      DX1X2_A=X1X2_D/2
      DX2_A=X2_D/2
      C(1,1)=DX1_A
      C(1,2)=DX1X2_A
      C(2,1)=DX1X2_A
      C(2,2)=DX2_A
      DETC=(C(1,1)*C(2,2))-(C(2,1)*C(1,2))
      RETURN
      END
```

DICRET UNIFORM

```
      FUNCTION DU(N)
      COMMON/SEED/IX,KK
      INTEGER N
      REAL T
      DOUBLE PRECISION IY
9     CALL RANDOM(IX,IY,RD)
      R=RD
      IF((R.EQ.0).OR.(R.EQ.1)) GOTO 9
      T=N*R
      IT=T
      IF(IT.EQ.0) THEN
      L=IT+1
      GOTO 7
      END IF
      IF(T-IT.GE.0.5) THEN
      L=IT+1
      ELSE
            L=IT
```

```
      END IF
7     DU=L
      RETURN
      END

      GENERATED  X1 X2 CASE THREE
      SUBROUTINE GENX1X2NOR(MEANX1,MEANX2,VARX1,VARX2,N,X)
      COMMON/SEED/IX,KK
      REAL X(100,3),X1K(100),X2K(100)
      INTEGER IX,KK,IKK1(100),IKK2(100)
      IKX1=0
929   IF(IKX1.EQ.1000)THEN
        WRITE(1,*)'STOP GEN NORMAL X1'
        STOP
      ELSE
              GOTO 925
      END IF
925   DO 911 I=1,N
      CALL NORMAL(MEANX1,VARX1,X1NOR)
      X(I,1)=X1NOR
911   CONTINUE
      DO 926 I=1,N
      X1K(I)=X(I,1)
      IKK1(I)=I
926   CONTINUE
      CALL XRANK(N,X1K,IKK1)
      CALL QUATILE(N,X1K,Q1KX1,Q3KX1)
      IQRKX1=Q3KX1-Q1KX1
      UKQ1=Q3KX1+3*IQRKX1
      LKQ1=Q1KX1-3*IQRKX1
      LKQ1I=Q1KX1-1.5*IQRKX1
      UKQ1I=Q3KX1+1.5*IQRKX1
      IKNOR1=0
      DO 927 I=1,N
      IF((X1K(I).GE.LKQ1I).AND.(X1K(I).LE.UKQ1I))    IKNOR1=IKNOR1+1
927   CONTINUE
      IF(IKNOR1.EQ.N)THEN
        GOTO 928
```

```
      ELSE
        IKX1=IKX1+1
              GOTO 929
      END IF
928   IKX2=0
829   IF(IKX2.EQ.1000)THEN
       WRITE(1,*)'STOP GEN NORMAL X2'
       STOP
      ELSE
              GOTO 825
      END IF
825   DO 811 I=1,N
      CALL NORMAL(MEANX2,VARX2,X2NOR)
      X(I,2)=X2NOR
811   CONTINUE
      DO 826 I=1,N
      X2K(I)=X(I,2)
      IKK2(I)=I
826   CONTINUE
      CALL XRANK(N,X2K,IKK2)
      CALL QUATILE(N,X2K,Q1KX2,Q3KX2)
      IQRKX2=Q3KX2-Q1KX2
      UKQ2=Q3KX2+3*IQRKX2
      LKQ2=Q1KX2-3*IQRKX2
      LKQ2I=Q1KX2-1.5*IQRKX2
      UKQ2I=Q3KX2+1.5*IQRKX2
      IKNOR2=0
      DO 827 I=1,N
      IF((X2K(I).GE.LKQ2I).AND.(X2K(I).LE.UKQ2I))    IKNOR2=IKNOR2+1
827   CONTINUE
      IF(IKNOR2.EQ.N)THEN
       GOTO 828
      ELSE
        IKX2=IKX2+1
              GOTO 829
      END IF
828   RETURN
      END
```

GENERATED X1 X2 CASE TWO

```
      SUBROUTINE GENX1X2CON(MEANX1,MEANX2,VARX1,VARX2,N,X)
      COMMON/SEED/IX,KK
      REAL X(100,3),CX1,PX1,CX2,PX2
      INTEGER IX,KK,NOX1,NOX2
      PRINT*,'SCALE FACTOR OF X1='
      READ*,CX1
      PRINT*,'SCALE FACTOR OF X2='
      READ*,CX2
      PRINT*,'PROPORTION OF CONTAMINATION X1='
      READ*,PX1
      PRINT*,'PROPORTION OF CONTAMINATION X2='
      READ*,PX2
      PRINT*,'AMOUNT OF OUTLIERS OF X1='
      READ*,NOX1
      PRINT*,'AMOUNT OF OUTLIERS OF X2='
      READ*,NOX2
      CALL GENX1(MEANX1,VARX1,CX1,PX1,NOX1,N,X)
      CALL GENX2(MEANX2,VARX2,CX2,PX2,NOX2,N,X)
      RETURN
      END
```

GENERATED X2 CONTAMINATED

```
      SUBROUTINE GENX2(MEANX2,VARX2,CX2,PX2,NOX2,N,X)
      COMMON/SEED/IX,KK
      REAL X(100,3),X2L(100),IQRX2
      INTEGER IO2(100)
      IRX2=0
512   IF(IRX2.EQ.1000)THEN
        STOP
      ELSE
        GOTO 513
      END IF
513   DO 514 I=1,N
      CALL CONTANOR(MEANX2,VARX2,CX2,PX2,X2CON)
      X(I,2)=X2CON
514   CONTINUE
      DO 515 I=1,N
```

```
      X2L(I)=X(I,2)
      IO2(I)=I
515   CONTINUE
      CALL XRANK(N,X2L,IO2)
      CALL QUATILE(N,X2L,Q1X2,Q3X2)
      IQRX2=Q3X2-Q1X2
      UQ2=Q3X2+3*IQRX2
      LQ2=Q1X2-3*IQRX2
      LQ2I=Q1X2-1.5*IQRX2
      UQ2I=Q3X2+1.5*IQRX2
      NOR2=N-NOX2
      IOUT2=0
      INOR2=0
      IF(CX2.EQ.5)THEN
       DO 516 I=1,N
IF(((X2L(I).GT.LQ2).AND.(X2L(I).LT.LQ2I)).OR.((X2L(I).GT.UQ2I).AND.(X2L(I).LT.UQ2)))   IOUT2=IOUT2+1
516    CONTINUE
       DO 530 I=1,N
       IF((X2L(I).GE.LQ2I).AND.(X2L(I).LE.UQ2I))   INOR2=INOR2+1
530    CONTINUE
       IF((IOUT2.EQ.NOX2).AND.(INOR2.EQ.NOR2))THEN
        GOTO 517
       ELSE
               IRX2=IRX2+1
               GOTO 512
       END IF
      ELSE IF(CX2.EQ.12)THEN
          DO 518 I=1,N
                IF((X2L(I).GT.UQ2).OR.(X2L(I).LT.LQ2))  IOUT2=IOUT2+1
518             CONTINUE
                DO 531 I=1,N
          IF((X2L(I).GE.LQ2I).AND.(X2L(I).LE.UQ2I))   INOR2=INOR2+1
531       CONTINUE
                IF((IOUT2.EQ.NOX2).AND.(INOR2.EQ.NOR2))THEN
                 GOTO 517
                ELSE
                  IRX2=IRX2+1
                  GOTO 512
```

```
         END IF
      ELSE
        WRITE(1,*)'OUT OF CASE LEVEL X2'
              STOP
      END IF
517   RETURN
      END
```

GENERATED X1 X2 CASE ONE

```
      SUBROUTINE GENX1CON(MEANX1,MEANX2,VARX1,VARX2,N,X)
      COMMON/SEED/IX,KK
      REAL X(100,3),CX1,PX1,X2B(100)
      INTEGER IX,KK,NOX1,IB2(100)
      PRINT*,'SCALE FACTOR OF X1='
      READ*,CX1
      PRINT*,'PROPORTION OF CONTAMINATION X1='
      READ*,PX1
      PRINT*,'AMOUNT OF OUTLIERS OF X1='
      READ*,NOX1
      CALL GENX1(MEANX1,VARX1,CX1,PX1,NOX1,N,X)
      IBX2=0
529   IF(IBX2.EQ.1000)THEN
       WRITE(1,*)'STOP GENERATED NORMAL X2'
       STOP
      ELSE
        GOTO 525
      END IF
525   DO 511 I=1,N
      CALL NORMAL(MEANX2,VARX2,X2NOR)
      X(I,2)=X2NOR
511   CONTINUE
      DO 526 I=1,N
      X2B(I)=X(I,2)
      IB2(I)=I
526   CONTINUE
      CALL XRANK(N,X2B,IB2)
      CALL QUATILE(N,X2B,Q1BX2,Q3BX2)
      IQRBX2=Q3BX2-Q1BX2
```

```
      UBQ2=Q3BX2+3*IQRBX2
      LBQ2=Q1BX2-3*IQRBX2
      LBQ2I=Q1BX2-1.5*IQRBX2
      UBQ2I=Q3BX2+1.5*IQRBX2
      IBNOR2=0
      DO 527 I=1,N
      IF((X2B(I).GE.LBQ2I).AND.(X2B(I).LE.UBQ2I))    IBNOR2=IBNOR2+1
527   CONTINUE
      IF(IBNOR2.EQ.N)THEN
       GOTO 528
      ELSE
        IBX2=IBX2+1
              GOTO 529
      END IF
528   RETURN
      END

      GENERATED  X1  CONTAMINATED
      SUBROUTINE GENX1(MEANX1,VARX1,CX1,PX1,NOX1,N,X)
      COMMON/SEED/IX,KK
      REAL X(100,3),X1L(100),IQRX1
      INTEGER IO1(100)
      IRX1=0
506   IF(IRX1.EQ.1000)THEN
       STOP
      ELSE
        GOTO 508
      END IF
508   DO 101 I=1,N
      CALL CONTANOR(MEANX1,VARX1,CX1,PX1,X1CON)
      X(I,1)=X1CON
101   CONTINUE
      DO 502 I=1,N
      X1L(I)=X(I,1)
      IO1(I)=I
502   CONTINUE
      CALL XRANK(N,X1L,IO1)
      CALL QUATILE(N,X1L,Q1X1,Q3X1)
```

```
      IQRX1=Q3X1-Q1X1
      UQ1=Q3X1+3*IQRX1
      LQ1=Q1X1-3*IQRX1
      LQ1I=Q1X1-1.5*IQRX1
      UQ1I=Q3X1+1.5*IQRX1
      NOR1=N-NOX1
      IOUT1=0
      INOR1=0
      IF(CX1.EQ.5)THEN
       DO 505 I=1,N
IF(((X1L(I).GT.LQ1).AND.(X1L(I).LT.LQ1I)).OR.((X1L(I).GT.UQ1I).AND.(X1L(I).LT.UQ1)))  IOUT1=IOUT1+1
505    CONTINUE
       DO 519 I=1,N
       IF((X1L(I).GE.LQ1I).AND.(X1L(I).LE.UQ1I))   INOR1=INOR1+1
519    CONTINUE
       IF((IOUT1.EQ.NOX1).AND.(INOR1.EQ.NOR1))THEN
        GOTO 510
       ELSE
               IRX1=IRX1+1
               GOTO 506
       END IF
      ELSE IF(CX1.EQ.12)THEN
          DO 507 I=1,N
                IF((X1L(I).GT.UQ1).OR.(X1L(I).LT.LQ1))  IOUT1=IOUT1+1
507             CONTINUE
                DO 520 I=1,N
          IF((X1L(I).GE.LQ1I).AND.(X1L(I).LE.UQ1I))   INOR1=INOR1+1
520       CONTINUE
                IF((IOUT1.EQ.NOX1).AND.(INOR1.EQ.NOR1))THEN
                 GOTO 510
                ELSE
                  IRX1=IRX1+1
                  GOTO 506
        END IF
      ELSE
        WRITE(1,*)'OUT OF CASE LEVEL X1'
              STOP
      END IF
```

```
510     RETURN
        END

        RANK X AND POSITION
        SUBROUTINE XRANK(N,RANKX,POX)
        REAL RANKX(100)
        INTEGER N,POX(100)
        N1=N-1
        DO 503 I=1,N1
        I1=I+1
        DO 504 K=I1,N
        IF(RANKX(I).LE.RANKX(K)) GOTO 504
          T=RANKX(I)
          RANKX(I)=RANKX(K)
          RANKX(K)=T
          HOLD=POX(I)
          POX(I)=POX(K)
          POX(K)=HOLD
504     CONTINUE
503     CONTINUE
        RETURN
        END

        FINE QUATILE Q1 Q3
        SUBROUTINE QUATILE(N,XL,Q1,Q3)
        REAL Q1,Q3,XL(100)
        INTEGER N
        VQ1=(N+1)*0.25
        IVQ1=VQ1
        AA=XL(IVQ1+1)-XL(IVQ1)
        VQ3=(N+1)*0.75
        IVQ3=VQ3
        BB=XL(IVQ3+1)-XL(IVQ3)
        IF(VQ1-IVQ1.EQ.0.25) THEN
          A1=AA*0.25
          Q1=XL(IVQ1)+A1
        ELSE
                IF(VQ1-IVQ1.EQ.0.75) THEN
```

```
    A1=AA*0.75
    Q1=XL(IVQ1)+A1
          END IF
END IF
IF(VQ3-IVQ3.EQ.0.25) THEN
  A3=BB*0.25
  Q3=XL(IVQ3)+A3
ELSE
          IF(VQ3-IVQ3.EQ.0.75) THEN
           A3=BB*0.75
    Q3=XL(IVQ3)+A3
          END IF
END IF
RETURN
END
```

CONTAMINATED NORMAL DISTRIBUTION

```
SUBROUTINE CONTANOR(MEAN,VAR,C,P,CON)
COMMON/SEED/IX,KK
REAL VAR,C,P,CVAR
INTEGER IX,KK,MEAN
DOUBLE PRECISION IY
CVAR=(C**2)*VAR
CALL RANDOM(IX,IY,RD)
IF(RD.LE.P)THEN
  CALL NORMAL(MEAN,CVAR,XNOR)
  CON=XNOR
ELSE
     CALL NORMAL(MEAN,VAR,XNOR)
     CON=XNOR
END IF
RETURN
END
```

NORMAL DISTRIBUTION

```
SUBROUTINE NORMAL(NMEAN,NVAR,XNOR)
COMMON/SEED/IX,KK
REAL PI,NVAR,SD
```

```
      INTEGER IX,KK,NMEAN
      DOUBLE PRECISION IY
      SD=SQRT(NVAR)
      PI=3.1415926
      IF(KK.EQ.1) GOTO 10
      CALL RANDOM(IX,IY,RD)
      RONE=RD
      CALL RANDOM(IX,IY,RD)
      RTWO=RD
      ZONE=SQRT(-2*ALOG(RONE))*cos(2*PI*RTWO)
      ZTWO=SQRT(-2*ALOG(RONE))*sin(2*PI*RTWO)
      XNOR=NMEAN+(ZONE*SD)
      KK=1
      GOTO 20
10    XNOR=NMEAN+(ZTWO*SD)
      KK=0
20    RETURN
      END
```

RANDOM UNIFROM (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
```

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวอรนิต เกตุสุข เกิดวันอาทิตย์ที่ 9 กันยายน พ.ศ. 2522 สำเร็จการศึกษาปริญญาตรีวิทยาศาสตรบัณฑิต สาขาวิชาสถิติ ภาควิชาสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ปีการศึกษา 2544 จากนั้นเข้าศึกษาต่อในหลักสูตรสถิติศาสตรมหาบัณฑิต สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี ที่จุฬาลงกรณ์มหาวิทยาลัย เมื่อ พ.ศ. 2545