การเปรียบเทียบวิธีการประมาณสำหรับตัวแบบการถดถอยเชิงเส้นพหุ
ที่ค่าความคลาดเคลื่อนมีความแปรปรวนไม่คงที่

นายพีรวัฒน์ เสรีวัฒนนุกูล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2555

COMPARISON OF THE ESTIMATION METHODS FOR THE MULTIPLE LINEAR REGRESSION MODEL WITH HETEROSCEDASTICITY ERROR

Mr.Peerawat Sereewattananukul

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Statistics
Department of Statistics
Faculty of Commerce and Accountancy
Chulalongkorn University
Academic Year 2012

หัวข้อวิทยานิพนธ์	การเปรียบเทียบวิธีการประมาณสำหรับตัวแบบการถดถอยเชิงเส้นพหุ ที่ค่าความคลาดเคลื่อนมีความแปรปรวนไม่คงที่
โดย	นายพีรวัฒน์ เสรีวัฒนนุกูล
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	อาจารย์ ดร.อนุภาพ สมบูรณ์สวัสดี

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

……………………………………คณบดีคณะพาณิชยศาสตร์และการบัญชี
(รองศาสตราจารย์ ดร.พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

……………………………………ประธานกรรมการ
(รองศาสตราจารย์ ดร.ธีระพร วีระถาวร)

……………………………………อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(อาจารย์ ดร.อนุภาพ สมบูรณ์สวัสดี)

……………………………………กรรมการ
(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

……………………………………กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ดร.มนต์ทิพย์ เทียนสุวรรณ)

พีรวัฒน์ เสรีวัฒนนุกูล: การเปรียบเทียบวิธีการประมาณสำหรับตัวแบบการถดถอยเชิงเส้นพหุที่ค่าความคลาดเคลื่อนมีความแปรปรวนไม่คงที่. (COMPARISON OF THE ESTIMATION METHODS FOR THE MULTIPLE LINEAR REGRESSION MODEL WITH HETEROSCEDASTICITY ERROR) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: อ. ดร.อนุภาพ สมบูรณ์สวัสดี, 139 หน้า.

ความคลาดเคลื่อนที่มีค่าความแปรปรวนไม่คงที่เป็นปัญหาพบบ่อยในการตัวแบบการถดถอย มีหลายวิธีการที่ใช้การแก้ปัญหาดังกล่าว ในการศึกษาครั้งนี้ผู้วิจัยมีความสนใจที่จะเปรียบเทียบวิธีการประมาณในสถานการณ์ที่ค่าความแปรปรวนของข้อมูลขึ้นอยู่กับค่าของตัวแปรอิสระและตัวแปรตาม ทั้งหมด 3 วิธีการ คือ วิธีกำลังสองน้อยที่สุด (Ordinary Least Square; OLS) วิธีการแปลงของ Box และ Cox (Box-Cox Transformation) และวิธี Iteratively Reweighted Least Square (IRWLS)

จากผลการการศึกษาแบบจำลองขนาด 5000 รอบ ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Average Mean Squared Error; AMSE) ของทุกวิธีการประมาณได้ถูกคำนวณและเปรียบเทียบ โดยค่าวิธีที่ดีกว่าจะให้ค่า AMSE ที่มีขนาดเล็กกว่า โดยรวมพบว่าวิธี IRWLS เป็นวิธีที่ดีที่สุด สำหรับวิธี Box-Cox Transformation มีประสิทธิภาพดีกว่าวิธี OLS โดยเฉพาะในกรณีที่ค่าความแปรปรวนของค่าความคลาดเคลื่อนมีขนาดใหญ่และขึ้นกับตัวแปรอิสระ

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

สาขาวิชา.......สถิติ......... ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก..............................

ปีการศึกษา.....2555.........

5481635626: MAJOR STATISTICS
KEYWORDS: BOX-COX TRANSFORMATION / ITERATIVELY REWEIGHTED LEAST SQUARE / HETEROSCEDASTICITY

PEERAWAT SEREEWATTANANUKUL: COMPARISON OF THE ESTIMATION METHODS FOR THE MULTIPLE LINEAR REGRESSION MODEL WITH HETEROSCEDASTICITY ERROR. ADVISOR: ANUPAP SOMBOONSAVATDEE, Ph.D., 139 pp.

Heteroscedasticity error are commonly known problem found in regression model. There have been several methods suggested to deal with such problem. In this study, researcher is interested in comparing the three estimation methods for dealing situations where the variances of error in the regression model depend on the value of the independent and dependent variables. These three methods are Ordinary Least Square (OLS), Box-Cox Transformation, and Iterative Reweighted Least Square (IRWLS).

Based on the simulation study of the simulation size of 5000, the average mean square errors (AMSEs) of all three estimation methods are computed and compared with smaller size of AMSE indicate better performance. Overall, the IRWLS performs best. Box-Cox transformation performs better than OLS especially when the variance of error is large and depends on independent variable.

Department: Statistics Student's Signature ....

Field of Study: Statistics Advisor's Signature ....

Academic Year: 2012

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดีเนื่องจากผู้วิจัยได้รับความช่วยเหลือ และ เอาใจใส่อย่างดียิ่งของอาจารย์ ดร. อนุภาพ สมบูรณ์สวัสดี อาจารย์ที่ปรึกษาวิทยานิพนธ์ ได้กรุณาสละเวลาให้ความช่วยเหลือ คำปรึกษา ข้อชี้แนะ ข้อคิดเห็น ผู้วิจัยขอกราบขอบพระคุณและสำนึกในพระคุณเป็นอย่างยิ่ง

ผู้วิจัย ขอขอบพระคุณท่านคณะกรรมการสอบวิทยานิพนธ์ ได้แก่ รองศาสตราจารย์ ดร. ธีระพร วีระถาวร ประธานกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา และ รองศาสตราจารย์ ดร.มนต์ทิพย์ เทียนสุวรรณ กรรมการสอบวิทยานิพนธ์ ที่กรุณาให้คำแนะนำ แนวคิดที่เป็นประโยชน์ รวมถึง การตรวจสอบ และแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์มากยิ่งขึ้น

ผู้วิจัยขอขอบพระคุณ คณาจารย์ทุกท่าน ของ ภาคสถิติ คณะพาณิชยศาสตร์ และ การบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ได้ถ่ายทอดความรู้ทางสถิติให้กับผู้วิจัย

ผู้วิจัยขอขอบคุณ เจ้าหน้าที่ทุกท่าน ของ ภาคสถิติ คณะพาณิชยศาสตร์ และ การบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่อำนวยความสะดวกต่างๆในเรื่องการทำวิจัย และขอบคุณเพื่อนๆนิสิต ภาควิชาสถิติ คณะพาณิชยศาสตร์ และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ช่วยให้กำลังใจและให้ความช่วยเหลือมาโดยตลอด

สุดท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณครอบครัว ที่ช่วยส่งเสริม ให้กำลังใจ และ สนับสนุนความสำเร็จของผู้วิจัย จนทำให้ผู้วิจัยสามารถประสบความสำเร็จไปด้วยดีอยู่เสมอ

สารบัญ

สารบัญตาราง

สารบัญภาพ

ภาพที่ หน้า

บทที่ 1

บทนำ

งานวิจัยนี้เป็นการศึกษาประสิทธิภาพของการพยากรณ์ตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ในแต่ละสถานการณ์ ซึ่งในบทนำนี้ ผู้วิจัยบรรยายถึงที่มาและความสำคัญของปัญหา วัตถุประสงค์ของการวิจัย ข้อตกลงเบื้องต้น คำจำกัดความที่ใช้ในการวิจัย เกณฑ์ที่ใช้ในการตัดสินใจ วิธีดำเนินการวิจัยโดยย่อ และประโยชน์ที่คาดว่าจะได้รับ ตามลำดับ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการศึกษาวิจัยด้านต่างๆ เช่น ด้านวิทยาศาสตร์ สังคมศาสตร์ เศรษฐศาสตร์ และทางด้านธุรกิจ ฯลฯ จำเป็นต้องอาศัยระเบียบวิธีทางสถิติเข้ามาช่วยในการศึกษา ค้นคว้าวิจัย และ การดำเนินงานอย่างเป็นระบบ เพื่อให้มีความถูกต้องและน่าเชื่อถือ โดยเฉพาะอย่างยิ่งการค้นคว้าหาคำตอบ เพื่อคาดคะเนเหตุการณ์ที่จะเกิดขึ้นในอนาคต หรือที่เรียกว่า การพยากรณ์ ซึ่งวิธีที่ใช้ในการพยากรณ์นั้นผู้วิจัย ส่วนใหญ่จะเลือกใช้การวิเคราะห์การถดถอย (Regression Analysis)

การวิเคราะห์การถดถอย (Regression Analysis) เป็นวิธีการวิเคราะห์ทางสถิติที่เกี่ยวข้องกับการสร้างตัวแบบทางคณิตศาสตร์ เพื่อแสดงความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระจำนวน 1 ตัวหรือมากกว่า 1 ตัวก็ได้ (วิชิต หล่อจีระชุณห์กุล และ จิราวัลย์ จิตรถเวช, 2548: 203) สำหรับการพยากรณ์หรือประมาณการตัวแปรด้วยการวิเคราะห์การถดถอยเชิงเส้น เมื่อทราบค่าของตัวแปรอิสระ จะมีข้อตกลงเบื้องต้นของตัวแปรสุ่มความคลาดเคลื่อน (random error) ซึ่งจะปรากฏในตัวแบบถดถอย ได้แก่ การเป็นอิสระจากกัน (independence), การที่ความแปรปรวนมีค่าคงที่ (homoscedasticity), การมีการแจกแจงปกติ (normality) ที่มีค่าเฉลี่ยเท่ากับศูนย์ (linearity) ทั้งนี้เพื่อให้การวิเคราะห์และอนุมานข้อมูลได้อย่างถูกต้อง (valid)

ปัญหาที่เกิดจากข้อกำหนดจากตัวแบบถดถอยเชิงเส้น ได้แก่ ความแปรปรวนของความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ (heteroscedasticity) ลักษณะเช่นนี้จะทำให้เกิดความ

คลาดเคลื่อนบางกลุ่มมีค่าสูงมากๆหรือต่ำมากๆ ซึ่งจะมีผลต่อค่าสังเกตจากตัวแปรตาม(Y) ส่งผลให้ตัวประมาณของค่าพารามิเตอร์ (β) ยังคงมีคุณสมบัติความไม่เอนเอียง (unbiasedness) มีความคงเส้นคงวา (consistency) แต่ความแปรปรวนจะขาดประสิทธิภาพ (efficiency) กล่าวคือ ความแปรปรวนของตัวประมาณอาจมีค่าสูงเกินไป ทำให้ขาดคุณสมบัติของตัวประมาณการเชิงเส้นตรงที่ไม่เอนเอียงดีที่สุด (Best Linear Unbiased Estimator: BLUE) ทำให้ช่วงความเชื่อมั่นของพารามิเตอร์ที่คำนวณได้กว้างเกินไป เนื่องจากค่าคลาดเคลื่อนของสัมประสิทธิ์การถดถอยมีค่าสูงเกินไป

ฉะนั้นเพื่อให้ปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่หมดไปก่อนการวิเคราะห์ข้อมูล จึงได้มีวิธีการพยากรณ์ตัวแปรตามจากการประมาณค่าสัมประสิทธิ์การถดถอย เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ ได้แก่ วิธีการแปลงของ Box และ Cox (Box-Cox Transformation) และ วิธี Iteratively Reweighted Least Square (IRWLS)

ดังนั้นในงานวิจัยครั้งนี้ ผู้วิจัยจึงได้ทำการเปรียบเทียบประสิทธิภาพของการพยากรณ์ตัวแปรตามในตัวแบบถดถอยเชิงเส้นพหุ โดยใช้เกณฑ์ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Average Mean Squared Error: AMSE) และ ค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency: RE) เป็นเกณฑ์เปรียบเทียบการพยากรณ์ที่ได้จากวิธีทางสถิติทั้งหมด 3 วิธี ได้แก่ วิธีกำลังสองน้อยที่สุด (Ordinary Least Square: OLS), วิธีการแปลงของ Box และ Cox และวิธี Iteratively Reweighted Least Square (IRWLS) ภายใต้ข้อมูลที่ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ทั้งหมด 2 รูปแบบ ได้แก่ ค่าความคลาดเคลื่อนที่สัมพันธ์กับค่าของตัวแปรอิสระกับ ค่าจริงของตัวแปรตาม

1.2 วัตถุประสงค์

ในงานวิจัย การเปรียบเทียบวิธีการประมาณสำหรับตัวแบบการถดถอยเชิงเส้นพหุที่ค่าความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ มีวัตถุประสงค์ของงานวิจัยดังต่อไปนี้

1.2.1 เพื่อศึกษาประสิทธิภาพของการพยากรณ์ตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ในแต่ละสถานการณ์ จากในแต่ละวิธี ทั้งหมด 3 วิธี ได้แก่ วิธีกำลังสองน้อยที่สุด (Ordinary Least Square: OLS) วิธีการแปลงของ Box

และ Cox (Box-Cox Transformation) และ วิธี Iteratively Reweighted Least Square (IRWLS)

1.2.2 เพื่อเสนอแนวทางที่เหมาะสมในแต่ละสถานการณ์ของความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ในตัวแบบถดถอยเชิงเส้นพหุ ให้มีการพยากรณ์ตัวแปรตามที่มีความถูกต้องมากขึ้น

1.3 ข้อตกลงเบื้องต้น

ตัวแบบที่ใช้ในการศึกษาครั้งนี้เป็นตัวแบบการถดถอยเชิงเส้นพหุ ที่เป็นตัวแบบความสัมพันธ์ระหว่างตัวแปรอิสระทั้งหมด 2 ตัวแปรและตัวแปรตามทั้งหมด 1 ตัวแปร จำนวนข้อมูลทั้งหมด n รายการ และเป็นตัวแบบเชิงเส้นของพารามิเตอร์ในตัวแบบ ไม่มีผลกระทบร่วมกัน (interaction) ระหว่างตัวแปรอิสระ ซึ่งสามารถเขียนสมการในรูปแบบดังกล่าวได้ดังนี้

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad ; \; i = 1, 2, ..., n$$

โดยที่ Y_i คือ ค่าของตัวแปรตาม จากค่าสังเกตที่ i

X_{i1} คือ ค่าสังเกตที่ i ของตัวแปรอิสระตัวที่ 1

X_{i2} คือ ค่าสังเกตที่ i ของตัวแปรอิสระตัวที่ 2

β_0 คือ พารามิเตอร์ ซึ่งเป็นจุดตัดบนแกน Y เมื่อค่าของตัวแปรอิสระทุกตัวเท่ากับศูนย์

β_1 และ β_2 คือ พารามิเตอร์ของตัวแบบการถดถอย หรือ สัมประสิทธิ์การถดถอย

ε_i คือ ความคลาดเคลื่อนที่เป็นอิสระซึ่งกันและกัน ที่มีการแจกแจงแบบปกติและมีค่าเฉลี่ยเท่ากับ ศูนย์ และ ความแปรปรวน ที่สัมพันธ์กับ ค่าตัวแปรอิสระตัวที่หนึ่ง เพียงตัวแปรเดียวเท่านั้น หรือ สัมพันธ์กับค่าจริงของตัวแปรตาม

1.4 คำจำกัดความที่ใช้ในการวิจัย

เนื่องจากมีศัพท์บางคำที่ผู้วิจัยเกรงว่า ผู้ศึกษางานวิจัยนี้อาจจะเข้าใจไม่ตรงกันกับผู้วิจัย ดังนั้นผู้วิจัยจึงมีคำจำกัดความที่ใช้ในการศึกษาของงานวิจัยนี้ ดังต่อไปนี้

1.4.1 ความแปรปรวนคงที่ (homoscedasticity) คือ ค่าความแปรปรวนความคลาดเคลื่อนที่ไม่เปลี่ยนแปลงสัมพันธ์กับค่าของตัวแปรอิสระ ไม่ว่าจะมีค่ามากหรือน้อยก็ตาม

1.4.2 ตัวแปรตาม (Dependent Variable) หมายถึง ตัวแปรที่ได้รับผลกระทบจากตัวแปรอื่น

1.4.3 ตัวแปรอิสระ (Independent Variable) หมายถึง ตัวแปรที่มีผลกระทบหรืออิทธิพลต่อตัวแปรตามอย่างมีนัยสำคัญ

1.5 ขอบเขตของการวิจัย

ผู้วิจัยได้มีการจำกัดขอบเขตในด้านต่างๆของงานวิจัยให้แคบลง เพื่อไม่ให้โครงการวิจัยมีขอบเขตทางการศึกษากว้างขวางจนเกินไปในการวิจัยครั้งนี้ ผู้วิจัยจึงได้กำหนดขอบเขตการวิจัยไว้ดังนี้

1.5.1 จำนวนตัวแปรอิสระที่ศึกษาในตัวแบบความถดถอย เท่ากับ 2

1.5.2 ตัวแปรอิสระแต่ละตัว มีการแจกแจงแบบปกติ และ มีความอิสระซึ่งกันและกัน

1.5.3 ขนาดตัวอย่าง (n) ที่ศึกษาเท่ากับ 25, 50 และ 100

1.5.4 ข้อมูลของตัวแปรตาม Y ที่ศึกษาในงานวิจัยครั้งนี้ เป็นข้อมูลที่มีค่าบวกเพียงอย่างเดียว

1.5.5 การจำลองข้อมูลใช้เทคนิคมอนติคาร์โล (Monte Carlo Simulation Technique) เพื่อจำลองข้อมูลสุ่มในสถานการณ์ต่างๆที่ต้องการ โดยใช้โปรแกรม R (R-Project for Statistical Computing) จะกระทำซ้ำจำนวน 5,000 ครั้งในแต่ละสถานการณ์ที่กำหนดขึ้นจากขนาดตัวอย่าง อัตราส่วนของความแปรปรวนของการแจกแจงของตัวแปรอิสระระหว่าง 2 ตัว และ รูปแบบความคลาดเคลื่อนที่ความแปรปรวนมีค่าสัมพันธ์กับตัวแปรอิสระ X_1 เพียงตัวแปรเดียวเท่านั้น และสัมพันธ์กับ ค่าจริงของตัวแปรตาม

1.6 เกณฑ์ที่ใช้ในการตัดสินใจ

เพื่อที่จะทราบว่าวิธีการประมาณค่าตัวแปรตามวิธีใดให้ค่าพยากรณ์ของตัวแปรตามใกล้เคียงกับค่าจริงมากที่สุด ภายใต้สถานการณ์ที่ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ เนื่องจากความแปรปรวนของค่าคลาดเคลื่อนสัมพันธ์กับค่าของตัวแปรอิสระตัวที่หนึ่ง หรือสัมพันธ์กับค่าจริงของตัวแปรตาม เกณฑ์ที่ใช้ในการตัดสินใจในงานวิจัยครั้งนี้จะพิจารณาจากค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Average Mean Squared Error: AMSE) ซึ่งวิธีการประมาณค่าตัวแปรตามที่ดีที่สุดจะมีค่า AMSE น้อยที่สุด และใช้ค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency: RE) ซึ่งเป็นอัตราส่วนระหว่างค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของวิธี OLS กับวิธีการประมาณค่าตัวแปรตามแต่ละวิธี ช่วยในการเปรียบเทียบและเพื่อดูแนวโน้มของแต่ละวิธี ซึ่งในการศึกษาครั้งนี้เลือกใช้วิธี OLS เป็นเกณฑ์การเปรียบเทียบเนื่องจากวิธี OLS เป็นวิธีประมาณค่าตัวแปรที่นิยมใช้มากที่สุดและเป็นวิธีพื้นฐานในการประมาณค่าตัวแปรตามในตัวแบบถดถอยเชิงเส้น ซึ่งวิธีใดให้ค่า RE มากกว่า 1 แสดงว่ามีประสิทธิภาพมากกว่าวิธี OLS โดยคำนวณได้จากสูตร

$$\mathrm{AMSE}_{*} = \frac{1}{5,000}\left[\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i')^2\right]$$

$$\mathrm{RE}_{*} = \frac{\mathrm{AMSE}_{\mathrm{OLS}}}{\mathrm{AMSE}_{*}} \;\; ; * \in \{\mathrm{OLS},\ \mathrm{Box\text{-}Cox},\ \mathrm{IRWLS}\}$$

โดยที่ $\hat{Y}_i$ แทน ค่าประมาณของตัวแปรตาม

Y_i' แทน ค่าจริงของตัวแปรตาม

n แทน ขนาดตัวอย่าง

AMSE_{*} แทน ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองระหว่างค่าพยากรณ์กับค่าจริงตัวแปรตามจากการทำซ้ำทั้งหมด 5,000 รอบ ของแต่ละวิธี ได้แก่ วิธี OLS, วิธีการแปลงของ Box และ Cox และวิธี IRWLS

$\mathrm{AMSE}_{\mathrm{OLS}}$ แทน ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองระหว่างค่าพยากรณ์กับค่าจริงของตัวแปรตามจากการทำซ้ำทั้งหมด 5,000 รอบของวิธี OLS

1.7 วิธีดำเนินการวิจัย

ในงานวิจัย การเปรียบเทียบวิธีการประมาณสำหรับตัวแบบการถดถอยเชิงเส้นพหุที่ค่าความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ มีวิธีดำเนินการวิจัยดังต่อไปนี้

1.7.1 สร้างข้อมูลของตัวแปรอิสระ X_1 และ X_2 ที่ต่างมีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ ศูนย์ และความแปรปรวนของตัวแปรอิสระแต่ละตัวที่มีอัตราส่วนระหว่างตัวแปรอิสระทั้งสองตัว ทั้งหมด 3 แบบ คือ 1:2, 1:1 และ 2:1

1.7.2 สร้างข้อมูลของค่าจริงของตัวแปรตาม โดยใช้ความสัมพันธ์เชิงเส้น คือ

$$\underset{\sim}{Y}' = X\underset{\sim}{\beta} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

ซึ่งโดยทั่วไปแล้ว β_0, β_1 และ β_2 จะเป็นพารามิเตอร์ที่ไม่ทราบค่า และในงานวิจัยนี้ผู้วิจัยได้กำหนด $\underset{\sim}{\beta}$ เป็นค่าพารามิเตอร์ที่มีค่าคงที่ โดยที่ $\beta_0 = 500$, $\beta_1 = \beta_2 = 1$ ทั้งนี้เนื่องจากต้องการเปรียบเทียบแบบของตัวแปรอิสระที่มีรูปแบบความแปรปรวนแตกต่างกัน ซึ่งเมื่อ β_1 หรือ β_2 เปลี่ยนไปจะทำให้ค่าตัวแปรตามที่ได้จากแบบของตัวแปรอิสระแต่ละแบบมีความแปรปรวนแตกต่างกันด้วย ฉะนั้นเพื่อให้ควบคุมค่าความแปรปรวนของค่าจริงของตัวแปรตามมีค่าเท่ากัน จึงกำหนดให้ β_1 และ β_2 มีค่าเท่ากับ 1 และ กำหนดให้ตัวแปรอิสระแต่ละตัวไม่มีความสัมพันธ์กัน นั่นคือกำหนดให้ค่าสหสัมพันธ์มีค่าเท่ากับศูนย์

1.7.3 สร้างข้อมูลของความคลาดเคลื่อนที่มีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนที่สัมพันธ์กับค่าของตัวแปรอิสระตัวที่หนึ่ง และ สัมพันธ์กับค่าจริงของตัวแปรตาม

1.7.4 สร้างข้อมูลของค่าสังเกตของตัวแปรตาม โดยที่ตัวแปรตามและตัวแปรอิสระที่ศึกษา มีความสัมพันธ์กันภายใต้การถดถอยเชิงเส้นพหุ (Multiple Linear Regression)

1.7.5 ประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุด้วยวิธีกำลังสองน้อยสุด (Ordinary Least Square: OLS), วิธีการแปลงของ Box และ Cox (Box-Cox Transformation) และวิธี Iteratively Reweighted Least Square (IRWLS)

1.7.6 นำค่าสัมประสิทธิ์การถดถอยที่ได้ มาสร้างสมการเชิงเส้นถดถอยพหุ เพื่อหาค่าพยากรณ์

1.7.7 คำนวณค่า AMSE และค่า RE ของแต่ละวิธีการประมาณค่าตัวแปรตาม เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าตัวแปรตามทั้งหมด 3 วิธี

1.7.8 สรุปผลที่ได้จากการทดลอง

1.8 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัยนี้ มีดังต่อไปนี้

1.8.1 ทำให้สามารถเปรียบเทียบแต่ละวิธีการพยากรณ์ตัวแปรตามที่อยู่ในตัวแบบถดถอยเชิงเส้นพหุ จากรูปแบบต่างๆที่ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ในแต่ละสถานการณ์ ทั้งหมด 3 วิธี ได้แก่ วิธีกำลังสองน้อยที่สุด (Ordinal Least Square: OLS) วิธีการแปลง Box และ Cox (Box-Cox Transformation) และ วิธี Iteratively Reweighted Least Square (IRWLS)

1.8.2 เพื่อทราบถึงปัจจัยที่มีผลต่อความแม่นยำในการพยากรณ์ของตัวแปรตามในตัวแบบถดถอยเชิงเส้นพหุ ภายใต้ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่

1.8.3 เพื่อเป็นแนวทางในการศึกษาวิจัยต่อไป

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ปัญหาที่เกิดจากข้อกำหนดของตัวแบบการถดถอยเชิงเส้นพหุ ได้แก่ ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ (heteroscedasticity) มีผลทำให้ความแปรปรวนของตัวประมาณอาจมีค่าสูงเกินไป ทำให้ขาดคุณสมบัติของตัวประมาณการเชิงเส้นตรงที่ไม่เอนเอียงดีที่สุด (Best Linear Unbiased Estimator: BLUE)

ฉะนั้นเพื่อให้ปัญหาความคลาดเคลื่อนที่ความแปรปรวนไม่คงที่หมดไปก่อนการวิเคราะห์ข้อมูล ถ้ามีวิธีการพยากรณ์ค่าตัวแปรตามที่น่าจะสามารถกำจัดปัญหาที่ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ได้ ค่าพยากรณ์ที่ได้น่าจะมีความถูกต้องและเที่ยงตรงมากกว่า ดังนั้นในบทนี้จะกล่าวถึง ทฤษฎีที่เกี่ยวข้อง ได้แก่ การวิเคราะห์การถดถอย (regression analysis), ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ (heteroscedasticity) และ การพยากรณ์ตัวแปรตามที่ใช้ในงานวิจัยชิ้นนี้ ทั้งหมด 3 วิธี ได้แก่ วิธีกำลังสองน้อยที่สุด (Ordinary Least Square: OLS) วิธีการแปลงของ Box และ Cox (Box-Cox Transformation) และ วิธี Iteratively Reweighted Least Square (IRWLS) และงานวิจัยที่เกี่ยวข้อง ซึ่งมีรายละเอียดดังต่อไปนี้

2.1 การวิเคราะห์การถดถอย (Regression Analysis)

การวิเคราะห์การถดถอยได้ชื่อมาจากงานวิจัยของ Sir Francis Galton งานวิจัยนี้ได้ศึกษาในเชิงเปรียบเทียบส่วนสูงของบุตร และ ส่วนสูงของบิดา นักวิจัยผู้นี้สรุปว่า บิดามารดาที่สูงผิดปกติจะมีบุตรที่เตี้ยกว่าบิดามารดา และ บิดามารดาที่เตี้ยผิดปกติก็จะมีบุตรที่สูงกว่าบิดามารดา ผลงานนี้ได้ตีพิมพ์เผยแพร่ในปี ค.ศ. 1885 ภายใต้ในชื่อเรื่อง “Regression Toward Mediocrity in Hereditary Stature” ความหมายของคำว่า “ถดถอย” คือ ส่วนสูงของบุตรจะถดถอยเข้าสู่ค่าเฉลี่ยของประชากรมากกว่าที่จะเข้าสู่ค่าที่สูงมากหรือค่าที่ต่ำมาก จะเห็นได้ว่า ตัวแปรตัวหนึ่ง คือ ส่วนสูงของบิดามารดา สามารถพยากรณ์ค่าของตัวแปรอีกตัวหนึ่งคือส่วนสูงของบุตรได้ การพยากรณ์ตัวแปรตัวหนึ่งโดยใช้ตัวแปรอื่นๆนั้น เรียกว่า การวิเคราะห์การถดถอย (วิชิต หล่อจีระชุณห์กุล และ จิราวัลย์ จิตรถเวช, 2548)

วิธีการสร้างตัวแบบการถดถอยสำหรับใช้ประมาณค่าหรือพยากรณ์ตัวแปรตาม (Dependent variable หรือ response variable) นิยมเขียนแทนด้วย Y จากตัวแปรอิสระหรือตัวแปรพยากรณ์ (Independent variable or predictor variable) นิยมเขียนแทนด้วย X ในกรณีตัวแปรอิสระมีหนึ่งตัว เรียกว่า การวิเคราะห์ตัวแปรเชิงเดียว (simple linear regression) แต่ในบางกรณีตัวแปรอิสระอาจมีตั้งแต่สองตัวขึ้นไป เรียกว่า การวิเคราะห์การถดถอยเชิงพหุ (Multiple regression)

1. ตัวแบบการถดถอยเชิงเส้นเชิงเดียว (Simple linear regression model)

เขียนได้เป็น $Y = \beta_0 + \beta_1 X + \varepsilon$ หรือ $\mu_{Y|X} = \beta_0 + \beta_1 X$

เมื่อ Y	แทน	ตัวแปรตามหรือตัวแปรที่สนใจศึกษา
X	แทน	ตัวแปรอิสระ
β_0, β_1	แทน	ค่าสัมประสิทธิ์การถดถอย
ε	แทน	ความคลาดเคลื่อนเชิงสุ่ม
$\mu_{Y\|X}$	แทน	เป็นค่าคาดหมายของ Y เมื่อกำหนดค่าหนึ่งของ X

เนื่องจากในทางปฏิบัติ ไม่สามารถทราบค่าที่แท้จริงของ β_0 และ β_1 ได้จึงจำเป็นต้องมีการประมาณค่าพารามิเตอร์ทั้งหมด 2 ตัว โดยใช้ข้อมูลที่ได้จากตัวอย่าง ซึ่งจะได้สมการถดถอย (regression equation) ซึ่งจะได้สมการถดถอยหรือสมการพยากรณ์ดังต่อไปนี้

$$\hat{Y} = b_0 + b_1 X$$

เมื่อ b_0, b_1 แทน ค่าประมาณสัมประสิทธิ์การถดถอยซึ่งเกิดจากการประมาณค่า β_0 และ β_1 ตามลำดับ

b_0 จะเป็นค่าคงที่เสมอ ไม่ว่าค่าของตัวแปรอิสระจะมีค่าเท่าใดก็ตาม ส่วน b_1 เป็นค่าที่แสดงว่าเมื่อค่าของตัวแปรอิสระ (X) เปลี่ยนไป 1 หน่วย จะทำให้ค่าของตัวแปรตามเปลี่ยนแปลงไป b_1 หน่วย

ถ้ารู้ค่าของ b_0 และ b_1 แล้วจะสามารถพยากรณ์ค่าต่างๆของตัวแปรตามจากค่าต่างๆของตัวแปรอิสระได้ ตัวอย่างเช่น ถ้า $b_0 = 4$ และ $b_1 = 2$ จะได้ $\hat{Y} = 4 + 2X$ ดังนั้น ยกตัวอย่างเช่น ถ้าทราบว่า $X = 3$ แล้วจะได้ว่า $\hat{Y} = 4 + 2(3) = 4 + 6 = 10$ เป็นต้น

แนวความคิดที่ใช้ค่าของตัวแปรอิสระมาพยากรณ์ตัวแปรตาม เนื่องจากว่าตัวแปรทั้ง 2 ประเภทนี้มีความสัมพันธ์กันในลักษณะใดลักษณะหนึ่ง ซึ่งอาจจะเป็นจะเป็นที่แน่ชัดหรือไม่เป็นที่แน่ชัดก็ได้ว่าตัวแปรใดคือเหตุและตัวแปรใดคือผล ลักษณะความสัมพันธ์อาจจะเป็นแบบเชิงเส้นหรือแบบไม่เชิงเส้นก็ได้ (วิชิต หล่อจีระชุณห์กุล และ จิราวัลย์ จิตรถเวช, 2548) ซึ่งในงานวิจัยนี้ความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม เป็นความสัมพันธ์เชิงเส้น

2. การวิเคราะห์การถดถอยเชิงเส้นพหุ (Multiple linear regression analysis)

ตัวแบบการถดถอยเชิงเส้น นั่นคือเป็นเชิงเส้นในพารามิเตอร์และในตัวแปรอิสระ ซึ่งมีรูปแบบทั่วไปคือ $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \varepsilon_i$; $i = 1,2,\ldots,n$ เมื่อ p เป็นจำนวนตัวแปรอิสระ

หรือเขียนในรูปเมทริกซ์ คือ $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ โดย

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots X_{1p} \\ 1 & X_{21} & X_{22} & \cdots X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots X_{np} \end{bmatrix}, \ \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

ส่วนข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเชิงพหุ ได้แก่ ε_i มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับศูนย์และมีความแปรปรวนคงที่ซึ่งเท่ากับ σ^2 และ ε_i และ ε_j สำหรับ $i \neq j$ มีการแจกแจงที่เหมือนกันและเป็นอิสระต่อกัน ซึ่งทำให้ $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0$ สำหรับ $i \neq j$

ต่าย เซี่ยงฉี่ (2525) ได้กล่าวถึงประเด็นข้อดีและข้อจำกัดของการวิเคราะห์การถดถอยเชิงเส้นพหุไว้ดังนี้

<u>ข้อดีของการวิเคราะห์การถดถอยเชิงเส้นพหุ</u>

1. เป็นการวิเคราะห์ที่ถูกหลักวิทยาศาสตร์ กล่าวคือ เป็นการวิเคราะห์ที่นำเอาตัวแปรอิสระทุกตัวมาวิเคราะห์พร้อมๆกัน ไม่ได้แยกวิเคราะห์ตัวแปรทีละตัว ซึ่งการแยกวิเคราะห์ตัวแปรอิสระทีละตัว จะไม่ทราบปฏิกิริยาร่วม (interaction) ที่เกิดจากตัวแปรตั้งแต่สองตัวขึ้นไป

2. เป็นการวิเคราะห์ที่มีการวิเคราะห์ทั้งในแง่ของความสัมพันธ์ และ ความแตกต่างในแง่ของความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันเอง และ ตัวแปรอิสระกับตัวแปรตาม

3. เป็นการวิเคราะห์ที่สามารถนำไปใช้ในการวิเคราะห์อย่างอื่น เช่น การวิเคราะห์ตัวแปรหลายตัว (Multivariate Analysis) การวิเคราะห์ความแปรปรวน (Analysis of Variance) การวิเคราะห์ปัจจัย (Factor Analysis)

ข้อจำกัดของการวิเคราะห์การถดถอยเชิงเส้นพหุ

1. ความไม่เที่ยงตรงของค่าสัมประสิทธิ์การถดถอยพหุ เกิดจาก ลำดับการนำตัวแปรอิสระก่อนหรือหลังเข้าตัวแบบถดถอยเชิงเส้นพหุ นั่นคือ ลำดับการนำตัวแปรอิสระก่อนหรือหลังดังกล่าวมีผลต่อค่าสัมประสิทธิ์การตัดสินใจ (R^2) (Coefficient of Determination)

2. สมการถดถอยจะเปลี่ยนไปตามขนาดของกลุ่มตัวอย่าง และ ลำดับก่อนหรือหลังการนำตัวแปรอิสระเข้าตัวแบบ

3. ในกรณีที่ตัวแปรอิสระมีความสัมพันธ์กันเองสูง (Multicollinearity) การวิเคราะห์การถดถอยเชิงเส้นพหุ จะทำให้เกิดการผิดพลาดในการแปลผล

2.2 **ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่** (Heteroscedasticity)

จากข้อสมมติพื้นฐานของตัวแบบการถดถอยไม่ว่าจะเป็นแบบจำลองเชิงถดถอยเชิงพหุ หรือ แบบจำลองถดถอยเชิงเดียว ที่กำหนดให้ความคลาดเคลื่อนมีความแปรปรวนมีค่าคงที่ (homoscedasticity) เท่ากับ σ^2 ในทุกๆตัวอย่าง ภายใต้ข้อสมมติข้างต้น เป็นส่วนหนึ่งที่ช่วยสนับสนุนให้ตัวประมาณการที่ได้รับจากวิธีกำลังสองน้อยที่สุดมีคุณสมบัติของตัวประมาณการเชิงเส้นตรงที่ไม่เอนเอียงที่ดีที่สุด (Best Linear Unbiased Estimators: BLUE) (บัณฑิต ชัยวิชญชาติ, 2550)

สาเหตุของการเกิดความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ ได้แก่

1. การที่ตัวอย่างเกิดการเรียนรู้ เช่น การเก็บข้อมูลความสัมพันธ์ระหว่างจำนวนชั่วโมงในการฝึกหัดพิมพ์คอมพิวเตอร์กับจำนวนครั้งในการพิมพ์ผิด เมื่อจำนวนชั่วโมงในการฝึกหัดเพิ่มขึ้นจะส่งผลให้จำนวนครั้งในการพิมพ์ลดลง หรือ การกระจายของจำนวนครั้งในการพิมพ์ผิดลดลง ทำให้ความแปรปรวนต่ำลง ไม่ได้มีค่าคงที่ตามข้อสมมติฐาน (Damodar, 1988)

2. การเปลี่ยนแปลงของตัวแปรภายนอก ส่งผลให้ขอบเขตของตัวแปรภายในกว้าง ยกตัวอย่างเช่น กรณีของแบบจำลองการบริโภค การเพิ่มขึ้นของรายได้ที่เป็นตัวแปรภายนอกส่งผล

ให้ผู้บริโภคมีทางเลือกในการบริโภคเพิ่มขึ้น การกระจายของการบริโภคที่เป็นตัวแปรภายในจึงมีการกระจายเพิ่มขึ้น ทำให้ความคลาดเคลื่อนมีความแปรปรวนเพิ่มขึ้นด้วย

3. การพัฒนาของเทคนิคในการเก็บข้อมูลที่ดีขึ้น ทำให้ข้อมูลที่ใช้ในแบบจำลองมีค่าตรงกับความเป็นจริงมากขึ้น ส่งผลให้ความคลาดเคลื่อนมีค่าต่ำลง และ แน่นอนย่อมทำให้ความคลาดเคลื่อนมีความแปรปรวนลดลง

4. การเปลี่ยนแปลงเชิงโครงสร้างต่างๆ เช่น การเปลี่ยนแปลงกฎหมาย การเปลี่ยนแปลงช่วงเวลา หรือ การเปลี่ยนแปลงระบบอัตราแลกเปลี่ยน เนื่องจากการเปลี่ยนแปลงเชิงโครงสร้างส่งผลต่อพฤติกรรมระหว่างตัวแปรภายนอกและตัวแปรภายใน ซึ่งผลของการเปลี่ยนแปลงเชิงโครงสร้างที่เกิดขึ้นแสดงออกมาที่ความคลาดเคลื่อน ซึ่งอาจทำให้รูปแบบการกระจายของความคลาดเคลื่อน โดยเฉพาะในส่วนของความแปรปรวนเปลี่ยนแปลงไป

5. การใช้ข้อมูลอนุกรมเวลา (Time series) ที่ได้มีการเก็บข้อมูลในแต่ละช่วงเวลาจากจำนวนตัวอย่างที่ไม่เท่ากัน เช่น การเก็บข้อมูลการบริโภค (C_{it}) และรายได้ (y_{it}) ของไทยดังนี้

ปีที่ 1 ใช้จำนวนตัวอย่าง n_1 จะได้ $\sum_{i=1}^{n_1} C_{i1}$ และ $\sum_{i=1}^{n_1} y_{i1}$

ปีที่ 2 ใช้จำนวนตัวอย่าง n_2 จะได้ $\sum_{i=1}^{n_2} C_{i2}$ และ $\sum_{i=1}^{n_2} y_{i2} : (n_1 \neq n_2)$

$\vdots$

ปีที่ t ใช้จำนวนตัวอย่าง n_t จะได้ $\sum_{i=1}^{n_t} C_{it}$ และ $\sum_{i=1}^{n_t} y_{it}$

เมื่อนำข้อมูลข้างต้นไปเขียนเป็นแบบจำลองถดถอยเชิงเดียว จะได้

$$y_t = \sum_{i=1}^{n_t} y_{it} \ , \ C_t = \beta_0 + \beta_1 y_t + \varepsilon_t$$

โดยที่ $C_t = \sum_{i=1}^{n_t} C_{it}$, $y_t = \sum_{i=1}^{n_t} y_{it}$ และ $\varepsilon_t = \sum_{i=1}^{n_t} \varepsilon_{it}$

พิจารณาความคลาดเคลื่อนมีความแปรปรวน หรือ

$$\mathrm{Var}(\varepsilon_i) = \mathrm{Var}\left(\sum_{i=1}^{n_t} \varepsilon_i\right) = \sum_{i=1}^{n_t} \mathrm{Var}(\varepsilon_i) = n_t \sigma^2$$

ซึ่งแสดงให้เห็นว่า ถ้ามีการใช้จำนวนตัวอย่างในแต่ละปี (n_t) เท่ากันจะทำให้ $n_1 = n_2 = ... = n_t$ และทำให้ $Var(\varepsilon_i)$ มีค่าคงที่ แต่ในกรณีที่จำนวนตัวอย่างในแต่ละปีไม่คงที่ก็จะทำให้ค่า $Var(\varepsilon_i)$ มีค่าไม่คงที่

6. การใช้ข้อมูลภาคตัดขวาง (Cross-sectional data) เนื่องจากการใช้ข้อมูลภาคตัดขวางจะมีความแตกต่างของตัวอย่างแต่ละตัวอย่างที่แสดงออกโดยความคลาดเคลื่อน ซึ่งจะทำให้ลักษณะการกระจายของความคลาดเคลื่อนของแต่ละตัวอย่างแตกต่างกันออกไป

7. ผลจากค่าข้อมูลผิดปกติ (outlier) กล่าวคือ ค่าที่ต่ำมากและค่าที่สูงมาก ซึ่งการนำค่าดังกล่าวเข้ามาหรือตัดออกไป มีผลต่อการประมาณสมการถดถอยอย่างมาก (นิติพงษ์ ส่งศรีโรจน์, 2553)

8. การละทิ้งตัวแปรสำคัญ จะทำให้ความแปรปรวนของการประมาณสมการไม่คงที่ ความเบ้ (skewness) ของข้อมูล เช่น การกระจายของรายได้ และความมั่งคั่งของกลุ่มคนบนสุดมักจะมีจำนวนน้อย โดยทั่วไปพบว่า ความแปรปรวนของค่าใช้จ่ายจะไม่คงที่แต่จะผันแปรค่าไปตามระดับค่าของรายได้ของครอบครัว (Kelyjian และ Oates, 1981)

9. ความคลาดเคลื่อนเกิดจากการวัด คือ การนำข้อมูลที่ค่าความคลาดเคลื่อนเกิดจากเครื่องมือวัดต่างประเภทกัน ซึ่งมีความแม่นยำต่างกันหรือความผิดพลาดจากการรวบรวมข้อมูล (ชิดชนก เชิงเชาว์, 2541) ซึ่งสาเหตุนี้สามารถควบคุมได้

10. สาเหตุอื่นๆ เช่นการแปลงข้อมูล (Transformation Data) ที่ไม่ถูกต้อง การกำหนดรูปแบบฟังก์ชันที่ไม่ถูกต้อง เป็นต้น นอกจากนี้ลักษณะที่เกี่ยวกับ ขนาดของตัวแปรที่อาจจะแบ่งเป็นขนาดเล็ก ขนาดกลาง ขนาดใหญ่ เช่น กลุ่มผู้มีขนาดรายได้สูง กลุ่มผู้มีขนาดรายได้ปานกลาง กลุ่มผู้มีขนาดรายได้ต่ำ กรณีผู้ประกอบการหรือธุรกิจอาจจะมีขนาดใหญ่ ขนาดกลาง ขนาดเล็ก หากลักษณะดังกล่าวเกิดขึ้นกับข้อมูลที่เก็บมาแบบรวมกัน ก็เป็นไปได้ว่าจะเกิดปัญหา Heteroscedasticity (นิติพงษ์ ส่งศรีโรจน์, 2553)

<u>ผลกระทบความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ (Consequence of Heteroscedasticity)</u>

สุพล ดุรงค์วัฒนา (2549) กล่าวไว้ว่า ในตัวแบบการถดถอยเชิงเส้นอย่างง่ายระหว่างตัวแปรอิสระกับตัวแปรตาม $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ โดย ε_i , $i = 1,2,...,n$ ณ ระดับของคงที่ของ

$X_i = x_i$ ใดๆ ต้องเป็นตัวแปรสุ่ม (Random Variable) ในกรณีที่ข้อมูลตัวแปรตาม ณ ระดับต่างๆ ของตัวแปรอิสระมีความแปรปรวนไม่คงที่ จะส่งผลกระทบกับการวิเคราะห์การถดถอยดังนี้

1. ตัวประมาณ $\hat{Y}$ และค่าพยากรณ์ $\hat{Y}$ จะเป็นตัวประมาณที่ไม่เอนเอียง (Unbiased) และคงเส้นคงวา (Consistent)

2. ตัวประมาณ b_0, b_1 และค่าพยากรณ์ $\hat{Y}$ จากวิธีกำลังสองน้อยที่สุด (Ordinary Least Square Method) จะไม่เป็นตัวประมาณที่เรียกว่า ตัวประมาณการเชิงเส้นตรงที่ไม่เอนเอียงที่ดีที่สุด (Best Linear Unbiased Estimators: BLUE) และตัวประมาณเหล่านี้จะไม่เป็นตัวประมาณที่มีประสิทธิภาพ (Inefficient)

3. ค่าประมาณของความแปรปรวนของตัวประมาณ b_0, b_1 และค่าประมาณของความแปรปรวนร่วม (Covariance) ระหว่าง b_0 และ b_1 จะเป็นค่าประมาณที่เอนเอียง (Biased) และไม่คงเส้นคงวา (Inconsistent) ดังนั้นผลของความเอนเอียงและความไม่คงเส้นคงวาดังกล่าวนี้ทำให้การทดสอบสมมติฐานเชิงสถิติไม่ถูกต้อง (Invalid)

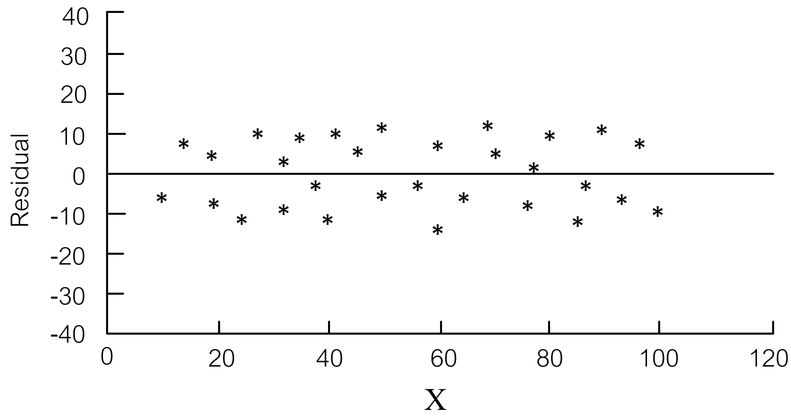
<u>วิธีการตรวจสอบความแปรปรวนไม่คงที่ของข้อมูล</u>

1. วิธีการตรวจสอบด้วยกราฟ

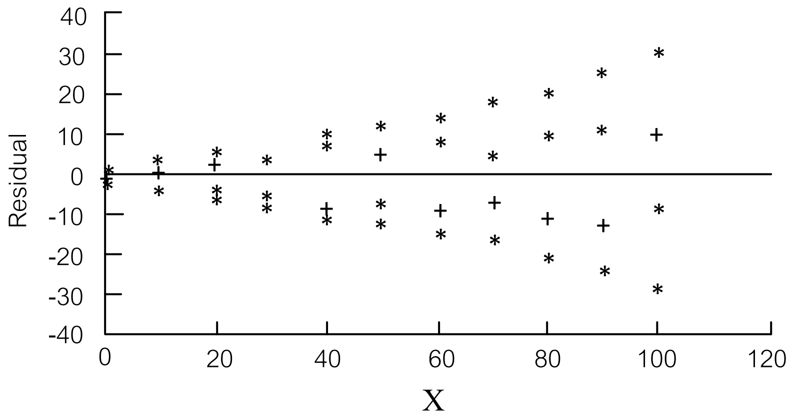
1.1 สำหรับการตรวจสอบด้วยวิธีเขียนกราฟนั้น จะใช้กราฟคู่ลำดับ (Ordered Pairs) ของข้อมูลตัวแปรอิสระกับข้อมูลเศษเหลือ (X_i, e_i); $i = 1, 2, ..., n$ หรือ

1.2 กราฟคู่ลำดับของข้อมูลพยากรณ์ (Predicted Value) กับข้อมูลเศษเหลือ ($\hat{Y}_i, e_i$); $i = 1,2,...,n$

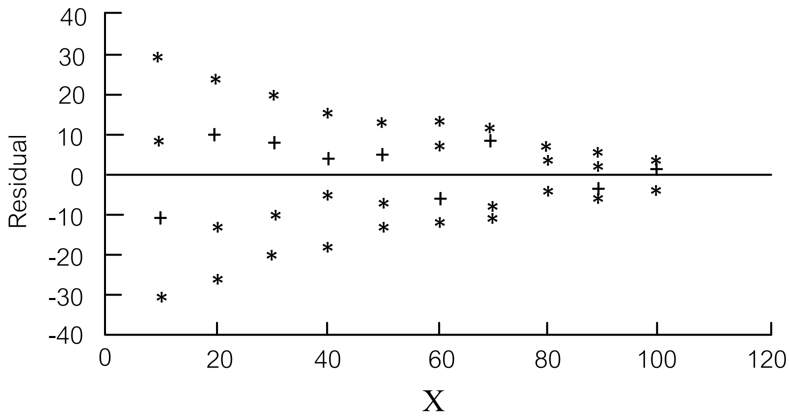
ค่าความคลาดเคลื่อน (errors) ควรจะปรากฏขึ้นโดยไม่กระจายอยู่รอบๆเส้นศูนย์แตกต่างกันโดยไม่คำนึงถึงค่าของ x ตัวอย่างเช่นถ้าเหลือจะกระจายเพิ่มเติมสำหรับค่ามากของ x กว่าสำหรับค่าขนาดเล็ก ลักษณะกราฟมีรูปลักษณะขยายหรือกระจายออกคล้ายปากลำโพงแล้วข้อสันนิษฐานของความแปรปรวนคงที่อาจถูกละเมิด คือกราฟที่มีลักษณะความแปรปรวนของข้อมูลไม่คงที่ (heteroscedasticity) ในขณะเดียวกัน กราฟที่มีลักษณะการกระจายอยู่ๆรอบเส้นศูนย์และขนานกับแกนของข้อมูลคือกราฟที่มีลักษณะความคลาดเคลื่อนมีความแปรปรวนของข้อมูลคงที่ (homoscedasticity) ดังรูปต่อไปนี้



ภาพที่ 2.1 แสดงรูปแบบของความคลาดเคลื่อนมีความแปรปรวนคงที่



ภาพที่ 2.2 แสดงรูปแบบของความคลาดเคลื่อนมีความแปรปรวนมากขึ้น เมื่อ X มากขึ้น



ภาพที่ 2.3 แสดงรูปแบบของความคลาดเคลื่อนมีความแปรปรวนลดลง เมื่อ X มากขึ้น

2. การใช้ค่าสถิติในการตรวจสอบปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่

การใช้ค่าสถิติในการตรวจสอบปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ วิธีที่นิยมใช้มีทั้งหมด 3 วิธี

1. วิธีการทดสอบด้วยตัวสถิติทดสอบ Szroeter

2. วิธีการทดสอบด้วยตัวสถิติทดสอบ Breusch’s และ Pagan’s (BP)

3. วิธีการทดสอบด้วยตัวสถิติทดสอบ White’s (W)

โดยที่สามารถเขียนสมมติฐานของการทดสอบดังนี้

$H_0: V(\varepsilon_i) = \sigma^2$; $\forall i = 1,2,...,n$ (H_0: ความคลาดเคลื่อนมีความแปรปรวนคงที่)

$H_a: V(\varepsilon_i) \neq \sigma^2$; $\exists i = 1,2,...,n$ (H_a: ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่)

<u>ความสำคัญของการแก้ไขปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่</u>

สุพล ดุรงค์วัฒนา (2549) ได้กล่าวถึง ประเด็นความสำคัญของการแก้ไขปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ว่า ทั้งปัญหาของความคลาดเคลื่อนที่ไม่เป็นอิสระจากกัน และ ปัญหาความคลาดเคลื่อนมีการแจกแจงที่ไม่ใช่การแจกแจงปกติ สามารถแก้ไขได้ง่ายกว่าปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ เนื่องจากปัญหาทั้ง 2 ปัญหาแรกดังกล่าวสามารถเก็บรวบรวมข้อมูลให้มากขึ้น ก็จะช่วยผ่อนเบาปัญหาดังกล่าวได้ ในขณะที่ปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ การเพิ่มรายการข้อมูลหรือขนาดตัวอย่างไม่สามารถช่วยได้ เพราะเป็นปัญหาซึ่งเกิดขึ้นจากความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามโดยตรง และ ยังพบอีกว่า งานวิจัยจำนวนมาก ได้กล่าวไว้ว่า การแก้ไขปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ อาจมีผลทำให้ข้อมูลความคลาดเคลื่อนมีการแจกแจงปกติได้

<u>โครงสร้างกับการแก้ไขความคลาดเคลื่อนที่มีความแปรปรวนไม่คงที่</u>

สุพล ดุรงค์วัฒนา (2549) ได้แบ่งลักษณะโครงสร้างของความคลาดเคลื่อนที่มีความแปรปรวนไม่คงที่ในตัวแบบการถดถอยเชิงเส้นพหุไว้ดังนี้

1. ความไม่คงที่ของความแปรปรวนของตัวแปรตามที่ไม่สัมพันธ์กับตัวแปรอิสระใดๆ

การแปลงให้ความแปรปรวนคงที่ (Variance Stabilized Method) โดยใช้ฟังก์ชันการแปลง มีวัตถุประสงค์หลัก 2 ประการ คือ เพื่อสามารถทำให้ตัวแบบการถดถอยเป็นตัวแบบเชิงเส้นของพารามิเตอร์ (Linear in parameters) อีกประการหนึ่งเพื่อที่จะแก้ไขปัญหาของความแปรปรวนของตัวแปรตามที่ไม่คงที่ ณ ระดับคงที่ต่างๆของตัวแปรอิสระ

สำหรับการแปลงข้อมูลของกรณีดังกล่าว มีงานวิจัยทางสถิติหลายงานวิจัย พบว่า ฟังก์ชันที่มักจะช่วยแก้ไขปัญหาดังกล่าวได้ คือ ฟังก์ชันรากที่สอง (Square root function) ฟังก์ชัน

ลอการิทึม (Logarithm function) และฟังก์ชันส่วนกลับ (Reciprocal function) ของตัวแปรตาม แต่บางครั้งการแปลงข้อมูลของตัวแปรตามอาจไม่ได้ผล กล่าวคือ ไม่สามารถแก้ไขปัญหาค่าความแปรปรวนของค่าคลาดเคลื่อนไม่คงที่ (heteroscedasticity) จากงานวิจัยพบว่า ให้ทำการแปลงข้อมูลของตัวแปรอิสระเพียงบางส่วนเดียวก่อน คือ ไม่ต้องแปลงข้อมูลตัวแปรตาม อาจแก้ไขปัญหาดังกล่าวได้ อย่างไรก็ตามกรณีของตัวแบบการถดถอยเชิงเส้นอย่างง่าย การแปลงข้อมูลของตัวแปรอิสระจะสามารถกระทำได้ง่ายกว่ากรณีตัวแบบการถดถอยเชิงเส้นพหุ ทั้งนี้เพราะไม่สามารถเลือกได้ว่าตัวแปรอิสระใดควรจะนำมาทำการแปลงข้อมูล ดังนั้น การแปลงข้อมูลของตัวแปรอิสระเหมาะสำหรับตัวแบบการถดถอยอย่างง่ายมากกว่า

2. ความไม่คงที่ของความแปรปรวนของตัวแปรตามที่สัมพันธ์กับตัวแปรอิสระบางตัวแปร

ในกรณีนี้จะพบว่าความแปรปรวนของตัวแปรตามจะมีความสัมพันธ์กับตัวแปรอิสระบางตัวแปร ซึ่งตัวแปรอิสระบางตัวเหล่านั้นอาจจะถูกพิจารณาอยู่ในตัวแบบการถดถอยนั้นหรือไม่ก็ได้ และ รูปแบบความสัมพันธ์ของความแปรปรวนตามกับตัวแปรอิสระ มีรูปแบบความสัมพันธ์ดังนี้

$$\sigma_i^2 = \sigma^2 h(x_i)\,;\, i = 1,2,...,n$$

สำหรับความสัมพันธ์ดังกล่าวจะขอกล่าวเพียงตัวแปรอิสระเพียงตัวเดียวเท่านั้น และความสัมพันธ์ดังกล่าวมีรูปแบบความสัมพันธ์ได้หลายรูปแบบตามฟังก์ชันของ $h(x_i)$ โดยที่ $i = 1,2,...,n$

ความสัมพันธ์ดังกล่าวเกิดขึ้นด้วยการวิเคราะห์ดังนี้ (พิจารณาจากตัวแบบการถดถอยเชิงเส้นอย่างง่าย)

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad ; \quad i = 1,2,...,n$$

เพราะว่า $Var(Y_i | X_i = x_i) = Var(\varepsilon_i | X_i = x_i) = \sigma^2 h(x_i) = \sigma_i^2\,;\ i = 1,2,...,n$

กรณี $h(x_i) = x_i$

จะได้ว่า $\sigma_i^2 = \sigma^2 x_i\ ; i = 1,2,...,n$

สามารถใช้การแปลงข้อมูลดังนี้

$$\frac{Y_i}{\sqrt{X_i}} = \frac{\beta_0}{\sqrt{X_i}} + \beta_1 \frac{X_i}{\sqrt{X_i}} + \frac{\varepsilon_i}{\sqrt{X_i}}$$

$$\frac{Y_i}{\sqrt{X_i}} = \beta_0 \frac{1}{\sqrt{X_i}} + \beta_1 \sqrt{X_i} + \frac{\varepsilon_i}{\sqrt{X_i}}$$

$$Y_i^* = \nu_1 V_i + \nu_2 W_i + \varepsilon_i^*$$

เมื่อ $Y_i^* = \frac{Y_i}{\sqrt{X_i}}$, $V_i = \frac{1}{\sqrt{X_i}}$, $W_i = \sqrt{X_i}$, $\varepsilon_i^* = \frac{\varepsilon_i}{\sqrt{X_i}}$

$$\mathrm{Var}(Y_i^* \,|X_i = x_i) = \frac{1}{x_i}\mathrm{Var}(\varepsilon_i \,|X_i = x_i) = \sigma^2$$

หมายความว่า การแปลงข้อมูลต้องทำการแปลงข้อมูลของตัวแปรอิสระและข้อมูลตัวแปรตามด้วยเช่นกัน โดยข้อมูลใหม่ของตัวแปรอิสระ ซึ่งขณะนี้มี 2 ตัวแปร คือ $V_i = \frac{1}{\sqrt{X_i}}$, $W_i = \sqrt{X_i}$ และข้อมูลใหม่ของตัวแปรตามคือ $Y_i^* = \frac{Y_i}{\sqrt{X_i}}$ โดยไม่สามารถแปลงข้อมูลของตัวแปรใดตัวแปรหนึ่งได้ และ ตัวแบบการถดถอยจะเป็นตัวแบบที่ไม่มีเทอมส่วนของการตัดแกนตั้ง (intercept term) เช่นกัน

อย่างไรก็ตามงานวิจัยมากมายพบว่าสามารถจะใช้การแปลงข้อมูลได้แสดงไว้ข้างล่างนี้ แม้ว่าการเปลี่ยนแปลงข้อมูลดังกล่าวจะเป็นการ Over-transformation ก็ตาม ดังแสดงไว้ดังนี้

กรณีที่ $h(x_i) = x_i^2$

จะได้ว่า $\sigma_i^2 = \sigma^2 x_i^2$; $i = 1,2,...,n$

จะสามารถใช้การแปลงข้อมูลดังนี้

$$\frac{Y_i}{X_i} = \frac{\beta_0}{X_i} + \beta_1 \frac{X_i}{X_i} + \frac{\varepsilon_i}{X_i}$$

$$Y_i^* = \nu_0 + \nu_1 X_i + \varepsilon_i^*$$

เมื่อ $Y_i^* = \frac{Y_i}{X_i}$, $X_i^* = \frac{1}{X_i}$, $\varepsilon_i^* = \frac{\varepsilon_i}{X_i}$

$$\nu_0 = \beta_1,\ \nu_1 = \beta_0$$

$$\mathrm{Var}(Y_i^* \,|X_i = x_i) = \frac{1}{x_i^2}\mathrm{Var}(\varepsilon_i \,|X_i = x_i) = \sigma^2$$

หมายความว่า การแปลงข้อมูล ต้องทำการแปลงข้อมูลของตัวแปรอิสระและข้อมูลตัวแปรตามด้วยเช่นกัน ไม่สามารถแปลงข้อมูลของตัวแปรใดตัวแปรหนึ่งได้ โดยการแปลงข้อมูลนั้นจะใช้ค่าสังเกตของตัวแปรอิสระไปหาข้อมูลทั้งของตัวแปรอิสระและตัวแปรตามในแต่ละรายการของค่า

สังเกตนั้นๆและกำหนดตัวแบบความถดถอยเป็นตัวแบบการถดถอยเชิงเส้นอย่างง่ายของข้อมูลตัวแปรอิสระใหม่ๆคือ $X_i^* = \frac{1}{X_i}$ และข้อมูลใหม่ของตัวแปรตาม $Y_i^* = \frac{Y_i}{X_i}$ และกำหนดตัวแบบการถดถอยเชิงเส้นอย่างง่ายในพารามิเตอร์ด้วยส่วนตัดแกน Y

<u>Heteroscedasticity Consistent Covariance Matrix (HCCM)</u>

เป็นตัวสถิติที่บรรเทาปัญหาค่าความคลาดเคลื่อนมีความแปรปรวนที่มีค่าไม่คงที่ (Heteroscedasticity) ซึ่งจะทำให้ค่าความคลาดเคลื่อนมาตรฐานของตัวประมาณสัมประสิทธิ์การถดถอยแตกต่างจากเดิม

Judge et al. (1985: 422-445) ได้พิจารณารูปแบบโครงสร้าง และค่าถ่วงน้ำหนักของแต่ละค่าสังเกตโดยใช้ส่วนกลับของส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนจากแต่ละค่าสังเกต พบว่า ตัวประมาณสัมประสิทธิ์การถดถอย เป็นตัวประมาณที่มีประสิทธิภาพ และ ไม่เอนเอียง แต่ในความเป็นจริง รูปแบบของความแปรปรวนของความคลาดเคลื่อนที่มีค่าไม่คงที่ ไม่สามารถทราบถึงรูปแบบได้ชัดเจน ซึ่งวิธีการที่ใช้ ได้แก่ the heteroscedasticity consistent covariance matrix ซึ่งต่อมาเรียกว่า HCCM ที่จะทำให้ค่าประมาณความแปรปรวนร่วมของสัมประสิทธิ์การถดถอยเป็นตัวประมาณที่คงเส้นคงวาได้ โดย White's (1980) ได้ต่อยอดจากงานวิจัยของ Eicker (1963, 1967), Huber (1967) และ Honn, Honn และ Duncan (1975)

จากตัวแบบการถดถอยเชิงเส้นที่สามารถเขียนได้ดังนี้

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip} + \varepsilon_i \ ; \ i = 1,2,...,n$$

โดยที่	Y	แทน	ตัวแปรตามหรือตัวแปรที่สนใจศึกษา
	$X_{i1}, X_{i2}, ..., X_{ip}$	แทน	ตัวแปรอิสระ
	$\beta_0, \beta_1, ..., \beta_p$	แทน	ค่าสัมประสิทธิ์การถดถอย
	ε	แทน	ความคลาดเคลื่อนเชิงสุ่ม

ในกรณีของวิธีกำลังสองน้อยที่สุด (OLS) จากข้อกำหนดของตัวแบบ ตัวประมาณของวิธี OLS คือ $\hat{\underset{\sim}{\beta}} = (X'X)^{-1} X'\underset{\sim}{Y}$ ซึ่งมีเมทริกซ์ความแปรปรวนร่วม คือ

$$Var(\hat{\beta}) = (X'X)^{-1} X'\Phi X (X'X)^{-1}$$

โดยที่ Φ คือเมทริกซ์แนวทแยง (diagonal matrix) ที่มีสมาชิกเป็น $\phi_{ii} = \mathrm{Var}(\varepsilon_i)$

ถ้าทราบว่าข้อมูลค่าความคลาดเคลื่อนมีความแปรปรวนคงที่ Φ จะสามารถเขียนได้เป็น $\Phi = \sigma^2 I$ และนิยามค่าเศษเหลือ (residuals) ที่ได้โดยจากวิธี OLS เป็น $e_i = Y_i - X_i\hat{\underset{\sim}{\beta}}$ ดังนั้น ค่าประมาณความแปรปรวนร่วมของสัมประสิทธิ์การถดถอยที่ได้จากวิธี OLS หรือเรียกสั้นๆว่า OLSCM จะเท่ากับ

$$\mathrm{OLSCM} = \frac{\sum e_i^2}{n-p}\left(X'X\right)^{-1} = s^2\left(X'X\right)^{-1}$$

เมื่อความแปรปรวนของความคลาดเคลื่อนมีค่าไม่คงที่ แล้ว OLSCM จะเป็นตัวประมาณที่เอนเอียง และทำให้ผลการทดสอบสมมติฐานผิดพลาด (invalid) ซึ่งถ้าทราบ Φ จะสามารถแก้ไขปัญหาค่าความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ได้ ซึ่งในความเป็นจริงจะไม่สามารถทราบรูปแบบของโครงสร้างความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ได้ชัดเจน แต่ถ้าไม่ทราบ Φ จะต้องหาตัวประมาณ Φ ที่คงเส้นคงวา ซึ่งใช้ค่าเศษเหลือกำลังสอง e_i^2 ที่ได้จากวิธี OLS ในการประมาณ ϕ_{ii} ซึ่งเมื่อประมาณความแปรปรวนของความคลาดเคลื่อนของแต่ละค่าสังเกตที่มีค่าไม่ซ้ำกัน จะได้ว่า

$$\hat{\phi}_{ii} = \frac{(e_i - 0)^2}{1} = e_i^2$$

แล้วให้ $\hat{\Phi} = \mathrm{diag}(e_i^2)$

ดังนั้นตัวประมาณที่ได้คือ

$$\mathrm{HC0} = (X'X)^{-1}X'\hat{\Phi}X(X'X)^{-1} = P\hat{\Phi}P'$$

โดยที่ $P = (X'X)^{-1}X'$ มาจากเมทริกซ์ H (Hat matrix) ที่ $H = X(X'X)^{-1}X' = XP$

ซึ่ง White's (1980) แสดงให้เห็นว่า $\mathrm{HC0}$ เป็นตัวประมาณที่คงเส้นคงวาของ $\mathrm{Var}(\hat{\underset{\sim}{\beta}})$ ที่มีข้อดีที่ไม่จำเป็นต้องทราบถึงรูปแบบความแปรปรวนของความคลาดเคลื่อนที่มีค่าไม่คงที่

ให้ h_i แทนสมาชิกในแนวทแยง (diagonal element) ของ hat matrix มีค่าอยู่ในช่วงตั้งแต่ 0 ถึง 1 (Belsley et al, 1990: 13-19) (ค่า h_i เป็นค่าที่ใช้ความผิดปกติของข้อมูลตัวแปรอิสระอันดับรายการที่ i แล้วจะทำให้ได้ว่า $\mathrm{Var}(e_i) = \sigma^2(1-h_i) \neq \sigma^2$ ซึ่งค่า $\mathrm{Var}(e_i)$ เป็นค่าประมาณที่ underestimate ค่า σ^2

Hinkley (1977) ได้ดัดแปลงจากตัวประมาณ HC0 ในการปรับค่าเศษเหลือแต่ละตัว โดยการใช้ค่าองศาอิสระ (degrees of freedom) ด้วยเทอม $\sqrt{\frac{n}{n-p}}$ โดยที่ p คือ จำนวนตัวแปรอิสระทั้งหมดที่อยู่ในตัวแบบ ดังนั้นตัวประมาณ HCCM ที่ได้ คือ

$$HC1 = \frac{n}{n-p}(X'X)^{-1}X'\hat{\Phi}X(X'X)^{-1} = \frac{n}{n-p}HC0 = PE_1\hat{\Phi}P'$$

โดยที่ $E_1 = \left(\frac{n}{n-p}\right)I$

แต่เนื่องจาก e_i^2 เป็นตัวประมาณที่เอนเอียง (biased estimator) ของ σ^2 ดังนั้นจาก $Var(e_i) = \sigma^2(1-h_i) \neq \sigma^2$ แสดงว่า $\frac{e_i^2}{1-h_i}$ ควรเป็นตัวประมาณที่มีความเอนเอียงน้อยกว่าเดิม ด้วยเหตุนี้ Mackinson และ White (1985) จึงเสนอตัวประมาณ

$$HC2 = (X'X)^{-1}X'\,\mathrm{diag}\left[\frac{e_i^2}{1-h_i}\right]X(X'X)^{-1} = PE_2\hat{\Phi}P'$$

โดยที่ $E_2 = \mathrm{diag}\left[\frac{1}{1-h_i}\right]$

Davidson และ Mackinson (1993) ได้พยายามทอนค่า e_i^2 ที่มีค่ามาก ให้มีค่าลดลงกว่าเดิม โดยการหารด้วย $(1-h_i)^2$ ทำให้ได้ตัวประมาณ

$$HC3 = (X'X)^{-1}X'\,\mathrm{diag}\left[\frac{e_i^2}{(1-h_i)^2}\right]X(X'X)^{-1} = PE_3\hat{\Phi}P'$$

โดยที่ $E_3 = \mathrm{diag}\left[\frac{1}{(1-h_i)^2}\right]$

ซึ่งสังเกตว่า ตัวประมาณ HC2 และ HC3 จะสามารถครอบคลุมถึงค่า leverage ของข้อมูลค่าสังเกตแต่ละตัว หากค่าสังเกตของตัวแปรอิสระมีความผิดปกติมาก จะทำให้ค่าเศษเหลือยกกำลังสอง มีค่าสูงกว่าที่ควรจะเป็น (inflated)

ต่อมา Cribari-Neto et al. (2007) ได้นำเสนอตัวประมาณ HCCM นั่นคือ

$$HC4 = (X'X)^{-1}X'\,\mathrm{diag}\left[\frac{e_i^2}{(1-h_i)^{\delta_i}}\right]X(X'X)^{-1} = PE_4\hat{\Phi}P'$$

เมื่อ $\delta_i = \min\left\{4, \frac{nh_i}{p}\right\}$ และ $E_4 = \mathrm{diag}\left[\frac{1}{(1-h_i)^{\delta_i}}\right]$

ในงานวิจัยของ Cribari-Neto และ Zarkos (2001) ได้อธิบายว่า ค่า leverage ที่มีค่ามากจะยังสามารถมีคุณสมบัติของตัวประมาณที่ดีความแปรปรวนร่วมของสัมประสิทธิ์การถดถอยภายใต้ขนาดตัวอย่างที่จำกัด ได้ดีกว่า กรณีที่เกิดปัญหาความคลาดเคลื่อนมีความแปรปรวนมีค่าไม่คงที่มีความรุนแรง การใช้ตัวประมาณ HC4 จะมีวัตถุประสงค์เพื่อลดอิทธิพลจากค่า leverage ของค่าสังเกตได้ดีกว่าตัวประมาณ HC2 และ HC3

และตัวประมาณ HC5 (Cribari-Neto et al. (2007) สามารถเขียนได้ดังนี้

$$HC5 = (X'X)^{-1}X' \,\mathrm{diag}\left[\frac{e_i^2}{\sqrt{(1-h_i)^{\delta_i}}}\right] X(X'X)^{-1} = PE_5\hat{\Phi}P'$$

เมื่อ $\delta_i = \min\left\{\frac{nh_i}{p}, \max\left\{4, \frac{nh_{max}}{p}\right\}\right\}$, $h_{max} = \max\{h_1,...,h_n\}$ และ $k \in [0,1]$

โดยค่า k ที่แนะนำ Cribari-Neto et al. (2010) คือเท่ากับ 0.7 ซึ่งจะเห็นได้ว่าเป็นการลดเทอมค่า e_i^2 โดยใช้ค่า leverage ที่มีค่ามากที่สุดจากข้อมูลค่าสังเกตทั้งหมด

2.3 วิธีกำลังสองน้อยที่สุด (Ordinary Least Square: OLS)

วิธีการหาสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด เป็นวิธีที่มีรากฐานมาจากทฤษฎีการประมาณค่าเชิงเส้น (Theory of linear Estimation) ซึ่งเป็นวิธีที่คิดค้นโดย คาร์ล เฟรดดิช เกาส์ (Karl Friedrich Gauss, 1777-1855) และ อังเดร แอนดรีวิช มาร์คอฟ (Andrie Andreevich Markov, 1856-1922) โดยมีหลักเกณฑ์ดังนี้ (ประชุม สุวัตถี, 2553) คือ หาค่าประมาณของพารามิเตอร์ที่ทำให้ผลบวกกำลังสองของผลต่างระหว่างค่าสังเกต กับ ค่าคาดหวังของตัวแปรมีค่าต่ำที่สุด

นุชรินทร์ ทิพยวรรณากร (2540) ได้อธิบายถึงนิยามเกี่ยวกับตัวประมาณกำลังสองน้อยที่สุดว่ากำหนดให้ $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ โดยที่ $\underset{\sim}{\varepsilon} \sim N_n\left(0, \sigma^2 I_n\right)$ จะได้ตัวประมาณกำลังสองน้อยสุดของ $\underset{\sim}{\beta}$ คือ $\hat{\underset{\sim}{\beta}}$ ที่จะทำให้ผลรวมกำลังสองของผลต่างระหว่างค่าสังเกต กับ ค่าคาดหวังของตัวแปร

มีค่าน้อยที่สุดโดยที่ค่าประมาณของ $\underset{\sim}{\beta}$ คือ $\hat{\underset{\sim}{\beta}} = (X'X)^{-1} X'\underset{\sim}{Y}$ และเมทริกซ์ความแปรปรวนร่วมของตัวประมาณ $\hat{\underset{\sim}{\beta}}$ คือ

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

ดังนั้นจากนิยาม จะได้ผลรวมกำลังสองของความคลาดเคลื่อนอยู่ในรูปแบบดังนี้

$$\begin{aligned} SSE\left(\underset{\sim}{\beta}\right) &= \underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon} = \left(\underset{\sim}{Y} - X\underset{\sim}{\beta}\right)'\left(\underset{\sim}{Y} - X\underset{\sim}{\beta}\right) \\ &= \underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{\beta}'X\underset{\sim}{Y} - \underset{\sim}{Y}'X\underset{\sim}{\beta} + \underset{\sim}{\beta}'X'X\underset{\sim}{\beta} \qquad (1) \end{aligned}$$

เนื่องจากวิธีกำลังสองน้อยสุดมีหลักการที่ทำให้ผลรวมของผลต่างระหว่างค่าสังเกต กับค่าคาดหวังของตัวแปรกำลังสองมีค่าต่ำที่สุด ดังนั้นเราจะได้ค่าประมาณสัมประสิทธิ์ความถดถอยจากการหาอนุพันธ์อันดับที่ 1 ของสมการที่ (1) เทียบกับ $\underset{\sim}{\beta}$ แล้วให้เท่ากับศูนย์ ซึ่งผลดังกล่าวอยู่ในรูปแบบดังนี้

$$\frac{\partial}{\partial \underset{\sim}{\beta}} SSE\left(\underset{\sim}{\beta}\right) = -2X^T\underset{\sim}{Y} - 2X^TX\hat{\underset{\sim}{\beta}} = 0$$

กล่าวคือ $X^TX\hat{\underset{\sim}{\beta}} = X'\underset{\sim}{Y}$

ถ้า X^TX เป็นเมทริกซ์ที่ไม่เอกฐาน *(*Non-singular matrix) แล้วจะได้ว่า

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1} X'\underset{\sim}{Y} \qquad (2)$$

ตัวประมาณในสมการที่ (2) มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงของ $\underset{\sim}{\beta}$ โดยกำลังสองน้อยสุดนี้เราไม่จำเป็นต้องทราบลักษณะการแจกแจงความน่าจะเป็นของความคลาดเคลื่อน $\underset{\sim}{\varepsilon}$ นอกจากนั้นตัวประมาณ $\hat{\underset{\sim}{\beta}}$ ยังเป็นตัวประมาณด้วยภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimator) ของ $\underset{\sim}{\beta}$ ด้วย รวมทั้งเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด (Best Linear Unbiased Estimator: BLUE) ทีความคงเส้นคงวา (consistent) และมีความพอเพียง (sufficient)

การประมาณค่า $\beta_0, \beta_1, \beta_2, ..., \beta_p$ ด้วย $b_0, b_1, b_2, ..., b_p$ โดยใช้วิธีกำลังสองน้อยที่สุด จะได้ว่า

1. ผลรวมของผลต่างระหว่างค่าสังเกตตัวแปรตาม (Y_i) กับ ค่าคาดหวังของตัวแปรตามในการประมาณค่าด้วย $\hat{Y}_i$ เป็นศูนย์ คือ $\sum_{i=1}^{n}(Y_i - \hat{Y}_i) = \sum_{i=1}^{n} e_i = 0$

2. จุด $(\bar{X}, \bar{Y})$ เป็นจุดที่อยู่บนเส้นความถดถอย Y

3. $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ มีค่าต่ำที่สุด

สุพล ดุรงค์วัฒนา (2549) ได้กล่าวถึงข้อควรระมัดระวังในการใช้สมการถดถอยเพื่อการพยากรณ์ไว้ว่า มีประเด็นที่ต้องระมัดระวังอยู่ 2 ประเด็น คือ

1. ปัญหาการพยากรณ์นอกกรอบค่าของตัวแปรอิสระ

ไม่ว่าจะเป็นการพยากรณ์ค่าเฉลี่ยหรือค่าแต่ละค่าของตัวแปรตาม เพราะจะทำให้ค่าคลาดเคลื่อนในการพยากรณ์สูง แต่ถ้ามีความจำเป็นที่จะต้องพยากรณ์นอกช่วงของตัวแปรอิสระ ผู้ทำพยากรณ์จะต้องกระทำด้วยความระมัดระวัง

2. ปัญหาการพยากรณ์แบบสวนกลับ

ในการใช้สมการถดถอยสำหรับการพยากรณ์นั้นมุ่งหวังที่จะพยากรณ์ตัวแปรตามโดยกำหนดหรือรู้ค่าของตัวแปรอิสระ บางครั้งการนำไปใช้ประโยชน์บ่อยครั้งที่ผู้นำไปใช้ ไม่ต้องการที่จะพยากรณ์ตัวแปรตาม แต่กลับต้องการพยากรณ์ตัวแปรอิสระเมื่อกำหนดหรือรู้ค่าของตัวแปรตามแทน ซึ่งการพยากรณ์แบบนี้ไม่สมควรกระทำ เนื่องจากผิดหลักการของการพยากรณ์ เพราะสมการถดถอยถูกกำหนดขึ้นจากตัวแปรตามและตัวแปรอิสระ โดยที่ตัวแปรตามเป็นเหตุการณ์ที่ต้องเกิดขึ้นหลังจากเหตุการณ์ของตัวแปรอิสระตามเงื่อนไข

2.4 วิธีการแปลง Box และ Cox (Box-Cox transformation)

Box และ Cox (1964) เสนอวิธีการแปลงข้อมูลโดยใช้หลักการประเมินค่า power ที่ใช้ในการแปลงด้วยวิธีการแบบภาวะความน่าจะเป็นสูงสุด (Maximum likelihood) จากสมการการแปลงค่าข้อมูลของ Box และ Cox ซึ่งมีรูปแบบเป็น

$$Y^* = \frac{Y^{\lambda} - 1}{\lambda} \quad \text{เมื่อ } \lambda \neq 0$$

$$Y^* = \log(Y) \quad \text{เมื่อ } \lambda = 0$$

ต่อมาหรือมีรูปแบบการแปลงดังนี้คือ

$$Y'' = Y^{\lambda} \quad \text{เมื่อ } \lambda \neq 0$$

$$Y'' = \log(Y) \quad \text{เมื่อ } \lambda = 0$$

ซึ่งรูปแบบการแปลงของ Box และ Cox จะมีรูปแบบเฉพาะดังนี้

เมื่อค่า $\lambda = 0$ คือรูปแบบ $Y^* = \log(Y)$ เรียกการแปลงดังกล่าวว่า การแปลงโดยใช้ลอการิทึม (Logarithm transformation)

เมื่อค่า $\lambda = \frac{1}{2}$ คือรูปแบบ $Y^* = \sqrt{Y}$ เรียกการแปลงดังกล่าวว่า การแปลงโดยใช้รากที่สอง (Square root transformation)

แต่ถ้าค่า $\lambda = 1$ แสดงว่า การแปลงข้อมูลของค่าตัวแปรตาม ไม่จำเป็นต้องมีการแปลงเกิดขึ้น

การคำนวณหาค่า λ กระทำได้โดยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimator: ML) ผสมกับ Iteration Method โดยค่อยๆเปลี่ยนค่า λ ไปทีละน้อยโดยในแต่ละครั้งที่เปลี่ยนค่า λ ไปให้คำนวณหา $\hat{\beta}$ และ $\hat{\sigma}^2$ ทุกครั้ง ซึ่งสามารถคำนวณได้โดย

$$\hat{\beta} = (X'X)^{-1}X'Y^*$$

$$\hat{\sigma}^2 = \frac{1}{n}(Y^{*\prime}(I-G)Y^*)$$

โดยที่ $G = X(X'X)^{-1}X'$

จากนั้นดูว่าค่า λ ค่าใดที่มีผลให้ $\hat{\sigma}^2$ มีค่าต่ำสุด ก็ถือว่า λ นั้นเป็นตัวประมาณภาวะน่าจะเป็นสูงสุด (maximum likelihood estimator) แล้ววิเคราะห์ด้วยสมการถดถอยตามปกติต่อไป

ซึ่งการแปลงรูปตัวแปรตามลักษณะนี้มีผลทำให้ตัวแปรตาม มีคุณสมบัติ 3 ประการ คือ

1. Y^* มีการแจกแจงแบบปกติ
2. ทำให้ความคลาดเคลื่อนมีความแปรปรวนมีค่าคงที่
3. Y^* เป็นฟังก์ชันเชิงเส้น (Linear function) ของพารามิเตอร์สัมประสิทธิ์การถดถอย (regression coefficient)

2.5 วิธี Iteratively Reweighted Least Square (IRWLS)

ก่อนที่ผู้วิจัยจะอธิบายถึง วิธีและขั้นตอน ของวิธี IRWLS ทั้งนี้เพื่อให้เกิดความเข้าใจมากขึ้น ผู้วิจัยจะกล่าวถึงเนื้อหาที่ควรรู้และเกี่ยวข้องกับวิธี IRWLS ดังนี้

2.5.1 ตัวแบบเชิงเส้นที่วางนัยทั่วไป (Generalized Linear Model: GLM)

หลักการของตัวแบบเชิงเส้นที่วางนัยทั่วไป ได้ถูกนำเสนอขึ้นโดย Nelder และ Wedderburn (1972) กับ Mc Cullagh และ Nelder (1989) แบ่งออกเป็นส่วนประกอบได้ทั้งหมด 3 ส่วน (วีรานันท์ พงศาภักดี, 2541) คือ

<u>ส่วนประกอบที่ 1</u> ส่วนประกอบเชิงสุ่ม (random experiment)

ส่วนประกอบเชิงสุ่มเป็นส่วนประกอบที่เกี่ยวข้องกับสมมติฐานของการแจกแจงของตัวแปรสุ่ม (Y) ที่เป็นตัวแปรตาม สมมติว่าค่าสังเกตของตัวแปรตามขนาด n หน่วยที่เป็นอิสระต่อกัน นั่นคือ $Y=(Y_1,Y_2,...,Y_n)$ แต่ละส่วนประกอบของ Y คือ Y_i , $i=1,2,...,n$ มีการแจกแจงในกลุ่มวงศ์ชี้กำลัง (Exponential family) ซึ่งอยู่ในรูปแบบของ ϕ

$$f(y|\theta,\phi)=\exp\left[\frac{y\theta-b(\theta)}{a(\phi)}+c(y,\phi)\right] \qquad (3)$$

โดยที่ $a(\cdot)$,$b(\cdot)$ และ $c(\cdot)$ แทนฟังก์ชันต่างๆ ถ้าทราบ ϕ แล้วสมการ (3) คือตัวแบบหนึ่งในกลุ่มวงศ์ชี้กำลังที่มีพารามิเตอร์ θ แต่ถ้าไม่ทราบ ϕ แล้วสมการ (3) อาจเป็นหรือไม่เป็นตัวแบบหนึ่งในกลุ่มวงศ์ชี้กำลัง ที่มีพารามิเตอร์จำนวน 2 ตัว (θ,ϕ) สำหรับพารามิเตอร์ θ เรียกว่า natural parameter หรือ canonical parameter ซึ่งแสดงถึง location parameter ส่วน ϕ มักเรียกว่า dispersion parameter ยกตัวอย่าง เช่น

1. การแจกแจงปกติ (Normal หรือ Gaussian Distribution)

$$f(y|\theta,\phi)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] =\exp\left[\frac{y\mu-\frac{\mu^2}{2}}{\sigma^2}-\frac{1}{2}\left(\frac{y^2}{\sigma^2}+\log(2\pi\sigma^2)\right)\right]$$

ซึ่งจากรูปแบบทั่วไปของกลุ่มวงศ์ชี้กำลัง (Exponential Family) จะได้ว่า

$$\theta=\mu\ ,\phi=\sigma^2,\ a(\phi)=\phi\ ,\ b(\theta)=\frac{\theta^2}{2}\ \text{ และ }\ c(y,\phi)=-\frac{1}{2}\left[\frac{y^2}{\phi}+\log(2\pi\phi)\right]$$

2. การแจกแจงปัวส์ซอง (Poisson Distribution)

$$f(y|\theta,\phi)=\frac{e^{-\mu}\mu^y}{y!} =\exp(y\ln\mu-\mu\ln y!)$$

ซึ่งจากรูปแบบทั่วไปของกลุ่มวงศ์ชี้กำลัง (Exponential Family) จะได้ว่า

$\theta = \ln(\mu)$, $\phi \equiv 1$, $a(\phi) = 1$, $b(\theta) = \exp(\theta)$ และ $c(y,\phi) = -\ln y!$

3. การแจกแจงทวินาม (Binomial Distribution)

$$f(y|\theta,\phi) = \binom{n}{y}\mu^{y}(1-\mu)^{n-y} = \exp\left(y\log\mu + (n-y)\log(1-\mu) + \log\binom{n}{y}\right)$$

ซึ่งจากรูปแบบทั่วไปของกลุ่มวงศ์ชี้กำลัง (Exponential Family) จะได้ว่า

$$\theta = \log\left(\frac{\mu}{1-\mu}\right) \text{ , } b(\theta) = -n\log(1-\mu) = n\log(1+\exp(\theta)) \text{ และ } c(y,\phi) = \log\binom{n}{y}$$

<u>ส่วนประกอบที่ 2</u> ส่วนประกอบแบบมีระบบ (systematic component)

ส่วนประกอบแบบมีระบบเป็นส่วนประกอบที่ทำหน้าที่เชื่อมเวกเตอร์ η โดยที่ $\eta = (\eta_1,\eta_2,...,\eta_n)'$ กับเซตของตัวแปรอิสระให้มีรูปแบบเชิงเส้นดังนี้

$$\eta = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p = X\underset{\sim}{\beta} \text{ , } \eta_i = \sum_j \beta_j X_{ij} \text{ ; } i = 1,2,...,n \text{ , } j = 1,2,...,p$$

โดยที่ X แทน เมทริกซ์ของตัวแปรอิสระ ที่ประกอบด้วยค่าสังเกตขนาด n อาจเรียกว่า design matrix

$\underset{\sim}{\beta}$ แทน เวกเตอร์ของพารามิเตอร์ $(\beta_1,\beta_2,...,\beta_p)'$

η แทน ตัวพยากรณ์เชิงเส้น (linear predictor)

<u>ส่วนประกอบที่ 3</u> ส่วนประกอบ link function

ส่วนประกอบที่ใช้อธิบายฟังก์ชันความสัมพันธ์ระหว่างส่วนประกอบแบบมีระบบกับค่าเฉลี่ยของส่วนประกอบเชิงสุ่ม เช่น

ให้ $\mu_i = E(Y_i)$, $i = 1,2,...,n$

ดังนั้น μ_i จะเกี่ยวข้องกับ η_i ด้วยฟังก์ชันของ $\eta_i = g(\mu_i)$ โดยที่ g แทน ฟังก์ชันแบบ monotonic differentiable function ทำให้ได้ว่า ตัวแบบที่ต้องการจะเชื่อมระหว่าง link function ที่ทำให้ $\eta = g(\mu) = X\underset{\sim}{\beta}$ และเป็น canonical link ที่ทำให้ $\eta = g(\mu) = \theta$

ถ้า $g(\mu)=\mu$ จะได้ว่า $\eta_i=\mu_i$ เป็น identity link

Family	Link
Normal	$\eta=\mu$
Poisson	$\eta=\ln\mu$
Binomial	$\eta=\log\left(\dfrac{\mu}{1-\mu}\right)$

2.5.2 **การประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุดแบบวนซ้ำ** (iterative maximum likelihood estimation)

วีรานันท์ พงศาภักดี (2541) ได้อธิบายถึง การประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุดแบบวนซ้ำ ว่าเป็นวิธีที่ใช้ในการประมาณค่าของพารามิเตอร์ของตัวแบบในกลุ่มวงศ์ชี้กำลัง (exponential family models) โดยสืบเนื่องมาจาก Nelder, J.A. และ Wedderburn, R.W.M. (1972) ได้ขยายวิธีการประมาณค่าแบบวนซ้ำ (iterative estimation) วิธีหนึ่งที่เรียกว่า วิธีฟิชเชอร์สกอริ่ง (Fisher Scoring) ให้ใช้ควบคู่กับวิธีภาวะน่าจะเป็นสูงสุด (maximum likelihood) ซึ่งเป็นตัวแบบที่หมายรวมถึง ตัวแบบเชิงเส้นที่มีพื้นฐานของการแจกแจงแบบปกติ และ ตัวแบบอื่นๆ ที่มีการแจกแจงในกลุ่มวงศ์ชี้กำลังด้วย

การประมาณค่าพารามิเตอร์ของตัวแบบเชิงเส้นทั่วไปในอดีต นิยมใช้วิธีภาวะน่าจะเป็นสูงสุด และ วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (Weighted Least Square: WLS) อย่างไรก็ตามในหลายสถานการณ์ พบว่า ไม่สามารถใช้วิธีการประมาณค่าดังกล่าวได้โดยตรง เนื่องจากสมการปกติที่พบอาจไม่เป็นลักษณะเชิงเส้น หรือ มีรูปแบบไม่เชิงเส้น (nonlinear) ในเทอมของพารามิเตอร์ การแก้สมการเหล่านี้ จึงจำเป็น ต้องมีการคำนวณเพิ่มเติมด้วยวิธีการวนซ้ำเชิงตัวเลข (numerical iteration) ซึ่งมีวิธีการวนซ้ำ (iterative procedures) หลายวิธี และ เมื่อนำมาใช้ร่วมกับวิธีภาวะน่าจะเป็นสูงสุดด้วย ทำให้เป็นวิธีที่มีประสิทธิภาพมากขึ้น และ เป็นที่นิยมใช้ในปัจจุบัน โดยสามารถใช้โปรแกรมสำเร็จรูปมาช่วยคำนวณได้ เช่น GLIM, SPSS, SAS, R เป็นต้น

การประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุดแบบวนซ้ำที่จะกล่าวต่อไปนี้ มีกระบวนการวนซ้ำ (iterative process) ที่เริ่มต้นจากวิธีภาวะน่าจะเป็นสูงสุดก่อน ส่วนสมการปกติภายใต้

กระบวนการวนซ้ำจะอยู่ในลักษณะสมการปกติของวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก ที่ทำให้ค่าของการถ่วงน้ำหนัก (weight matrix) เปลี่ยนไปในแต่ละรอบ (cycle) หรือแต่ละ iteration ของการวนซ้ำต่างๆ กระบวนการวนซ้ำจะจบลงต่อเมื่อค่าประมาณของพารามิเตอร์ลู่เข้า (converge) สู่ค่าประมาณภาวะน่าจะเป็นสูงสุด (maximum likelihood estimates) หรือ ทำให้ค่าผลต่างของค่าประมาณจากกระบวนการวนซ้ำนั้นมีค่าน้อย หรือ เล็กเพียงพอ (sufficient small) จึงเรียกกระบวนการดังกล่าวว่า การประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุดแบบวนซ้ำ

การประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุดแบบวนซ้ำใช้ได้สำหรับการประมาณค่าพารามิเตอร์ทั้งของตัวแบบเชิงเส้น ตัวแบบเชิงเส้นที่วางนัยทั่วไป และ ตัวแบบไม่เชิงเส้นอื่นๆ โดยสามารถใช้หลักเกณฑ์ของวิธีภาวะน่าจะเป็นสูงสุด ร่วมกับ การใช้กระบวนการวนซ้ำ จากวิธีใดวิธีหนึ่งใน 3 วิธีต่อไปนี้

1. วิธีนิวตัน-รัฟสัน (Newton-Raphson method)
2. วิธีฟิชเชอร์-สกอริ่ง (Fisher-Scoring หรือ Fisher's Scoring หรือ the method of scoring)
3. วิธีเดมมิ่ง-สตีเฟน (Deming-Stephen Iterative Proportional Fitting method: IDF)

ในเรื่องการประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุดแบบวนซ้ำนี้ ประกอบด้วยหัวข้อที่เป็นลำดับ คือ ทบทวนวิธีประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุด และ วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก หลักเกณฑ์ของการประมาณค่าด้วยกระบวนการวนซ้ำ ซึ่งในงานวิจัยนี้จะกล่าวเพียง 2 วิธีเท่านั้น คือ วิธีนิวตัน-รัฟสัน และวิธีฟิชเชอร์-สกอริ่ง ก่อนที่จะเข้าสู่วิธี IRWLS ในหัวข้อถัดไป

(i) วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood method หรือ ML

ให้ $Y_1, Y_2, ..., Y_n$ แทนตัวแปรสุ่ม n ตัวที่มีฟังก์ชันความหนาแน่นน่าจะเป็น (joint probability density function) คือ

$$f(Y;\theta) = f(Y_1, ..., Y_n; \theta_1, ..., \theta_p)$$

ขึ้นอยู่กับพารามิเตอร์ $\theta_1, ..., \theta_p$ โดยที่ Y แทน $(Y_1, ..., Y_n)'$ และ θ แทน $(\theta_1, ..., \theta_p)'$ ส่วน p แทนจำนวนของพารามิเตอร์ สำหรับฟังก์ชัน $f(Y;\theta)$ นั้น Y เป็นตัวแปรสุ่ม และ θ เป็นค่าคงที่

ในกรณีที่มีค่าสังเกตค่าของตัวแปรสุ่ม Y จำนวน n ค่าที่เป็นอิสระต่อกัน คือ $y_1,...,y_n$ ภายใต้พารามิเตอร์ θ ฟังก์ชันภาวะน่าจะเป็น (likelihood function) คือ

$$f(y;\theta) = f(y_1,...,y_n;\theta_1,...,\theta_p) = f(y_1;\theta)\cdot f(y_2;\theta)\cdots f(y_n;\theta)$$

เป็นที่น่าสังเกต y ในฟังก์ชันภาวะน่าจะเป็นจะเหมือนกับ Y ในฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม เพียงแต่ y เป็นค่าสังเกตของ Y

ให้ $L = \log f(y;\theta) = \sum_i \log f(y_i;\theta) = \sum_i l_i$

ให้ φ แทนเซตที่เป็นไปได้ทั้งหมดของเวกเตอร์พารามิเตอร์ θ

ดังนั้น ตัวประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimator หรือ MLE) ของ θ คือ $\hat{\theta}$ ที่ทำให้ฟังก์ชันภาวะน่าจะเป็น $f(Y;\theta)$ มีค่าสูงสุด นั่นคือ

$$f(y;\hat{\theta}) \geq f(y;\theta) \quad \text{สำหรับทุก } \theta \text{ ใน } \varphi$$

ตัวประมาณภาวะน่าจะเป็นสูงสุด หรือ $\hat{\theta}$ นี้เป็นค่าที่ทำให้ $L(\theta)$ สูงสุดด้วย เนื่องจากทำให้ฟังก์ชันลอการิทึม เป็นฟังก์ชันทางเดียว (monotone function) ดังนั้น

$$\log f(y;\hat{\theta}) \geq \log f(y;\theta) \quad \text{สำหรับทุก } \theta \text{ ใน } \varphi$$

$$\text{หรือ } L(\hat{\theta}) \geq L(\theta) \qquad \text{สำหรับทุก } \theta \text{ ใน } \varphi$$

โดยทั่วไป นิยมใช้ฟังก์ชันของล็อกภาวะน่าจะเป็น (log-likelihood) มากกว่าฟังก์ชันของ likelihood โดยตรง เนื่องจากช่วยให้การคำนวณสะดวกขึ้น สำหรับตัวประมาณ MLE สามารถหาได้จากอนุพันธ์ฟังก์ชัน L เทียบกับพารามิเตอร์ทีละตัว แล้วแก้สมการปกติพร้อมกัน คือ

$$\frac{\partial L}{\partial \theta_i} = 0 \quad \text{สำหรับ } i = 1,2,...,n$$

การตรวจว่าฟังก์ชัน L มีค่าสูงสุดหรือไม่ สามารถทำได้โดยการหาเมทริกซ์ของอนุพันธ์อันดับที่ 2 (second derivative) ของ L คือ $\frac{\partial^2 L}{\partial \theta_j \theta_k}$ ว่ามีค่า ไม่เป็นลบแน่นอน ณ ค่า $\hat{\theta}$ หรือไม่ เช่น กรณีถ้ามีพารามิเตอร์ θ เพียงตัวเดียว การตรวจสอบฟังก์ชัน L ว่ามีค่าสูงสุดหรือไม่ จะสามารถกระทำได้โดยตรวจสอบว่า $\frac{\partial^2 L}{\partial \theta_j \theta_k}$ มีค่าเป็นลบ ณ ค่าของ $\hat{\theta}$ หรือไม่

คุณสมบัติที่สำคัญอย่างหนึ่งของตัวประมาณภาวะน่าจะเป็นสูงสุด คือ คุณสมบัติที่เรียกว่า “invariance property of MLE” คือ ถ้า $f(\theta)$ เป็นฟังก์ชันใดๆของพารามิเตอร์ θ แล้วจะได้ว่า $f(\hat{\theta})$ เป็นตัวประมาณภาวะน่าจะเป็นสูงสุดของ $f(\theta)$ ด้วย โดยที่ $\hat{\theta}$ เป็น MLE ของ θ จากคุณสมบัติข้างต้น ทำให้สามารถหาตัวประมาณภาวะน่าจะเป็นสูงสุดของฟังก์ชันของ MLE ได้ นอกจากนี้ ตัวประมาณภาวะน่าจะเป็นสูงสุด ยังคงมีคุณสมบัติของคงเส้นคงวา (consistency), เพียงพอ (sufficiency) และ asymptotic efficiency ด้วย

(ii) วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (Weighted Least Square method หรือ WLS method

ให้ $Y_1, Y_2, ..., Y_n$ แทนตัวแปรสุ่มที่มีค่าคาดหวังเป็น

$$E(Y_i) = \mu_i \quad \text{สำหรับ } i = 1, 2, ..., n$$

ให้ μ_i's เป็นฟังก์ชันของพารามิเตอร์ $\beta_1, \beta_2, ..., \beta_p$ $(p \le n)$ และต้องการประมาณค่าของด้วย $\hat{\underset{\sim}{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_p)'$ ตามลำดับ

ให้ $Y_i = \mu_i + e_i$ สำหรับ $i = 1, 2, ..., n$ โดยที่ e_i แทนความคลาดเคลื่อนเชิงสุ่ม

จากวิธีกำลังสองน้อยที่สุด (OLS) ซึ่งมีวิธีการคำนวณสำหรับการหาค่าประมาณของ β ที่ทำให้ผลรวมกำลังสองของ e_i น้อยที่สุด คือ

ให้ $SSE = \sum_{i}^{n} e_i^2 = \sum_{i}^{n} (Y_i - \mu_i)^2 = (\underset{\sim}{Y} - \underset{\sim}{\mu})'(\underset{\sim}{Y} - \underset{\sim}{\mu})$

โดยที่ $\underset{\sim}{Y} = (Y_1, ..., Y_n)'$, $\mu = (\mu_1, ..., \mu_n)'$

โดยทั่วไปตัวประมาณของ β หรือ $\hat{\beta}$ สามารถหาได้จากการอนุพันธ์เทอม SSE เทียบกับ β_j ของ β แล้วแก้สมการ

$$\frac{\partial}{\partial \beta_j} SSE = 0 \quad \text{สำหรับ } j = 1, 2, ..., p$$

การตรวจสอบค่าต่ำสุดของ SSE ที่ใช้ประมาณค่าของ β ทำได้โดยการตรวจสอบเมทริกซ์ของอนุพันธ์อันดับที่สองว่าเป็นบวกแน่นอน (positive definite) หรือไม่ ถ้าใช่ ค่าประมาณที่ได้ก็ตรงกับที่ต้องการ

ในทางปฏิบัติ อาจมี Y_i บางตัวที่มีค่าสังเกตที่เชื่อถือได้น้อย กล่าวคือ อาจมีความแปรปรวนมากกว่า Y_i's ตัวอื่นๆ กรณีเช่นนี้อาจจำเป็นต้องมีการถ่วงน้ำหนักในเทอม SSE และใช้ SSE_w แทนเทอม SSE โดยที่

$$SSE_w = \sum_i V_i[Y_i - \mu_i]^2$$

โดยที่ $V_i = \dfrac{1}{Var(Y_i)}$

นอกจากนี้ Y_i's แต่ละตัวยังอาจไม่เป็นตัวอิสระต่อกัน กรณีเช่นนี้จึงควรใช้ $V = \dfrac{1}{W}$ เมื่อ W แทน เมทริกซ์ความแปรปรวนร่วม ของ Y_i's ดังนั้นวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก หรือ วิธี WLS สามารถคำนวณได้จากการทำให้ SSE_W มีค่าต่ำสุด โดยที่

$$SSE_W = (\underset{\sim}{Y} - \underset{\sim}{\mu})' W^{-1} (\underset{\sim}{Y} - \underset{\sim}{\mu})$$

แต่เนื่องจาก $\underset{\sim}{\mu}$ เป็นฟังก์ชันของ $\underset{\sim}{\beta}$ ซึ่งในกรณีที่ μ_i เป็นผลรวมเชิงเส้น (linear combination) ของ β_i , $i = 1,2,...,p$ เช่น ถ้า $\underset{\sim}{\mu} = X\underset{\sim}{\beta}$ โดยที่ X แทนเมทริกซ์ที่มีมิติขนาด $n \times p$ แล้วจะได้ว่า

$$SSE_W = (\underset{\sim}{Y} - X\underset{\sim}{\beta})' W^{-1} (\underset{\sim}{Y} - X\underset{\sim}{\beta})$$

$$\frac{\partial}{\partial \beta} SSE_W = (\underset{\sim}{Y} - X\underset{\sim}{\beta})' W^{-1} (\underset{\sim}{Y} - X\underset{\sim}{\beta})$$

ดังนั้น ค่าประมาณ $\hat{\underset{\sim}{\beta}}$ จากวิธี WLS สามารถคำนวณจากสมการปกติ

$$X' W^{-1} (\underset{\sim}{Y} - X\underset{\sim}{\beta}) = 0$$

$$X' W^{-1} X\underset{\sim}{\beta} = X' W^{-1} \underset{\sim}{Y}$$

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

<u>ข้อสังเกต</u>

1. สามารถตรวจสอบเมทริกซ์อนุพันธ์อันดับที่สอง ของ SSE_W ว่าเป็นบวกแน่นอน (positive definite) หรือไม่
2. วิธี WLS สามารถใช้ได้โดยไม่ต้องมีข้อตกลง (assumptions) เกี่ยวกับการแจกแจงของ Y_i นอกจากการกำหนดค่าคาดหวังและโครงสร้างของความแปรปรวนและความ

แปรปรวนร่วมของ Y_i's เท่านั้น อย่างไรก็ตาม ถ้าต้องการอนุมานเกี่ยวกับ $\underset{\sim}{\beta}$ จำเป็นต้องมีข้อตกลงเกี่ยวกับ Y_i's เพิ่มเติม ในขณะที่วิธีภาวะน่าจะเป็นสูงสุดจำเป็นต้องทราบการแจกแจงของ Y_i's เนื่องจากต้องกำหนดฟังก์ชันหนาแน่นความน่าจะเป็นร่วม (joint probability density function) ของ Y_i's

<u>ประโยชน์ของวิธี WLS</u>

วิธีกำลังสองน้อยที่สุด (OLS) และวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (WLS) เป็นวิธีที่ใช้สำหรับกรณีที่ความแปรปรวนของ Y_i คงที่และไม่คงที่ หรือ Y_i ไม่เป็นอิสระต่อกัน ตามลำดับ ถ้าค่าความแปรปรวน Y_i คงที่หรือเท่ากันทุก Y_i และเป็นอิสระต่อกัน แล้ววิธี OLS และ WLS จะให้ผลเหมือนกัน ทั้งสองวิธีนี้ต่างเป็นทางเลือกของวิธีภาวะน่าจะเป็นสูงสุด

1. การคำนวณของวิธี WLS มีรูปแบบมาตรฐาน (standard form) ที่ไม่ซับซ้อน และ สามารถนำไปประยุกต์สำหรับการแก้สมการในตัวแบบอื่นๆได้
2. การคำนวณของวิธี ML สามารถไปประยุกต์กับกระบวนการวนซ้ำ (iterative process) นั้นอาศัยประโยชน์ของการวนซ้ำด้วยวิธี WLS โดยในแต่ละรอบซ้ำ จะมีการถ่วงน้ำหนักด้วยเทอมที่มีค่าเปลี่ยนไปเรื่อยๆ เช่น การใช้วิธี Newton-Raphson และ วิธี Fisher Scoring ร่วมกับวิธี ML เพื่อประมาณพารามิเตอร์ในตัวแบบเชิงเส้นที่วางนัยทั่วไป และตัวแบบล็อกลิเนียร์
3. สำหรับตัวแบบที่เหมาะสม ตัวประมาณของวิธี WLS และ ML เป็นตัวประมาณที่สมมูลกัน และต่างก็เป็นตัวประมาณแบบ Best Asymptotical Normal เมื่อตัวอย่างมีขนาดใหญ่มากขึ้น มีการแจกแจงเข้าใกล้การแจกแจงแบบปกติ และอัตราส่วนของความแปรปรวนของทั้งสองวิธี จะลู่เข้าสู่ค่า 1 ด้วย
4. สำหรับการประมาณค่าแบบจุด (point estimation) ของพารามิเตอร์ด้วยวิธี WLS ไม่จำเป็นต้องมีข้อตกลงเกี่ยวกับลักษณะการแจกแจงของ Y แต่การประมาณค่าแบบช่วง (interval estimation) และการทดสอบสมมติฐานเกี่ยวกับตัวแบบ จำเป็นต้องทราบการแจกแจงเพิ่มเติม

<u>ข้อจำกัดของวิธี WLS เมื่อเปรียบเทียบกับวิธี ML</u>

1. สำหรับข้อมูลเชิงกลุ่ม ในการใช้วิธี WLS ต้องมีการประมาณค่าความแปรปรวนร่วมกับโครงสร้างความแปรปรวนร่วมแบบพหุนาม (multinomial covariance structure) ของข้อมูลตัวอย่างในแต่ละชุดของตัวแปรอิสระ
2. ในกรณีที่จำนวนของตัวแปรอิสระมีมาก วิธี WLS ไม่เหมาะสม เพราะมีปัญหาเกี่ยวกับค่าถ่วงน้ำหนักมีค่าน้อยมากหรือเท่ากับศูนย์ ซึ่งปัญหานี้จะไม่มีผลกระทบต่อการใช้วิธี ML ถ้าค่าถ่วงน้ำหนักเท่ากับศูนย์ แล้ว วิธี ML อาจแทนค่าคงที่ที่มีค่าน้อยมาก เพื่อให้สามารถคำนวณค่าประมาณจากวิธี ML พร้อมกับใช้วิธีวนซ้ำปรับค่าถ่วงน้ำหนักไปต่างๆ ส่วนวิธี WLS โดยตรงที่ไม่มีการวนซ้ำ เมื่อแทนค่าถ่วงน้ำหนักที่เท่ากับศูนย์ด้วย เช่น ค่า 0.5 หรือค่าน้อยกว่านี้อีกมากๆ อาจทำให้เทอมความแปรปรวนมีค่าสูงหรือต่ำกว่าปกติ จนกระทั่งมีผลกระทบอย่างมาก (strong influence) ต่อการวิเคราะห์ถ่วงน้ำหนัก ทำให้อาจลดความเชื่อถือทั้งผลลัพธ์และข้อมูล กรณีเช่นนี้จึงควรใช้วิธี ML แทนวิธี WLS นอกจากนี้โปรแกรมสำเร็จรูปทางสถิติต่างๆ เช่น SPSS, GLIM, Minitab, R ซึ่งมีวิธี ML อยู่จึงมีประโยชน์ต่อการใช้ในหลายสถานการณ์

(iii) วิธีนิวตัน-รัฟสัน (Newton-Raphson method)

วิธีนิวตัน-รัฟสัน เป็นวิธีการแก้สมการไม่เชิงเส้น (nonlinear equations) แบบมีการวนซ้ำ และ สามารถนำมาใช้ร่วมกับการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด ทำให้เป็นวิธีการประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุดแบบวนซ้ำวิธีหนึ่ง สมการที่ใช้ในกรณีนี้ คือ สมการภาวะน่าจะเป็น (likelihood equation) หรือ ฟังก์ชันล็อกภาวะน่าจะเป็น (log-likelihood) มีค่าสูงสุด (maximized) สมการเหล่านี้อาจมีลักษณะเชิงเส้นหรือไม่เชิงเส้นในเทอมของพารามิเตอร์

การเริ่มต้น ต้องมีการเดาค่าประมาณ ที่ทำให้ ฟังก์ชันล็อกภาวะน่าจะเป็น ซึ่งแทนด้วย L มีค่าสูงสุด และนำไปใช้ในกระบวนการวนซ้ำครั้งที่ 1 โดยมีหลักเกณฑ์ดังนี้

ให้ $U = \left(\frac{\partial L}{\partial \beta_1}, \frac{\partial L}{\partial \beta_2}, \ldots \right)$

ให้ H แทน เมทริกซ์ของอนุพันธ์อันดับที่สองของ L และมีสมาชิกเป็น $h_{ij} = \frac{\partial^2 L}{\partial \beta_i \partial \beta_j}$

ให้ $L^{(S)}$ และ $H^{(S)}$ แทนเทอมของ L และ H ของการประมาณค่า β ในครั้งที่ s หรือ $\beta^{(s)}$ ตามลำดับ เมื่อ $s = 0,1,2,...$ และในกระบวนการวนซ้ำของขั้นตอนที่ s พิจารณาในเทอม $L(\beta)$ ใกล้เทอม $L(\beta^{(s)})$ ณ $\beta^{(s)}$ ในลักษณะของการกระจายแบบอนุกรมเทย์เลอร์ (Taylor series expansion) อันดับที่ 2 (2[nd] order) รอบๆค่า $\beta = \beta^{(s)}$ ซึ่งมีรูปแบบดังนี้

$$L(\beta) = L(\beta^{(s)} + (\beta - \beta^{(s)}))$$

$$\therefore L(\beta) = L(\beta^{(s)}) + U^{(s)}(\beta - \beta^{(s)}) + \frac{1}{2}(\beta - \beta^{(s)})' H^{(s)}(\beta - \beta^{(s)})$$

และแก้สมการปกติจากสมการ

$$\frac{\partial}{\partial \beta} L(\beta) = U^{(s)} + H^{(s)}(\beta - \beta^{(s)}) = 0$$

$$\hat{\beta} = \beta^{(s)} - \left(H^{(s)}\right)^{-1} U^{(s)} \qquad (4)$$

นำค่า $\beta^{(0)}$ ซึ่งเป็นค่าของการเดาค่าเริ่มต้น (an initial guess) ของ $\beta^{(s)}$ ในสมการ (4) ไปแทนและคำนวณค่าทางขวามือของ สมการ (4) ใหม่ เพื่อใช้สำหรับการวนซ้ำครั้งต่อไป นั่นคือ การวนซ้ำครั้งต่อไปจะมีลักษณะเป็น

$$\beta^{(s+1)} = \beta^{(s)} - \left(H^{(s)}\right)^{-1} U^{(s)}$$

$$\beta^{(s+1)} = \beta^{(s)} - \left(\frac{\partial^2 L}{\partial \beta_i \partial \beta_j}\right)^{-1}_{\beta=\beta^{(s)}} U^{(s)} \qquad (5)$$

โดยที่ $H = \dfrac{\partial^2 L}{\partial \beta_i \partial \beta_j}$ เป็นเมทริกซ์เอกฐาน (singular matrix) และ $s = 0,1,2,...$

กระบวนการวนซ้ำจะทำเช่นนี้ไปเรื่อยๆ จนได้ค่าประมาณที่ลู่เข้า จึงสรุปได้ว่าหลังจากการเดาค่าเริ่มต้น $\beta^{(0)}$ หรือแทนค่า $\beta^{(0)}$ ในสมการที่ (5) ณ $s = 0$ เพื่อหา $\beta^{(1)}$ และทำครั้งต่อไป ณ $s = 1,2,...$ จนกระทั่งได้ค่า $\beta^{(s)}$ ลู่เข้าหาค่า MLE ซึ่งทำให้ $\beta^{(s+1)}$ มีผลต่างระหว่างค่า $\beta^{(s+1)}$ และ $\beta^{(s)}$ น้อยมาก (บางวิธีใช้ผลต่างเท่ากับศูนย์) เช่นนี้กระบวนการวนซ้ำจึงเสร็จสิ้น และ ตัวประมาณ MLE ที่ได้คือ $\beta^{(s+1)}$

สำหรับรายละเอียดเพิ่มเติมเกี่ยวกับวิธี Newton-Raphson สามารถอ่านเพิ่มเติมได้ที่ Bock (1973) และ Haberman (1978)

(iv) วิธีฟิชเชอร์-สกอริง (Fisher-Scoring)

วิธีการวนซ้ำแบบฟิชเชอร์-สกอริง คล้ายคลึงกันกับวิธีการวนซ้ำแบบ Newton-Raphson โดยส่วนที่แตกต่างกันนั้น วิธีฟิชเชอร์-สกอริง ใช้ค่าคาดหวัง (expected value) ของ เมทริกซ์ H คือ $E(H)$ แทนเมทริกซ์ H ซึ่งใช้ในวิธีนิวตัน-รัฟสัน ดังนั้นสมาชิกของเมทริกซ์ H จากวิธีฟิชเชอร์-สกอริง คือ

$$h_{ij} = E\left(\frac{\partial^2 L}{\partial\beta_i \partial\beta_j}\right) = E\left(\frac{-\partial^2 l_i}{\partial\beta_i \partial\beta_j}\right)$$

และสมการวนซ้ำในครั้งที่ $s+1$ ของวิธีฟิชเชอร์-สกอริง คือ

$$\beta^{(s+1)} = \beta^{(s)} - \left(M^{(s)}\right)^{-1} U^{(s)}$$

โดยที่ $M = E(H)$ และ H แทน singular matrix ในสมการ (5)

2.5.3 **วิธี** Iteratively Reweighted Least Square (IRWLS)

จากที่กล่าวมาข้างต้นในหัวข้อ 2.5.2 การหาค่าประมาณพารามิเตอร์แบบวนซ้ำทั้งวิธีนิวตัน-รัฟสัน และ วิธีฟิชเชอร์-สกอริ่ง สามารถเริ่มจากวิธีภาวะน่าจะเป็นสูงสุดภายใต้ลักษณะการแจกแจงของตัวแปรจากตัวแบบที่สนใจ และในทำนองเดียวกัน เมื่อพิจารณาถึงตัวแบบเชิงเส้นที่วางนัยทั่วไป สำหรับค่าสังเกตที่เป็นอิสระต่อกันจำนวน n ตัวอย่างนั้น ฟังก์ชันล็อกภาวะน่าจะเป็นของตัวแปรสุ่ม Y_i คือ

$$\therefore\ \theta_i = \sum_j \beta_j X_{ij}$$

โดยที่ θ แทน natural parameter ที่ขึ้นกับพารามิเตอร์ของตัวแบบ เช่น β ส่วน $a(\cdot)$ แทน dispersion parameter และ

$$f(y_i;\theta_i,\phi) = \exp\left[\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right]$$

ดังนั้น $l_i = \log\left\{\exp\left[\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right]\right\}$

$$= \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)$$

เนื่องจากตัวแบบเชิงเส้นที่วางนัยทั่วไปประกอบด้วยส่วนประกอบทั้งหมด 3 ส่วน ได้แก่ ส่วนประกอบเชิงสุ่ม คือ $f(y_i;\theta_i,\phi)$ ส่วนประกอบแบบมีระบบ คือ $\eta_i = X\beta$ และส่วนประกอบฟังก์ชันเชื่อมโยง (link function) คือ $\eta_i = g(\mu_i)$ เช่น $\eta = \mu$ หรือ $E(Y)$ ในกรณีของการแจกแจงแบบปกติ จะได้ว่า $g(\mu_i) = X\beta = \sum_j \beta_j X_{ij} = \eta_i$ โดยที่ฟังก์ชัน g ซึ่งทำให้ $g(\mu_i) = \theta_i$ เรียกว่า canonical link function และ g เป็น monotone differentiable function

$$\therefore\ \theta_i = \sum_j \beta_j X_{ij}$$

นั่นคือ พารามิเตอร์ θ_i ในส่วนประกอบเชิงสุ่มข้างต้น ขึ้นอยู่กับ พารามิเตอร์ β

ดังนั้นการหาตัวประมาณภาวะน่าจะเป็นสูงสุด จากฟังก์ชันล็อกภาวะน่าจะเป็น หรือ L ข้างต้นสามารถใช้วิธีอนุพันธ์แบบกฎลูกโซ่ (chain rule) คือ

$$\frac{\partial l_i}{\partial \beta_j} = \frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j}$$

โดยที่ $\dfrac{\partial l_i}{\partial \theta_i} = \dfrac{(y_i - b'(\theta_i))}{a(\phi)}$

แต่จาก Cox และ Hinkley (1974) พบว่า

(i) $E\left[\dfrac{\partial l}{\partial \theta}\right] = 0$ และ $-E\dfrac{\partial^2 l}{\partial \theta^2} = -E\left[\dfrac{\partial l}{\partial \theta}\right]^2$

ดังนั้น $b'(\theta_i) = E(Y_i) = \mu_i$

$$\frac{\partial l_i}{\partial \theta_i} = \frac{y_i - \mu_i}{a(\phi)} \qquad (7)$$

$$\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i)$$

และ $\dfrac{\partial \mu_i}{\partial \theta^2} = -\dfrac{b''(\theta_i)}{a(\phi)}$

(ii) $\dfrac{b''(\theta_i)}{a(\phi)} = E\left(\dfrac{(Y-\mu)}{a(\phi)}\right)^2 = \dfrac{Var(Y_i)}{\left(a(\phi)\right)^2}$

ดังนั้น $b''(\theta_i) = \dfrac{Var(Y_i)}{a(\phi)}$

$$\therefore \; \frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{b''(\theta_i)} = \frac{a(\phi)}{Var(Y_i)} \;\therefore\; \frac{\partial \eta_i}{\partial \beta_j} = X_{ij} \tag{8}$$

เนื่องจาก $\eta_i = \sum_j \beta_j X_{ij}$

$$\therefore \; \frac{\partial \eta_i}{\partial \beta_j} = X_{ij} \tag{9}$$

นำค่าในสมการที่ (7) ถึง (9) แทนในสมการ (6) จะได้ว่า

$$\frac{\partial \, l_i}{\partial \beta_j} = \frac{(y_i - \mu_i)}{a(\phi)} \cdot \frac{a(\phi)}{Var(Y_i)} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot X_{ij}$$

$$\therefore \; \frac{\partial \, l_i}{\partial \beta_j} = \frac{(y_i - \mu_i) X_{ij}}{Var(Y_i)} \cdot \frac{\partial \mu_i}{\partial \eta_i}$$

ดังนั้นสมการภาวะน่าจะเป็นหรือสมการปกติ จะอยู่ในรูปของ

$$\frac{\partial \, l_i}{\partial \beta_j} = U_j = \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \cdot \sum_i^n \frac{(y_i - \mu_i) X_{ij}}{Var(Y_i)} = 0 \tag{10}$$

อย่างไรก็ตาม สมการที่ (10) เป็นฟังก์ชันที่ไม่ได้อยู่ในรูปแบบเชิงเส้นของ β ดังนั้นสมการที่ (10) จำเป็นต้องใช้การวนซ้ำของวิธีนิวตัน-รัฟสัน หรือ วิธีพิชเชอร์-สกอริ่ง มาช่วยโดยอัตราลู่เข้าขึ้นอยู่กับเมทริกซ์ข้อสนเทศ (information matrix) หรือ $E\left[-\dfrac{\partial^2 L(\beta)}{\partial \beta_i \, \partial \beta_j} \right]$

$$\therefore \; E\left[\frac{\partial^2 \, l_i}{\partial \beta_a \, \partial \beta_b} \right] = -E\left[\left(\frac{\partial \, l_i}{\partial \beta_a} \right) \left(\frac{\partial \, l_i}{\partial \beta_b} \right) \right]$$

$$= -E\left[\frac{(y_i - \mu_i) X_{ia}}{Var(Y_i)} \cdot \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \cdot \frac{(y_i - \mu_i) X_{ib}}{Var(Y_i)} \cdot \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right]$$

$$= -\frac{X_{ia} X_{ib}}{Var(Y_i)} \cdot \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

$$\therefore\ E\left[\frac{\partial^2 L}{\partial\beta_a\ \partial\beta_b}\right] = -\sum_i^n \frac{X_{ia}X_{ib}}{\mathrm{Var}(Y_i)}\cdot\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2$$

ถ้าเมทริกซ์ข้อสนเทศ (information matrix) มีสมาชิกเป็น $E\left[\frac{\partial^2 L(\beta)}{\partial\beta_a\ \partial\beta_b}\right]$ และเขียนให้อยู่ในรูปเมทริกซ์ คือ $M = X'WX$ โดยที่ W แทนเมทริกซ์ ซึ่งมีสมาชิกในแนวทแยงหลัก เป็น

$$W_i = \frac{\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2}{\mathrm{Var}(Y_i)}$$

ในการประมาณค่าของ β ด้วยวิธี Newton-Raphson ทำได้โดยใช้การวนซ้ำครั้งที่ s ใดๆ คือ

$$\beta^{(s+1)} = \beta^{(s)} - \left(H^{(s)}\right)^{-1} U^{(s)}$$

โดยที่ H แทนเมทริกซ์ ซึ่งมีสมาชิกเป็น $\frac{\partial^2 L(\beta)}{\partial\beta_a\ \partial\beta_b}$

และ U แทน เวกเตอร์ซึ่งมีสมาชิกเป็น $\frac{\partial L(\beta)}{\partial\beta_j}$

ทั้ง $H^{(s)}$ และ $U^{(s)}$ คือ ค่าของ H และ U ในครั้งที่ s ณ การคำนวณค่าประมาณของ $\beta = \beta^{(s)}$

ดังนั้น $\beta^{(s+1)} = \beta^{(s)} - \left(\frac{\partial^2 L}{\partial\beta_a\partial\beta_b}\right)^{-1}_{\beta=\beta^{(s)}} U^{(s)}$

อนึ่งสำหรับเมทริกซ์ H นั้นอาจเขียนให้อยู่ในรูปเมทริกซ์ M โดยที่ $M = E(H)$ นั่นคือใช้สูตรสำหรับ Fisher Scoring และจะได้ว่า

$$\therefore\ X'W^{(s)}X\beta^{(s+1)} = X'W^{(s)}X\beta^{(s)} + U^{(s)}\ \beta^{(s+1)} = \beta^{(s)} - \left(M^{(s)}\right)^{-1} U^{(s)}$$

หรือ $$M^{(s)}\beta^{(s+1)} = \left(M^{(s)}\right)\beta^{(s)} + U^{(s)} \qquad (11)$$

โดยที่ $M = X'WX$ และมีสมาชิกเป็น $-E\left(\frac{\partial^2 L(\beta)}{\partial\beta_a\ \partial\beta_b}\right)$

$$\therefore\ X'W^{(s)}X\beta^{(s+1)} = X'W^{(s)}X\beta^{(s)} + U^{(s)}$$

จะเห็นว่าสมการที่ (11) เป็นวิธีประมาณภาวะน่าจะเป็นสูงสุดที่มีการวนซ้ำแบบ Fisher Scoring ซึ่งมีการถ่วงน้ำหนักเมทริกซ์ $M = X'WX$ หรือเปรียบเทียบการประมาณแบบ IWLS (Iterative Weighted Least Square) ซึ่งจะเห็นได้ชัดเจนจากด้านขวามือของสมการที่ (11) ซึ่งอาจเขียนอยู่ในรูป

$$\sum_j \left[\sum_i \frac{X_{ia}X_{ij}}{\mathrm{Var}(Y_i)} \cdot \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \beta_j^{(s)} \right] + \sum_i \frac{\left(y_i - \mu_i^{(s)} \right) X_{ia}}{\mathrm{Var}(Y_i)} \cdot \left(\frac{\partial \mu_i}{\partial \eta_i} \right)$$

โดยที่ μ_i และ $\frac{\partial \mu_i}{\partial \eta_i}$ คำนวณ ณ $\beta^{(s)}$

$\therefore\ M^{(s)}\beta^{(s)} + U^{(s)} = X'W^{(s)}\psi^{(s)}$

โดยที่ $W^{(s)}$ คือ W ที่มีสมาชิกเป็น $\frac{\left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2}{\mathrm{Var}(Y_i)}$ ณ $\beta^{(s)}$ และ $\psi^{(s)}$ คือ ψ ที่มีสมาชิกเป็น

$$\psi_i^{(s)} = \sum_j X_{ij}\beta_j^{(s)} + (y_i - \mu_i^{(s)}) \left(\frac{\partial \eta_i^{(s)}}{\partial \mu_i^{(s)}} \right)$$

$$= \eta_i^{(s)} + (y_i - \mu_i^{(s)}) \left(\frac{\partial \eta_i^{(s)}}{\partial \mu_i^{(s)}} \right) \qquad (12)$$

ดังนั้นวิธี Fisher Scoring ในสมการที่ (11) เขียนใหม่ได้เป็น

$$M^{(s)}\beta^{(s+1)} = X'W^{(s)}\psi^{(s)}$$

$$(X'W^{(s)}X)\beta^{(s+1)} = X'W^{(s)}\psi^{(s)} \qquad (13)$$

ซึ่งมีรูปแบบของสมการปกติคล้ายวิธีกำลังสองน้อยที่สุดสำหรับตัวแบบเชิงเส้นที่มีตัวแปรตามเป็น $\psi^{(s)}$ ผลลัพธ์ที่ได้ คือ

$$\beta^{(s+1)} = (X'W^{(s)}X)^{-1}X'W^{(s)}\psi^{(s)}$$

ข้อสังเกตที่สำคัญคือ เวกเตอร์ $\psi^{(s)}$ ในสมการที่ (12) กับสมการที่ (13) มีรูปแบบเชิงเส้นของ link function ณ μ ที่คำนวณภายใต้ตัวแปรตาม Y ณ y นั่นคือ

$$g(y) = g(\mu_i) + (y_i - \mu_i)g'(\mu) = \eta + (y_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i} \right) = \psi_i$$

จะเห็นได้ชัดเจนว่ากระบวนการวนซ้ำข้างต้น ขึ้นอยู่กับค่าตัวแปรตาม ψ ซึ่งมีสมาชิกที่ i ประมาณด้วย $\psi^{(s)}$ ของการวนซ้ำครั้งที่ s การวนซ้ำรอบที่ s เป็นการถดถอยของ $\psi^{(s)}$ บน X พร้อมกับถ่วงน้ำหนักด้วย $W^{(s)}$ ซึ่งทำให้ได้ผลลัพธ์ตัวใหม่ คือ ตัวประมาณของ $\beta^{(s+1)}$ โดยตัวประมาณค่านี้จะให้ผลลัพธ์ใหม่ $\eta^{(s+1)}=X\beta^{(s+1)}$ ตลอดจนตัวแปรใหม่คือ $\psi^{(s+1)}$ ซึ่งใช้สำหรับรอบถัดไป

ดังนั้นตัวประมาณภาวะน่าจะเป็นสูงสุด คือ ลิมิตของ $\beta^{(s)}$ ในขณะที่ $s \to \infty$ กล่าวอีกนัยหนึ่ง คือ ตัวประมาณภาวะน่าจะเป็นของตัวแบบเชิงเส้นที่วางนัยทั่วไป เป็นผลของการวนซ้ำที่ใช้การถ่วงน้ำหนักของวิธีกำลังสองน้อยที่สุด โดยเมทริกซ์ของการถ่วงน้ำหนักเปลี่ยนค่าไปทุกๆครั้งของการวนซ้ำ green (1984) ซึ่งต่อมาเรียกในชื่อว่า “IRLS Process” หรือ “IRWLS Process” (Iteratively Reweighted Least Square process) กระบวนการวนซ้ำทั้งหมดจะเสร็จสิ้น ต่อเมื่อค่าประมาณ β ที่มาจากการวนซ้ำสองครั้งติดต่อกัน ให้ค่าผลต่างที่เล็กพอเพียงหรือเป็นศูนย์ ซึ่งสามารถสรุปออกเป็นขั้นตอนได้ดังนี้

ขั้นตอนที่ 1 ประมาณค่า $\underset{\sim}{\mu}^{(s)}$ โดยที่

$\mu^{(s)} = g^{-1}(X\beta^{(s)})$ เมื่อ g เป็น link function ที่ทำให้ $g(b'(\theta)) = \theta$

ขั้นตอนที่ 2 คำนวณค่า working responses (adjusted dependent variate) $\psi^{(s)}$

$$\psi^{(s)} = \eta^{(s)} + \frac{\partial\eta}{\partial\mu}(Y - \mu^{(s)})$$

ขั้นตอนที่ 3 คำนวณค่า working weight โดยที่

$$W^{(s)} = \left(\frac{\partial\mu}{\partial\eta}\right)^2 \cdot \frac{1}{Var(Y_i)}$$

ขั้นตอนที่ 4 จาก iteration ของ fisher scoring โดยที่

$$\underset{\sim}{\beta}^{(s+1)} = (X'W^{(s)}X)^{-1}X'W^{(s)}\psi^{(s)}$$

ขั้นตอนที่ 5 ทำขั้นตอนที่ 1 ถึงขั้นตอนที่ 4 ไปเรื่อยๆ จนกระทั่งค่าสัมประสิทธิ์การถดถอย β ลู่เข้า (ค่าสัมบูรณ์ของผลต่างค่าสัมประสิทธิ์การถดถอยทุกค่าน้อยกว่า 0.001)

รายละเอียดต่างๆของการประมาณค่าพารามิเตอร์ของตัวแบบเชิงเส้นที่วางนัยทั่วไป หรือ IRWLS process อาจศึกษาเพิ่มเติมได้หลายแห่ง เช่น Nelder และ Wedderburn (1972), McCullagh และ Nelder (1989), Agresti (1990), Dobson (1990) และ Fahrmeir และ Tutz (1994)

2.6 **ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง** (Average Mean Squared Error)

ในงานวิจัยชิ้นนี้ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Squared Error: MSE) เป็นตัวประมาณทางสถิติที่ใช้วัดค่าความแตกต่างระหว่างค่าจริงกับค่าประมาณโดยเฉลี่ย

<u>นิยาม</u> ถ้า $\hat{Y}$ เป็นค่าประมาณของตัวแปรตาม และ Y' เป็นค่าจริง (True value) แล้ว ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง จะเท่ากับ $MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i')^2$

วิธีการประมาณค่าที่ดี จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อย

ดังนั้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง เป็นการประเมินคุณภาพของตัวประมาณในแง่ของความไม่เอนเอียง (unbiasedness) แสดงว่า ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง เท่ากับศูนย์ จะหมายความว่า ตัวประมาณของตัวแปรตาม ($\hat{Y}$) สามารถพยากรณ์ค่าจริงของตัวแปรตาม (Y) ได้ถูกต้องสมบูรณ์ (perfect accuracy) ซึ่งมีโอกาสเกิดขึ้นได้น้อยมาก

ในการวิจัยครั้งนี้ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองทั้งหมด 5,000 ค่าในแต่ละวิธีตามจำนวนรอบของการจำลองซ้ำ ดังนั้นจึงต้องนำค่าทั้งหมดมารวมกันแล้วนำมาเฉลี่ยโดยการหารด้วย 5,000 ซึ่งจะได้ “ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง” ที่ใช้ในการเปรียบเทียบตัวแบบ

ดังนั้นจะได้ว่า

$$AMSE_* = \frac{1}{5,000}\left[\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i')^2\right] \quad ;* \in \{OLS, Box\text{-}Cox, IRWLS\}$$

และ $i = 1, 2, ..., n$

โดยที่ $\hat{Y}_i$ คือ ค่าประมาณของตัวแปรตาม

Y_i' คือ ค่าจริงของค่าตัวแปรตาม

n คือ ขนาดตัวอย่าง

ถ้าค่าของ $\hat{Y}_i$ มีค่าเท่ากับ $\bar{Y}$ ทุกๆค่า $i = 1,2,...,n$ จะหมายความว่า สมการถดถอยพยากรณ์ค่าของ Y ได้เท่ากับ $\bar{Y}$ ไม่ว่าค่าของตัวแปรอิสระจะมีค่าเท่าใด แสดงว่า สมการถดถอยไม่มีประโยชน์ในการพยากรณ์ เนื่องจากทำให้ผลรวมกำลังสองของการถดถอย $\left(\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2\right)$ มีค่าเท่ากับศูนย์ ตัวแปรอิสระไม่สามารถพยากรณ์ตัวแปรตามได้ ซึ่งจะทำให้ค่า F มีค่าเท่ากับศูนย์ ทำให้สรุปได้ว่า ไม่มีหลักฐานเพียงพอที่จะปฏิเสธสมมติฐานหลัก

2.7 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการแก้ไขปัญหาค่าความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ ซึ่งเกี่ยวข้องกับเรื่องที่ผู้วิจัยสนใจศึกษา มีดังต่อไปนี้

2.7.1 งานวิจัยที่เป็นภาษาอังกฤษ

อาเมมิยา (Amemiya, 1973) พิจารณาสมการถดถอยเมื่อความแปรปรวนของตัวแปรตามเป็นสัดส่วนกับกำลังสองของค่าคาดหวังของตัวแปรตามนั้น แล้วเปรียบเทียบประสิทธิภาพของตัวประมาณที่ได้จาก วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักเทียบกับวิธีการแบบภาวะน่าจะเป็นสูงสุด โดยที่ตัวแปรตามมีการแจกแจงแบบต่างๆดังนี้ คือ การแจกแจงแบบปกติ การแจกแจงแบบล็อคนอร์มอล (Lognormal Distribution) และการแจกแจงแบบแกมมา (Gamma Distribution) พบว่าทั้ง 2 วิธี มีประสิทธิภาพเท่าเทียมกัน เมื่อตัวแปรตามมีการแจกแจงแบบแกมมา และ วิธีการแบบภาวะน่าจะเป็นสูงสุดมีประสิทธิภาพมากกว่า เมื่อตัวแปรตามมีการแจกแจงแบบปกติและการแจกแจงแบบล็อคนอร์มอล

คาร์โรล และ รัพเพอร์ท (Carroll และ Rupport, 1982) พิจารณาสมการถดถอยอย่างง่ายในกรณีที่ความแปรปรวนของความคลาดเคลื่อนไม่คงที่ คือ

$$Y_i = \Gamma_i + \sigma_i\varepsilon_i \text{ ; } \Gamma_i = X_i\beta_i \text{ , } i = 1,2,...,N$$

โดยที่ ε_i มีค่าเฉลี่ยเท่ากับศูนย์ และ มีการแจกแจงแบบสมมาตรที่เป็นอิสระกัน แต่ไม่ทราบการแจกแจง σ_i เป็นส่วนเบี่ยงเบนมาตรฐานขนาดคงที่ใดๆ ณ ตำแหน่ง i และมีรูปแบบการแจกแจงเป็น $\sigma_i = \sigma(1+|\Gamma_i|)^{\lambda}$ พบว่า เมื่อ λ มีค่าเป็น 0.5 และ 1.0 ตัวประมาณโดยวิธีถ่วงน้ำหนักที่ใช้การ

ประมาณน้ำหนักโดยวิธีที่นำเสนอ ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Squared Error: MSE) ต่ำกว่า วิธีกำลังสองน้อยที่สุด วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก และ วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักเมื่อทราบน้ำหนักที่เหมาะสม

โคเฮน ดาร์ลาล และ ตูกี (1983) ได้เสนอวิธีการหาตัวประมาณค่าสัมประสิทธิ์การถดถอยอย่างง่ายในกรณีที่ความแปรปรวนของความคลาดเคลื่อนไม่คงที่ คือ วิธีนิว พบว่า วิธีนิวมีกระบวนการตรวจสอบค่าสังเกตที่มีค่าสูงๆที่เกิดขึ้นจากความแปรปรวนไม่คงที่ได้ดี แต่วิธีนี้มีประสิทธิภาพต่ำ เมื่อความแปรปรวนของความคลาดเคลื่อนผันแปรเพียงเล็กน้อย

บรอซ และ โมเสส (Bloch และ Moses, 1988) พิจารณาสมการถดถอยอย่างง่าย ในกรณีที่ความแปรปรวนของความคลาดเคลื่อนไม่คงที่ คือ

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

โดยที่ ε_i เป็นความคลาดเคลื่อนที่เป็นอิสระกัน และ $E(\varepsilon_i) = 0$; $\mathrm{Var}(\varepsilon_i) = \sigma_i^2$

พิจารณาตัวประมาณค่าพารามิเตอร์ 2 วิธี คือ วิธีกำลังสองน้อยที่สุด

$$\bar{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})Y_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \text{ โดยที่ } \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \text{ และ } E(\bar{\beta}) = \beta$$

กับ วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักที่ให้น้ำหนักไม่เท่ากัน คือ

$$\beta = \frac{\sum_{i=1}^{n} \dfrac{1}{\sigma_i^2 (X_i - \bar{X}_w) Y_i}}{\sum_{i=1}^{n} \dfrac{1}{\sigma_i^2 (X_i - \bar{X}_w)^2}} \text{ โดยที่ } \bar{X}_w = \frac{\sum_{i=1}^{n} \dfrac{1}{\sigma_i^2} X_i}{\sum_{i=1}^{n} \dfrac{1}{\sigma_i^2}} \text{ , } \hat{\beta} \text{ คือ ตัวประมาณค่า } \beta$$

แล้วเปรียบเทียบประสิทธิภาพจากความแปรปรวนที่ได้จากตัวประมาณค่าความชันในสมการการถดถอยทั้ง 2 กรณี พบว่า ตัวประมาณค่าที่ให้น้ำหนักไม่เท่ากันมีประสิทธิภาพสูงกว่าตัวประมาณที่ให้น้ำหนักเท่ากัน โดยมีค่าประสิทธิภาพสัมพัทธ์เป็น $\frac{9}{8}$

โคเฮน ดาร์ลาล และ ตูกี (1993) ได้พัฒนาตัวประมาณค่าสัมประสิทธิ์การถดถอยอย่างง่ายขึ้นใหม่ เรียกว่า วิธีออร์โค โดยพัฒนาจาก วิธีนิว ซึ่งทั้งสามได้นำเสนอ ใน ปี ค.ศ.1983

ผสมผสานกับวิธีไบเวท ของ มอสเทเลอร์ และตูกี ซึ่งได้นำเสนอไว้ในปี ค.ศ.1977 ในการศึกษาครั้งนี้ผู้ศึกษาได้พิจารณาในกรณีที่ตัวอย่างมีขนาดใหญ่ (n = 100) ความแปรปรวนของค่าความคลาดเคลื่อนมีรูปแบบพาราโบลาหงาย รูปแบบพาราโบลาคว่ำ และ รูปแบบเส้นตรงที่ลดลง (\) ตามค่าของตัวแปรอิสระ (X_i) และพิจารณาการแจกแจง 2 แบบ คือ การแจกแจงแบบปกติ และ การแจกแจงแบบปกติที่ปลอมปน โดยเปรียบเทียบประสิทธิภาพกับตัวประมาณที่ได้จากวิธีกำลังสองน้อยที่สุด วิธีไบเวท วิธีนิว และวิธีออร์โต โดยใช้ตัวประมาณเอ็มเคลื่อนที่ (Moving M-estimate) เป็น 25% ของข้อมูลทั้งหมด ในการหาตัวประมาณค่าโดยวิธีนิว พบว่า วิธีออร์โตมีประสิทธิภาพสัมพัทธ์สูงเมื่อเทียบกับวิธีอื่นๆ

2.7.2 งานวิจัยที่เป็นภาษาไทย

ชูใจ คูหารัตนไชย (2531) ได้ศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณพารามิเตอร์ β_0 และ β_1 ของวิธีการ 5 วิธี ได้แก่ วิธีกำลังสองน้อยที่สุด วิธีกำลังน้อยที่สุดแบบทั่วไปในการแก้ปัญหาเฉพาะลักษณะของความแปรปรวนของค่าความคลาดเคลื่อนไม่คงที่ วิธีกำลังสองน้อยที่สุดแบบทั่วไปในการแก้ปัญหาเฉพาะลักษณะของค่าความคลาดเคลื่อนมีสหสัมพันธ์กัน วิธีกำลังสองน้อยที่สุดแบบทั่วไปในการแก้ปัญหาเฉพาะลักษณะของความแปรปรวนของค่าความคลาดเคลื่อนไม่คงที่และค่าความคลาดเคลื่อนมีสหสัมพันธ์กันโดยอาศัยการแปลงข้อมูล กำหนดให้ค่าความคลาดเคลื่อนมีสหสัมพันธ์กันและความแปรปรวนไม่คงที่ โดยความแปรปรวนของค่าความคลาดเคลื่อนผันแปรตามตัวแปรอิสระ ตัวแปรตามและแบบสุ่ม สำหรับรูปแบบของสหสัมพันธ์นั้นจะทำการศึกษา เมื่อค่าสหสัมพันธ์มีค่าเป็น 0.3, 0.5, 0.7 และ 0.9 ขนาดตัวอย่างที่ใช้ขนาด 15, 30, 45 และ 60 ณ ระดับนัยสำคัญ 0.05 และ 0.01 ใช้ในการทดสอบเกี่ยวกับค่าความคลาดเคลื่อนของข้อมูลมีสหสัมพันธ์และความแปรปรวนไม่คงที่ สำหรับการเปรียบเทียบประสิทธิภาพของตัวประมาณจะพิจารณาจากค่าความแปรปรวนของตัวประมาณ ผลจากการศึกษาการเปรียบเทียบประสิทธิภาพของตัวประมาณไม่สามารถสรุปได้ว่าวิธีการใดมีประสิทธิภาพดีที่สุด วิธีการหนึ่งๆจะให้ผลดีภายใต้เงื่อนไขหนึ่งๆเท่านั้น โดยผลการวิจัยสรุปได้ดังนี้

1. การเปรียบเทียบประสิทธิภาพของตัวประมาณ เมื่อความแปรปรวนของค่าความคลาดเคลื่อนผันแปรตามตัวแปรอิสระ พบว่า เมื่อค่าสหสัมพันธ์เท่ากับ 0.3 วิธีกำลังสองน้อยที่สุด

แบบทั่วไปในการแก้ปัญหาเฉพาะลักษณะของความแปรปรวนของค่าความคลาดเคลื่อนไม่คงที่และค่าความคลาดเคลื่อนมีสหสัมพันธ์กันจะมีประสิทธิภาพดี แต่เมื่อค่าสหสัมพันธ์เท่ากับ 0.5, 0.7 และ 0.9 วิธีกำลังสองน้อยที่สุดแบบทั่วไปในการแก้ปัญหาเฉพาะลักษณะของความแปรปรวนของค่าความคลาดเคลื่อนไม่คงที่ และค่าความคลาดเคลื่อนมีสหสัมพันธ์กัน และ วิธีกำลังสองน้อยที่สุดแบบทั่วไปในการแก้ปัญหาเฉพาะลักษณะของความแปรปรวนของค่าความคลาดเคลื่อนไม่คงที่และค่าความคลาดเคลื่อนมีสหสัมพันธ์กันโดยอาศัยการแปลงข้อมูล จะมีประสิทธิภาพดี

2. การเปรียบเทียบประสิทธิภาพของตัวประมาณ เมื่อความแปรปรวนของค่าความคลาดเคลื่อนผันแปรตามตัวแปรตามพบว่า วิธีกำลังสองน้อยที่สุดแบบทั่วไปในการแก้ปัญหาเฉพาะลักษณะของค่าความคลาดเคลื่อนมีสหสัมพันธ์กัน จะมีประสิทธิภาพดี สำหรับทุกค่าสหสัมพันธ์

3. การเปรียบเทียบประสิทธิภาพของตัวประมาณ เมื่อความแปรปรวนของค่าความคลาดเคลื่อนผันแปรแบบสุ่มพบว่า วิธีกำลังสองน้อยที่สุดแบบทั่วไปในการแก้ปัญหาเฉพาะลักษณะของความแปรปรวนของค่าความคลาดเคลื่อนไม่คงที่และค่าความคลาดเคลื่อนมีสหสัมพันธ์กัน และ วิธีกำลังสองน้อยที่สุดแบบทั่วไปในการแก้ปัญหาเฉพาะลักษณะของความแปรปรวนของค่าความคลาดเคลื่อนไม่คงที่และค่าความคลาดเคลื่อนมีสหสัมพันธ์กันโดยอาศัยการแปลงข้อมูล จะมีประสิทธิภาพดีสำหรับทุกค่าสหสัมพันธ์

ไพโรจน์ ขาวสิทธิวงษ์ (2539) ได้ศึกษาเปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบสัมประสิทธิ์การถดถอยของตัวแบบการถดถอยเชิงเส้นที่มีจำนวนซ้ำในแต่ละระดับของตัวแปรอิสระ เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่โดยตัวสถิติทดสอบ 3 วิธีได้แก่ วิธีถ่วงน้ำหนักกำลังสองน้อยที่สุด (t_w) วิธีปรับแก้แจคไนฟ์แบบตัดทีละค่าสังเกต (t_M) และวิธีแจคไนฟ์แบบตัดเป็นกลุ่ม (t_j) พบว่ากรณีความคลาดเคลื่อนภายในระดับของตัวแปรอิสระไม่มีความสัมพันธ์กัน ตัวสถิติทดสอบ t_w และ t_M ไม่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ ถ้าความแปรปรวนของความคลาดเคลื่อนระหว่างระดับตัวแปรอิสระมีความแตกต่างกันมากและจำนวนซ้ำในแต่ละระดับของตัวแปรอิสระน้อย กรณีความคลาดเคลื่อนภายในระดับของตัวแปรอิสระมีความสัมพันธ์กัน ตัวแปรสถิติทดสอบ t_w และ t_M สามารถ

ควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ และกรณีความคลาดเคลื่อนภายในระดับของตัวแปรอิสระมีความสัมพันธ์กันน้อย ตัวสถิติทดสอบ t_j สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ทั้งกรณีความคลาดเคลื่อนภายในระดับของตัวแปรอิสระไม่มีความสัมพันธ์กันและมีความสัมพันธ์กัน ความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ลดลงเมื่อความแตกต่างของจำนวนซ้ำระหว่างระดับของตัวแปรอิสระเพิ่มขึ้น และ พบว่าทั้งกรณีความคลาดเคลื่อนภายในระดับของตัวแปรอิสระทั้งไม่มีความสัมพันธ์กัน มีความสัมพันธ์กัน และ มีความสัมพันธ์กันน้อย ตัวสถิติทดสอบ t_w มีอำนาจการทดสอบสูงกว่าตัวสถิติทดสอบ t_M กรณีความคลาดเคลื่อนภายในระดับของตัวแปรอิสระไม่มีความสัมพันธ์กัน ตัวสถิติทดสอบ t_w และ t_M มีอำนาจการทดสอบสูงกว่าตัวสถิติทดสอบ t_j และถ้าจำนวนซ้ำในแต่ละระดับของตัวแปรอิสระมีมาก กรณีที่ความคลาดเคลื่อนภายในระดับของตัวแปรอิสระมีความสัมพันธ์กันน้อย ตัวสถิติทดสอบ t_j มีอำนาจการทดสอบสูงกว่าตัวสถิติทดสอบ t_w และ t_M ถ้าความแปรปรวนของความคลาดเคลื่อนระหว่างระดับของตัวแปรอิสระมีความแตกต่างกันมากและจำนวนซ้ำในแต่ละระดับของตัวแปรอิสระน้อย กรณีความคลาดเคลื่อนภายในระดับของตัวแปรอิสระมีความสัมพันธ์ปานกลางและมาก ตัวสถิติทดสอบ t_j มีอำนาจการทดสอบสูงสุด

วิธิดา สุนทรวิภาต (2547) ได้ทำการศึกษาเปรียบเทียบวิธีการแก้ปัญหาความแปรปรวนของค่าความคลาดเคลื่อนไม่คงที่ 4 วิธี ได้แก่ วิธีแปลงข้อมูลและวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก 3 วิธี คือ วิธีกำลังสองน้อยที่สุดถ่วงน้ำหนักเมื่อทราบค่าน้ำหนัก วิธีกำลังสองน้อยที่สุดถ่วงน้ำหนักเมื่อไม่ทราบค่าน้ำหนักโดยแบ่งการประมาณค่าน้ำหนักออกเป็น 3 กลุ่ม และวิธีกำลังสองน้อยที่สุดถ่วงน้ำหนักเมื่อไม่ทราบค่าน้ำหนักโดยออกเป็น 5 กลุ่มภายใต้สถานการณ์ต่างกัน 48 สถานการณ์ ตามสาเหตุการเกิดปัญหาความแปรปรวนของค่าความคลาดเคลื่อนไม่คงที่ 4 สาเหตุ ได้แก่ กรณีที่ใช้ค่าเฉลี่ยของแต่ละกลุ่มเป็นค่าสังเกต กรณีการสุ่มตัวอย่างแบบชั้นภูมิ กรณีข้อมูลภาคตัดขวาง และ กรณีตัวแบบการถดถอยเกี่ยวกับพฤติกรรมของมนุษย์ โดยใช้เทคนิคมอนติคาร์โลทำการจำลองข้อมูลขนาดตัวอย่าง (n) ที่ใช้ศึกษาเท่ากับ 15, 30, 60 และ 90 เกณฑ์ที่ใช้เปรียบเทียบความสามารถในการแก้ปัญหาความแปรปรวนของค่าความคลาดเคลื่อนไม่คงที่ คือ ร้อยละของชุดข้อมูลที่ยอมรับสมมติฐานหลักว่าความแปรปรวนของค่าความคลาดเคลื่อนคงที่ ที่มี

ค่ามากที่สุด ผลการวิจัยสรุปได้ว่า วิธีกำลังสองน้อยที่สุดถ่วงน้ำหนักเมื่อไม่ทราบค่าน้ำหนักแบ่งการประมาณค่าน้ำหนักออกเป็น 3 กลุ่ม และ 5 กลุ่ม สามารถแก้ปัญหาความแปรปรวนของค่าความคลาดเคลื่อนไม่คงที่ได้มากที่สุดและใกล้เคียงกันทุกสาเหตุการเกิดปัญหาและขนาดตัวอย่าง วิธีกำลังสองน้อยที่สุดถ่วงน้ำหนักเมื่อทราบค่าน้ำหนักสามารถแก้ไขปัญหาได้ดีในกรณีข้อมูลภาคตัดขวางและกรณีตัวแบบการถดถอยเกี่ยวกับพฤติกรรมของมนุษย์เมื่อตัวอย่างมีขนาดปานกลางและขนาดใหญ่ สำหรับวิธีการแปลงข้อมูล สามารถแก้ปัญหาได้ดีเพียงเล็กน้อย เมื่อเทียบกับวิธีกำลังสองน้อยที่สุดถ่วงน้ำหนักเมื่อไม่ทราบค่าน้ำหนักแบ่งการประมาณค่าน้ำหนักออกเป็น 3 กลุ่มและ 5 กลุ่ม ผลการวิจัยยังพบว่า สาเหตุการเกิดปัญหามีผลต่อความสามารถในการแก้ปัญหา นั่นคือ กรณีข้อมูลภาคตัดขวาง สามารถแก้ปัญหาได้มากที่สุด รองลงมาคือ กรณีตัวแบบการถดถอยเกี่ยวกับพฤติกรรมของมนุษย์ กรณีที่ใช้ค่าเฉลี่ยของแต่ละกลุ่มเป็นค่าสังเกตและกรณีการสุ่มตัวอย่างแบบชั้นภูมิ ตามลำดับ

บทที่ 3

วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นการศึกษาและเปรียบเทียบวิธีการพยากรณ์ตัวแปรตาม ในตัวแบบการถดถอยเชิงเส้นพหุ เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ ที่สัมพันธ์กับตัวแปรอิสระ หรือสัมพันธ์กับค่าจริงของตัวแปรตาม ซึ่งวิธีการพยากรณ์ตัวแปรตามที่ใช้ในงานวิจัยครั้งนี้มีทั้งหมด 3 วิธีได้แก่ วิธีกำลังสองน้อยที่สุด (Ordinary Least Square: OLS), วิธีการแปลงของ Box และ Cox (Box-Cox Transformation) และวิธี Iteratively Reweighted Least Square (IRWLS) โดยการเปรียบเทียบแต่ละวิธีการพยากรณ์ตัวแปรตาม เพื่อจะทราบว่าวิธีใดเป็นวิธีที่ดีที่สุด โดยพิจารณาจาก ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Average Mean Squared Error: AMSE) ระหว่างค่าพยากรณ์กับค่าจริงของตัวแปรตาม และค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency: RE) และในการศึกษาครั้งนี้จะทำการจำลองข้อมูลให้มีสถานการณ์ที่แตกต่างกันด้วยเทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) โดยใช้โปรแกรม R เวอร์ชัน 2.15.3 ซึ่งในบทนี้ ผู้วิจัยจะกล่าวถึงแผนการจำลองข้อมูลและขั้นตอนในการวิจัยดังนี้

3.1 แผนการจำลองข้อมูล

ในงานวิจัยนี้ ผู้วิจัยจะทำการศึกษาสถานการณ์จำลองที่มีความแตกต่างกันทั้งหมด 216 สถานการณ์ เพื่อใช้เปรียบเทียบวิธีการประมาณค่าหรือการพยากรณ์ค่าตัวแปรตาม โดยมีรูปแบบการจำลอง มีเงื่อนไขดังต่อไปนี้

3.1.1 ตัวแปรอิสระที่ศึกษามีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวน ทั้งหมด 3 แบบดังต่อไปนี้

อัตราส่วนของ $\sigma_1^2 : \sigma_2^2$	σ_1^2	σ_2^2
1:2	300	600
1:1	450	450
2:1	600	300

3.1.2 ความคลาดเคลื่อนมีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวน ที่มีรูปแบบทั้งหมด 2 รูปแบบ

รูปแบบที่ 1 สัมพันธ์กับค่าของตัวแปรอิสระตัวที่หนึ่ง (X_1) เท่านั้น ซึ่งมีรูปแบบความสัมพันธ์ คือ

$$\mathrm{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2 \left|X_{i1}\right|^{\delta}$$

โดยที่ $\sigma^2 > 0$, δ เป็นค่าพารามิเตอร์

รูปแบบที่ 2 สัมพันธ์กับค่าจริงของตัวแปรตาม ซึ่งค่าจริงของตัวแปรตามเกิดจากความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระ นั่นคือ

$$Y_i' = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} \quad : \ i = 1, 2, ..., n$$

ซึ่งมีรูปแบบความสัมพันธ์คือ

$$\mathrm{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2 \left|\beta_0 + \sum_{j=1}^{2} \beta_j X_{ij}\right|^{\delta} = \sigma^2 \left|Y_i'\right|^{\delta}$$

โดยที่ $\sigma^2 > 0$, δ เป็นค่าพารามิเตอร์

ความคลาดเคลื่อนมีความแปรปรวนทั้ง 2 รูปแบบ จะมาจากการกำหนด $\underset{\sim}{\beta}$ เป็นค่าพารามิเตอร์ที่มีค่าคงที่ โดยที่ $\beta_0 = 500$, $\beta_1 = 1$ และ $\beta_2 = 1$

ซึ่งโดยทั่วไปแล้ว β_0, β_1 และ β_2 จะเป็นพารามิเตอร์ที่ไม่ทราบค่า และในงานวิจัยนี้ผู้วิจัยได้กำหนด $\underset{\sim}{\beta}$ เป็นค่าพารามิเตอร์ที่มีค่าคงที่ โดยที่ $\beta_0 = 500$, $\beta_1 = \beta_2 = 1$ ทั้งนี้เนื่องจากต้องการเปรียบเทียบแบบของตัวแปรอิสระที่มีรูปแบบความแปรปรวนแตกต่างกัน ซึ่งเมื่อ β_1 หรือ β_2 เปลี่ยนไปจะทำให้ค่าตัวแปรตามที่ได้จากแบบของตัวแปรอิสระแต่ละแบบมีความแปรปรวนแตกต่างกันด้วย ฉะนั้นเพื่อให้ควบคุมค่าความแปรปรวนของค่าจริงของตัวแปรตามมีค่าเท่ากัน จึงกำหนดให้ β_1 และ β_2 มีค่าเท่ากับ 1 และ กำหนดให้ตัวแปรอิสระแต่ละตัวไม่มีความสัมพันธ์กัน นั่นคือกำหนดให้ค่าสหสัมพันธ์มีค่าเท่ากับศูนย์

3.1.3 พารามิเตอร์ที่เกี่ยวข้องกับความแปรปรวนของความคลาดเคลื่อนที่ผู้วิจัยได้กำหนดและศึกษา จากในขั้นตอนที่ 3.1.2 ได้แก่

3.1.3.1 พารามิเตอร์ δ ที่เป็นค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความคลาดเคลื่อนต่อตัวแปรอิสระหรือค่าจริงของตัวแปรตามในรูปแบบเลขชี้กำลัง โดยศึกษาทั้งหมด 4 ค่า ได้แก่ 0, 0.1, 0.3 และ 0.5

3.1.3.2 พารามิเตอร์ σ^2 โดยในงานวิจัยชิ้นนี้ ผู้วิจัยศึกษาทั้งหมด 3 ค่าได้แก่ 450, 900 และ 1,800 โดยมาจากค่าความแปรปรวนของตัวแปรอิสระในแต่ละแบบ ซึ่งเมื่อรวมกันจะมีค่าเท่ากับ 900 ซึ่งจะได้อัตราส่วนระหว่าง $\sigma^2 : (\sigma_1^2 + \sigma_2^2)$ เท่ากับ 1:2, 1:1 และ 2:1 ตามลำดับ

3.1.4 ค่าสังเกตของตัวแปรตามและตัวแปรอิสระมีความสัมพันธ์กันภายใต้การถดถอยเชิงเส้นพหุ (Multiple Linear Regression) นั่นคือ

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i = Y_i' + \varepsilon_i \qquad : \ i = 1, 2, ..., n$$

โดยกำหนดให้ ตัวแปรอิสระแต่ละตัวไม่มีความสัมพันธ์กัน นั่นคือ ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระ เท่ากับศูนย์

3.1.5 ขนาดตัวอย่างที่ใช้ในการศึกษา มีทั้งหมด 3 ระดับ ได้แก่ 25, 50 และ 100

3.1.6 ทำการจำลองสถานการณ์แต่ละสถานการณ์ทั้งหมด 5,000 รอบ

3.2 วิธีการดำเนินงานวิจัย

ในงานวิจัย การเปรียบเทียบวิธีการประมาณสำหรับตัวแบบการถดถอยเชิงเส้นพหุที่ค่าความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ มีวิธีการดำเนินการวิจัยดังต่อไปนี้

3.2.1 สร้างข้อมูลตัวแปรอิสระ X_1 และ X_2 ที่ต่างมีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนของตัวแปรอิสระทั้งสองตัวเป็นอัตราส่วนกันระหว่างทั้งหมด 3 แบบ คือ 1:2, 1:1 และ 2:1 ซึ่งเป็นไปดังต่อไปนี้

อัตราส่วนของ $\sigma_1^2 : \sigma_2^2$	σ_1^2	σ_2^2
1:2	300	600
1:1	450	450
2:1	600	300

โดยมีขนาดตัวอย่าง เท่ากับ n และตัวแปรอิสระทุกตัวต่างไม่มีความสัมพันธ์กัน

3.2.2 สร้างข้อมูลค่าจริงของตัวแปรตาม โดยใช้ความสัมพันธ์เชิงเส้น คือ $\underset{\sim}{Y}' = X\underset{\sim}{\beta}$ โดยกำหนด $\underset{\sim}{\beta}$ เป็นค่าพารามิเตอร์ที่มีค่าคงที่ โดยที่ $\beta_0 = 500$, $\beta_1 = 1$ และ $\beta_2 = 1$

3.2.3 สร้างข้อมูลความคลาดเคลื่อนที่มีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนที่สัมพันธ์กับค่าของตัวแปรอิสระตัวที่หนึ่ง และค่าจริงของตัวแปรตาม ตามรูปแบบ ค่าพารามิเตอร์ที่กำหนด และมีขนาดตัวอย่างเท่ากับ n

3.2.4 สร้างข้อมูลค่าสังเกตของตัวแปรตาม โดยใช้ตัวแบบเชิงเส้น $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} = Y' + \underset{\sim}{\varepsilon}$

3.2.5 นำข้อมูลค่าสังเกตของตัวแปรตาม จากขั้นตอนที่ 4 และ ค่าของตัวแปรอิสระแต่ละตัวที่จำลองได้ จากขั้นตอนที่ 1 มาหาสัมประสิทธิ์การถดถอยเชิงเส้นพหุ ทั้งหมด 3 วิธี ได้แก่ วิธีกำลังสองน้อยที่สุด (Ordinary Least Square: OLS) วิธีการแปลงของ Box และ Cox (Box-Cox Transformation) และ วิธี Iteratively Reweighted Least Square (IRWLS) ซึ่งแต่ละวิธีมีรายละเอียดดังต่อไปนี้

1. วิธีกำลังสองน้อยที่สุด (Ordinary Least Square: OLS)

ตัวประมาณกำลังสองน้อยที่สุดที่ทำให้ผลบวกของค่าความคลาดเคลื่อนกำลังสอง (Sum Squared Error: SSE) น้อยที่สุด คือ $\hat{\underset{\sim}{\beta}} = (X'X)^{-1} X'\underset{\sim}{Y}$ โดยที่ $X'X$ เป็น เมทริกซ์ที่ไม่เอกฐาน *(*non-singular matrix)

2. วิธีการแปลงของ Box **และ** Cox (Box-Cox Transformation)

คำนวณหาค่า λ ที่เหมาะสม ซึ่งค่า λ ดังกล่าวจะอยู่ในช่วง -3 ถึง 3 และมาจากตัวแบบ

$$Y_i^* = \frac{Y_i^\lambda - 1}{\lambda} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} \text{ เมื่อ } \lambda \neq 0 \text{ ที่ } i = 1, 2, ..., n$$

$$Y_i^* = \log Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} \text{ เมื่อ } \lambda = 0 \text{ ที่ } i = 1, 2, ..., n$$

ที่ทำให้ค่าล็อกภาวะน่าจะเป็น log likelihood มากที่สุด (ในที่นี้กำหนดให้การแจกแจง $\varepsilon_i \sim N(0, Var(\varepsilon_i))$) ซึ่งคำนวณมาจาก

$$\log\varsigma = -\frac{n}{2}\log(2\pi) - n\log(\sqrt{\mathrm{Var}(\varepsilon_i)})$$
$$-\frac{1}{2\mathrm{Var}(\varepsilon_i)}\sum_{i=1}^{n}[Y_i^* - (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2})]^2 + (\lambda-1)\sum_{i=1}^{n}\log Y_i$$

ซึ่งค่า λ ที่ได้ในที่นี้คือ ตัวประมาณภาวะน่าจะเป็นสูงสุด (Maximum likelihood estimator) ของ λ นั่นเอง จากนั้นนำค่า λ ดังกล่าวมาเข้าวิเคราะห์ในตัวแบบ

$$Y'' = Y^{\lambda} \qquad \text{เมื่อ } \lambda \neq 0$$
$$Y'' = \log(Y) \qquad \text{เมื่อ } \lambda = 0$$

จากนั้นหาค่าของสัมประสิทธิ์การถดถอย โดยใช้วิธีกำลังสองน้อยที่สุด

3. **วิธี** Iteratively Reweighted Least Square (IRWLS)

เนื่องจากทั้งความคลาดเคลื่อนและตัวแปรอิสระทุกตัวต่างมีการแจกแจงปกติ และต่างมีค่าเฉลี่ยเท่ากับศูนย์ และเนื่องจากตัวแปรตามที่นำมาใช้วิเคราะห์เกิดจากความสัมพันธ์เชิงเส้น $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$ โดยที่ $i = 1, 2, ..., n$ ดังนั้นตัวแปรตามที่ได้จึงมีการแจกแจงปกติ ที่มีค่าเฉลี่ยเท่ากับ β_0 ซึ่งในงานวิจัยครั้งนี้กำหนดให้ $\beta_0 = 500$ ดังนั้นค่าเฉลี่ยของตัวแปรตามจึงได้เท่ากับ 500 และความแปรปรวนที่ตามรูปแบบและพารามิเตอร์ในแต่ละสถานการณ์ของความคลาดเคลื่อน

จากความสัมพันธ์เชิงเส้น

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \ ; \ i = 1, 2, ..., n$$

ยังแสดงให้เห็นว่า ผลกระทบของตัวแปรอิสระ (predictors) บนตัวแปรตาม ผ่านตัวพยากรณ์เชิงเส้น (linear predictor) โดยที่

$$\eta = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p = X'\underset{\sim}{\beta}$$

ดังนั้น ถ้ากำหนดให้ฟังก์ชัน g เป็นฟังก์ชันเชื่อมโยง (link function) ที่ทำให้ $\eta = g(\mu) = X'\underset{\sim}{\beta}$ จะได้ว่าฟังก์ชันเชื่อมโยงที่ใช้ในงานวิจัยครั้งนี้ คือ identity link นั่นคือ $g(\mu) = \mu$

ในแต่ละรอบ Iteration การคำนวณพารามิเตอร์รอบเริ่มต้นสุดในที่นี้จะใช้ $\hat{\eta}_0 = \hat{Y}$ หรือค่าประมาณตัวแปรตามที่ได้จากวิธี OLS และ $\hat{\mu}_0 = Y$ หรือค่าสังเกตของตัวแปรตาม ส่วนค่าน้ำหนักที่ให้เริ่มต้นจะเท่ากับ $w_i\sigma^{-2}$ โดยที่ w_i คือ ค่าน้ำหนักที่ทำให้ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ ซึ่งจะสิ้นสุดการทำงานใน Iteration เมื่อค่าความแตกต่างสัมประสิทธิ์การถดถอยระหว่างรอบก่อนหน้าและรอบล่าสุด ลู่เข้า (converge)

3.2.6 นำค่าสัมประสิทธิ์การถดถอยที่ได้ จากขั้นตอนที่ 5 มาสร้างสมการเชิงเส้นพหุ เพื่อหาค่าพยากรณ์ตัวแปรตาม ซึ่งมีรูปแบบดังนี้

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} \ ; \ i = 1, 2, ..., n$$

3.2.7 คำนวณค่า AMSE ของแต่ละวิธีการประมาณค่าตัวแปรตาม เพื่อเปรียบเทียบแต่ละวิธีการประมาณค่าตัวแปรตามทั้งหมด 3 วิธี ซึ่งคำนวณได้จากสูตร

$$AMSE_* = \frac{1}{5,000}\left[\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i')^2\right] \quad ; \ * \in \{OLS, Box\text{-}Cox, IRWLS\}$$

และ $i = 1, 2, ..., n$

โดยที่ $\hat{Y}_i$ คือ ค่าประมาณของตัวแปรตาม

Y_i' คือ ค่าจริงของตัวแปรตาม

n คือ ขนาดตัวอย่าง

ในงานวิจัยนี้ใช้ค่าจริงของตัวแปรตาม ในการหาค่า AMSE แทนการใช้ค่าสังเกตของตัวแปรตาม เนื่องจากต้องการเปรียบเทียบว่าวิธีการพยากรณ์ตัวแปรตามวิธีใดดีกว่า ซึ่งจะพิจารณาจากผลต่างระหว่างค่าพยากรณ์ กับค่าจริงของตัวแปรตามที่ไม่มีความคลาดเคลื่อนมารบกวน

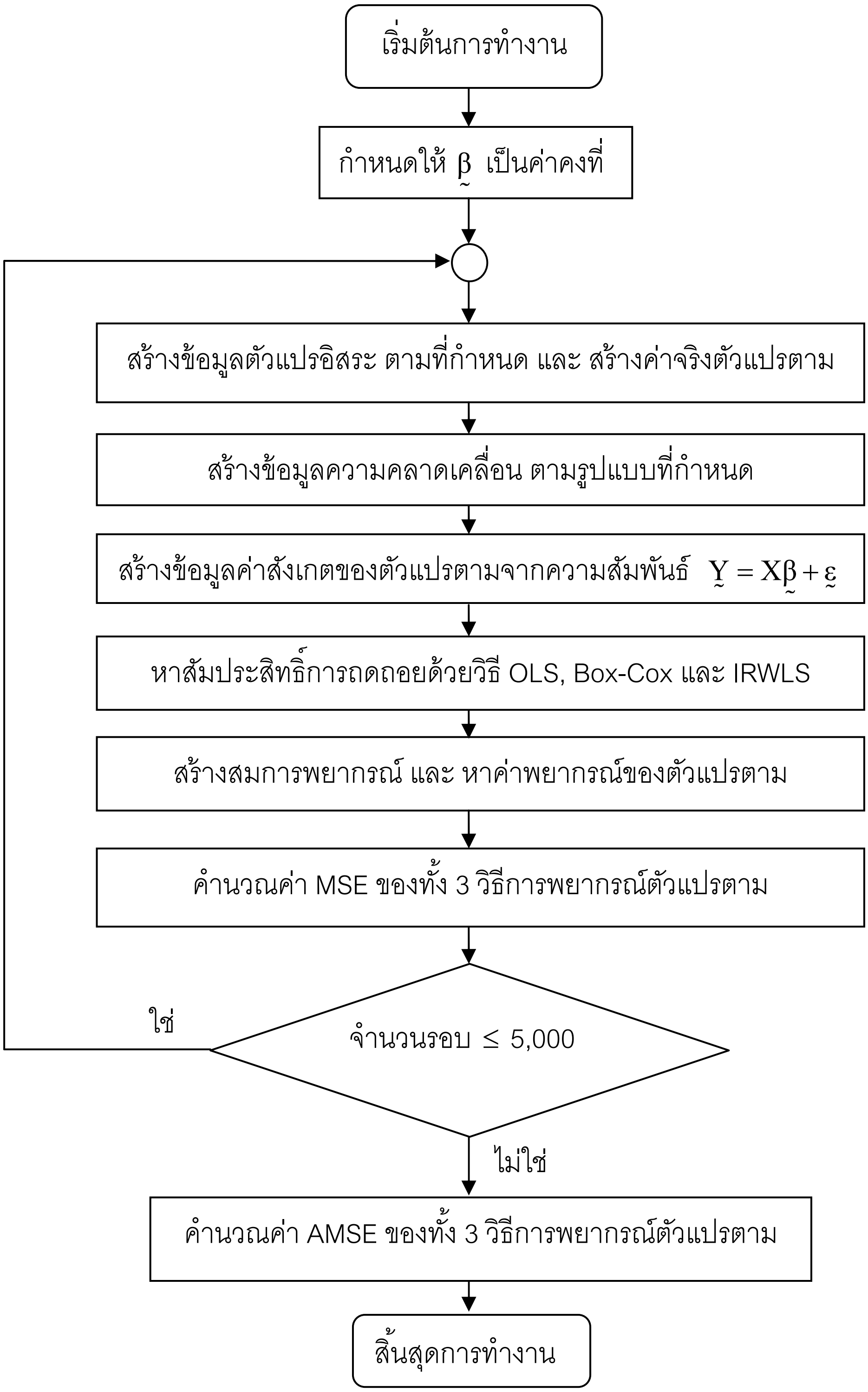
3.2.8 คำนวณค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency: RE) ซึ่งเป็นอัตราส่วนระหว่าง ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ของวิธี OLS กับวิธีการประมาณค่าตัวแปรตามแต่ละวิธี ซึ่งในการศึกษาครั้งนี้เลือกใช้วิธี OLS เป็นเกณฑ์ในการเปรียบเทียบ เนื่องจากวิธี OLS เป็นวิธีประมาณค่าตัวแปรตามที่นิยมใช้มากที่สุด

และเป็นวิธีพื้นฐานในการประมาณค่าตัวแปรตามในตัวแบบการถดถอยเชิงเส้น ซึ่งวิธีใดให้ค่า RE มากกว่า 1 แสดงว่ามีประสิทธิภาพมากกว่าวิธี OLS โดยคำนวณได้จากสูตร

$$\mathrm{RE}_{*} = \frac{\mathrm{AMSE}_{\mathrm{OLS}}}{\mathrm{AMSE}_{*}} \;;\; * \in \{\mathrm{OLS, Box\text{-}Cox, IRWLS}\}$$

โดยที่ AMSE_{*} แทน ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองระหว่างค่าจริงกับค่าพยากรณ์ของตัวแปรตามจากการทำซ้ำทั้งหมด 5,000 รอบ ของแต่ละวิธี ได้แก่ วิธี OLS, วิธีการแปลงของ Box และ Cox หรือวิธี IRWLS

$\mathrm{AMSE}_{\mathrm{OLS}}$ แทน ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองระหว่างค่าจริงกับค่าพยากรณ์ของตัวแปรตามจากการทำซ้ำทั้งหมด 5,000 รอบของวิธี OLS



ภาพที่ 3.1 แผนผังการเขียนโปรแกรม

บทที่ 4

ผลการวิจัย

ในงานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาประสิทธิภาพของการพยากรณ์ตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ในแต่ละสถานการณ์ ในตัวแบบการถดถอยเชิงเส้นพหุทั้งหมด 3 วิธี ได้แก่ วิธีกำลังสองน้อยที่สุด (Ordinary Least Square: OLS) วิธีการแปลงของ Box และ Cox (Box-Cox Transformation) และ วิธี Iteratively Reweighted Least Square (IRWLS) เกณฑ์การเปรียบเทียบที่ใช้พิจารณาว่าวิธีการประมาณหรือวิธีพยากรณ์ตัวแปรตามวิธีใดเป็นวิธีดีกว่า คือ ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Average Mean Square Error: AMSE) ระหว่างค่าพยากรณ์กับค่าจริงของตัวแปรตาม และค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency: RE)

ในบทนี้ การนำเสนอผลการวิจัย จะแสดงในรูปแบบของตารางและกราฟ โดยมีสัญลักษณ์ที่ใช้ซึ่งแทนความหมายต่างๆดังนี้

σ_1^2 แทน ระดับความแปรปรวนของตัวแปรอิสระตัวที่หนึ่ง (X_1)

σ_2^2 แทน ระดับความแปรปรวนของตัวแปรอิสระตัวที่สอง (X_2)

n แทน ขนาดตัวอย่าง

σ^2 แทน พารามิเตอร์ที่เกี่ยวข้องกับค่าความแปรปรวนของความคลาดเคลื่อนที่บ่งบอกถึงขนาดระดับความแปรปรวนของความคลาดเคลื่อน

δ (delta) แทน พารามิเตอร์ที่เกี่ยวข้องกับค่าความแปรปรวนของความคลาดเคลื่อนเป็นค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความคลาดเคลื่อนต่อตัวแปรอิสระหรือค่าจริงของตัวแปรตามในรูปแบบเลขชี้กำลัง

e_X_1 แทน กรณีที่ความคลาดเคลื่อนมีความแปรปรวนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง

e_Y' แทน กรณีที่ความคลาดเคลื่อนมีความแปรปรวนสัมพันธ์กับค่าจริงของตัวแปรตาม

OLS	แทน	วิธีการประมาณค่าตัวแปรตาม ด้วย วิธี OLS
Box-Cox	แทน	วิธีการประมาณค่าตัวแปรตาม ด้วย วิธีการแปลงของ Box และ Cox
IRWLS	แทน	วิธีการประมาณค่าตัวแปรตาม ด้วย วิธี IRWLS
AMSE	แทน	ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองระหว่างค่าพยากรณ์กับค่าจริงของตัวแปรตาม
RE	แทน	ค่าประสิทธิภาพสัมพัทธ์ ซึ่งเป็นอัตราส่วนระหว่างค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองระหว่างค่าจริงกับค่าพยากรณ์ของวิธีกำลังสองน้อยที่สุด (OLS) กับวิธีการประมาณค่าตัวแปรตามแต่ละวิธี

ในการแสดงผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตามทั้ง 3 วิธี จะแบ่งออกเป็น 2 ส่วน คือ ใช้ลักษณะรูปแบบความสัมพันธ์ของการที่ทำให้ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ เป็นเกณฑ์ในการแบ่ง และในแต่ละส่วนจะนำเสนอแบ่งออกเป็น 3 ส่วนย่อย กล่าวคือ ในแต่ละส่วนย่อย จะใช้เกณฑ์อัตราส่วนความแปรปรวนระหว่างตัวแปรอิสระทั้ง 2 ตัว และ ในส่วนสุดท้ายจะเป็นการเปรียบเทียบผลการวิจัยที่ได้จากส่วนที่ 1 และ ส่วนที่ 2 ซึ่งมีรายละเอียดการนำเสนอดังนี้

<u>ส่วนที่ 1</u> ผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1)

1.1 อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2 กล่าวคือ ค่า σ_1^2 เท่ากับ 300 และ ค่า σ_2^2 เท่ากับ 600

1.2 อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1 กล่าวคือ ค่า σ_1^2 เท่ากับ 450 และ ค่า σ_2^2 เท่ากับ 450

1.3 อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 2:1 กล่าวคือ ค่า σ_1^2 เท่ากับ 600 และ ค่า σ_2^2 เท่ากับ 300

<u>ส่วนที่ 2</u> ผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับค่าจริงของตัวแปรตาม (Y')

2.1 อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2 กล่าวคือ ค่า σ_1^2 เท่ากับ 300 และ ค่า σ_2^2 เท่ากับ 600

2.2 อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1 กล่าวคือ ค่า σ_1^2 เท่ากับ 450 และ ค่า σ_2^2 เท่ากับ 450

2.3 อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 2:1 กล่าวคือ ค่า σ_1^2 เท่ากับ 600 และ ค่า σ_2^2 เท่ากับ 300

ส่วนที่ 3 การเปรียบเทียบผลการวิจัยของวิธีการประมาณค่าหรือการพยากรณ์ค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่จากส่วนที่ 1 ที่สัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง และจากส่วนที่ 2 ที่สัมพันธ์กับค่าจริงของตัวแปรตาม

การนำเสนอผล จะประกอบด้วย ตารางแสดงผล รูปภาพ และการอธิบายผลที่ได้จากตารางและรูปภาพที่แสดง เมื่อจบการแสดงผลของแต่ละส่วนแล้วจะมีการอธิบายผลการวิจัยภาครวมของแต่ละส่วน และในส่วนของท้ายสุดในบทนี้หลังจากที่แสดงผลการวิจัยครบทุกส่วน จะมีการสรุปผลการวิจัยทั้งหมดอีกครั้ง

4.1 ผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1)

ผู้วิจัยทำการศึกษาในกรณีที่ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2, 1:1 และ 2:1 ขนาดตัวอย่าง เท่ากับ 25, 50 และ 100 เทอมพารามิเตอร์ที่เกี่ยวข้องกับความแปรปรวนของความคลาดเคลื่อน σ^2 เท่ากับ 450, 900 และ 1,800 และ เทอมพารามิเตอร์ δ เท่ากับ 0, 0.1, 0.3 และ 0.5 ซึ่งผลการวิจัยในส่วนนี้จะแสดงในตารางที่ 4.1.1 ถึง ตารางที่ 4.1.3

ตารางที่	ความสัมพันธ์ของความคลาดเคลื่อน	อัตราส่วนความแปรปรวนของตัวแปรอิสระระหว่าง X_1 และ X_2
4.1.1	ตัวแปรอิสระตัวที่หนึ่ง (X_1)	1:2
4.1.2	ตัวแปรอิสระตัวที่หนึ่ง (X_1)	1:1
4.1.3	ตัวแปรอิสระตัวที่หนึ่ง (X_1)	2:1

ตารางที่ 4.1.1 ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2

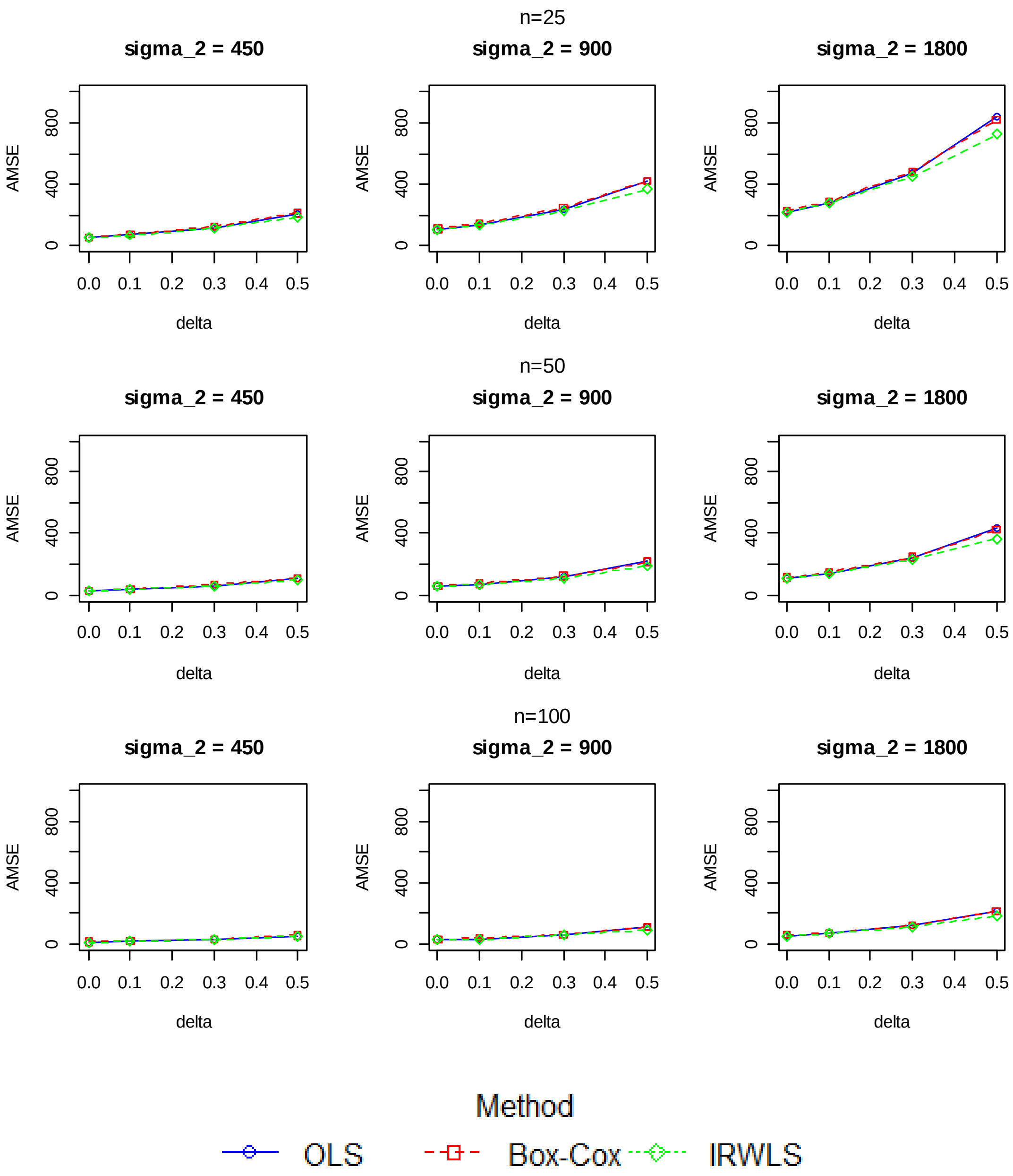
σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
300	600	25	450	0	AMSE	53.62	57.65	53.62
					RE	1.0000	0.9301	1.0000
				0.1	AMSE	69.37	73.80	69.15
					RE	1.0000	0.9400	1.0032
				0.3	AMSE	119.62	124.21	115.24
					RE	1.0000	0.9630	1.0380
				0.5	AMSE	211.98	213.20	188.77
					RE	1.0000	0.9943	1.1230
			900	0	AMSE	105.71	111.11	105.71
					RE	1.0000	0.9514	1.0000
				0.1	AMSE	139.52	144.95	138.64
					RE	1.0000	0.9625	1.0063
				0.3	AMSE	239.46	244.84	229.24
					RE	1.0000	0.9780	1.0446
				0.5	AMSE	423.60	420.12	371.13
					RE	1.0000	1.0083	1.1414
			1,800	0	AMSE	215.32	224.29	215.32
					RE	1.0000	0.9600	1.0000
				0.1	AMSE	280.71	290.15	278.91
					RE	1.0000	0.9675	1.0065
				0.3	AMSE	476.03	481.56	449.89
					RE	1.0000	0.9885	1.0581
				0.5	AMSE	843.06	823.60	730.75
					RE	1.0000	1.0236	1.1537

ตารางที่ 4.1.1 (ต่อ) ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2

σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
300	600	50	450	0	AMSE	**27.15**	29.61	**27.15**
					RE	**1.0000**	0.9169	**1.0000**
				0.1	AMSE	34.91	37.54	**34.75**
					RE	1.0000	0.9299	**1.0046**
				0.3	AMSE	60.88	63.70	**58.20**
					RE	1.0000	0.9557	**1.0460**
				0.5	AMSE	109.74	110.81	**97.12**
					RE	1.0000	0.9903	**1.1299**
			900	0	AMSE	**55.17**	58.54	**55.17**
					RE	**1.0000**	0.9424	**1.0000**
				0.1	AMSE	69.95	73.51	**69.57**
					RE	1.0000	0.9516	**1.0055**
				0.3	AMSE	119.68	123.63	**113.17**
					RE	1.0000	0.9680	**1.0575**
				0.5	AMSE	215.72	215.12	**185.23**
					RE	1.0000	1.0028	**1.1646**
			1,800	0	AMSE	**107.65**	112.80	**107.65**
					RE	**1.0000**	0.9543	**1.0000**
				0.1	AMSE	140.59	145.82	**139.35**
					RE	1.0000	0.9641	**1.0089**
				0.3	AMSE	242.12	246.01	**227.81**
					RE	1.0000	0.9842	**1.0628**
				0.5	AMSE	430.47	426.81	**366.70**
					RE	1.0000	1.0086	**1.1739**

ตารางที่ 4.1.1 (ต่อ) ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2

σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
300	600	100	450	0	AMSE	**13.53**	14.92	**13.53**
					RE	**1.0000**	0.9068	**1.0000**
				0.1	AMSE	17.43	18.86	**17.35**
					RE	1.0000	0.9242	**1.0046**
				0.3	AMSE	30.57	32.03	**29.13**
					RE	1.0000	0.9544	**1.0494**
				0.5	AMSE	54.92	55.91	**48.49**
					RE	1.0000	0.9823	**1.1326**
			900	0	AMSE	**27.04**	28.68	**27.04**
					RE	**1.0000**	0.9428	**1.0000**
				0.1	AMSE	34.60	36.59	**34.33**
					RE	1.0000	0.9456	**1.0079**
				0.3	AMSE	60.80	62.76	**57.37**
					RE	1.0000	0.9688	**1.0598**
				0.5	AMSE	107.57	108.48	**91.73**
					RE	1.0000	0.9916	**1.1727**
			1,800	0	AMSE	**54.31**	56.81	**54.31**
					RE	**1.0000**	0.9560	**1.0000**
				0.1	AMSE	70.95	74.09	**70.36**
					RE	1.0000	0.9576	**1.0084**
				0.3	AMSE	119.32	122.31	**111.60**
					RE	1.0000	0.9756	**1.0692**
				0.5	AMSE	215.65	216.20	**182.16**
					RE	1.0000	0.9975	**1.1838**



ภาพที่ 4.1.1 การเปรียบเทียบวิธีการประมาณค่าตัวแปรตามระหว่างค่า AMSE กับค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความแปรปรวนความคลาดเคลื่อน ในรูปแบบเลขชี้กำลัง (δ) เมื่อความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2

จากตารางที่ 4.1.1 ซึ่งแสดงผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2 พบว่า โดยส่วนใหญ่ เมื่อเปรียบเทียบแต่วิธีการประมาณค่าตัวแปรตาม พบว่าค่า AMSE ของแต่ละวิธีเมื่อเรียงจากน้อยที่สุดไปหามากที่สุด คือ วิธี IRWLS วิธี OLS และ วิธีการแปลง Box และ Cox ตามลำดับ แต่ยกเว้นกรณีที่ระดับ $n = 25$ และ 50 , $\sigma^2 = 900$ หรือเท่ากับ 1,800 และ $\delta = 0.5$ วิธีที่มีค่า AMSE มากที่สุด คือ วิธี OLS และวิธีที่มีค่า AMSE น้อยที่สุดคือ วิธี IRWLS และกรณีที่ δ เท่ากับศูนย์ ค่า AMSE ระหว่างวิธี OLS กับ IRWLS จะให้ค่า AMSE เท่ากัน และวิธีการแปลง Box และ Cox จะมีค่า AMSE มากที่สุด ซึ่งแสดงให้เห็นว่าวิธี IRWLS เป็นวิธีที่ดีที่สุด เมื่อ δ ไม่เท่ากับศูนย์ และระหว่างวิธีการแปลง Box และ Cox กับ วิธี OLS วิธีใดจะดีกว่า ก็ขึ้นอยู่กับ ความแปรปรวนของความคลาดเคลื่อนในเทอมพารามิเตอร์ σ^2 และเทอมชี้กำลัง δ มากน้อยเพียงใด จากภาพที่ 4.1.1 จะเห็นว่า แนวโน้มค่า AMSE ของแต่ละวิธีการประมาณค่าตัวแปรตาม พบว่า เมื่อ δ เพิ่มขึ้นหรือ σ^2 เพิ่มขึ้น จะทำให้ค่า AMSE เพิ่มขึ้น ซึ่งการเพิ่มขึ้นของค่า σ^2 จะแปรผันตามกับค่า AMSE กล่าวคือ เมื่อ σ^2 เพิ่มขึ้นจำนวน 1 เท่า จะทำให้ค่า AMSE เพิ่มขึ้นโดยประมาณ 1 เท่า ในขณะที่เมื่อขนาดตัวอย่าง (n) เพิ่มขึ้นจะทำให้ค่า AMSE มีค่าน้อยลง โดยที่เมื่อเพิ่ม n จำนวน 1 เท่าจากเดิมจะทำให้ค่า AMSE ลดลงโดยประมาณ 0.5 เท่าของค่า AMSE จากขนาดตัวอย่างเดิม เพราะขนาดตัวอย่างที่เพิ่มขึ้น จะทำให้ค่าความคลาดเคลื่อนในการประมาณหรือพยากรณ์ลดลง

เมื่อพิจารณาค่า RE ซึ่งเป็นอัตราส่วนระหว่างค่า AMSE ของวิธี OLS กับค่า AMSE ของวิธีการประมาณค่าตัวแปรตามแต่ละวิธี พบว่า สำหรับทุกระดับขนาดตัวอย่าง ค่า RE ของวิธีการแปลง Box และ Cox และวิธี IRWLS มีแนวโน้มที่จะเพิ่มขึ้นเรื่อยๆ ถ้าค่า σ^2 หรือ δ มีค่าสูงขึ้น โดยค่า RE ของวิธี IRWLS จะมีค่ามากกว่าวิธีการแปลงของ Box และ Cox โดยที่ค่า RE ของวิธีการแปลงของ Box และ Cox จะมีค่าน้อยกว่า 1 และจะเริ่มมีค่าเข้าใกล้หรือมากกว่า 1 เมื่อ $\delta = 0.5$ และ $\sigma^2 = 900$ หรือ $1,800$ และหากพิจารณาที่ขนาดตัวอย่างต่างกันพบว่าค่า RE ของวิธีการแปลง Box และ Cox กับวิธี IRWLS จะน้อยลงเพียงเล็กน้อย และ เพิ่มขึ้นเพียงเล็กน้อยตามลำดับ เมื่อ n มีค่าเพิ่มขึ้น ดังนั้นปัจจัยที่มีอิทธิพลต่อค่า RE ของแต่ละวิธีการประมาณคือ σ^2, δ และ n

ตารางที่ 4.1.2 ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1

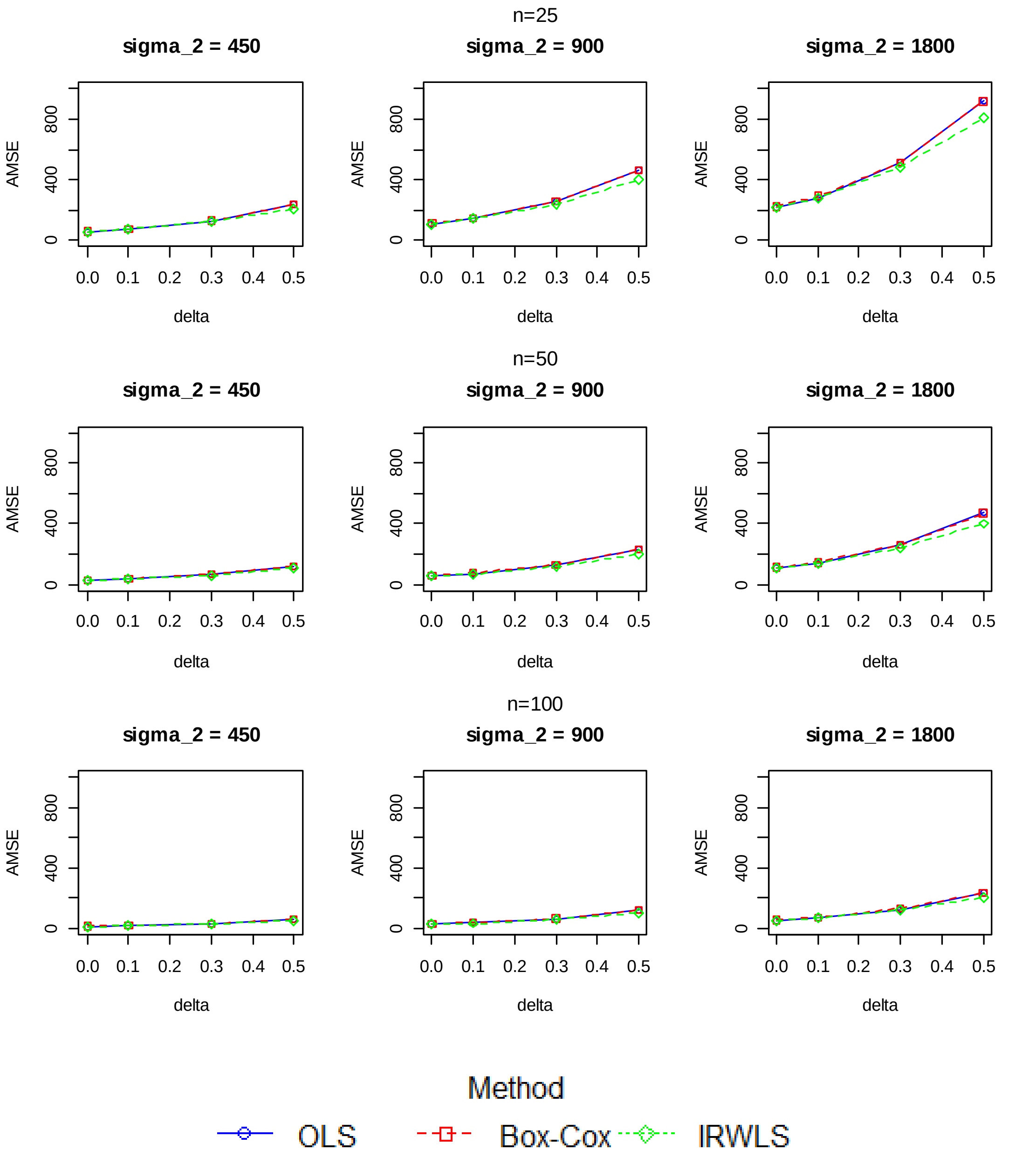
σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
450	450	25	450	0	AMSE	54.17	58.10	54.17
					RE	1.0000	0.9324	1.0000
				0.1	AMSE	70.82	75.33	70.52
					RE	1.0000	0.9401	1.0043
				0.3	AMSE	126.61	130.50	121.19
					RE	1.0000	0.9702	1.0447
				0.5	AMSE	237.61	239.92	212.03
					RE	1.0000	0.9904	1.1206
			900	0	AMSE	109.05	114.93	109.05
					RE	1.0000	0.9488	1.0000
				0.1	AMSE	142.05	147.65	141.20
					RE	1.0000	0.9621	1.0060
				0.3	AMSE	254.14	258.84	241.29
					RE	1.0000	0.9818	1.0524
				0.5	AMSE	463.11	458.89	402.53
					RE	1.0000	1.0092	1.1505
			1,800	0	AMSE	215.70	223.25	215.70
					RE	1.0000	0.9662	1.0000
				0.1	AMSE	282.24	293.65	280.32
					RE	1.0000	0.9611	1.0068
				0.3	AMSE	511.02	517.47	482.96
					RE	1.0000	0.9875	1.0581
				0.5	AMSE	926.29	917.72	805.30
					RE	1.0000	1.0093	1.1502

ตารางที่ 4.1.2 (ต่อ) ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1

σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
450	450	50	450	0	AMSE	**26.47**	28.94	**26.47**
450	450	50	450	0	RE	**1.0000**	0.9147	**1.0000**
450	450	50	450	0.1	AMSE	35.76	38.34	**35.58**
450	450	50	450	0.1	RE	1.0000	0.9327	**1.0051**
450	450	50	450	0.3	AMSE	63.44	66.45	**60.67**
450	450	50	450	0.3	RE	1.0000	0.9547	**1.0457**
450	450	50	450	0.5	AMSE	118.32	120.37	**104.82**
450	450	50	450	0.5	RE	1.0000	0.9830	**1.1288**
450	450	50	900	0	AMSE	**53.64**	57.07	**53.64**
450	450	50	900	0	RE	**1.0000**	0.9399	**1.0000**
450	450	50	900	0.1	AMSE	71.26	74.99	**70.79**
450	450	50	900	0.1	RE	1.0000	0.9503	**1.0066**
450	450	50	900	0.3	AMSE	128.87	132.33	**121.84**
450	450	50	900	0.3	RE	1.0000	0.9739	**1.0577**
450	450	50	900	0.5	AMSE	232.00	231.75	**200.68**
450	450	50	900	0.5	RE	1.0000	1.0011	**1.1561**
450	450	50	1,800	0	AMSE	**107.03**	112.81	**107.03**
450	450	50	1,800	0	RE	**1.0000**	0.9488	**1.0000**
450	450	50	1,800	0.1	AMSE	138.66	144.52	**137.59**
450	450	50	1,800	0.1	RE	1.0000	0.9595	**1.0078**
450	450	50	1,800	0.3	AMSE	258.10	262.36	**242.02**
450	450	50	1,800	0.3	RE	1.0000	0.9838	**1.0664**
450	450	50	1,800	0.5	AMSE	473.67	471.00	**399.57**
450	450	50	1,800	0.5	RE	1.0000	1.0057	**1.1854**

ตารางที่ 4.1.2 (ต่อ) ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1

σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
450	450	100	450	0	AMSE	**13.66**	15.01	**13.66**
					RE	**1.0000**	0.9101	**1.0000**
				0.1	AMSE	18.07	19.47	**17.99**
					RE	1.0000	0.9281	**1.0044**
				0.3	AMSE	31.80	33.50	**30.30**
					RE	1.0000	0.9493	**1.0495**
				0.5	AMSE	59.96	61.05	**52.41**
					RE	1.0000	0.9821	**1.1441**
			900	0	AMSE	**26.69**	28.50	**26.69**
					RE	**1.0000**	0.9365	**1.0000**
				0.1	AMSE	35.35	37.31	**35.11**
					RE	1.0000	0.9475	**1.0068**
				0.3	AMSE	64.07	65.64	**60.04**
					RE	1.0000	0.9761	**1.0671**
				0.5	AMSE	118.51	118.73	**101.23**
					RE	1.0000	0.9981	**1.1707**
			1,800	0	AMSE	**53.61**	56.33	**53.61**
					RE	**1.0000**	0.9517	**1.0000**
				0.1	AMSE	71.67	74.75	**71.28**
					RE	1.0000	0.9588	**1.0055**
				0.3	AMSE	127.00	130.00	**118.92**
					RE	1.0000	0.9769	**1.0679**
				0.5	AMSE	236.82	235.36	**201.05**
					RE	1.0000	1.0062	**1.1779**



ภาพที่ 4.1.2 การเปรียบเทียบวิธีการประมาณค่าตัวแปรตามระหว่างค่า AMSE กับค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความแปรปรวนความคลาดเคลื่อน ในรูปแบบเลขชี้กำลัง (δ) เมื่อความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1

จากตารางที่ 4.1.2 ซึ่งแสดงผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1 พบว่า โดยส่วนใหญ่ เมื่อเปรียบเทียบแต่วิธีการประมาณค่าตัวแปรตาม พบว่าค่า AMSE ของแต่ละวิธีเมื่อเรียงจากน้อยที่สุดไปหามากที่สุด คือ วิธี IRWLS วิธี OLS และ วิธีการแปลง Box และ Cox ตามลำดับ แต่ยกเว้นกรณีที่ทุกระดับขนาดตัวอย่าง $\sigma^2 = 900$ หรือเท่ากับ 1,800 และ $\delta = 0.5$ วิธีที่มีค่า AMSE มากที่สุด คือ วิธี OLS ส่วนวิธีที่มีค่า AMSE น้อยที่สุดคือ วิธี IRWLS และกรณีที่ δ เท่ากับศูนย์ ค่า AMSE ระหว่างวิธี OLS กับ IRWLS จะให้ค่า AMSE เท่ากัน และวิธีการแปลง Box และ Cox จะมีค่า AMSE มากที่สุด ซึ่งแสดงให้เห็นว่าวิธี IRWLS เป็นวิธีที่ดีที่สุด เมื่อ δ ไม่เท่ากับศูนย์และระหว่างวิธีการแปลง Box และ Cox กับ วิธี OLS วิธีใดจะดีกว่า ก็ขึ้นอยู่กับ ความแปรปรวนของความคลาดเคลื่อนในเทอมพารามิเตอร์ σ^2 และเทอมชี้กำลัง δ มากน้อยเพียงใด จากภาพที่ 4.1.2 จะเห็นว่า แนวโน้มค่า AMSE ของแต่ละวิธีการประมาณค่าตัวแปรตาม พบว่า เมื่อ δ เพิ่มขึ้นหรือ σ^2 เพิ่มขึ้น จะทำให้ค่า AMSE เพิ่มขึ้น ซึ่งการเพิ่มขึ้นของค่า σ^2 จะแปรผันตามกับค่า AMSE กล่าวคือ เมื่อ σ^2 เพิ่มขึ้นจำนวน 1 เท่า จะทำให้ค่า AMSE เพิ่มขึ้นโดยประมาณ 1 เท่า ในขณะที่เมื่อขนาดตัวอย่าง (n) เพิ่มขึ้นจะทำให้ค่า AMSE มีค่าน้อยลง โดยที่เมื่อเพิ่ม n จำนวน 1 เท่าจากเดิมจะทำให้ค่า AMSE ลดลงโดยประมาณ 0.5 เท่าของค่า AMSE จากขนาดตัวอย่างเดิม เพราะขนาดตัวอย่างที่เพิ่มขึ้น จะทำให้ค่าความคลาดเคลื่อนในการประมาณหรือพยากรณ์ลดลง

เมื่อพิจารณาค่า RE ซึ่งเป็นอัตราส่วนระหว่างค่า AMSE ของวิธี OLS กับค่า AMSE ของวิธีการประมาณค่าตัวแปรตามแต่ละวิธี พบว่า สำหรับทุกระดับขนาดตัวอย่าง ค่า RE ของวิธีการแปลง Box และ Cox และวิธี IRWLS มีแนวโน้มที่จะเพิ่มขึ้นเรื่อยๆ ถ้าค่า σ^2 หรือ δ มีค่าสูงขึ้น โดยค่า RE ของวิธี IRWLS จะมีค่ามากกว่าวิธีการแปลงของ Box และ Cox โดยที่ค่า RE ของวิธีการแปลงของ Box และ Cox จะมีค่าน้อยกว่า 1 และจะเริ่มมีค่าเข้าใกล้หรือมากกว่า 1 เมื่อ $\delta = 0.5$ และ $\sigma^2 = 900$ หรือ $1,800$ และหากพิจารณาที่ขนาดตัวอย่างต่างกันพบว่าค่า RE ของวิธีการแปลง Box และ Cox กับวิธี IRWLS จะน้อยลงเพียงเล็กน้อย และ เพิ่มขึ้นเพียงเล็กน้อยตามลำดับ เมื่อ n มีค่าเพิ่มขึ้น ดังนั้นปัจจัยที่มีอิทธิพลต่อค่า RE ของแต่ละวิธีการประมาณคือ σ^2, δ และ n

ตารางที่ 4.1.3 ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 2:1

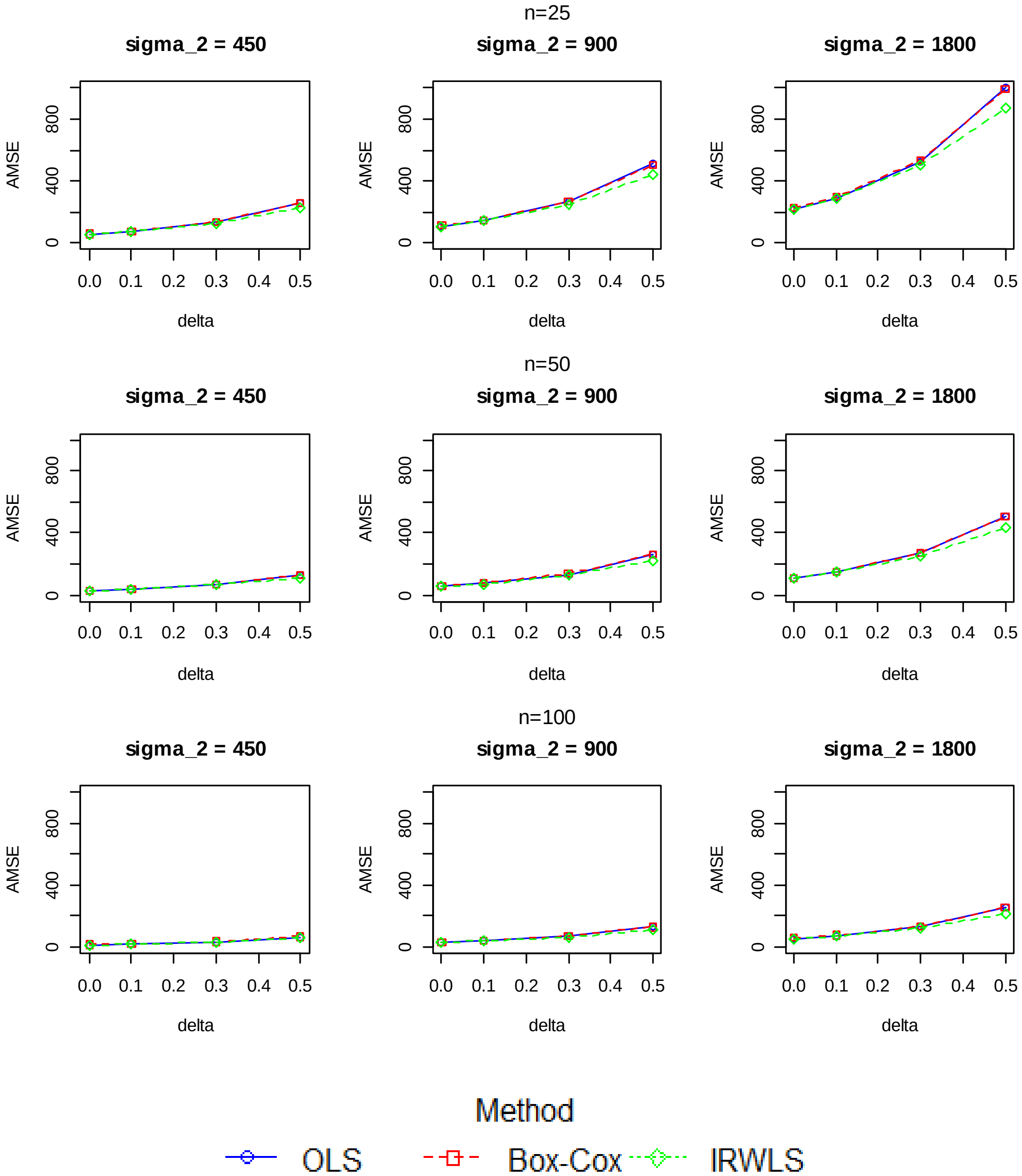
σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
600	300	25	450	0	AMSE	54.07	58.14	54.07
					RE	1.0000	0.9300	1.0000
				0.1	AMSE	72.67	77.19	72.38
					RE	1.0000	0.9414	1.0040
				0.3	AMSE	133.79	138.11	128.14
					RE	1.0000	0.9687	1.0441
				0.5	AMSE	255.07	257.16	226.30
					RE	1.0000	0.9919	1.1271
			900	0	AMSE	108.90	114.32	108.90
					RE	1.0000	0.9526	1.0000
				0.1	AMSE	142.25	148.10	141.49
					RE	1.0000	0.9605	1.0054
				0.3	AMSE	264.27	269.78	249.88
					RE	1.0000	0.9796	1.0576
				0.5	AMSE	510.44	506.02	441.35
					RE	1.0000	1.0087	1.1565
			1,800	0	AMSE	217.95	225.73	217.95
					RE	1.0000	0.9655	1.0000
				0.1	AMSE	293.49	301.76	291.27
					RE	1.0000	0.9726	1.0076
				0.3	AMSE	527.59	535.32	500.11
					RE	1.0000	0.9856	1.0549
				0.5	AMSE	999.60	989.68	870.31
					RE	1.0000	1.0100	1.1486

ตารางที่ 4.1.3 (ต่อ) ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 2:1

σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
600	300	50	450	0	AMSE	**27.12**	29.68	**27.12**
					RE	**1.0000**	0.9137	**1.0000**
				0.1	AMSE	36.04	38.63	**35.83**
					RE	1.0000	0.9330	**1.0059**
				0.3	AMSE	66.95	69.89	**64.29**
					RE	1.0000	0.9579	**1.0414**
				0.5	AMSE	127.01	128.44	**111.82**
					RE	1.0000	0.9889	**1.1358**
			900	0	AMSE	**52.56**	55.72	**52.56**
					RE	**1.0000**	0.9433	**1.0000**
				0.1	AMSE	72.93	76.59	**72.51**
					RE	1.0000	0.9522	**1.0058**
				0.3	AMSE	132.35	135.94	**124.93**
					RE	1.0000	0.9736	**1.0594**
				0.5	AMSE	258.40	259.27	**222.60**
					RE	1.0000	0.9966	**1.1608**
			1,800	0	AMSE	**106.66**	111.61	**106.66**
					RE	**1.0000**	0.9556	**1.0000**
				0.1	AMSE	146.63	151.65	**145.42**
					RE	1.0000	0.9669	**1.0083**
				0.3	AMSE	269.33	273.60	**251.65**
					RE	1.0000	0.9844	**1.0703**
				0.5	AMSE	506.84	504.99	**432.24**
					RE	1.0000	1.0037	**1.1726**

ตารางที่ 4.1.3 (ต่อ) ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 2:1

σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
600	300	100	450	0	AMSE	**13.56**	14.89	**13.56**
					RE	**1.0000**	0.9107	**1.0000**
				0.1	AMSE	18.43	19.80	**18.32**
					RE	1.0000	0.9308	**1.0060**
				0.3	AMSE	33.27	34.75	**31.72**
					RE	1.0000	0.9574	**1.0489**
				0.5	AMSE	65.31	66.26	**56.71**
					RE	1.0000	0.9857	**1.1516**
			900	0	AMSE	**26.77**	28.56	**26.77**
					RE	**1.0000**	0.9373	**1.0000**
				0.1	AMSE	36.08	38.02	**35.83**
					RE	1.0000	0.9490	**1.0070**
				0.3	AMSE	66.97	68.84	**63.18**
					RE	1.0000	0.9728	**1.0600**
				0.5	AMSE	128.00	128.26	**109.43**
					RE	1.0000	0.9980	**1.1697**
			1,800	0	AMSE	**54.26**	56.92	**54.26**
					RE	**1.0000**	0.9533	**1.0000**
				0.1	AMSE	72.83	75.80	**72.15**
					RE	1.0000	0.9608	**1.0094**
				0.3	AMSE	130.88	134.02	**122.57**
					RE	1.0000	0.9766	**1.0678**
				0.5	AMSE	257.24	256.67	**215.83**
					RE	1.0000	1.0022	**1.1919**



ภาพที่ 4.1.3 การเปรียบเทียบวิธีการประมาณค่าตัวแปรตามระหว่างค่า AMSE กับค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความแปรปรวนความคลาดเคลื่อน ในรูปแบบเลขชี้กำลัง (δ) เมื่อความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 2:1

จากตารางที่ 4.1.3 ซึ่งแสดงผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 2:1 พบว่า โดยส่วนใหญ่ เมื่อเปรียบเทียบแต่วิธีการประมาณค่าตัวแปรตาม พบว่าค่า AMSE ของแต่ละวิธีเมื่อเรียงจากน้อยที่สุดไปหามากที่สุด คือ วิธี IRWLS วิธี OLS และ วิธีการแปลง Box และ Cox ตามลำดับ แต่ยกเว้นกรณีที่ทุกระดับขนาดตัวอย่าง ค่า σ^2 เท่ากับ 1,800 และ $\delta = 0.5$ วิธีที่มีค่า AMSE มากที่สุด คือ วิธี OLS ส่วนวิธีที่มีค่า AMSE น้อยที่สุดคือ วิธี IRWLS และกรณีที่ δ เท่ากับศูนย์ ค่า AMSE ระหว่างวิธี OLS กับ IRWLS จะให้ค่า AMSE เท่ากัน และวิธีการแปลง Box และ Cox จะมีค่า AMSE มากที่สุด ซึ่งแสดงให้เห็นว่าวิธี IRWLS เป็นวิธีที่ดีที่สุด เมื่อ δ ไม่เท่ากับศูนย์ และระหว่างวิธีการแปลง Box และ Cox กับ วิธี OLS วิธีใดจะดีกว่า ก็ขึ้นอยู่กับ ความแปรปรวนของความคลาดเคลื่อนในเทอมพารามิเตอร์ σ^2 และเทอมชี้กำลัง δ มากน้อยเพียงใด จากภาพที่ 4.1.3 จะเห็นว่า แนวโน้มค่า AMSE ของแต่ละวิธีการประมาณค่าตัวแปรตาม พบว่า เมื่อ δ เพิ่มขึ้นหรือ σ^2 เพิ่มขึ้น จะทำให้ค่า AMSE เพิ่มขึ้น ซึ่งการเพิ่มขึ้นของค่า σ^2 จะแปรผันตามกับค่า AMSE กล่าวคือ เมื่อ σ^2 เพิ่มขึ้นจำนวน 1 เท่า จะทำให้ค่า AMSE เพิ่มขึ้นโดยประมาณ 1 เท่า ในขณะที่เมื่อขนาดตัวอย่าง (n) เพิ่มขึ้นจะทำให้ค่า AMSE มีค่าน้อยลง โดยที่เมื่อเพิ่ม n จำนวน 1 เท่าจากเดิมจะทำให้ค่า AMSE ลดลงโดยประมาณ 0.5 เท่าของค่า AMSE จากขนาดตัวอย่างเดิม เพราะขนาดตัวอย่างที่เพิ่มขึ้น จะทำให้ค่าความคลาดเคลื่อนในการประมาณหรือพยากรณ์ลดลง

เมื่อพิจารณาค่า RE ซึ่งเป็นอัตราส่วนระหว่างค่า AMSE ของวิธี OLS กับค่า AMSE ของวิธีการประมาณค่าตัวแปรตามแต่ละวิธี พบว่า สำหรับทุกระดับขนาดตัวอย่าง ค่า RE ของวิธีการแปลง Box และ Cox และวิธี IRWLS มีแนวโน้มที่จะเพิ่มขึ้นเรื่อยๆ ถ้าค่า σ^2 หรือ δ มีค่าสูงขึ้น โดยค่า RE ของวิธี IRWLS จะมีค่ามากกว่าวิธีการแปลงของ Box และ Cox โดยที่ค่า RE ของวิธีการแปลงของ Box และ Cox จะมีค่าน้อยกว่า 1 และจะเริ่มมีค่าเข้าใกล้ หรือมากกว่า 1 เมื่อ $\delta = 0.5$ และ $\sigma^2 = 900$ หรือ $1,800$ และหากพิจารณาที่ขนาดตัวอย่างต่างกันพบว่าค่า RE ของวิธีการแปลง Box และ Cox กับวิธี IRWLS จะน้อยลงเพียงเล็กน้อย และ เพิ่มขึ้นเพียงเล็กน้อยตามลำดับ เมื่อ n มีค่าเพิ่มขึ้น ดังนั้นปัจจัยที่มีอิทธิพลต่อค่า RE ของแต่ละวิธีการประมาณคือ σ^2, δ และ n

การสรุปผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) (ในหัวข้อ 4.1)

เมื่อพิจารณาผลการวิจัยจากตารางที่ 4.1.1 ถึง 4.1.3 ซึ่งเป็นการแสดงผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) พบว่า ที่ δ ไม่เท่ากับศูนย์ ในทุกขนาดตัวอย่าง และ ทุกค่าของ σ^2 ที่ศึกษา เมื่ออัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เปลี่ยนแปลงเป็น 1:2 1:1 และ 2:1 ตามลำดับ จะทำให้ค่า AMSE มีแนวโน้มเพิ่มขึ้นของแต่ละวิธีการประมาณหรือการพยากรณ์ตัวแปรตาม

ในกรณีที่อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2 หรือ 1:1 เมื่อเปรียบเทียบค่า AMSE แต่ละวิธีการประมาณค่าตัวแปรตาม พบว่าเมื่อเรียงค่า AMSE ของแต่ละวิธีจากน้อยที่สุดไปยังมากที่สุดคือ วิธี IRWLS วิธี OLS และ วิธีการแปลง Box และ Cox ตามลำดับ ยกเว้นกรณีที่

$$n = 25,50 \quad \sigma^2 = 900,1800 \text{ และ } \delta = 0.5$$

$$n = 100 \quad \sigma^2 = 1,800 \text{ และ } \delta = 0.5$$

ซึ่งทั้งสองกรณี วิธี IRWLS ยังเป็นวิธีที่มีค่า AMSE น้อยที่สุด แต่วิธีที่มีค่า AMSE มากที่สุดคือวิธี OLS

แต่ในกรณีที่อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 2:1 เมื่อเปรียบเทียบค่า AMSE แต่ละวิธีการประมาณค่าตัวแปรตาม พบว่าเมื่อเรียงค่า AMSE ของแต่ละวิธีจากน้อยที่สุดไปยังมากที่สุดคือ วิธี IRWLS วิธี OLS และ วิธีการแปลง Box และ Cox ตามลำดับ ยกเว้นกรณีที่

$$n = 25, \ \sigma^2 = 900,1800 \text{ และ } \delta = 0.5$$

$$n = 50,100, \ \sigma^2 = 1,800 \text{ และ } \delta = 0.5$$

ซึ่งทั้งสองกรณี วิธี IRWLS ยังเป็นวิธีที่มีค่า AMSE น้อยที่สุด แต่วิธีที่มีค่า AMSE มากที่สุดคือวิธี OLS

สรุปปัจจัยที่มีผลการเปลี่ยนแปลงค่า AMSE (ในหัวข้อ 4.1)

1. ขนาดตัวอย่าง (n)

เมื่อขนาดตัวอย่าง (n) เพิ่มขึ้น จะทำให้ค่า AMSE น้อยลง ในอัตราส่วนที่แปรผกผัน เนื่องจากการเพิ่มขึ้นของขนาดตัวอย่างจะทำให้ความคลาดเคลื่อนในการประมาณหรือพยากรณ์ตัวแปรตามน้อยลง

2. ความแปรปรวนของความคลาดเคลื่อน

ซึ่งมีพารามิเตอร์ที่เกี่ยวข้องในงานวิจัยนี้ ดังต่อไปนี้

2.1 พารามิเตอร์ σ^2 ที่บอกระดับขนาดของความแปรปรวนของความคลาดเคลื่อน

เมื่อค่า σ^2 เพิ่มขึ้น จะทำให้ค่า AMSE เพิ่มขึ้น ในอัตราส่วนที่แปรผันตาม เนื่องจากเมื่อเพิ่มค่า σ^2 จะทำให้ความแปรปรวนของความคลาดเคลื่อนสูงขึ้น ส่งผลให้การประมาณค่าตัวแปรตามมีความคลาดเคลื่อนมากขึ้น

2.2 พารามิเตอร์ δ เป็นค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความคลาดเคลื่อนในรูปแบบเลขชี้กำลัง

เมื่อค่า δ เพิ่มขึ้น จะทำให้ค่า AMSE เพิ่มขึ้น เนื่องจากการเพิ่มค่า δ จะทำให้ความแปรปรวนของความคลาดเคลื่อนสูงขึ้น ส่งผลให้การประมาณค่าตัวแปรตามมีความคลาดเคลื่อนมากขึ้น

3. อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2

ที่ δ ไม่เท่ากับศูนย์ เมื่อสัดส่วนของความแปรปรวนของตัวแปรอิสระตัวที่หนึ่งมีค่ามากขึ้น จะมีผลทำให้ค่า AMSE เพิ่มขึ้น เนื่องจากการเพิ่มความแปรปรวนของตัวแปรอิสระตัวที่หนึ่ง จะมีผลทำให้ความแปรปรวนของความคลาดเคลื่อนในตัวแบบการถดถอยเชิงเส้นพหุมากขึ้น ส่งผลการประมาณค่าตัวแปรตามมีความคลาดเคลื่อนมากขึ้น

4.2 ผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับค่าจริงของตัวแปรตาม

ผู้วิจัยทำการศึกษาในกรณีที่ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2, 1:1 และ 2:1 ขนาดตัวอย่าง เท่ากับ 25, 50 และ 100 เทอมพารามิเตอร์ที่เกี่ยวข้องกับความแปรปรวนของความคลาดเคลื่อน σ^2 เท่ากับ 450, 900 และ 1,800 และ เทอมพารามิเตอร์ δ เท่ากับ 0, 0.1, 0.3 และ 0.5 ซึ่งผลการวิจัยในส่วนนี้จะแสดงในตารางที่ 4.2.1 ถึง ตารางที่ 4.2.3

ตารางที่	ความสัมพันธ์ของความคลาดเคลื่อน	อัตราส่วนความแปรปรวนของตัวแปรอิสระระหว่าง X_1 และ X_2
4.2.1	ค่าจริงของตัวแปรตาม (Y')	1:2
4.2.2	ค่าจริงของตัวแปรตาม (Y')	1:1
4.2.3	ค่าจริงของตัวแปรตาม (Y')	2:1

ตารางที่ 4.2.1 ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2

σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
300	600	25	450	0	AMSE	53.79	57.97	53.79
					RE	1.0000	0.9279	1.0000
				0.1	AMSE	99.53*	105.41	99.53*
					RE	1.0000*	0.9442	1.0000*
				0.3	AMSE	343.51	356.41	343.44
					RE	1.0000	0.9635	1.0002
				0.5	AMSE	1210.68	1252.99	1208.39
					RE	1.0000	0.9662	1.0019
			900	0	AMSE	110.45	116.19	110.45
					RE	1.0000	0.9506	1.0000
				0.1	AMSE	198.57*	206.01	198.57*
					RE	1.0000*	0.9639	1.0000*
				0.3	AMSE	715.17	738.81	714.83
					RE	1.0000	0.9680	1.0005
				0.5	AMSE	2410.04	2505.67	2408.88
					RE	1.0000	0.9618	1.0005
			1,800	0	AMSE	214.64	223.42	214.64
					RE	1.0000	0.9607	1.0000
				0.1	AMSE	400.63*	415.18	400.64*
					RE	1.0000*	0.9650	1.0000*
				0.3	AMSE	1394.82	1438.08	1394.42
					RE	1.0000	0.9699	1.0003
				0.5	AMSE	4879.00	4944.99	4876.45
					RE	1.0000	0.9867	1.0005

***หมายเหตุ** ค่า AMSE ระหว่างวิธี OLS กับวิธี IRWLS มีค่าแตกต่างกันน้อยกว่า 10^{-2}

ตารางที่ 4.2.1 (ต่อ) ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2

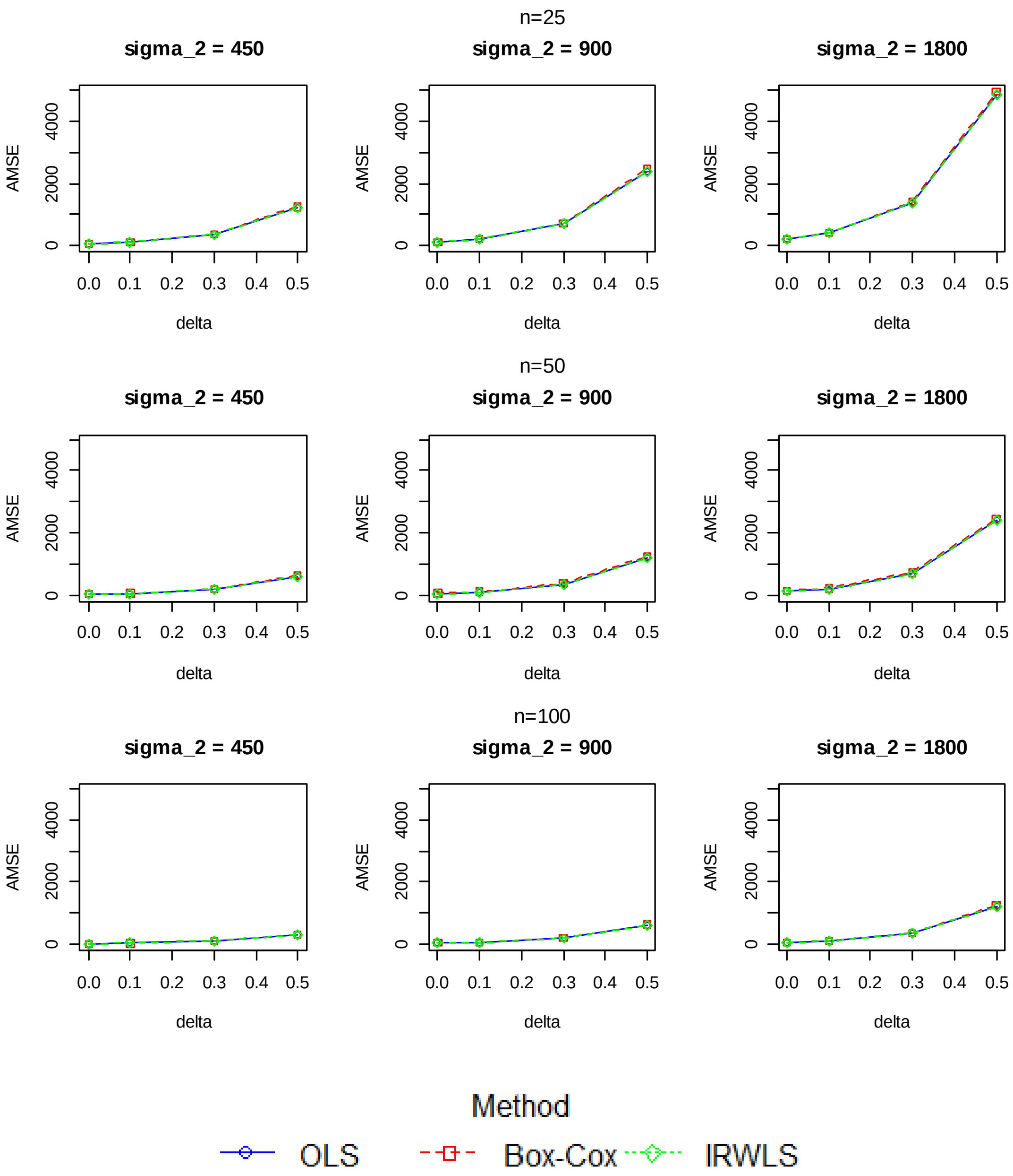
σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
300	600	50	450	0	AMSE	26.77	29.32	26.77
					RE	1.0000	0.9130	1.0000
				0.1	AMSE	49.82*	52.89	49.82*
					RE	1.0000*	0.9420	1.0000*
				0.3	AMSE	173.00	180.46	173.03
					RE	1.0000	0.9587	0.9998
				0.5	AMSE	608.05	634.59	607.58
					RE	1.0000	0.9582	1.0008
			900	0	AMSE	54.38	57.78	54.38
					RE	1.0000	0.9412	1.0000
				0.1	AMSE	99.72*	104.95	99.72*
					RE	1.0000*	0.9502	1.0000*
				0.3	AMSE	353.17	366.50	353.04
					RE	1.0000	0.9636	1.0004
				0.5	AMSE	1194.57	1248.78	1193.75
					RE	1.0000	0.9566	1.0007
			1,800	0	AMSE	107.95	112.82	107.95
					RE	1.0000	0.9568	1.0000
				0.1	AMSE	201.28*	209.61	201.29*
					RE	1.0000*	0.9603	1.0000*
				0.3	AMSE	689.43	715.95	688.95
					RE	1.0000	0.9630	1.0007
				0.5	AMSE	2385.12	2453.42	2383.56
					RE	1.0000	0.9722	1.0007

***หมายเหตุ** ค่า AMSE ระหว่างวิธี OLS กับวิธี IRWLS มีค่าแตกต่างกันน้อยกว่า 10^{-2}

ตารางที่ 4.2.1 (ต่อ) ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2

σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
300	600	100	450	0	AMSE	13.36	14.67	13.36
					RE	1.0000	0.9107	1.0000
				0.1	AMSE	24.49*	26.10	24.49*
					RE	1.0000*	0.9383	1.0000*
				0.3	AMSE	86.38*	90.53	86.38*
					RE	1.0000	0.9542	1.0000
				0.5	AMSE	299.93	313.10	299.73
					RE	1.0000	0.9579	1.0007
			900	0	AMSE	26.89	28.49	26.89
					RE	1.0000	0.9438	1.0000
				0.1	AMSE	49.12*	51.98	49.12*
					RE	1.0000*	0.9450	1.0000*
				0.3	AMSE	174.71	181.86	174.65
					RE	1.0000	0.9607	1.0003
				0.5	AMSE	600.03	625.25	599.22
					RE	1.0000	0.9597	1.0014
			1,800	0	AMSE	53.57	56.08	53.57
					RE	1.0000	0.9552	1.0000
				0.1	AMSE	101.98*	106.42	101.98*
					RE	1.0000*	0.9583	1.0000*
				0.3	AMSE	344.48	360.30	344.46
					RE	1.0000	0.9561	1.0001
				0.5	AMSE	1231.67	1262.54	1230.94
					RE	1.0000	0.9755	1.0006

***หมายเหตุ** ค่า AMSE ระหว่างวิธี OLS กับวิธี IRWLS มีค่าแตกต่างกันน้อยกว่า 10^{-2}



ภาพที่ 4.2.1 การเปรียบเทียบวิธีการประมาณค่าตัวแปรตามระหว่างค่า AMSE กับค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความแปรปรวนความคลาดเคลื่อน ในรูปแบบเลขชี้กำลัง (δ) เมื่อความคลาดเคลื่อนสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2

จากตารางที่ 4.2.1 ซึ่งแสดงผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2 พบว่า โดยส่วนใหญ่เมื่อเปรียบเทียบแต่วิธีการประมาณค่าตัวแปรตาม พบว่า ที่ทุกระดับขนาดตัวอย่าง และทุกค่าพารามิเตอร์ σ^2 ค่า AMSE ของแต่ละวิธีเมื่อเรียงจากน้อยที่สุดไปหามากที่สุด คือ วิธี IRWLS วิธี OLS และ วิธีการแปลง Box และ Cox ตามลำดับ แต่ยกเว้นกรณีที่ δ เท่ากับศูนย์ ค่า AMSE ระหว่างวิธี OLS กับ IRWLS จะให้ค่า AMSE เท่ากัน และวิธีการแปลง Box และ Cox จะมีค่า AMSE มากที่สุด ซึ่งแสดงให้เห็นว่าวิธี IRWLS เป็นวิธีที่ดีที่สุด เมื่อ δ ไม่เท่ากับศูนย์

จากภาพที่ 4.2.1 จะเห็นว่า แนวโน้มค่า AMSE ของแต่ละวิธีการประมาณค่าตัวแปรตาม พบว่า เมื่อ δ เพิ่มขึ้นหรือ σ^2 เพิ่มขึ้น จะทำให้ค่า AMSE เพิ่มขึ้น ซึ่งการเพิ่มขึ้นของค่า σ^2 จะแปรผันตามกับค่า AMSE กล่าวคือ เมื่อ σ^2 เพิ่มขึ้นจำนวน 1 เท่า จะทำให้ค่า AMSE เพิ่มขึ้นโดยประมาณ 1 เท่า ในขณะที่เมื่อขนาดตัวอย่าง (n) เพิ่มขึ้นจะทำให้ค่า AMSE มีค่าน้อยลง โดยที่เมื่อเพิ่ม n จำนวน 1 เท่าจากเดิมจะทำให้ค่า AMSE ลดลงโดยประมาณ 0.5 เท่าของค่า AMSE จากขนาดตัวอย่างเดิม เพราะขนาดตัวอย่างที่เพิ่มขึ้น จะทำให้ค่าความคลาดเคลื่อนในการประมาณหรือพยากรณ์ลดลง

เมื่อพิจารณาค่า RE ซึ่งเป็นอัตราส่วนระหว่างค่า AMSE ของวิธี OLS กับค่า AMSE ของวิธีการประมาณค่าตัวแปรตามแต่ละวิธี พบว่า สำหรับทุกระดับขนาดตัวอย่าง ค่า RE ของวิธีการแปลง Box และ Cox ไม่มีความแตกต่างกันมาก และค่า RE ของวิธี IRWLS มีแนวโน้มที่จะเพิ่มขึ้นเพียงเล็กน้อย ถ้าค่า σ^2 หรือ δ มีค่าสูงขึ้น ซึ่งค่า RE ของวิธี IRWLS จะมีค่ามากกว่าวิธีการแปลงของ Box และ Cox โดยที่ค่า RE ของวิธี IRWLS จะมีค่ามากกว่า 1 เพียงเล็กน้อยในตำแหน่งทศนิยมที่ 4 และค่า RE ของวิธีการแปลงของ Box และ Cox จะมีค่าน้อยกว่า 1 และหากพิจารณาที่ขนาดตัวอย่างต่างกันพบว่าค่า RE ของทั้งวิธีการแปลง Box และ Cox กับวิธี IRWLS จะไม่มีความแตกต่างกันมาก เมื่อ n มีค่าเพิ่มขึ้น ดังนั้นปัจจัยที่มีอิทธิพลต่อค่า RE ของแต่ละวิธีการประมาณคือ σ^2 และ δ

ตารางที่ 4.2.2 ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1

σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
450	450	25	450	0	AMSE	54.15	58.37	54.15
					RE	1.0000	0.9277	1.0000
				0.1	AMSE	101.43*	106.58	101.43*
					RE	1.0000*	0.9517	1.0000*
				0.3	AMSE	357.58	371.40	357.38
					RE	1.0000	0.9628	1.0006
				0.5	AMSE	1231.85	1279.47	1230.97
					RE	1.0000	0.9628	1.0007
			900	0	AMSE	109.45	115.10	109.45
					RE	1.0000	0.9509	1.0000
				0.1	AMSE	198.53*	206.04	198.53*
					RE	1.0000*	0.9636	1.0000*
				0.3	AMSE	684.89	709.97	684.52
					RE	1.0000	0.9647	1.0005
				0.5	AMSE	2451.13	2544.99	2449.52
					RE	1.0000	0.9631	1.0007
			1,800	0	AMSE	216.69	225.70	216.69
					RE	1.0000	0.9601	1.0000
				0.1	AMSE	402.62*	416.46	402.61*
					RE	1.0000*	0.9668	1.0000*
				0.3	AMSE	1376.39	1418.08	1376.70
					RE	1.0000	0.9706	0.9998
				0.5	AMSE	4793.76	4928.36	4792.85
					RE	1.0000	0.9727	1.0002

***หมายเหตุ** ค่า AMSE ระหว่างวิธี OLS กับวิธี IRWLS มีค่าแตกต่างกันน้อยกว่า 10^{-2}

ตารางที่ 4.2.2 (ต่อ) ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1

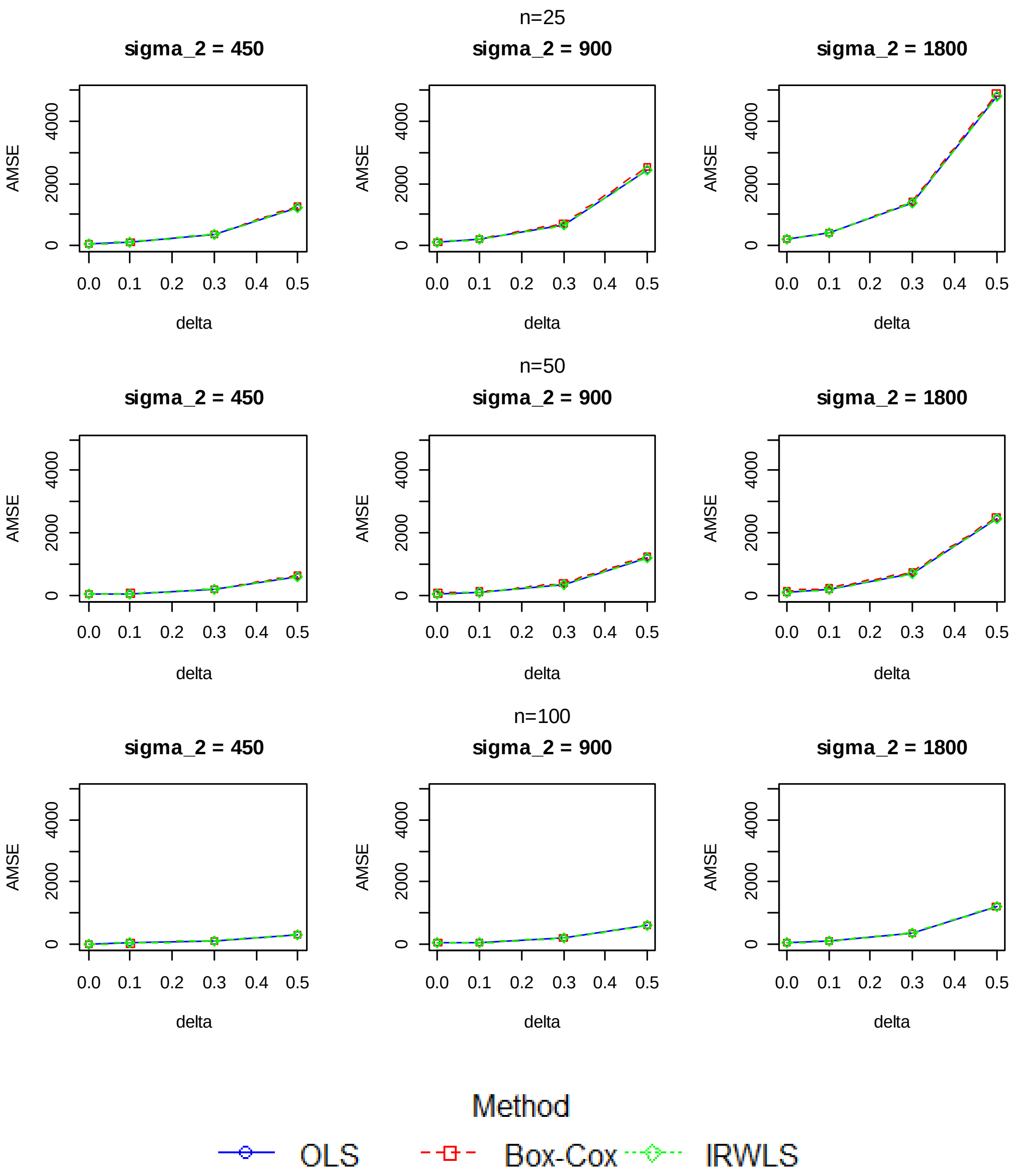
σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
450	450	50	450	0	AMSE	26.81	29.36	26.81
					RE	1.0000	0.9131	1.0000
				0.1	AMSE	50.60*	53.72	50.60*
					RE	1.0000*	0.9419	1.0000*
				0.3	AMSE	173.63	180.52	173.54
					RE	1.0000	0.9618	1.0005
				0.5	AMSE	592.03	618.42	591.63
					RE	1.0000	0.9573	1.0007
			900	0	AMSE	53.72	56.86	53.72
					RE	1.0000	0.9448	1.0000
				0.1	AMSE	100.08*	104.48	100.08*
					RE	1.0000*	0.9579	1.0000*
				0.3	AMSE	350.44	366.07	350.42
					RE	1.0000	0.9573	1.0001
				0.5	AMSE	1192.97	1247.22	1191.20
					RE	1.0000	0.9565	1.0015
			1,800	0	AMSE	106.01	111.33	106.01
					RE	1.0000	0.9522	1.0000
				0.1	AMSE	201.20*	209.36	201.21*
					RE	1.0000*	0.9610	1.0000*
				0.3	AMSE	688.16	720.03	687.97
					RE	1.0000	0.9557	1.0003
				0.5	AMSE	2457.87	2512.74	2456.50
					RE	1.0000	0.9782	1.0006

***หมายเหตุ** ค่า AMSE ระหว่างวิธี OLS กับวิธี IRWLS มีค่าแตกต่างกันน้อยกว่า 10^{-2}

ตารางที่ 4.2.2 (ต่อ) ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1

σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
450	450	100	450	0	AMSE	**13.50**	14.84	**13.50**
					RE	**1.0000**	0.9097	**1.0000**
				0.1	AMSE	**25.26***	26.95	**25.26***
					RE	**1.0000***	0.9373	**1.0000***
				0.3	AMSE	87.97	91.56	**87.91**
					RE	1.0000	0.9608	**1.0007**
				0.5	AMSE	300.82	313.42	**300.33**
					RE	1.0000	0.9598	**1.0016**
			900	0	AMSE	**27.24**	28.87	**27.24**
					RE	**1.0000**	0.9435	**1.0000**
				0.1	AMSE	**51.15***	53.29	**51.15***
					RE	**1.0000***	0.9598	**1.0000***
				0.3	AMSE	175.99	184.98	**175.84**
					RE	1.0000	0.9514	**1.0009**
				0.5	AMSE	595.08	620.12	**594.56**
					RE	1.0000	0.9596	**1.0009**
			1,800	0	AMSE	**54.45**	57.19	**54.45**
					RE	**1.0000**	0.9521	**1.0000**
				0.1	AMSE	**99.94***	103.96	**99.93***
					RE	**1.0000***	0.9613	**1.0001***
				0.3	AMSE	**347.13**	361.58	347.19
					RE	**1.0000**	0.9600	0.9998
				0.5	AMSE	1193.17	1220.99	**1191.95**
					RE	1.0000	0.9772	**1.0010**

***หมายเหตุ** ค่า AMSE ระหว่างวิธี OLS กับวิธี IRWLS มีค่าแตกต่างกันน้อยกว่า 10^{-2}



ภาพที่ 4.2.2 การเปรียบเทียบวิธีการประมาณค่าตัวแปรตามระหว่างค่า AMSE กับค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความแปรปรวนความคลาดเคลื่อน ในรูปแบบเลขชี้กำลัง (δ) เมื่อความคลาดเคลื่อนสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1

จากตารางที่ 4.2.2 ซึ่งแสดงผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1 พบว่า โดยส่วนใหญ่เมื่อเปรียบเทียบแต่วิธีการประมาณค่าตัวแปรตาม พบว่า ที่ทุกระดับขนาดตัวอย่าง และทุกค่าพารามิเตอร์ σ^2 ค่า AMSE ของแต่ละวิธีเมื่อเรียงจากน้อยที่สุดไปหามากที่สุด คือ วิธี IRWLS วิธี OLS และ วิธีการแปลง Box และ Cox ตามลำดับ แต่ยกเว้นกรณีที่ δ เท่ากับศูนย์ ค่า AMSE ระหว่างวิธี OLS กับ IRWLS จะให้ค่า AMSE เท่ากัน และวิธีการแปลง Box และ Cox จะมีค่า AMSE มากที่สุด ซึ่งแสดงให้เห็นว่าวิธี IRWLS เป็นวิธีที่ดีที่สุด เมื่อ δ ไม่เท่ากับศูนย์

จากภาพที่ 4.2.2 จะเห็นว่า แนวโน้มค่า AMSE ของแต่ละวิธีการประมาณค่าตัวแปรตาม พบว่า เมื่อ δ เพิ่มขึ้นหรือ σ^2 เพิ่มขึ้น จะทำให้ค่า AMSE เพิ่มขึ้น ซึ่งการเพิ่มขึ้นของค่า σ^2 จะแปรผันตามกับค่า AMSE กล่าวคือ เมื่อ σ^2 เพิ่มขึ้นจำนวน 1 เท่า จะทำให้ค่า AMSE เพิ่มขึ้นโดยประมาณ 1 เท่า ในขณะที่เมื่อขนาดตัวอย่าง (n) เพิ่มขึ้นจะทำให้ค่า AMSE มีค่าน้อยลง โดยที่เมื่อเพิ่ม n จำนวน 1 เท่าจากเดิมจะทำให้ค่า AMSE ลดลงโดยประมาณ 0.5 เท่าของค่า AMSE จากขนาดตัวอย่างเดิม เพราะขนาดตัวอย่างที่เพิ่มขึ้น จะทำให้ค่าความคลาดเคลื่อนในการประมาณหรือพยากรณ์ลดลง

เมื่อพิจารณาค่า RE ซึ่งเป็นอัตราส่วนระหว่างค่า AMSE ของวิธี OLS กับค่า AMSE ของวิธีการประมาณค่าตัวแปรตามแต่ละวิธี พบว่า สำหรับทุกระดับขนาดตัวอย่าง ค่า RE ของวิธีการแปลง Box และ Cox ไม่มีความแตกต่างกันมาก และค่า RE ของวิธี IRWLS มีแนวโน้มที่จะเพิ่มขึ้นเพียงเล็กน้อย ถ้าค่า σ^2 หรือ δ มีค่าสูงขึ้น ซึ่งค่า RE ของวิธี IRWLS จะมีค่ามากกว่าวิธีการแปลงของ Box และ Cox โดยที่ค่า RE ของวิธี IRWLS จะมีค่ามากกว่า 1 เพียงเล็กน้อยในตำแหน่งทศนิยมที่ 4 และค่า RE ของวิธีการแปลงของ Box และ Cox จะมีค่าน้อยกว่า 1 และหากพิจารณาที่ขนาดตัวอย่างต่างกันพบว่าค่า RE ของทั้งวิธีการแปลง Box และ Cox กับวิธี IRWLS จะไม่มีความแตกต่างกันมาก เมื่อ n มีค่าเพิ่มขึ้น ดังนั้นปัจจัยที่มีอิทธิพลต่อค่า RE ของแต่ละวิธีการประมาณคือ σ^2 และ δ

ตารางที่ 4.2.3 ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 2:1

σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
600	300	25	450	0	AMSE	54.11	58.17	54.11
					RE	1.0000	0.9302	1.0000
				0.1	AMSE	99.43*	105.42	99.43*
					RE	1.0000*	0.9432	1.0000*
				0.3	AMSE	342.62	356.67	342.52
					RE	1.0000	0.9606	1.0003
				0.5	AMSE	1198.88	1246.33	1197.74
					RE	1.0000	0.9619	1.0010
			900	0	AMSE	107.09	112.46	107.09
					RE	1.0000	0.9522	1.0000
				0.1	AMSE	201.46*	209.86	201.45*
					RE	1.0000*	0.9600	1.0000*
				0.3	AMSE	695.64	718.34	695.44
					RE	1.0000	0.9684	1.0003
				0.5	AMSE	2397.85	2487.60	2395.74
					RE	1.0000	0.9639	1.0009
			1,800	0	AMSE	215.92	224.81	215.92
					RE	1.0000	0.9605	1.0000
				0.1	AMSE	400.75	413.98	400.79
					RE	1.0000	0.9680	0.9999
				0.3	AMSE	1395.49	1439.86	1395.03
					RE	1.0000	0.9692	1.0003
				0.5	AMSE	4774.96	4912.14	4769.47
					RE	1.0000	0.9721	1.0012

***หมายเหตุ** ค่า AMSE ระหว่างวิธี OLS กับวิธี IRWLS มีค่าแตกต่างกันน้อยกว่า 10^{-2}

ตารางที่ 4.2.3 (ต่อ) ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 2:1

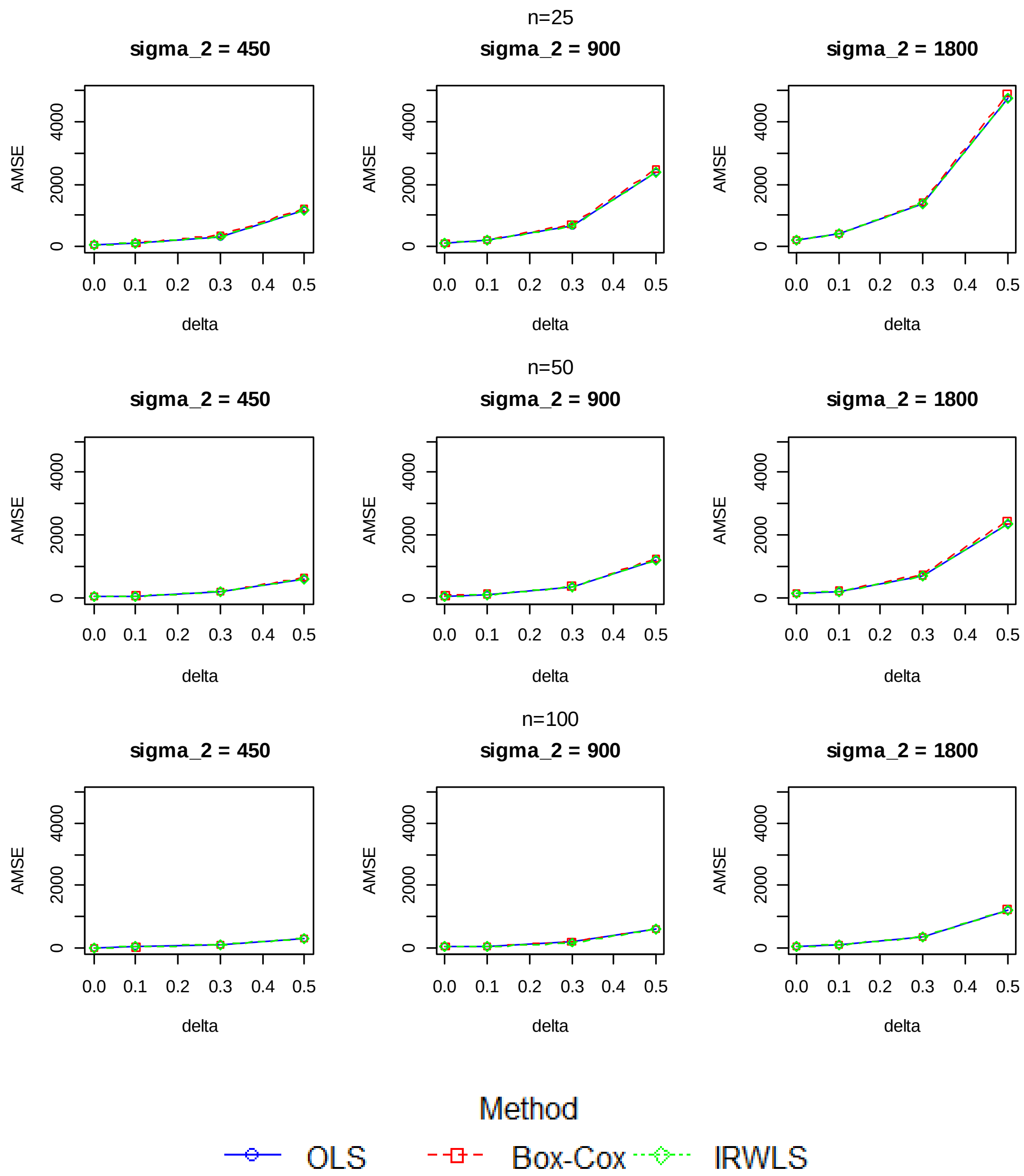
σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
600	300	50	450	0	AMSE	26.76	29.38	26.76
					RE	1.0000	0.9108	1.0000
				0.1	AMSE	50.68*	53.98	50.68*
					RE	1.0000*	0.9389	1.0000*
				0.3	AMSE	173.08	181.50	173.07
					RE	1.0000	0.9536	1.0001
				0.5	AMSE	602.95	628.87	602.47
					RE	1.0000	0.9588	1.0008
			900	0	AMSE	55.11	58.23	55.11
					RE	1.0000	0.9464	1.0000
				0.1	AMSE	99.55	104.48	99.56
					RE	1.0000	0.9528	0.9999
				0.3	AMSE	345.47	359.07	345.51
					RE	1.0000	0.9621	0.9999
				0.5	AMSE	1187.52	1239.37	1187.34
					RE	1.0000	0.9582	1.0002
			1,800	0	AMSE	107.59	112.81	107.59
					RE	1.0000	0.9537	1.0000
				0.1	AMSE	198.27	206.90	198.26
					RE	1.0000	0.9583	1.0001
				0.3	AMSE	686.17	715.02	685.93
					RE	1.0000	0.9597	1.0003
				0.5	AMSE	2367.44	2441.26	2365.91
					RE	1.0000	0.9698	1.0006

***หมายเหตุ** ค่า AMSE ระหว่างวิธี OLS กับวิธี IRWLS มีค่าแตกต่างกันน้อยกว่า 10^{-2}

ตารางที่ 4.2.3 (ต่อ) ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) เมื่อความคลาดเคลื่อนสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 2:1

σ_1^2	σ_2^2	n	σ^2	δ		OLS	Box-Cox	IRWLS
600	300	100	450	0	AMSE	**13.54**	14.85	**13.54**
					RE	**1.0000**	0.9118	**1.0000**
				0.1	AMSE	**25.06***	26.75	**25.05***
					RE	**1.0000***	0.9368	**1.0000***
				0.3	AMSE	**86.95**	90.68	86.96
					RE	**1.0000**	0.9589	0.9999
				0.5	AMSE	303.35	317.21	**302.88**
					RE	1.0000	0.9563	**1.0016**
			900	0	AMSE	**27.45**	29.13	**27.45**
					RE	**1.0000**	0.9423	**1.0000**
				0.1	AMSE	**49.62***	52.06	**49.62***
					RE	**1.0000***	0.9531	**1.0000***
				0.3	AMSE	174.55	182.25	**174.46**
					RE	1.0000	0.9578	**1.0005**
				0.5	AMSE	593.81	621.73	**593.64**
					RE	1.0000	0.9551	**1.0003**
			1,800	0	AMSE	**54.41**	57.11	**54.41**
					RE	**1.0000**	0.9527	**1.0000**
				0.1	AMSE	**100.87***	105.02	**100.87***
					RE	**1.0000***	0.9605	**1.0000***
				0.3	AMSE	342.87	356.24	**342.72**
					RE	1.0000	0.9625	**1.0004**
				0.5	AMSE	**1194.63**	1231.37	1194.84
					RE	**1.0000**	0.9702	0.9998

***หมายเหตุ** ค่า AMSE ระหว่างวิธี OLS กับวิธี IRWLS มีค่าแตกต่างกันน้อยกว่า 10^{-2}



ภาพที่ 4.2.3 การเปรียบเทียบวิธีการประมาณค่าตัวแปรตามระหว่างค่า AMSE กับค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความแปรปรวนความคลาดเคลื่อน ในรูปแบบเลขชี้กำลัง (δ) เมื่อความคลาดเคลื่อนสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 2:1

จากตารางที่ 4.2.3 ซึ่งแสดงผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 2:1 พบว่า โดยส่วนใหญ่เมื่อเปรียบเทียบแต่วิธีการประมาณค่าตัวแปรตาม พบว่า ที่ทุกระดับขนาดตัวอย่าง และทุกค่าพารามิเตอร์ σ^2 ค่า AMSE ของแต่ละวิธีเมื่อเรียงจากน้อยที่สุดไปหามากที่สุด คือ วิธี IRWLS วิธี OLS และ วิธีการแปลง Box และ Cox ตามลำดับ แต่ยกเว้นกรณีที่ δ เท่ากับศูนย์ ค่า AMSE ระหว่างวิธี OLS กับ IRWLS จะให้ค่า AMSE เท่ากัน และวิธีการแปลง Box และ Cox จะมีค่า AMSE มากที่สุด ซึ่งแสดงให้เห็นว่าวิธี IRWLS เป็นวิธีที่ดีที่สุด เมื่อ δ ไม่เท่ากับศูนย์

จากภาพที่ 4.2.3 จะเห็นว่า แนวโน้มค่า AMSE ของแต่ละวิธีการประมาณค่าตัวแปรตาม พบว่า เมื่อ δ เพิ่มขึ้นหรือ σ^2 เพิ่มขึ้น จะทำให้ค่า AMSE เพิ่มขึ้น ซึ่งการเพิ่มขึ้นของค่า σ^2 จะแปรผันตามกับค่า AMSE กล่าวคือ เมื่อ σ^2 เพิ่มขึ้นจำนวน 1 เท่า จะทำให้ค่า AMSE เพิ่มขึ้นโดยประมาณ 1 เท่า ในขณะที่เมื่อขนาดตัวอย่าง (n) เพิ่มขึ้นจะทำให้ค่า AMSE มีค่าน้อยลง โดยที่เมื่อเพิ่ม n จำนวน 1 เท่าจากเดิมจะทำให้ค่า AMSE ลดลงโดยประมาณ 0.5 เท่าของค่า AMSE จากขนาดตัวอย่างเดิม เพราะขนาดตัวอย่างที่เพิ่มขึ้น จะทำให้ค่าความคลาดเคลื่อนในการประมาณหรือพยากรณ์ลดลง

เมื่อพิจารณาค่า RE ซึ่งเป็นอัตราส่วนระหว่างค่า AMSE ของวิธี OLS กับค่า AMSE ของวิธีการประมาณค่าตัวแปรตามแต่ละวิธี พบว่า สำหรับทุกระดับขนาดตัวอย่าง ค่า RE ของวิธีการแปลง Box และ Cox ไม่มีความแตกต่างกันมาก และค่า RE ของวิธี IRWLS มีแนวโน้มที่จะเพิ่มขึ้นเพียงเล็กน้อย ถ้าค่า σ^2 หรือ δ มีค่าสูงขึ้น ซึ่งค่า RE ของวิธี IRWLS จะมีค่ามากกว่าวิธีการแปลงของ Box และ Cox โดยที่ค่า RE ของวิธี IRWLS จะมีค่าใกล้เคียงประมาณ 1 ในตำแหน่งทศนิยมที่ 4 และค่า RE ของวิธีการแปลงของ Box และ Cox จะมีค่าน้อยกว่า 1 และหากพิจารณาที่ขนาดตัวอย่างต่างกันพบว่าค่า RE ของทั้งวิธีการแปลง Box และ Cox กับวิธี IRWLS จะไม่มีความแตกต่างกันมาก เมื่อ n มีค่าเพิ่มขึ้น ดังนั้นปัจจัยที่มีอิทธิพลต่อค่า RE ของแต่ละวิธีการประมาณคือ σ^2 และ δ

การสรุปผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') (ในหัวข้อ 4.2)

เมื่อพิจารณาผลการวิจัยจากตารางที่ 4.2.1 ถึง 4.2.3 ซึ่งเป็นการแสดงผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') พบว่า ที่ δ เท่ากับ 0 หรือ 0.1 และ σ^2 เท่ากับ 450, 900 หรือ 1,800 ในทุกขนาดตัวอย่าง ค่า AMSE มีแนวโน้มที่จะเพิ่มขึ้นของแต่ละวิธีการประมาณหรือการพยากรณ์ตัวแปรตาม เมื่ออัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เปลี่ยนแปลงเป็น 1:2 2:1 และ 1:1 ตามลำดับ ในขณะที่ δ เท่ากับ 0.3 หรือ 0.5 และ σ^2 เท่ากับ 900 หรือ 1,800 ในทุกขนาดตัวอย่าง ค่า AMSE จะมีค่ามากที่สุด เมื่ออัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เปลี่ยนแปลงเป็น 1:2

เมื่อเปรียบเทียบค่า AMSE แต่ละวิธีการประมาณค่าตัวแปรตาม ทำให้พบว่าเมื่อเรียงค่า AMSE ของแต่ละวิธีจากน้อยที่สุดไปยังมากที่สุดคือ วิธี IRWLS วิธี OLS และ วิธีการแปลง Box และ Cox ตามลำดับ ยกเว้นกรณีที่ δ เท่ากับ ศูนย์ ที่ค่า AMSE ของวิธี OLS กับ IRWLS มีค่าเท่ากัน และค่า AMSE ของวิธีการแปลง Box และ Cox มีค่ามากที่สุด ไม่ว่าอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 จะมีค่าเท่าใด

เมื่อพิจารณาค่า RE ซึ่งเป็นอัตราส่วนระหว่างค่า AMSE ของวิธี OLS กับค่า AMSE ของวิธีการประมาณค่าตัวแปรตามวิธีอื่นๆ พบว่า ที่ระดับขนาดตัวอย่างเดียวกันทุกขนาดตัวอย่าง และค่า RE ของวิธีการแปลง Box และ Cox ไม่มีความแตกต่างกันมาก และค่า RE ของวิธี IRWLS มีแนวโน้มที่จะเพิ่มขึ้นเพียงเล็กน้อย เมื่อค่า σ^2 หรือ δ มีค่าสูงขึ้น ซึ่งค่า RE ของวิธี IRWLS จะมีค่ามากกว่า 1 เพียงเล็กน้อยตรงทศนิยมตำแหน่งที่ 4 และมากกว่า ค่าของวิธีการแปลงของ Box และ Cox ที่น้อยกว่า 1 และหากพิจารณาที่ขนาดตัวอย่างต่างกัน และ อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 ต่างกันพบว่าค่า RE ของทั้งวิธีการแปลง Box และ Cox กับวิธี IRWLS จะไม่มีความแตกต่างกันมาก เมื่อ n มีค่าเพิ่มขึ้น ดังนั้นปัจจัยที่มีอิทธิพลต่อค่า RE ของแต่ละวิธีการประมาณคือ σ^2 และ δ

สรุปปัจจัยที่มีผลการเปลี่ยนแปลงค่า AMSE (ในหัวข้อ 4.2)

1. ขนาดตัวอย่าง (n)

เมื่อขนาดตัวอย่างเพิ่มขึ้น จะทำให้ค่า AMSE น้อยลง ในอัตราส่วนที่แปรผกผัน เนื่องจากการเพิ่มขึ้นของขนาดตัวอย่างจะทำให้ความคลาดเคลื่อนในการประมาณหรือพยากรณ์ตัวแปรตามน้อยลง

2. ความแปรปรวนของความคลาดเคลื่อน

ซึ่งมีพารามิเตอร์ที่เกี่ยวข้องในงานวิจัยนี้ ดังต่อไปนี้

2.1 พารามิเตอร์ σ^2 ที่บอกระดับขนาดของความแปรปรวนของความคลาดเคลื่อน

เมื่อค่า σ^2 เพิ่มขึ้น จะทำให้ค่า AMSE เพิ่มขึ้น ในอัตราส่วนที่แปรผันตาม เนื่องจากเมื่อเพิ่มค่า σ^2 จะทำให้ความแปรปรวนของความคลาดเคลื่อนสูงขึ้น ส่งผลให้การประมาณค่าตัวแปรตามมีความคลาดเคลื่อนมากขึ้น

2.2 พารามิเตอร์ δ เป็นค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความคลาดเคลื่อนในรูปแบบเลขชี้กำลัง

เมื่อค่า δ เพิ่มขึ้น จะทำให้ค่า AMSE เพิ่มขึ้น เนื่องจากการเพิ่มค่า δ จะทำให้ความแปรปรวนของความคลาดเคลื่อนสูงขึ้น ส่งผลให้การประมาณค่าตัวแปรตามมีความคลาดเคลื่อนมากขึ้น

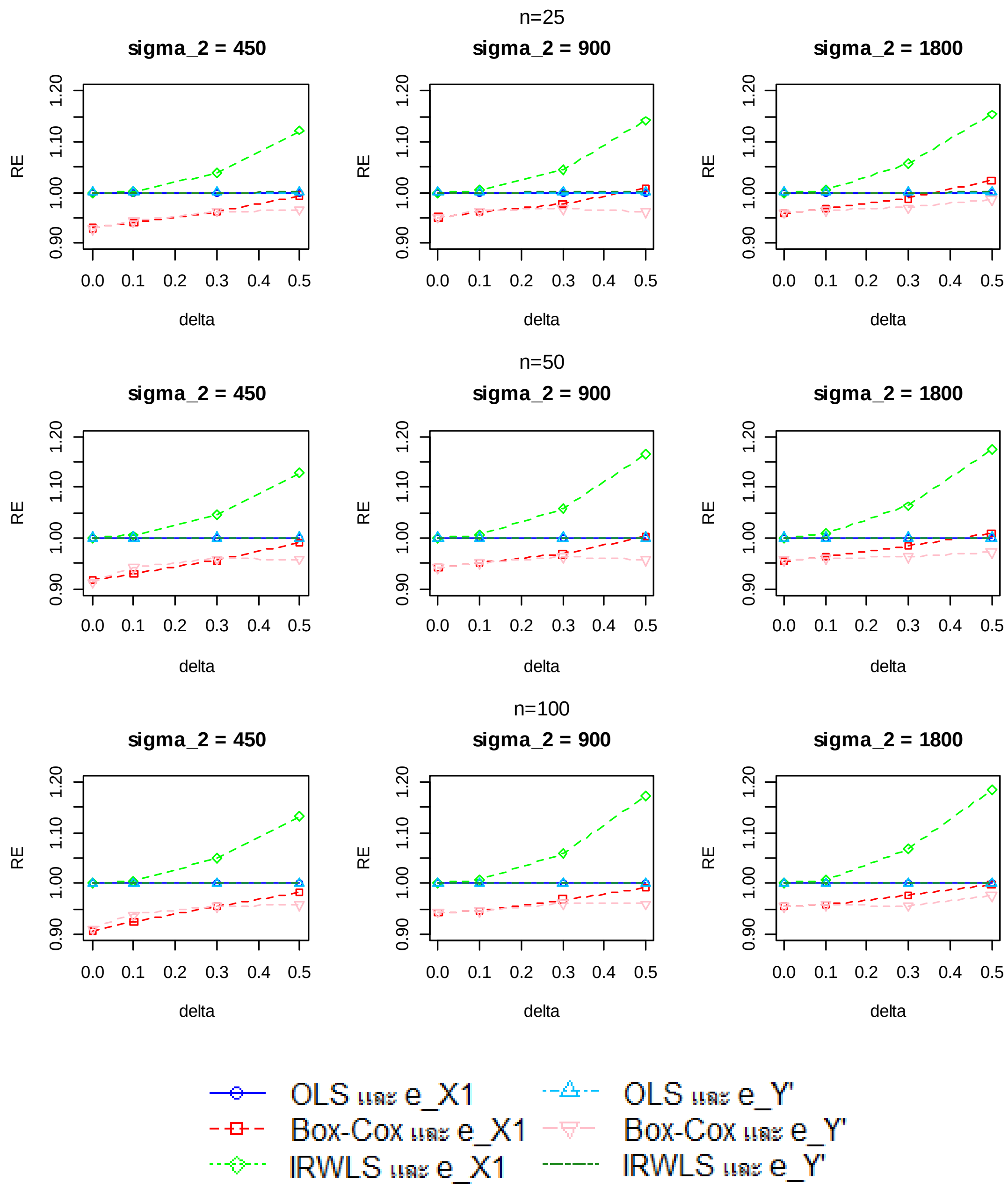
3. อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2

อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 ที่ต่างกัน ค่า AMSE มีแนวโน้มเปลี่ยนแปลงตามช่วงระดับความแปรปรวนของความคลาดเคลื่อนที่ต่างกัน ซึ่งเป็นผลจากการเปลี่ยนแปลงพารามิเตอร์ δ และ σ^2 โดยค่า AMSE จะมีค่าโดยส่วนใหญ่มากที่สุด ค่อนข้างเห็นได้ชัด เมื่ออัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2 และ δ เท่ากับ 0.3 หรือ 0.5 และ σ^2 เท่ากับ 900 หรือ 1,800 ในขณะที่ δ เท่ากับ 0 หรือ 0.1 ค่า AMSE จะมีค่ามากที่สุด เมื่ออัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1

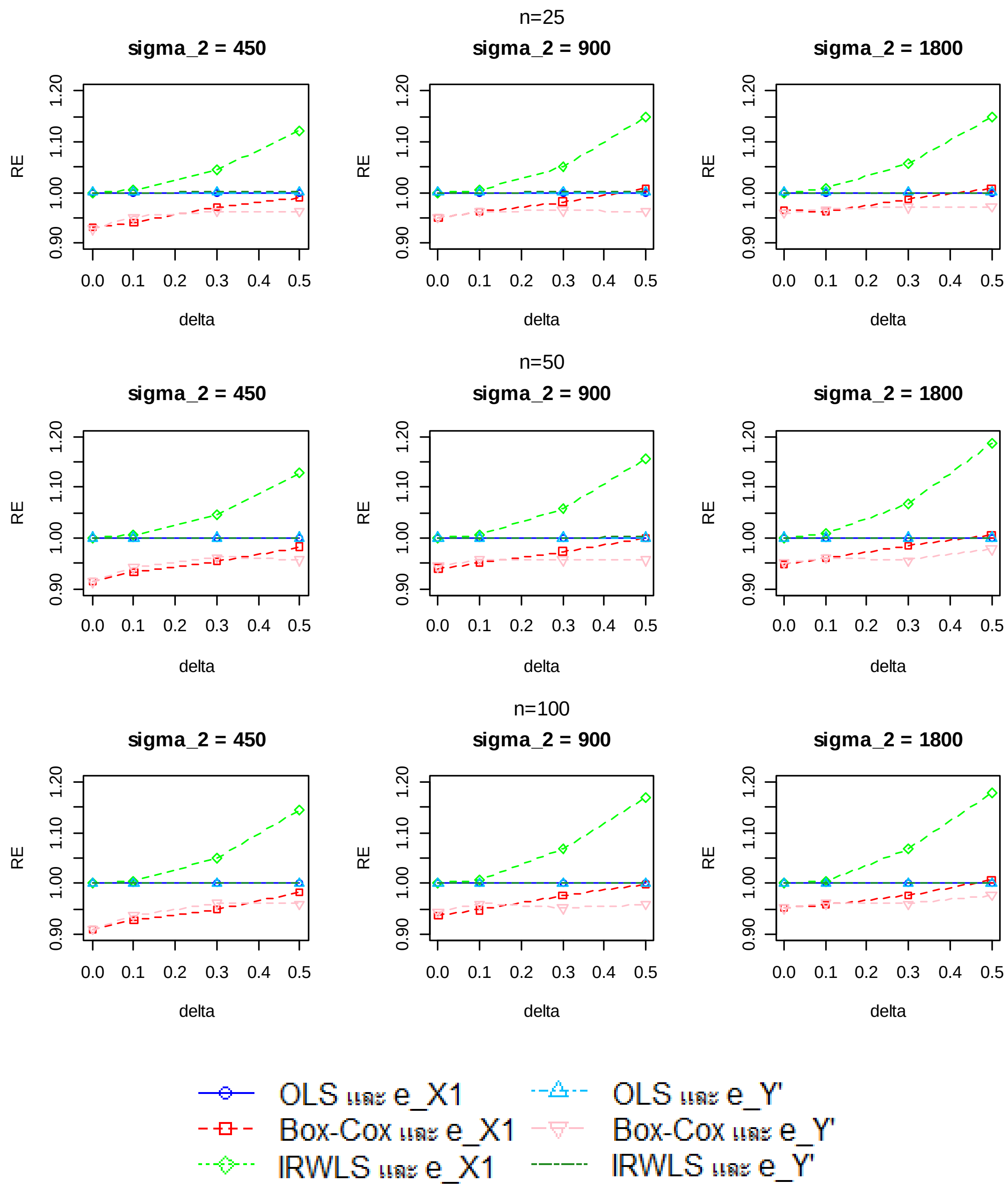
4.3 การเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ระหว่างที่สัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) ค่าจริงของตัวแปรตาม (Y')

ผู้วิจัยทำการเปรียบเทียบในกรณีที่ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เนื่องจากสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) กับ ค่าจริงของตัวแปรตาม (Y') อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2, 1:1 และ 2:1 ขนาดตัวอย่าง เท่ากับ 25, 50 และ 100 ระดับความแปรปรวนของความคลาดเคลื่อน ในส่วนพารามิเตอร์ σ^2 ที่ เท่ากับ 450, 900 และ 1,800 และ ส่วนเทอมพารามิเตอร์ δ เท่ากับ 0, 0.1, 0.3 และ 0.5 ซึ่งการวิจัยในส่วนนี้จะนำผลการวิจัยที่แสดงในตารางที่ 4.1.1 – 4.1.3 เปรียบเทียบกับ 4.2.1 – 4.2.3 ตามลำดับ

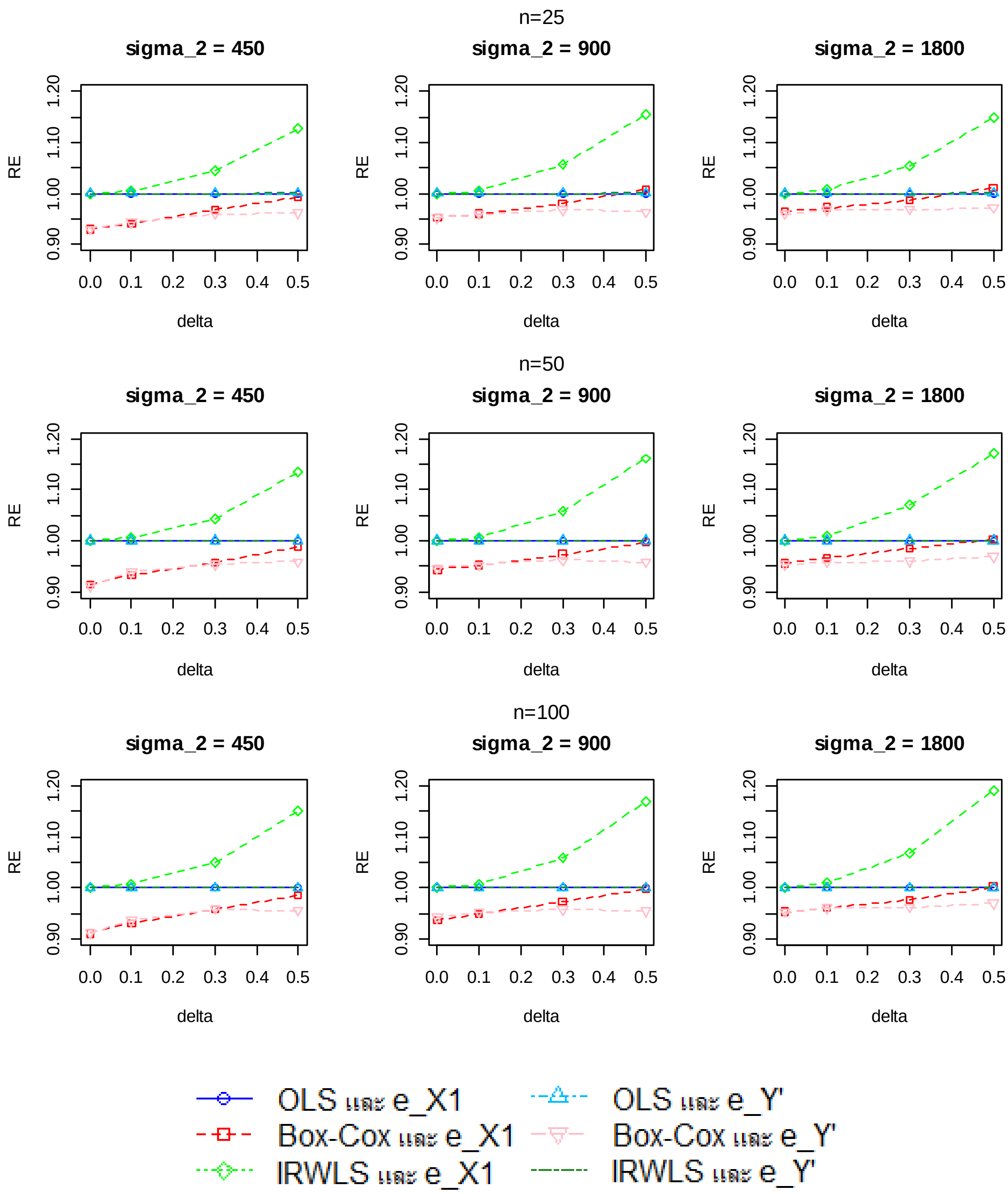
ตารางที่ 4.1.1	เปรียบเทียบกับ	ตารางที่ 4.2.1
ตารางที่ 4.1.2	เปรียบเทียบกับ	ตารางที่ 4.2.2
ตารางที่ 4.1.3	เปรียบเทียบกับ	ตารางที่ 4.2.3



ภาพที่ 4.3.1 การเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม ระหว่างค่า RE กับค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความแปรปรวนความคลาดเคลื่อน ในรูปแบบเลขชี้กำลัง (δ) เมื่อความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และ ค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2



ภาพที่ 4.3.2 การเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม ระหว่างค่า RE กับค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความแปรปรวนความคลาดเคลื่อน ในรูปแบบเลขชี้กำลัง (δ) เมื่อความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และ ค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1



ภาพที่ 4.3.3 การเปรียบเทียบวิธีการประมาณค่าตัวแปรตาม ระหว่างค่า RE กับค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความแปรปรวนความคลาดเคลื่อน ในรูปแบบเลขชี้กำลัง (δ) เมื่อความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) และ ค่าจริงของตัวแปรตาม (Y') และมีอัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 2:1

เมื่อพิจารณาผลการวิจัยที่แสดงในตารางที่ 4.1.1 ถึงตารางที่ 4.1.3 ซึ่งเป็นผลเปรียบเทียบวิธีการประมาณค่าตัวแปรตามที่ความคลาดเคลื่อนมีความแปรปรวนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) เปรียบเทียบกับ ผลการวิจัยที่แสดงในตารางที่ 4.2.1 ถึงตารางที่ 4.2.3 ซึ่งเป็นผลการเปรียบเทียบวิธีการประมาณค่าตัวแปรตามที่ความคลาดเคลื่อนมีความแปรปรวนสัมพันธ์กับค่าจริงของตัวแปรตาม (Y') พบว่า ที่อัตราส่วนความแปรปรวนระหว่างตัวแปรอิสระ 2 ตัวในระดับอัตราส่วนเดียวกัน ขนาดตัวอย่าง และ พารามิเตอร์ค่า σ^2 และ δ เดียวกัน ค่า RE ของวิธีการแปลง Box และ Cox และวิธี IRWLS ที่ความแปรปรวนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) มีแนวโน้มมีค่ามากกว่า ค่า RE ที่ความแปรปรวนสัมพันธ์กับค่าจริงของตัวแปรตาม (Y')

จากภาพที่ 4.3.1 ถึงภาพที่ 4.3.3 ซึ่งเป็นภาพแสดงค่าเปรียบเทียบค่า RE ของแต่ละวิธีเมื่อความแปรปรวนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง (X_1) หรือ ค่าจริงของตัวแปรตาม (Y') ในแต่ละค่าอัตราส่วนความแปรปรวนของตัวแปรอิสระระหว่าง X_1 และ X_2 เดียวกัน ขนาดตัวอย่างเดียวกันและค่า σ^2 เดียวกันแต่ δ ต่างกัน พบว่า ค่า RE ของวิธีการแปลง Box และ Cox ที่ได้ความสัมพันธ์ของความแปรปรวนความคลาดเคลื่อนระหว่าง 2 รูปแบบ ในช่วง δ เท่ากับศูนย์และหนึ่ง มีค่าใกล้เคียงกัน แต่เมื่อเพิ่ม δ เท่ากับช่วง 0.3 ถึง 0.5 ค่า RE ของวิธีการแปลง Box และ Cox ที่ความคลาดเคลื่อนสัมพันธ์กับ X_1 ซึ่งมีค่าน้อยกว่า 0.1 มีค่ามากกว่า ค่า RE ของวิธีการแปลง Box และ Cox ที่ความคลาดเคลื่อนสัมพันธ์กับ Y' และที่ระดับ Y' และที่ δ เท่ากับ 0.5 ค่า RE ของวิธีการแปลง Box และ Cox ที่ความคลาดเคลื่อนสัมพันธ์กับ X_1 จะมีค่ามากกว่า 1 หรือมีค่าใกล้เคียงเท่ากับ 1 มากๆ แสดงว่า ถ้าความแปรปรวนของความคลาดเคลื่อนมีค่าสูงมากในระดับหนึ่ง การประมาณค่าตัวแปรตามด้วยวิธีการแปลง Box และ Cox จะมีประสิทธิภาพมากกว่าหรือใกล้เคียงเทียบเท่ากับการประมาณค่าตัวแปรตามด้วยวิธี OLS

เมื่อพิจารณาถึงอัตราส่วนความแปรปรวนของตัวแปรอิสระระหว่าง X_1 และ X_2 เดียวกันขนาดตัวอย่างเดียวกันแต่ σ^2 มีค่าต่างกัน พบว่าค่า RE ของแต่ละวิธีมีแนวโน้มเพิ่มขึ้น เมื่อค่า σ^2 มีค่าเพิ่มขึ้น และเมื่อ δ เพิ่มขึ้น ค่า RE ของแต่ละวิธีจะมีแนวโน้มเพิ่มขึ้นเช่นเดียวกัน โดยเฉพาะค่า RE ของวิธี IRWLS ที่ความแปรปรวนสัมพันธ์กับ X_1 จะมีค่าเพิ่มสูงขึ้นและมีค่ามากกว่า 1 เพิ่มขึ้นอย่างเห็นได้ชัด แสดงว่า ประสิทธิภาพของวิธี IRWLS จะมากกว่าประสิทธิภาพวิธี OLS เพิ่มขึ้นเรื่อยๆ เมื่อ δ เพิ่มขึ้น

เมื่อพิจารณาถึงขนาดตัวอย่าง พารามิเตอร์ค่า σ^2 และ δ มีค่าเดียวกัน แต่อัตราส่วนความแปรปรวนของตัวแปรอิสระระหว่าง X_1 และ X_2 ต่างกัน พบว่าค่า RE ของแต่ละวิธีการประมาณตัวแปรตามเมื่อความคลาดเคลื่อนสัมพันธ์กับ X_1 หรือ Y' จะมีค่าแตกต่างกันไม่มาก และเมื่อ อัตราส่วนความแปรปรวนของตัวแปรอิสระระหว่าง X_1 และ X_2 เดียวกัน และ พารามิเตอร์ค่า σ^2 และ δ มีค่าเดียวกัน แต่ขนาดตัวอย่างต่างกัน พบว่า ค่า RE ของแต่ละวิธีการประมาณตัวแปรตาม ไม่เห็นถึงความแตกต่างได้อย่างเห็นชัด เช่นเดียวกัน

จากผลการวิจัยในส่วนที่ 4.3 สามารถสรุปได้ดังนี้

เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ มีรูปแบบความสัมพันธ์ในตัวแบบการถดถอยเชิงเส้นพหุ แตกต่างกัน กล่าวคือ สัมพันธ์กับ ตัวแปรอิสระตัวที่หนึ่ง หรือ ค่าจริงของตัวแปรตาม จะมีผลทำให้ค่า RE ของแต่ละวิธี ที่ความคลาดเคลื่อนสัมพันธ์กับ X_1 มีค่าน้อยกว่า ค่า RE ที่ความคลาดเคลื่อนสัมพันธ์กับค่าจริงของตัวแปรตาม ค่า δ และ σ^2 เพิ่มขึ้น จะทำให้ค่า RE ของแต่ละวิธี เพิ่มขึ้น ในขณะที่ขนาดตัวอย่างหรืออัตราส่วนความแปรปรวนของตัวแปรอิสระระหว่าง X_1 และ X_2 มีค่าเปลี่ยนแปลง จะทำให้ค่า RE ของแต่ละวิธีไม่เปลี่ยนแปลงจากเดิมมาก

วิธีที่มีค่า RE มากที่สุดทุกกรณีได้เห็นเด่นชัดคือ วิธี IRWLS ที่ความคลาดเคลื่อนมีความสัมพันธ์กับ X_1 ส่วนวิธีที่มีค่า RE น้อยที่สุด คือ วิธีการแปลงของ Box และ Cox หรือจะเป็นวิธี OLS ขึ้นอยู่กับค่าความแปรปรวนของความคลาดเคลื่อนว่ามีค่ามากหรือมีค่าน้อย และ รูปแบบความสัมพันธ์ของความคลาดเคลื่อน โดยที่ ถ้าความแปรปรวนของความคลาดเคลื่อนว่ามีค่าสูง และ ความคลาดเคลื่อนมีความสัมพันธ์กับ X_1 วิธีที่มีค่า RE น้อยที่สุด คือ วิธี OLS

โดยสรุปปัจจัยต่างๆที่มีผลต่อค่า AMSE มีดังนี้ (จากผลการวิจัยในส่วนที่ 4.1 ถึง 4.3)

1. ขนาดตัวอย่าง (n)

เมื่อขนาดตัวอย่างเพิ่มขึ้น จะทำให้ค่า AMSE น้อยลง ในอัตราส่วนที่แปรผกผัน เนื่องจากการเพิ่มขึ้นของขนาดตัวอย่างจะทำให้ความคลาดเคลื่อนในการประมาณหรือพยากรณ์ตัวแปรตามน้อยลง

2. ความแปรปรวนของความคลาดเคลื่อน

ซึ่งมีพารามิเตอร์ที่เกี่ยวข้องในงานวิจัยนี้ ดังต่อไปนี้

2.1 พารามิเตอร์ σ^2 ที่บอกระดับขนาดของความแปรปรวนของความคลาดเคลื่อน

เมื่อค่า σ^2 เพิ่มขึ้น จะทำให้ค่า AMSE เพิ่มขึ้น ในอัตราส่วนที่แปรผันตาม เนื่องจากเมื่อเพิ่มค่า σ^2 จะทำให้ความแปรปรวนของความคลาดเคลื่อนสูงขึ้น ส่งผลให้การประมาณค่าตัวแปรตามมีความคลาดเคลื่อนมากขึ้น

2.2 พารามิเตอร์ δ ที่ใช้เป็นบอกค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความคลาดเคลื่อนในรูปแบบเลขชี้กำลัง

เมื่อค่า δ เพิ่มขึ้น จะทำให้ค่า AMSE เพิ่มขึ้น เนื่องจากการเพิ่มค่า δ จะทำให้ความแปรปรวนของความคลาดเคลื่อนสูงขึ้น ส่งผลให้การประมาณค่าตัวแปรตามมีความคลาดเคลื่อนมากขึ้น

3. อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 และ รูปแบบความแปรปรวนของความคลาดเคลื่อนที่ไม่คงที่ (สัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง หรือ ค่าจริงของตัวแปรตาม)

อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 และ รูปแบบความแปรปรวนของความคลาดเคลื่อนที่ไม่คงที่ที่แตกต่างกัน ส่งผลต่อค่า AMSE ดังนี้

ก. เมื่อความแปรปรวนของความคลาดเคลื่อนที่ไม่คงที่ เนื่องจากสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง และ δ ไม่เท่ากับศูนย์ การเพิ่มสัดส่วนของความแปรปรวนของตัวแปรอิสระ

ตัวที่หนึ่งให้มากขึ้น จะมีผลทำให้ค่า AMSE เพิ่มขึ้น เนื่องจากการเพิ่มความแปรปรวนของตัวแปรอิสระตัวที่หนึ่ง จะมีผลทำให้ความแปรปรวนของความคลาดเคลื่อนในตัวแบบการถดถอยเชิงเส้นพหุมากขึ้น ส่งผลการประมาณค่าตัวแปรตามมีความคลาดเคลื่อนมากขึ้น แต่เพิ่มมากขึ้นในระดับที่ไม่เห็นชัด

ข. เมื่อความแปรปรวนของความคลาดเคลื่อนที่ไม่คงที่ เนื่องจากสัมพันธ์กับค่าจริงของตัวแปรตาม อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 ที่ต่างกัน ค่า AMSE มีแนวโน้มเปลี่ยนแปลงตามช่วงระดับความแปรปรวนของความคลาดเคลื่อนที่ต่างกัน ซึ่งเป็นผลจากการเปลี่ยนแปลงพารามิเตอร์ δ และ σ^2 โดยค่า AMSE จะมีค่าโดยส่วนใหญ่มากที่สุด เมื่ออัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2 และ δ เท่ากับ 0.3 หรือ 0.5 และ σ^2 เท่ากับ 900 หรือ 1,800 ในขณะที่ δ เท่ากับ 0 หรือ 0.1 ค่า AMSE จะมีค่ามากที่สุด เมื่ออัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1

บทที่ 5

สรุปผลการวิจัย และ ข้อเสนอแนะ

งานวิจัยนี้เป็นการศึกษาและเปรียบเทียบวิธีการพยากรณ์หรือประมาณค่าตัวแปรตาม ที่ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ ในตัวแบบการถดถอยเชิงเส้นพหุ โดยวิธีการพยากรณ์หรือประมาณค่าตัวแปรตามมีจำนวนวิธีทั้งหมด 3 วิธี ได้แก่ วิธีกำลังสองน้อยที่สุด (Ordinary Least Square: OLS), วิธีการแปลงของ Box และ Cox (Box-Cox Transformation), และวิธี Iteratively Reweighted Least Square (IRWLS) เกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการพยากรณ์ตัวแปรตามจะพิจารณาจาก ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Average Mean Square Error: AMSE) ระหว่างค่าพยากรณ์กับค่าจริงของตัวแปรตาม และค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency: RE) โดยมีการจำลองสถานการณ์ทั้งหมด 216 สถานการณ์ ตามเงื่อนไขดังต่อไปนี้

1. ตัวแปรอิสระที่ศึกษามีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนทั้งหมด 3 แบบดังต่อไปนี้

อัตราส่วนของ $\sigma_1^2 : \sigma_2^2$	σ_1^2	σ_2^2
1:2	300	600
1:1	450	450
2:1	600	300

2. ความคลาดเคลื่อนมีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนที่มีรูปแบบทั้งหมด 2 รูปแบบ

   2.1 สัมพันธ์กับค่าของตัวแปรอิสระตัวที่หนึ่ง (X_1) ซึ่งมีรูปแบบความสัมพันธ์คือ

   $\mathrm{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2 \left|X_{i1}\right|^{\delta}$ โดยที่ $\sigma^2 > 0$, δ เป็นค่าพารามิเตอร์

   2.2 สัมพันธ์กับค่าจริงของตัวแปรตาม ซึ่งค่าจริงของตัวแปรตาม เกิดจากความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระ นั่นคือ

$$Y_i' = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} \quad : \ i = 1, 2, ..., n$$

ซึ่งมีรูปแบบความสัมพันธ์คือ

$$\mathrm{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2 \left| \beta_0 + \sum_{j=1}^{2} \beta_j X_{ij} \right|^{\delta} = \sigma^2 \left| Y_i' \right|^{\delta}$$

โดยที่ $\sigma^2 > 0$, δ เป็นค่าพารามิเตอร์

ความแปรปรวนของความคลาดเคลื่อนทั้ง 2 รูปแบบ จะมาจากการกำหนด $\underset{\sim}{\beta}$ เป็นพารามิเตอร์ที่มีค่าคงที่ โดยที่ $\beta_0 = 500$, $\beta_1 = 1$ และ $\beta_2 = 1$

ซึ่งโดยทั่วไปแล้ว β_0, β_1 และ β_2 จะเป็นพารามิเตอร์ที่ไม่ทราบค่า และในงานวิจัยนี้ผู้วิจัยได้กำหนด $\underset{\sim}{\beta}$ เป็นพารามิเตอร์ที่มีค่าคงที่ โดยที่ $\beta_0 = 500$, $\beta_1 = \beta_2 = 1$ ทั้งนี้เนื่องจาก ต้องการเปรียบเทียบแบบของตัวแปรอิสระที่มี รูปแบบความแปรปรวนต่างกัน ซึ่งเมื่อ β_1 หรือ β_2 เปลี่ยนไปจะทำให้ค่าตัวแปรตามที่ได้จากแบบของตัวแปรอิสระแต่ละแบบมีความแปรปรวนแตกต่างกันด้วย ฉะนั้นเพื่อให้ควบคุมค่าความแปรปรวนของค่าจริงของตัวแปรตามมีค่าเท่ากัน จึงกำหนดให้ β_1 และ β_2 มีค่าเท่ากับ 1 และ กำหนดให้ตัวแปรอิสระแต่ละตัวไม่มีความสัมพันธ์กัน นั่นคือกำหนดให้ค่าสหสัมพันธ์มีค่าเท่ากับศูนย์

3. ความแปรปรวนของความคลาดเคลื่อน

ซึ่งมีพารามิเตอร์ที่เกี่ยวข้องในงานวิจัยนี้ตามที่ผู้วิจัยกำหนดในขั้นตอนที่ 2 ดังต่อไปนี้

3.1 *พารามิเตอร์ δ เป็นค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความคลาดเคลื่อนในรูปแบบเลขชี้กำลัง*

โดยในงานวิจัยนี้ พารามิเตอร์ δ ที่ศึกษา มีทั้งหมด 4 ระดับ ได้แก่ 0, 0.1, 0.3 และ 0.5 ซึ่งค่า δ เท่ากับศูนย์ จะหมายถึง ความคลาดเคลื่อน เป็นอิสระจากตัวแปรอิสระ หรือ ค่าจริงของตัวแปรตาม

3.2 *พารามิเตอร์ σ^2 ซึ่งบอกระดับขนาดของความแปรปรวนของความคลาดเคลื่อน*

โดยในงานวิจัยนี้ พารามิเตอร์ σ^2 ที่ศึกษามีทั้งหมด 3 ระดับได้แก่ 450, 900 และ 1,800 ซึ่งมาจากทุกระดับอัตราส่วนของ σ^2 กับผลรวมของความ

แปรปรวนระหว่างตัวแปรอิสระทั้ง X_1 และ X_2 รวมกัน $(\sigma_1^2+\sigma_2^2)$ ที่มีค่าเท่ากับ 900 ทำให้ได้อัตราส่วนระหว่างค่า $\sigma^2:(\sigma_1^2+\sigma_2^2)$ ทั้งหมด 3 ระดับ ได้แก่ 0.5, 1.0 และ 2.0 ตามลำดับ

4. ค่าสังเกตของตัวแปรตามและตัวแปรอิสระมีความสัมพันธ์กันภายใต้การถดถอยเชิงเส้นพหุ (Multiple Linear Regression) นั่นคือ

$$Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\varepsilon_i=Y_i'+\varepsilon_i \qquad : \ i=1,2,...,n$$

โดยกำหนดให้ ตัวแปรอิสระแต่ละตัวไม่มีความสัมพันธ์กัน นั่นคือ ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระ เท่ากับศูนย์

5. ขนาดตัวอย่างที่ใช้ในการศึกษา มีทั้งหมด 3 ระดับ ได้แก่ 25, 50 และ 100

6. ในการจำลองสถานการณ์จะใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) โดยใช้โปรแกรม R และในแต่ละสถานการณ์จะทำการจำลองซ้ำอีกจำนวนทั้งหมด 5,000 รอบ

5.1 ผลการเปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

ในการเปรียบเทียบวิธีการประมาณค่าตัวแปรตามทั้ง 3 วิธีด้วยการพิจารณาจากค่า AMSE พบว่า เมื่อค่าความแปรปรวนของความคลาดเคลื่อน สัมพันธ์กับ

ก. ตัวแปรอิสระตัวที่หนึ่ง (X_1)

โดยส่วนใหญ่ เมื่อเรียงค่า AMSE ของแต่ละวิธีจากน้อยที่สุดไปยังมากที่สุดคือ วิธี IRWLS วิธี OLS และ วิธีการแปลง Box และ Cox ตามลำดับ แต่กรณีดังต่อไปนี้

$\sigma_1^2:\sigma_2^2=1:2$ หรือ $1:1$ $\quad n=25,50 \quad \sigma^2=900,1800$ และ $\delta=0.5$

$\sigma_1^2:\sigma_2^2=1:2$ หรือ $1:1$ $\quad n=100 \quad \sigma^2=1,800$ และ $\delta=0.5$

$\sigma_1^2:\sigma_2^2=2:1$ $\quad n=25 \quad \sigma^2=900,1800$ และ $\delta=0.5$

$\sigma_1^2:\sigma_2^2=2:1$ $\quad n=50,100 \quad \sigma^2=1,800$ และ $\delta=0.5$

วิธี IRWLS ยังเป็นวิธีที่มีค่า AMSE น้อยที่สุด แต่วิธีที่มีค่า AMSE มากที่สุดคือวิธี OLS

ข. ค่าจริงของตัวแปรตาม (Y)

เมื่อเรียงค่า AMSE ของแต่ละวิธีจากน้อยที่สุดไปยังมากที่สุด แนวโน้มที่ได้คือ วิธี IRWLS วิธี OLS และ วิธีการแปลง Box และ Cox ยกเว้นกรณีที่ δ เท่ากับศูนย์ ที่ค่า AMSE ของวิธี OLS กับ IRWLS มีค่าเท่ากัน

5.2 ปัจจัยต่าง ๆที่มีผลต่อค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

1. ขนาดตัวอย่าง (n)

   เมื่อขนาดตัวอย่าง (n) เพิ่มขึ้น จะทำให้ค่า AMSE น้อยลง ในอัตราส่วนที่แปรผกผัน เนื่องจากการเพิ่มขึ้นของขนาดตัวอย่างจะทำให้ความคลาดเคลื่อนในการประมาณหรือพยากรณ์ตัวแปรตามน้อยลง

2. ความแปรปรวนของความคลาดเคลื่อน

   ซึ่งมีพารามิเตอร์ที่เกี่ยวข้องในงานวิจัยนี้ ดังต่อไปนี้

   2.1 พารามิเตอร์ σ^2 ที่บอกระดับขนาดของความแปรปรวนของความคลาดเคลื่อน

   เมื่อค่า σ^2 เพิ่มขึ้น จะทำให้ค่า AMSE เพิ่มขึ้น ในอัตราส่วนที่แปรผันตาม เนื่องจากเมื่อเพิ่มค่า σ^2 จะทำให้ความแปรปรวนของความคลาดเคลื่อนสูงขึ้น ส่งผลให้การประมาณค่าตัวแปรตามมีความคลาดเคลื่อนมากขึ้น

   2.2 พารามิเตอร์ δ เป็นค่าสัมประสิทธิ์ที่ใช้ในการระบุความสัมพันธ์ของความคลาดเคลื่อนในรูปแบบเลขชี้กำลัง

   เมื่อค่า δ เพิ่มขึ้น จะทำให้ค่า AMSE เพิ่มขึ้น เนื่องจากการเพิ่มค่า δ จะทำให้ความแปรปรวนของความคลาดเคลื่อนสูงขึ้น ส่งผลให้การประมาณค่าตัวแปรตามมีความคลาดเคลื่อนมากขึ้น

3. อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 และ รูปแบบความแปรปรวนของความคลาดเคลื่อนที่ไม่คงที่ (สัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง หรือ ค่าจริงของตัวแปรตาม)

   อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 และ รูปแบบความแปรปรวนของความคลาดเคลื่อนที่ไม่คงที่ที่แตกต่างกัน ส่งผลต่อค่า AMSE ดังนี้

ก. เมื่อความแปรปรวนของความคลาดเคลื่อนที่ไม่คงที่ เนื่องจากสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง และ δ ไม่เท่ากับศูนย์ การเพิ่มสัดส่วนของความแปรปรวนของตัวแปรอิสระตัวที่หนึ่งให้มากขึ้น จะมีผลทำให้ค่า AMSE เพิ่มขึ้น เนื่องจากการเพิ่มความแปรปรวนของตัวแปรอิสระตัวที่หนึ่ง จะมีผลทำให้ความแปรปรวนของความคลาดเคลื่อนในตัวแบบถดถอยเชิงเส้นพหุมากขึ้น ส่งผลการประมาณค่าตัวแปรตามมีความคลาดเคลื่อนมากขึ้น แต่เพิ่มมากขึ้นในระดับน้อย

ข. เมื่อความแปรปรวนของความคลาดเคลื่อนที่ไม่คงที่ เนื่องจากสัมพันธ์กับค่าจริงของตัวแปรตาม อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 ที่ต่างกัน ค่า AMSE มีแนวโน้มเปลี่ยนแปลงตามช่วงระดับความแปรปรวนของความคลาดเคลื่อนที่ต่างกัน ซึ่งเป็นผลจากการเปลี่ยนแปลงพารามิเตอร์ δ และ σ^2 โดยค่า AMSE จะมีค่าโดยส่วนใหญ่มากที่สุด เมื่ออัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:2 และ δ เท่ากับ 0.3 หรือ 0.5 และ σ^2 เท่ากับ 900 หรือ 1,800 ในขณะที่ δ เท่ากับ 0 หรือ 0.1 ค่า AMSE จะมีค่ามากที่สุด เมื่ออัตราส่วนความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 เท่ากับ 1:1

5.3 การอภิปรายผลการเปรียบเทียบที่ได้ของแต่ละวิธีการประมาณค่าตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่

จากผลการวิจัยในครั้งนี้วิธี OLS เป็นวิธีการประมาณค่าตัวแปรตามที่มีความเหมาะสมน้อยกว่าวิธี IRWLS เกือบทุกกรณี ยกเว้นกรณีที่ค่า δ เท่ากับศูนย์ ค่า AMSE ของวิธี OLS จะให้ผลที่เหมือนกับวิธี IRWLS ซึ่งสอดคล้องกับความเป็นจริงเนื่องจากรูปแบบความแปรปรวนของความคลาดเคลื่อนที่ผู้วิจัยกำหนด หากค่า δ เท่ากับศูนย์จะทำให้เทอมในส่วนที่ไม่เป็นค่าคงที่หรือเทอมถ่วงน้ำหนักของค่า σ^2 มีค่าเท่ากับ 1 ซึ่งแสดงว่า ความคลาดเคลื่อนมีความแปรปรวนคงที่ นั่นหมายความว่า ค่าตัวแปรอิสระตัวที่หนึ่งหรือค่าจริงของตัวแปรตามจะไม่มีผลหรือสัมพันธ์กับความแปรปรวนของความคลาดเคลื่อน ดังนั้นทำให้สัมประสิทธิ์การถดถอยที่ได้ในระหว่างวิธี OLS กับวิธี IRWLS มีค่าเหมือนกัน จึงทำให้ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Squared Error: MSE) ของทั้งสองวิธีมีค่าเท่ากันด้วย ทำให้หลังจากการทำซ้ำทั้งหมด 5,000 รอบ ค่า AMSE

ของวิธี OLS จึงเท่ากับค่า AMSE ของวิธี IRWLS การที่วิธี IRWLS มีความเหมาะสมมากกว่าวิธี OLS เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ สอดคล้องตามความเป็นจริง เนื่องจากวิธี OLS จะมีความเหมาะสมในการใช้ประมาณค่าก็ต่อเมื่อข้อตกลงของตัวแบบการถดถอยเชิงเส้นพหุในทุกข้อเป็นจริง ในขณะที่วิธี IRWLS จะเป็นวิธีที่ fit กับตัวแบบการถดถอยเชิงเส้นพหุที่ความคลาดเคลื่อนมีความแปรปรวนมีค่าไม่คงที่ โดยการใช้ค่าถ่วงน้ำหนักที่เหมาะสมที่สุดของข้อมูลจากการวนซ้ำ iteration เพื่อปรับเปลี่ยนค่าสัมประสิทธิ์การถดถอยของตัวแบบที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยที่สุดจนค่าสัมประสิทธิ์การถดถอยที่ทำในรอบใหม่มีความแตกต่างกับของรอบก่อนหน้านี้น้อยมากจนแน่ใจว่า ค่าสัมประสิทธิ์การถดถอยที่ได้เหมาะสมที่สุดและลู่เข้า

ในขณะที่เมื่อเปรียบเทียบวิธี OLS กับวิธีการแปลง Box และ Cox ด้วยค่า AMSE พบว่า ค่า AMSE ของวิธีการแปลง Box และ Cox มีค่ามากกว่า วิธี OLS ในกรณีที่ความแปรปรวนของความคลาดเคลื่อนสัมพันธ์กับค่าจริงของตัวแปรตาม ในทุกระดับพารามิเตอร์อื่นที่ศึกษา และในกรณีที่ความแปรปรวนของความคลาดเคลื่อนสัมพันธ์กับตัวแปรอิสระตัวที่หนึ่ง โดยส่วนใหญ่วิธีการแปลง Box และ Cox จะมีความเหมาะสมน้อยกว่าวิธี OLS แต่จะเริ่มมีความเหมาะสมมากกว่าวิธี OLS เมื่อความแปรปรวนของความคลาดเคลื่อนมีค่ามาก กล่าวคือ พารามิเตอร์ σ^2 และ δ มีค่าเท่ากับ 1,800 และ 0.5 ตามลำดับ ที่ไม่ว่าขนาดตัวอย่างและอัตราส่วนทุกระดับความแปรปรวนตัวแปรอิสระระหว่าง X_1 และ X_2 จะมีค่าเท่าใด เพราะ วิธีการแปลงของ Box และ Cox สามารถแก้ปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ได้ในบางสถานการณ์หรือบางกรณี ซึ่งสอดคล้องกับผลงานวิจัยของพรพล คงอิ่ม (พรพล คงอิ่ม, 2548)

และเมื่อเปรียบเทียบวิธีการแปลงของ Box และ Cox กับวิธี IRWLS ด้วยค่า AMSE พบว่าทุกสถานการณ์หรือทุกกรณีที่ศึกษา ค่า AMSE ของวิธีการแปลงของ Box และ Cox มีค่ามากกว่าวิธี IRWLS ซึ่งอาจจะเป็นเพราะว่า ค่าประมาณ power ที่ได้มาจากเทคนิคความน่าจะเป็นสูงสุด (Maximum Likelihood) ยังให้ค่าที่ไม่ละเอียดมากพอเมื่อเทียบกับกับวิธี IRWLS อีกทั้งวิธีการแปลงของ Box และ Cox เป็นวิธีการแปลงค่าตัวแปรตาม ซึ่งไม่ได้แก้ไขปัญหาความคลาดเคลื่อนมีค่าความแปรปรวนไม่คงที่ จากสาเหตุที่ค่าความคลาดเคลื่อนโดยตรง ทำให้ค่าพยากรณ์ตัวแปรตามที่ได้มีความถูกต้องน้อยกว่าที่ควรจะเป็น ต่างจากวิธี IRWLS เป็นวิธีที่สามารถ fit กับตัวแบบ

การถดถอยเชิงเส้นพหุที่ความคลาดเคลื่อนมีความแปรปรวนมีค่าไม่คงที่ โดยการใช้ค่าถ่วงน้ำหนักที่เหมาะสมที่สุดของข้อมูลจากการวนซ้ำ iteration เพื่อปรับเปลี่ยนค่าสัมประสิทธิ์การถดถอยของตัวแบบที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยที่สุดจนค่าสัมประสิทธิ์การถดถอยที่ทำในรอบใหม่มีความแตกต่างกับของรอบก่อนหน้านี้น้อยมากจนแน่ใจว่า ค่าสัมประสิทธิ์การถดถอยที่ได้เหมาะสมที่สุดและลู่เข้า ซึ่งเป็นการแก้ไขปัญหาความคลาดเคลื่อนมีค่าความแปรปรวนไม่คงที่จากสาเหตุที่ค่าความคลาดเคลื่อนโดยตรง ดังนั้นวิธีการพยากรณ์ค่าตัวแปรตามด้วยวิธี IRWLS จึงให้ผลที่มีความคลาดเคลื่อนน้อยกว่าวิธีวิธีการพยากรณ์ค่าตัวแปรตามด้วยวิธีการแปลงของ Box และ Cox เสมอทุกกรณี

5.4 การสรุปการเลือกใช้วิธีการประมาณค่าตัวแปรตามเมื่อความคลาดเคลื่อนในตัวแบบการถดถอยเชิงเส้นพหุ มีความแปรปรวนไม่คงที่

จากผลการวิจัยที่ได้ในครั้งนี้พบว่า เนื่องจากการประมาณหรือพยากรณ์ค่าตัวแปรตามด้วยวิธี IRWLS เมื่อค่าความคลาดเคลื่อนมีความแปรปรวนที่มีค่าไม่คงที่ หรือ ที่ค่าพารามิเตอร์ δ ไม่เท่ากับศูนย์ ดังนั้นในการวิเคราะห์การถดถอยด้วยตัวแบบการถดถอยเชิงเส้นพหุ เมื่อความแปรปรวนของความคลาดเคลื่อนมีค่าไม่คงที่ วิธีการประมาณหรือพยากรณ์ค่าตัวแปรตามที่ควรใช้มากที่สุด คือ วิธี IRWLS ส่วนวิธีการแปลงของ Box และ Cox ควรเริ่มใช้เมื่อความแปรปรวนของความคลาดเคลื่อนมีค่ามาก และสัมพันธ์กับตัวแปรอิสระ (ตัวแปรอิสระบางตัวในตัวแบบ) แต่ความถูกต้องของการประมาณหรือพยากรณ์จากวิธีการแปลงของ Box และ Cox จะอยู่ในระดับที่พอใช้ได้ และควรเริ่มใช้วิธี OLS เมื่อความแปรปรวนของความคลาดเคลื่อนมีค่าคงที่ หรือ ที่ค่าพารามิเตอร์ δ เท่ากับศูนย์ หรือควรเริ่มใช้เมื่อค่าความคลาดเคลื่อนมีความแปรปรวนที่มีค่าน้อย หรือที่ค่าพารามิเตอร์ δ เท่ากับ 0.1 และ สัมพันธ์กับค่าจริงของตัวแปรตาม เนื่องจากในสถานการณ์ดังกล่าว ค่า AMSE หรือ ค่า RE ระหว่างวิธี OLS และวิธี IRWLS มีค่าใกล้เคียงหรือถือว่าแตกต่างกันน้อยมาก เพราะค่า AMSE ของวิธีทั้งสองเริ่มแตกต่างกันในทศนิยมในตำแหน่งที่ 3 และ ค่า RE ของวิธีทั้งสองเริ่มแตกต่างกันในทศนิยมในตำแหน่งที่ 6

5.5 ข้อเสนอแนะ

ผู้วิจัยได้แบ่งข้อเสนอแนะของงานวิจัยครั้งนี้ออกเป็น 2 ด้านดังนี้

5.5.1 ด้านการนำไปใช้ประโยชน์

งานวิจัยนี้มีข้อเสนอแนะในด้านการนำไปใช้ประโยชน์ดังต่อไปนี้

1. เมื่อข้อมูลที่นำมาพยากรณ์ไม่ได้อยู่ในข้อตกลงของการวิเคราะห์การถดถอย ในกรณีที่ของความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ในตัวแบบการถดถอยเชิงเส้นพหุ วิธีการพยากรณ์ที่ควรใช้มากที่สุดคือ วิธี IRWLS ในการประมาณค่าตัวแปรตาม ส่วนวิธีการแปลงของ Box และ Cox จะมีความเหมาะสมในการพยากรณ์ตัวแปรตามก็ต่อเมื่อข้อมูล ความแปรปรวนของความคลาดเคลื่อนที่สัมพันธ์กับตัวแปรอิสระ มีพารามิเตอร์ที่เกี่ยวข้องที่ทำให้ระดับความแปรปรวนของความคลาดเคลื่อนมีค่าสูง และต้องการความถูกต้องในระดับพอใช้ได้ แต่ในกรณีที่ความแปรปรวนของความคลาดเคลื่อนที่สัมพันธ์กับค่าจริงของตัวแปรตาม ไม่ว่าจะมีขนาดตัวอย่างเท่าใด และความแปรปรวนของความคลาดเคลื่อนมีค่าที่ไม่มากนัก วิธี OLS น่าจะเป็นวิธีที่เหมาะสมที่สุด เนื่องจากผู้วิจัยเห็นว่าจากผลการวิจัยค่า AMSE ของวิธี OLS กับ วิธี IRWLS ให้ผลที่ใกล้เคียงกันหรือแตกต่างกันน้อยมาก อีกทั้งวิธี OLS ยังเป็นวิธีมีการคำนวณที่ง่าย และ สะดวกกว่าเมื่อเทียบกับวิธี IRWLS

2. สำหรับข้อมูลจริงที่ไม่ได้จากการจำลองข้อมูล เมื่อจะนำข้อมูลจริงดังกล่าววิเคราะห์ด้วยตัวแบบการถดถอยเชิงเส้นพหุที่มีจำนวนตัวแปรอิสระที่อยู่ในตัวแบบทั้งหมด 2 ตัวและตัวแปรอิสระแต่ละตัวต่างเป็นอิสระกัน และต่างมีผลต่อการเปลี่ยนแปลงค่าตัวแปรตามอย่างมีนัยสำคัญในทางปฏิบัติควรตรวจสอบข้อมูลความคลาดเคลื่อนว่ามีค่าความแปรปรวนที่ไม่คงที่หรือไม่ เนื่องจากสาเหตุที่สัมพันธ์กับตัวแปรอิสระ อาจเป็นตัวแปรอิสระตัวใดตัวหนึ่งที่อยู่ในตัวแบบ หรือตัวแปรอิสระทั้ง 2 ตัวที่อยู่ในตัวแบบ และค่าความคลาดเคลื่อนมีขนาดความแปรปรวนโดยประมาณอยู่ในระดับใด ก่อนการวิเคราะห์ข้อมูลตามแนวทางการสรุปผลการเปรียบเทียบแต่ละวิธีการพยากรณ์ตัวแปรตาม จากงานวิจัยชิ้นนี้

3. ในกรณีที่ตัวแปรอิสระที่ใช้ในตัวแบบการถดถอยเชิงเส้นพหุ ที่จำนวนตัวแปรอิสระที่มากกว่า 2 ตัวขึ้นไป สามารถใช้วิธีการทางสถิติ เช่น การวิเคราะห์ปัจจัย (factor analysis) โดยใช้เทคนิคการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis: PCA) และหมุนแกนปัจจัย (Factor Rotation) มาตั้งฉากกัน ทำให้แต่ละปัจจัยเป็นอิสระกัน เพื่อทำการจัดกลุ่มตัวแปร

อิสระที่อยู่ในตัวแบบทั้งหมด ให้เหลือเพียง 2 กลุ่ม หรือ 2 ปัจจัย ที่อิสระต่อกัน และตรวจสอบว่า เมื่อนำข้อมูลตัวแปรอิสระทั้ง 2 ปัจจัยมาวิเคราะห์กับตัวแปรตามในตัวแบบการถดถอยเชิงเส้นพหุ จะทำให้ค่าความคลาดเคลื่อนมีความแปรปรวนไม่คงที่หรือไม่ ซึ่งถ้าค่าความคลาดเคลื่อนมีความแปรปรวนไม่คงที่เป็นจริง แล้วจะสามารถนำข้อสรุปผลจากงานวิจัยชิ้นนี้มาเป็นหลักการแนวทางวิธีการประมาณหรือพยากรณ์ค่าตัวแปรตาม เนื่องจากผลที่ได้จากข้อมูลนำมาวิเคราะห์จะให้ผลในทิศทางเดียวกันกับข้อสรุปผลจากงานวิจัยชิ้นนี้

5.5.2 ด้านการศึกษาวิจัย

สำหรับผู้ที่สนใจที่จะศึกษาเพิ่มเติม ในการศึกษาครั้งต่อไป อาจทำการศึกษา ในแนวทางหรือกรณีต่างๆดังต่อไปนี้

1. ในงานวิจัยชิ้นนี้ ผู้วิจัยได้ศึกษาเฉพาะกรณีที่ตัวแปรอิสระมีการแจกแจงปกติ และ ไม่มีความสัมพันธ์กัน และ กรณีที่ตัวแปรตามมีความสัมพันธ์กันกับตัวแปรอิสระในรูปแบบเชิงเส้น สำหรับงานวิจัยครั้งต่อไปอาจทำการศึกษากรณีที่ข้อมูลตัวแปรอิสระมีรูปแบบการแจกแจงอื่นๆ หรือ ความสัมพันธ์ของตัวแปรตามกับตัวแปรอิสระ ที่ไม่ได้อยู่ในรูปแบบเชิงเส้น หรือ กรณีที่ตัวแปรอิสระแต่ละตัวมีความสัมพันธ์กัน ซึ่งการที่ตัวแปรอิสระและตัวแปรตามมีลักษณะที่ต่างจากเดิม อาจจะทำให้ผลการวิจัยที่แตกต่างจากเดิม

2. ทำการศึกษาเพิ่มเติมในกรณีที่รูปแบบความแปรปรวนของความคลาดเคลื่อนที่แตกต่างออกไป หรือ/และ อาจศึกษาปัญหาที่ความคลาดเคลื่อนมีอัตตสหสัมพันธ์ด้วยเพิ่มเติม

3. ทำการศึกษาเพิ่มเติมเกี่ยวกับวิธีการหาสัมประสิทธิ์การถดถอยเพื่อประมาณค่าตัวแปรตามวิธีอื่นหรือวิธีใหม่ๆ ในกรณีที่ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ เพื่อนำมาเปรียบเทียบวิธีเพิ่มเติมและหาวิธีที่เหมาะสมที่สุดในแต่ละสถานการณ์

4. ควรทำการศึกษาเพิ่มเติมในกรณีข้อมูลที่เป็นข้อมูลอนุกรมเวลา เพื่อศึกษาว่าการเปลี่ยนแปลงของเวลา จำนวนข้อมูลที่อยู่ในแต่ละช่วงเวลา จะส่งผลกระทบต่อความแปรปรวนของความคลาดเคลื่อนมีค่าไม่คงที่ได้อย่างไร

5. ในงานวิจัยนี้ วิธีการแปลง Box และ Cox เป็นวิธีการประมาณค่าตัวแปรตามที่ไม่มีความเหมาะสมในทุกกรณี ทั้งนี้อาจเป็นเพราะว่า ค่าประมาณ power ที่ได้มาจากเทคนิคภาวะ

น่าจะเป็นสูงสุด (Maximum Likelihood) ยังให้ค่าที่ไม่ละเอียดมากพอเมื่อเทียบกับกับวิธี IRWLS หรือข้อมูลค่าตัวแปรตามอาจจะมีค่ามากเกินไป ทำให้การแปลงโดยใช้การประมาณ power เพื่อหาค่าพยากรณ์ เมื่อทำการคำนวณหาความคลาดเคลื่อนที่เกิดจากความแตกต่างระหว่างค่าพยากรณ์กับค่าจริงของตัวแปรตาม ค่าความคลาดเคลื่อนที่ได้อาจมีค่ามากกว่าที่ควรจะเป็นกว่าในกรณีที่ข้อมูลตัวแปรตามมีค่าน้อย อีกทั้งวิธีการแปลงของ Box และ Cox เป็นวิธีการแปลงค่าตัวแปรตาม ซึ่งไม่ได้แก้ไขปัญหาความคลาดเคลื่อนมีค่าความแปรปรวนไม่คงที่ จากสาเหตุที่ค่าความคลาดเคลื่อนโดยตรง ทำให้ค่าพยากรณ์ตัวแปรตามที่ได้มีความถูกต้องน้อยกว่าที่ควรจะเป็น

รายการอ้างอิง

ภาษาไทย

ชิดชนก เชิงเชาว์. การวิเคราะห์การถดถอยสำหรับการวิจัยทางการศึกษา. มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี, 2541.

ชูใจ คูหารัตนไชย. การวิเคราะห์ความถดถอยเมื่อค่าความคลาดเคลื่อนของข้อมูลมีสหสัมพันธ์และความแปรปรวนไม่คงที่. วิทยานิพนธ์สถิติศาสตรมหาบัณฑิต บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2531.

ต่าย เซี่ยงฉี่. การวิเคราะห์การถดถอยพหุคูณ. พิมพ์ครั้งที่ 1. มหาวิทยาลัยเชียงใหม่, 2529.

นิติพงษ์ ส่งศรีโรจน์. การละเมิดข้อสมมติของการวิเคราะห์ความถดถอย สาเหตุ ผล การตรวจสอบและการแก้ไข. [ออนไลน์]. 2553. แหล่งที่มา: http://www.nitiphong.com [24 มกราคม 2556]

นุชรินทร์ ทิพยวรรณากร. การเปรียบเทียบค่าพยากรณ์ที่ได้จากตัวแบบที่คัดเลือกด้วยวิธีเบส์เซียน วิธีการกำจัดตัวแปรแบบถอยหลังและวิธีการถดถอยแบบขั้นบันได ในการวิเคราะห์ความถดถอยพหุนามแบบลำดับขั้น. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย, 2540.

บัณฑิต ชัยวิชญชาติ. เอกสารประกอบการเรียนการสอน วิชา 107411 เศรษฐมิติ 1. กรุงเทพมหานคร: มหาวิทยาลัยเกษตรศาสตร์, 2550.

ประชุม สุวัตถี. ทฤษฎีการอนุมานเชิงสถิติ. ปรับปรุงแก้ไขครั้งที่ 3. กรุงเทพมหานคร: โครงการส่งเสริมและพัฒนา เอกสารวิชาการสถาบันพัฒนบริหารศาสตร์, 2553.

พรพล คงอิ่ม. การแก้ไขปัญหาเกี่ยวกับความไม่เป็นเอกภาพของความแปรปรวนในแผนแบบการทดลองสุ่มตลอด. วิทยานิพนธ์ปริญญามหาบัณฑิต จุฬาลงกรณ์มหาวิทาลัย, 2548.

ไพโรจน์ ขาวสิทธิวงษ์. การเปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบสัมประสิทธิ์การถดถอยของตัวแบบเชิงเส้น เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่เท่ากัน. วิทยานิพนธ์ปริญญามหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย, 2539.

วิชิต หล่อจีระชุณห์กุล และ จิราวัลย์ จิตรถเวช. เทคนิคการพยากรณ์. พิมพ์ครั้งที่ 3. กรุงเทพมหานคร: โครงการส่งเสริมและพัฒนาเอกสารวิชาการ สถาบันบัณฑิตพัฒนบริหารศาสตร์, 2548.

วิธิดา สุนทรวิภาต. การศึกษาเปรียบเทียบวิธีการแก้ปัญหาความแปรปรวนของค่าคลาดเคลื่อนไม่คงที่ จากสาเหตุต่างๆในการถดถอยเชิงเส้นอย่างง่าย. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ มหาวิทยาลัยเกษตรศาสตร์, 2547.

วีรานันท์ พงศาภักดี. การวิเคราะห์ข้อมูลเชิงกลุ่ม: ทฤษฎีการประยุกต์. พิมพ์ครั้งที่ 2. นครปฐม: ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร, 2541.

สุพล ดุรงค์วัฒนา. การวิเคราะห์ความถดถอย สำหรับงานวิจัยขั้นสูง. พิมพ์สวย, 2549.

ภาษาอังกฤษ

Agresti, A. Categorical Data Analysis. New York: Willey, 1990.

Amemiya, T. Regression analysis when the variance of the dependent variable is proportional to the squares of its expectation. Journal of the American Statistical Association. 68 (1973): 928-934.

Belsley, D.A., Kuh, E., and Welsch, R.E. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley, 1980.

Bloch, D.A. and Moses, L.E. Non-optimally weighted least squares. the American Statistician. 42: (1988): 50-53.

Bock, R.D. and Yates, G. Log-linear Analysis of Nominal or Ordinal Qualitative Data by the Method of Maximum Likelihood. Chicago: International Educational Services, 1973.

Box, G.E.P. and Cox, D.R. An Analysis of Transformation. Journal of the Royal Statistical Society, Series B. 26 (1964): 211-252.

Carroll, R. and Ruppert, D. Robust Estimation in Heteroscedastic Linear Models. the American Statistician. 10 (1982): 429-441.

Cohen, M., Dalal, S. and Tukey, J.W. Robust, Smoothly Heterogeneous Variance regression. Technical Report. Murray Hill: Bell Laboratories, 1983.

Cohen, M., Dalal, S. and Tukey, J.W. Robust, smoothly heterogeneous variance regression. Appl.Statist. 42 (1993): 339-353.

Cox, D.R. and Hinkley, D.V. Theoretical Statistics. London: Chapman and Hall, 1974.

Cribari-Neto, F., Zarkos, S.G.: Heteroskedasticity-consistent covariance matrix estimation: White's estimator and the bootstrap. Journal of Statistical Computation and Simulation. 68 (2001): 391–411.

Cribari-Neto, F., Souza, T.C., Vasconcellos, K.L.P.: Inference under heteroskedasticity and leveraged data. Communication in Statistics- Theory and Methods. 36 (2007): 1877–1888.

Cribari-Neto, F., Lima, M.G.A.: Approximate inference in heteroskedastic regressions: a numerical evaluation. Journal of Applied Statistics. 37 (2010): 591–615.

Damodar N. Gujarati. Basic Econometrics. New York: Mcgraw-Hill International Editions, 1988.

Davidson, R., MacKinnon, J.G. Estimation and Inference in Econometrics. New York: Oxford University Press, 1993.

Dobson, A.J. An introduction to Generalized Linear Models. London: Chapman and Hall, 1990.

Eicker, F. Asymptotic normality and consistency of the least squares estimator for families of linear regressions. The Annals of Mathematical Statistics, 34 (1963): 447-456.

Eicker, F. Limit theorems for regressions with unequal and dependent errors. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1 (1967): 59-82

Fahrmeir,L. and Tutz,G. Multivariate Statistical Modelling Based on Generalized Linear Model. New York: Springer Berlog, 1994.

Haberman, S.J. Analysis of Qualitative Data. Vol. I. Introductory Topics. New York: Academic Press, 1978.

Hinkley, D.V. Jackknifing in unbalanced situations. Technometrics, 19 (1977): 285-292.

Horn, S.D., R.A. Horn, and D.B. Duncan. Estimating heteroscedastic variances in linear model. Journal of the American Statistical Association, 70 (1975): 380-385.

Huber, P.J. The behavior of maximum likelihood estimates under nonstandard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1 (1967): 221-233.

Judge, G.G., W.E. Griffiths, R.C. Hill, H. L¨utkepohl, and T-C. Lee. The Theory and Practice of Econometrics. 2nd Ed. New York: Wiley, 1985.

Kelejian, H.H. and Oates, W. Introduction to Econometrics Principles and application. 2nd ed. New York: Harper& Row Publishers, 1981.

MacKinnon, J.G. and H. White. Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties. Journal of Econometrics, 29 (1985): 53-57.

McCullagh, P. and Nelder, FRS J.A. Generalized Linear Models. 2nd ed. New York: Chapman and Hall, 1989.

Nelder, J. and R. Wedderburn. Generalized linear models. Journal of the Royal Statistical Society, Series A. 132 (1972): 370-384.

White, H. A heteroskedastic-consistent covariance matrix estimator and a direct test of heteroskedasticity. Econometrica, 48 (1980): 817-838.

บรรณานุกรม

ภาษาไทย

กัลยา วานิชย์บัญชา. การวิเคราะห์สถิติ: สถิติสำหรับการบริหารและวิจัย. พิมพ์ครั้งที่ 11. กรุงเทพมหานคร: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2551.

ชูใจ คูหารัตนไชย. การวิเคราะห์ความถดถอยเมื่อค่าความคลาดเคลื่อนของข้อมูลมีสหสัมพันธ์และความแปรปรวนไม่คงที่. วิทยานิพนธ์สถิติศาสตรมหาบัณฑิต บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2531.

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

ภาษาอังกฤษ

Harvey, A.C. Estimating regression models with multiplicative heteroscedasticity. Econometrica, 44 (1976): 461-465.

Rutemiller, H.C. and D.A. Bowers. Estimation in a heteroscedastic regression model. Journal of the American Statistical Association. 63 (1968): 552-557.

ภาคผนวก

รายละเอียดของโปรแกรมที่ใช้ในงานวิจัย

ในการวิจัยครั้งนี้ได้ใช้โปรแกรม R เวอร์ชัน 2.15.3 ในการจำลองข้อมูลและการประมาณค่าตัวแปรตามของแต่ละวิธี ในตัวแบบถดถอยเชิงเส้นพหุ เมื่อความคลาดเคลื่อนมีความแปรปรวนมีค่าไม่คงที่ ซึ่งมีคำสั่งดังต่อไปนี้

```
### 5000 run loop delta loop sigma loop n everycase Error_t1


MMMTOTAL_mse<-c()
nloop=5000


### loop Total #####
for ( k in 1:nloop){
        MMTOTAL_mse<-c()
        ### loop delta ####
        ddelta<-c(0,0.1,0.3,0.5) #set delta
        for ( d in 1:length(ddelta)){
        delta<-ddelta[d]
        ### loop sigma2 ####
        MTOTAL_mse<-c()
        sigma2<-c(450,900,1800) # set sigma_2
        for ( m in 1:length(sigma2)) {
                sigma_2<-sigma2[m]
                mse_M<-c() ### loop n
                ### loop n ####
                nsample<-c(25,50,100) # set n of sample
                for ( j in 1:length(nsample)) {


                n<-nsample[j]
```

```
#### simulate independent variable ##############
var_x1<-c(300,450,600)
var_x2<-c(600,450,300)
X1<-c()
X2<-c()
for ( i in 1:3){
	x1<-rnorm(n,0,sqrt(var_x1[i]))
	x2<-rnorm(n,0,sqrt(var_x2[i]))
	X1<-c(X1,x1)
	X2<-c(X2,x2)
}
XX1<-matrix(X1,c(n,3))
XX2<-matrix(X2,c(n,3))
#####Design matrix#####
vone<-rep(1,n)
x_C1<-cbind(vone,XX1[ ,1],XX2[,1])
x_C2<-cbind(vone,XX1[ ,2],XX2[,2])
x_C3<-cbind(vone,XX1[ ,3],XX2[,3])
x<-cbind(x_C1,x_C2,x_C3)

#####simulate error#####
#error_t1  is error depend x1
#error_t2  is error depend Y
V_e1<-c()
for ( i in 1:3){
	v_e1<-sigma_2*((abs(XX1[ ,i]))^delta)
	V_e1<-c(V_e1,v_e1)
}
V_error1<-matrix(V_e1,c(n,3))
Error_t1<-c()
```

```
for ( i in 1:3){
error_t1<-rnorm(n,0,sqrt(V_error1[, i]))
        Error_t1<-c(Error_t1,error_t1)
}
ERROR_t1<-matrix(Error_t1,c(n,3))

b0=500 ###set intercept

ytrue<-c()
for ( i in 1:3){
        YTRUE<-b0+XX1[ ,i]+XX2[ ,i]
        ytrue<-c(ytrue,YTRUE)
}
ytrue1<-matrix(ytrue,c(n,3))

#####simulate dependent variables#####
y1<-ytrue1+ERROR_t1

#####find reg coefficients#####
####1stmethod OLS ####
##error_t1###
OLS_C1<-lm(y1[ ,1] ~XX1[ ,1]+XX2[ ,1])
OLS_C2<-lm(y1[ ,2] ~XX1[ ,2]+XX2[ ,2])
OLS_C3<-lm(y1[ ,3] ~XX1[ ,3]+XX2[ ,3])
yhat_C1<-OLS_C1$fit
yhat_C2<-OLS_C2$fit
yhat_C3<-OLS_C3$fit
yhat1<-cbind(yhat_C1 ,yhat_C2 ,yhat_C3)
######MSE_OLS#######
```

```
MSE_ols<-c()
for ( i in 1:3){
MSE_OLS<-sum((ytrue1[ ,i]-yhat1[ ,i])^2)/n
MSE_ols<-c(MSE_ols,MSE_OLS)
}
#MSE_ols

#####2ndmethod IRWLS1######
yyhat_C1<-OLS_C1$fit
betastart<-c(500,1,1)
w_C1<-(abs(XX1[ ,1]))^delta
out1_C1<-glm.fit(x_C1,y1[ ,1],weights=1/(450*(2+(sigma_2/450)*w_C1)), _
        start=betastart,etastart=yyhat_C1,mustart=y1[ ,1])
yhat3_C1<-out1_C1$fit

yyhat_C2<-OLS_C2$fit
betastart<-c(500,1,1)
w_C2<-(abs(XX1[ ,2]))^delta
out1_C2<-glm.fit(x_C2,y1[ ,2],weights=1/(450*(2+(sigma_2/450)*w_C2)), _
        start=betastart,etastart=yyhat_C2,mustart=y1[ ,2])
yhat3_C2<-out1_C2$fit

yyhat_C3<-OLS_C3$fit
betastart<-c(500,1,1)
w_C3<-(abs(XX1[ ,3]))^delta
out1_C3<-glm.fit(x_C3,y1[ ,3],weights=1/(450*(2+(sigma_2/450)*w_C3)), _
        start=betastart,etastart=yyhat_C3,mustart=y1[ ,3])

yhat3_C3<-out1_C3$fit
```

```
yhat3<-cbind(yhat3_C1,yhat3_C2,yhat3_C3)
######MSE_IRWLS1#######
MSE_Irwls1<-c()
for ( i in 1:3){
		MSE_IRWLS1<-sum((ytrue1[ ,i]-yhat3[ ,i])^2)/n
		MSE_Irwls1<-c(MSE_Irwls1,MSE_IRWLS1)
}
#MSE_Irwls1


####3rdmethod Boxcox ####
##error_t1##
aa<-c()
bb<-c()
XXX1<-c()
XXX2<-c()
yyy1<-c()
yytrue1<-c()

for( ii in 1:length(y1)){
 #####condition1#####
 if(y1[ii]<0 & ii>=1 & ii<=n ){
   repeat{
 XXX1[ii]<-rnorm(1,0,sqrt(300))
 XXX2[ii]<-rnorm(1,0,sqrt(600))
 yytrue1[ii]<-b0+XXX1[ii]+XXX2[ii]
 V_error1[ii]<-sigma_2*abs(XXX1[ii])^delta
 ERROR_t1[ii]<-rnorm(1,0,sqrt(V_error1[ii]))
 yyy1[ii]<-yytrue1[ii]+ERROR_t1[ii]
   if(yyy1[ii]>=0){
```

```
     break
      }
  } ##end loop do while
bb[ii]<-yyy1[ii]
aa<-c(aa,bb[ii])
} else
#####condition2#####
if(y1[ii]<0 & ii>n & ii<=2*n ){
    repeat{
XXX1[ii]<-rnorm(1,0,sqrt(450))
XXX2[ii]<-rnorm(1,0,sqrt(450))
yytrue1[ii]<-b0+XXX1[ii]+XXX2[ii]
V_error1[ii]<-sigma_2*abs(XXX1[ii])^delta
ERROR_t1[ii]<-rnorm(1,0,sqrt(V_error1[ii]))
yyy1[ii]<-yytrue1[ii]+ERROR_t1[ii]
   if(yyy1[ii]>=0){
     break
      }
  } ##end loop do while
bb[ii]<-yyy1[ii]
aa<-c(aa,bb[ii])
} else

#####condition3#####
if(y1[ii]<0 & ii>2*n & ii<=3*n ){
    repeat{
XXX1[ii]<-rnorm(1,0,sqrt(600))
XXX2[ii]<-rnorm(1,0,sqrt(300))
yytrue1[ii]<-b0+XXX1[ii]+XXX2[ii]
V_error1[ii]<-sigma_2*abs(XXX1[ii])^delta
```

```
        ERROR_t1[ii]<-rnorm(1,0,sqrt(V_error1[ii]))
        yyy1[ii]<-yytrue1[ii]+ERROR_t1[ii]
          if(yyy1[ii]>=0){
            break
             }
        } ##end loop while
        bb[ii]<-yyy1[ii]
        aa<-c(aa,bb[ii])
        } else

        #####condition4#####
        if(y1[ii]>=0){
        XXX1[ii]<-XX1[ii]
        XXX2[ii]<-XX2[ii]
        yytrue1[ii]<-ytrue1[ii]
        yyy1[ii]<-y1[ii]
        bb[ii]<-yyy1[ii]
        aa<-c(aa,bb[ii])
        }
      }
      XXXX1<-matrix(XXX1,c(n,3))
      XXXX2<-matrix(XXX2,c(n,3))
      yyyy1<-matrix(aa,c(n,3))
      yyytrue1<-matrix(yytrue1,c(n,3))


library(MASS)
b_C1<-boxcox(lm(yyyy1[ ,1]~XXXX1[ ,1]+XXXX2[ ,1]),lambda=seq(-3,3,1/1000))
lamda_C1=b_C1$x
lik_C1=b_C1$y
```

```
bc_C1=cbind(lamda_C1,lik_C1)
bcorder_C1<-bc_C1[order(-lik_C1),]
la_C1<-bcorder_C1[1,1]          #lambda
BOX_C1<-lm(((yyyy1[ ,1]^la_C1-1)/la_C1)~XXXX1[ ,1]+XXXX2[ ,1])
yhat2trans_C1<-BOX_C1$fit
yhat2_C1<-(la_C1*yhat2trans_C1+1)^(1/la_C1)

b_C2<-boxcox(lm(yyyy1[ ,2]~XXXX1[ ,2]+XXXX2[ ,2]),lambda=seq(-3,3,1/1000))
lamda_C2=b_C2$x
lik_C2=b_C2$y
bc_C2=cbind(lamda_C2,lik_C2)
bcorder_C2<-bc_C2[order(-lik_C2),]
la_C2<-bcorder_C2[1,1]          #lambda
BOX_C2<-lm(((yyyy1[ ,2]^la_C2-1)/la_C2)~XXXX1[ ,2]+XXXX2[ ,2])
yhat2trans_C2<-BOX_C2$fit
yhat2_C2<-(la_C2*yhat2trans_C2+1)^(1/la_C2)

b_C3<-boxcox(lm(yyyy1[ ,3]~XXXX1[ ,3]+XXXX2[ ,3]),lambda=seq(-3,3,1/1000))
lamda_C3=b_C3$x
lik_C3=b_C3$y
bc_C3=cbind(lamda_C3,lik_C3)
bcorder_C3<-bc_C3[order(-lik_C3),]
la_C3<-bcorder_C3[1,1]          #lambda
BOX_C3<-lm(((yyyy1[ ,3]^la_C3-1)/la_C3)~XXXX1[ ,3]+XXXX2[ ,3])
yhat2trans_C3<-BOX_C3$fit
yhat2_C3<-(la_C3*yhat2trans_C3+1)^(1/la_C3)

yhat2<-cbind(yhat2_C1,yhat2_C2,yhat2_C3)

          #MSE_BOX-COX#
```

```
        MSE_box<-c()
        for ( i in 1:3){
        MSE_BOX<-sum((yyytrue1[ ,i]-yhat2[ ,i])^2)/n
        MSE_box<-c(MSE_box,MSE_BOX)
        }
        #MSE_box

        MSE<-cbind(MSE_ols,MSE_box,MSE_Irwls1)
        MSE_M<-c()
        MSE_M<-c(MSE_M,MSE_ols[1],MSE_ols[2],MSE_ols[3], _
                MSE_box[1],MSE_box[2],MSE_box[3], _
                MSE_Irwls1[1],MSE_Irwls1[2],MSE_Irwls1[3])
        mse_M<-c(mse_M,MSE_M)
        } # end loop n
        MTOTAL_mse<-c(MTOTAL_mse,mse_M)
    } # end loop sigma_2
    MMTOTAL_mse<-c(MMTOTAL_mse,MTOTAL_mse)
    } # end loop delta
    MMMTOTAL_mse<-c(MMMTOTAL_mse,MMTOTAL_mse)

    #####COUNT TIME#####
    pie(c(k,nloop-k),radius=1,clockwise=T)

} # end loop total

mse_Matrix<-matrix(MMMTOTAL_mse,c(324,nloop))
AMSE<-c()

for ( i in 1:324){
```

```
        amse<-mean(mse_Matrix[i,])
        AMSE<-c(AMSE,amse)
}

AMSE_M<-matrix(AMSE,c(9,36))
bb<-c("OLS_V12","OLS_V11","OLS_V21", _
        "BOX_V12","BOX_V11","BOX_V21", _
        "IRWLS_V12","IRWLS_V11","IRWLS_V21")
rownames(AMSE_M)<-bb
AMSE_M

#### 5000 run loop delta loop sigma loop n everycase Error_t2

MMMTOTAL_mse<-c()
nloop=5000
### loop Total #####
for ( k in 1:nloop){
          MMTOTAL_mse<-c()
          ### loop delta ####
          ddelta<-c(0,0.1,0.3,0.5) #set delta
          for ( d in 1:length(ddelta)){
          delta<-ddelta[d]
          ### loop sigma2 ####
          MTOTAL_mse<-c()
          sigma2<-c(450,900,1800) # set sigma_2
          for ( m in 1:length(sigma2)) {
                    sigma_2<-sigma2[m]

                    mse_M<-c() ### loop n
```

```
### loop n ####
nsample<-c(25,50,100) # set n of sample
for ( j in 1:length(nsample)) {
        n<-nsample[j]
######simulate independent variables######
var_x1<-c(300,450,600)
var_x2<-c(600,450,300)
X1<-c()
X2<-c()
#set.seed(112233)
for ( i in 1:3){
x1<-rnorm(n,0,sqrt(var_x1[i]))
x2<-rnorm(n,0,sqrt(var_x2[i]))
X1<-c(X1,x1)
X2<-c(X2,x2)
}
XX1<-matrix(X1,c(n,3))
XX2<-matrix(X2,c(n,3))

#####Design matrix#####
vone<-rep(1,n)
x_C1<-cbind(vone,XX1[ ,1],XX2[,1])
x_C2<-cbind(vone,XX1[ ,2],XX2[,2])
x_C3<-cbind(vone,XX1[ ,3],XX2[,3])
x<-cbind(x_C1,x_C2,x_C3)

#####error#####
#error_t1  is error depend x1
#error_t2  is error depend Y
b0=500 #####set intercept
```

```
ytrue<-c()
for ( i in 1:3){
YTRUE<-b0+XX1[ ,i]+XX2[ ,i]
ytrue<-c(ytrue,YTRUE)
}
ytrue1<-matrix(ytrue,c(n,3))

V_e2<-c()
for ( i in 1:3){
v_e2<-sigma_2*(abs(ytrue1[ ,i])^delta)
V_e2<-c(V_e2,v_e2)
}
V_error2<-matrix(V_e2,c(n,3))

Error_t2<-c()
for ( i in 1:3){
        error_t2<-rnorm(n,0,sqrt(V_error2[, i]))
        Error_t2<-c(Error_t2,error_t2)
}
ERROR_t2<-matrix(Error_t2,c(n,3))

#####simulate dependent variables#####
y2<-ytrue1+ERROR_t2

#####find reg coefficients#####
####1stmethod OLS ####
##error_t2###
OLS_C1<-lm(y2[ ,1] ~XX1[ ,1]+XX2[ ,1])
OLS_C2<-lm(y2[ ,2] ~XX1[ ,2]+XX2[ ,2])
```

```
OLS_C3<-lm(y2[ ,3] ~XX1[ ,3]+XX2[ ,3])
yhat_C1<-OLS_C1$fit
yhat_C2<-OLS_C2$fit
yhat_C3<-OLS_C3$fit
yhat1<-cbind(yhat_C1 ,yhat_C2 ,yhat_C3)

######MSE_OLS#######
MSE_ols<-c()
for ( i in 1:3){
MSE_OLS<-sum((ytrue1[ ,i]-yhat1[ ,i])^2)/n
MSE_ols<-c(MSE_ols,MSE_OLS)
}
#MSE_ols

#####2ndmethod IRWLS1######
yyhat_C1<-OLS_C1$fit
betastart<-c(500,1,1)
w_C1<-(abs(ytrue1[ ,1]))^delta
out1_C1<-glm.fit(x_C1,y2[ ,1],weights=1/(450*(2+(sigma_2/450)*w_C1)),
            _ start=betastart,etastart=yyhat_C1,mustart=y2[ ,1])
yhat3_C1<-out1_C1$fit

yyhat_C2<-OLS_C2$fit
betastart<-c(500,1,1)
w_C2<-(abs(ytrue1[ ,2]))^delta
out1_C2<-glm.fit(x_C2,y2[ ,2],weights=1/(450*(2+(sigma_2/450)*w_C2)),
            _ start=betastart,etastart=yyhat_C2,mustart=y2[ ,2])
yhat3_C2<-out1_C2$fit
yyhat_C3<-OLS_C3$fit
```

```
betastart<-c(500,1,1)
w_C3<-(abs(ytrue1[ ,3]))^delta
out1_C3<-glm.fit(x_C3,y2[ ,3],weights=1/(450*(2+(sigma_2/450)*w_C3)),
        _ start=betastart,etastart=yyhat_C3,mustart=y2[ ,3])
yhat3_C3<-out1_C3$fit

yhat3<-cbind(yhat3_C1,yhat3_C2,yhat3_C3)

######MSE_IRWLS1#######
MSE_Irwls1<-c()
for ( i in 1:3){
MSE_IRWLS1<-sum((ytrue1[ ,i]-yhat3[ ,i])^2)/n
MSE_Irwls1<-c(MSE_Irwls1,MSE_IRWLS1)
}
#MSE_Irwls1

####3rdmethod Boxcox ####
##error_t2##
aa<-c()
bb<-c()
XXX1<-c()
XXX2<-c()
yyy2<-c()
yytrue1<-c()

for( ii in 1:length(y2)){
 #####condition1#####
 if(y2[ii]<0 & ii>=1 & ii<=n ){
  repeat{
```

```
XXX1[ii]<-rnorm(1,0,sqrt(300))
XXX2[ii]<-rnorm(1,0,sqrt(600))
yytrue1[ii]<-b0+XXX1[ii]+XXX2[ii]
V_error2[ii]<-sigma_2*abs(yytrue1[ii])^delta
ERROR_t2[ii]<-rnorm(1,0,sqrt(V_error2[ii]))
yyy2[ii]<-yytrue1[ii]+ERROR_t2[ii]
    if(yyy2[ii]>=0){
    break
     }
  } ## end loop do while
bb[ii]<-yyy2[ii]
aa<-c(aa,bb[ii])
} else
#####condition2#####
if(y2[ii]<0 & ii>n & ii<=2*n ){
  repeat{
XXX1[ii]<-rnorm(1,0,sqrt(450))
XXX2[ii]<-rnorm(1,0,sqrt(450))
yytrue1[ii]<-b0+XXX1[ii]+XXX2[ii]
V_error2[ii]<-sigma_2*abs(yytrue1[ii])^delta
ERROR_t2[ii]<-rnorm(1,0,sqrt(V_error2[ii]))
yyy2[ii]<-yytrue1[ii]+ERROR_t2[ii]
    if(yyy2[ii]>=0){
    break
     }
  } ##end loop do while
bb[ii]<-yyy2[ii]
aa<-c(aa,bb[ii])
} else
```

```
  #####condition3#####
  if(y2[ii]<0 & ii>2*n & ii<=3*n ){
    repeat{
  XXX1[ii]<-rnorm(1,0,sqrt(600))
  XXX2[ii]<-rnorm(1,0,sqrt(300))
  yytrue1[ii]<-b0+XXX1[ii]+XXX2[ii]
  V_error2[ii]<-sigma_2*abs(yytrue1[ii])^delta
  ERROR_t2[ii]<-rnorm(1,0,sqrt(V_error2[ii]))
  yyy2[ii]<-yytrue1[ii]+ERROR_t2[ii]
      if(yyy2[ii]>=0){
      break
       }
   } ##end loop do while
  bb[ii]<-yyy2[ii]
  aa<-c(aa,bb[ii])
  } else
  #####condition4#####
  if(y2[ii]>=0){
  XXX1[ii]<-XX1[ii]
  XXX2[ii]<-XX2[ii]
  yytrue1[ii]<-ytrue1[ii]
  yyy2[ii]<-y2[ii]
  bb[ii]<-yyy2[ii]
  aa<-c(aa,bb[ii])
  }
}
XXXX1<-matrix(XXX1,c(n,3))
XXXX2<-matrix(XXX2,c(n,3))
yyyy2<-matrix(aa,c(n,3))
```

```
        yyytrue1<-matrix(yytrue1,c(n,3))

library(MASS)
b_C1<-boxcox(lm(yyyy2[ ,1]~XXXX1[ ,1]+XXXX2[ ,1]),lambda=seq(-3,3,1/1000))
lamda_C1=b_C1$x
lik_C1=b_C1$y
bc_C1=cbind(lamda_C1,lik_C1)
bcorder_C1<-bc_C1[order(-lik_C1),]
la_C1<-bcorder_C1[1,1]          #lambda
BOX_C1<-lm(((yyyy2[ ,1]^la_C1-1)/la_C1)~XXXX1[ ,1]+XXXX2[ ,1])
yhat2trans_C1<-BOX_C1$fit
yhat2_C1<-(la_C1*yhat2trans_C1+1)^(1/la_C1)

b_C2<-boxcox(lm(yyyy2[ ,2]~XXXX1[ ,2]+XXXX2[ ,2]),lambda=seq(-3,3,1/1000))
lamda_C2=b_C2$x
lik_C2=b_C2$y
bc_C2=cbind(lamda_C2,lik_C2)
bcorder_C2<-bc_C2[order(-lik_C2),]
la_C2<-bcorder_C2[1,1]          #lambda
BOX_C2<-lm(((yyyy2[ ,2]^la_C2-1)/la_C2)~XXXX1[ ,2]+XXXX2[ ,2])
yhat2trans_C2<-BOX_C2$fit
yhat2_C2<-(la_C2*yhat2trans_C2+1)^(1/la_C2)

b_C3<-boxcox(lm(yyyy2[ ,3]~XXXX1[ ,3]+XXXX2[ ,3]),lambda=seq(-3,3,1/1000))
lamda_C3=b_C3$x
lik_C3=b_C3$y
bc_C3=cbind(lamda_C3,lik_C3)
bcorder_C3<-bc_C3[order(-lik_C3),]
la_C3<-bcorder_C3[1,1]          #lambda
```

```
BOX_C3<-lm(((yyyy2[ ,3]^la_C3-1)/la_C3)~XXXX1[ ,3]+XXXX2[ ,3])
yhat2trans_C3<-BOX_C3$fit
yhat2_C3<-(la_C3*yhat2trans_C3+1)^(1/la_C3)

yhat2<-cbind(yhat2_C1,yhat2_C2,yhat2_C3)

                ####MSE_BOX-COX####
                MSE_box<-c()
                for ( i in 1:3){
                      MSE_BOX<-sum((yyytrue1[ ,i]-yhat2[ ,i])^2)/n
                      MSE_box<-c(MSE_box,MSE_BOX)
                }
                #### pick MSE ####
                MSE<-cbind(MSE_ols,MSE_box,MSE_lrwls1)
                MSE_M<-c()
                MSE_M<-c(MSE_M,MSE_ols[1],MSE_ols[2],MSE_ols[3], _
                          MSE_box[1],MSE_box[2],MSE_box[3], _
                          MSE_lrwls1[1],MSE_lrwls1[2],MSE_lrwls1[3])
                mse_M<-c(mse_M,MSE_M)
                } # end loop n
                MTOTAL_mse<-c(MTOTAL_mse,mse_M)
        } # end loop sigma_2
        MMTOTAL_mse<-c(MMTOTAL_mse,MTOTAL_mse)
        } # end loop delta
        MMMTOTAL_mse<-c(MMMTOTAL_mse,MMTOTAL_mse)

        #####COUNT TIME#####
        pie(c(k,nloop-k),radius=1,clockwise=T)
} # end loop total
```

```
mse_Matrix<-matrix(MMMTOTAL_mse,c(324,nloop))
AMSE<-c()
        for ( i in 1:324){
        amse<-mean(mse_Matrix[i,])
        AMSE<-c(AMSE,amse)
}
AMSE_M<-matrix(AMSE,c(9,36))
bb<-c("OLS_V12","OLS_V11","OLS_V21", _
         "BOX_V12","BOX_V11","BOX_V21", _
        "IRWLS_V12","IRWLS_V11","IRWLS_V21")
rownames(AMSE_M)<-bb
AMSE_M
```

ประวัติผู้เขียนวิทยานิพนธ์

นายพีรวัฒน์ เสรีวัฒนนุกูล เกิดวันพุธที่ 14 กันยายน พ.ศ.2531 สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาคณิตศาสตร์ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล ในปีการศึกษา 2553 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต (วท.ม.) สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2554