

## การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์จากตัวแบบการถดถอย สำหรับข้อมูลที่ถูกตัดปลายทางขวาแบบที่ 1 ที่มีการแจกแจงล็อกนอร์มอล

### A Comparison on Parameter Estimation from Regression Model for Type-I Right-Censored Data from Lognormal Distribution

ศิวพร ทิพย์พันธุ์<sup>1\*</sup> และ อนุนภาพ สมบูรณ์สวัสดิ์<sup>2</sup>

Siwaporn Tippan<sup>1\*</sup> and Anupap Somboonsawatdee<sup>2</sup>

Received: 23 December 2018; Revised: 17 April 2019; Accepted: 6 May 2019

#### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์จากตัวแบบการถดถอยเมื่อตัวแปรตามมีการแจกแจงล็อกนอร์มอลและตัวแปรตามบางค่าเป็นข้อมูลที่ถูกตัดปลายทางขวาแบบที่ 1 ด้วยวิธีกำลังสองต่ำสุด (Ordinary Least Squares Method; OLS) วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation; MLE) วิธีของแชตเทอร์จีและแมคลีช (Chatterjee and McLeish Method; CM) และวิธีภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็ม (Maximum Likelihood Estimation using the EM algorithm; MLE\_EM) ข้อมูลที่ใช้ในการศึกษาครั้งนี้ได้จำลองข้อมูลจำนวน 81 สถานการณ์ สถานการณ์ละ 10,000 รอบ ขนาดตัวอย่าง ( $n$ ) เท่ากับ 30, 50, 100 และเปอร์เซ็นต์การถูกตัดปลายทางขวาของตัวแปรตาม ( $r$ ) เท่ากับ 10%, 20%, 30% และอัตราส่วนความแปรปรวนของตัวแปรอิสระตัวที่ 1 ต่อตัวแปรอิสระตัวที่ 2 เท่ากับ 1:1, 1:2, 1:5 และอัตราส่วนความแปรปรวนรวมของตัวแปรอิสระต่อความคลาดเคลื่อน เท่ากับ 2:1, 1:1, 1:2 และเกณฑ์ที่ใช้ในการเปรียบเทียบ คือ ค่าเฉลี่ยของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Average Mean Square Error; AMSE) และค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency; RE) จากการศึกษาพบว่า 1) วิธี MLE และวิธี MLE\_EM เป็นวิธีที่มีประสิทธิภาพสูงสุดเมื่อตัวอย่างมีขนาดใหญ่ ( $n = 100$ ) หรือตัวแปรตามถูกตัดปลายทางขวามาก ( $r = 30\%$ ) ในทางกลับกันนั้น พบว่า 2) วิธี OLS เป็นวิธีที่มีประสิทธิภาพสูงสุดเมื่อตัวอย่างมีขนาดเล็ก ( $n = 30$ ) หรือตัวแปรตามถูกตัดปลายทางขวาน้อย ( $r = 10\%$ ) และ 3) วิธี CM นั้นมีประสิทธิภาพสูงสุดในสถานการณ์ที่เหลือ กล่าวคือ เมื่อตัวอย่างมีขนาดปานกลาง ( $n = 50$ ) หรือตัวแปรตามถูกตัดปลายทางขวาปานกลาง ( $r = 20\%$ )

**คำสำคัญ:** ตัวแบบการถดถอย, ข้อมูลที่ถูกตัดปลายทางขวาแบบที่ 1, ความแปรปรวน

\*Corresponding author: E-mail: siwaporn.t@student.chula.ac.th

<sup>1,2</sup> ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

<sup>1,2</sup> Faculty of Commerce and Accountancy, Chulalongkorn University

## Abstract

The objective of this research is to compare the estimation methods for log-linear regression model with dependent variable under type-I right-censoring: 1) Ordinary Least Squares Method (OLS); 2) Maximum Likelihood Estimation (MLE); 3) Chatterjee and McLeish Method (CM); and 4) Maximum Likelihood Estimation using the EM Algorithm (MLE\_EM). The results on this research are from 81 simulated scenarios with simulation size of 10,000, the sample sizes ( $n$ ) are 30, 50, 100; the censoring proportions of data ( $r$ ) of 10%, 20%, 30%; the ratio of variances of two independent variables of 1:1, 1:2, 1:5; the ratio of the sum of variances of two independent variables to error variance of 2:1, 1:1, 1:2. The criterion is Average Mean Square Error (AMSE) and Relative Efficiency (RE). The findings are: 1) MLE\_EM and MLE perform the best at large sample size ( $n = 100$ ) or high censoring proportion ( $r = 30\%$ ); on the other hand, 2) OLS performs the best at small sample size ( $n = 30$ ) or low censoring proportion ( $r = 10\%$ ); and 3) CM performs the best generally for the rest of scenarios such that the sample size is moderate ( $n = 50$ ) or censoring proportion is moderate ( $r = 20\%$ ).

**Keywords:** regression model, type-I right-censored data, variance

## 1. บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันการวิเคราะห์การถดถอยได้รับความนิยมเป็นอย่างมาก และในสถานการณ์จริงอาจพบตัวแปรตามบางค่าเป็นข้อมูลถูกตัดปลายทางขวาประเภทที่ 1 ซึ่งเป็นข้อมูลเกี่ยวกับระยะเวลา จึงต้องใช้ตัวแปรสุ่มที่มีการแจกแจงที่มีค่ามากกว่า 0 โดยตัวแปรสุ่มที่มีการแจกแจงล็อกนอร์มอลมีคุณสมบัติสอดคล้องกับข้อมูลในสถานการณ์ดังกล่าว ซึ่ง Smith and Merceret (2000) ได้ยกตัวอย่างข้อมูลจริงที่มีการแจกแจงล็อกนอร์มอล เช่น ระยะเวลาแพร่กระจายของเชื้อโรคในผู้ป่วย ระยะเวลาฟื้นตัวของผู้ป่วย ระยะเวลาจ้างงาน ระยะเวลาคุยโทรศัพท์ เป็นต้น และสำหรับการประมาณค่าพารามิเตอร์จากตัวแบบการถดถอยในสถานการณ์ดังกล่าวนี้ สามารถทำได้หลายวิธี ซึ่งวิธีที่ง่ายที่สุด คือ การตัดข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 นั้นทิ้งไป แต่วิธีนี้จะทำให้ค่าเฉลี่ยของค่าประมาณของตัวแปรตามที่ได้ต่ำกว่าค่าเฉลี่ยที่แท้จริงและความคลาดเคลื่อนของการประมาณจะมีค่าสูง ดังนั้น ผู้วิจัยจึงเกิดความสนใจในการหาวิธีประมาณค่าพารามิเตอร์ที่ทำให้ตัวแปรตามที่ได้จากการประมาณมีค่าใกล้เคียงกับค่าที่แท้จริงมากที่สุด และจากการทบทวนงานวิจัยที่ผ่านมาพบว่า

บังอร กุมพล (2539) ได้ศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในสมการถดถอย เมื่อค่าสังเกตของตัวแปรตามและความคลาดเคลื่อนมีการแจกแจงปกติ และค่าสังเกตของตัวแปรตามบางค่าเป็นข้อมูลที่ถูกตัดทิ้งทางขวาประเภทที่ 1 วิธีประมาณที่ทำการศึกษา คือ วิธีการของสมิธ วิธีโมดิไฟด์แอกซ์เรียล และวิธีภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็ม ขนาดตัวอย่างเท่ากับ 10, 20, 30, 50, 100 ข้อมูลถูกตัดปลายทางขวาร้อยละ 10, 20, 30, 40, 50 และจำนวนตัวแปรอิสระเท่ากับ 3 ตัว, 5 ตัว และ 7 ตัว เกณฑ์การเปรียบเทียบ คือ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Root Mean Square Error; RMSE) โดยคำนวณมาจากการเปรียบเทียบระหว่างค่าประมาณ

ของตัวแปรตามเทียบกับค่าสังเกตของตัวแปรตามเฉพาะในส่วนของคุณค่าที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 เท่านั้น และจากการศึกษาพบว่า เมื่อข้อมูลถูกตัดปลายทางขวาร้อยละ 10, 20, 30 ในทุกขนาดตัวอย่างและจำนวนตัวแปรที่ศึกษา พบว่า ค่า RMSE ที่ได้จากการประมาณค่าด้วยวิธีการของสมิธมีค่าต่ำที่สุด ยกเว้นในกรณีที่ข้อมูลถูกตัดปลายทางขวาเท่ากับร้อยละ 10 ขนาดตัวอย่างเท่ากับ 10 และจำนวนตัวแปรอิสระเท่ากับ 5 ตัว ซึ่งค่า RMSE ที่ได้จากการประมาณค่าด้วยวิธีโมดิไฟด์แอกซ์เรียลมีค่าต่ำกว่าวิธีอื่น และในกรณีที่เมื่อข้อมูลถูกตัดปลายทางขวาเป็นร้อยละ 20 และ 30 จำนวนตัวแปรอิสระเท่ากับ 7 ตัว ซึ่งค่า RMSE ที่ได้จากการประมาณค่าด้วยวิธีด้วยภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็มมีค่าต่ำกว่าวิธีอื่น และเมื่อข้อมูลถูกตัดปลายทางขวาเท่ากับร้อยละ 40, 50 พบว่า ค่า RMSE ที่ได้จากการประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็มมีค่าต่ำที่สุด นอกจากนั้นพบว่า ค่า RMSE แปรผกผันกับขนาดตัวอย่าง และแปรผันตรงกับร้อยละของข้อมูลที่ถูกตัดปลายทางขวา

จำเนียร จ้างรงค์รักษ์ (2539) ศึกษาเปรียบเทียบการประมาณค่าตัวแปรตามในสมการถดถอย เมื่อตัวแปรตามบางค่าเป็นข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 และความคลาดเคลื่อนมีการแจกแจงดับเบิลเอ็กซ์โพเนนเชียล ปกติ และล็อกนอร์มอล จำนวนตัวแปรอิสระเท่ากับ 2 ตัว โดยที่ตัวแปรอิสระตัวที่ 1 มีการแจกแจงไวบูลล์และตัวแปรอิสระตัวที่ 2 มีการแจกแจงปกติ ขนาดตัวอย่างเท่ากับ 10, 20, 30, 50, 100 ข้อมูลถูกตัดปลายทางขวาร้อยละ 10, 20, 30, 40 และ 50 วิธีการประมาณที่ใช้ในการศึกษา คือ วิธีการกำลังสองต่ำสุด วิธีการของแซตเทอร์เจย์และแมคลิช วิธีการของบัคเลย์และเจมส์ และวิธีภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็ม เกณฑ์การเปรียบเทียบ คือ ค่า RMSE โดยคำนวณมาจากการเปรียบเทียบระหว่างค่าประมาณของตัวแปรตามเทียบกับค่าสังเกตของตัวแปรตามเฉพาะในส่วนของคุณค่าที่ถูกตัดปลายทางขวาประเภทที่ 1 เท่านั้น และจากการศึกษาพบว่า ค่า RMSE ที่ได้จากการประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็มมีค่าต่ำกว่าวิธีอื่นในทุกสถานการณ์ที่ทำการศึกษา และกรณีที่ความคลาดเคลื่อนมีการแจกแจงล็อกนอร์มอล ค่า RMSE ของทุกวิธีที่ขนาดเดียวกันจะเพิ่มขึ้นเมื่อการตัดทิ้งของข้อมูลเพิ่มขึ้น

ในการวิจัยครั้งนี้สนใจเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์จากตัวแบบการถดถอย เมื่อตัวแปรตามบางค่าเป็นข้อมูลที่ถูกตัดปลายทางขวาแบบที่ 1 ที่มีความสัมพันธ์กับตัวแปรอิสระและความคลาดเคลื่อนที่แจกแจงล็อกนอร์มอล ด้วยวิธีการประมาณค่าแบบกำลังสองต่ำสุด วิธีภาวะน่าจะเป็นสูงสุด วิธีของแซตเทอร์เจย์และแมคลิช และวิธีภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็ม ส่วนการออกแบบการทดลองในการวิจัยครั้งนี้แตกต่างจากงานวิจัยที่ผ่านมา กล่าวคือ งานวิจัยที่ผ่านมาได้ทำศึกษาผลของขนาดตัวอย่าง สัดส่วนของคุณค่าที่ถูกตัดปลายทางขวา จำนวนของตัวแปรอิสระ และการแจกแจงของความคลาดเคลื่อน แต่ในวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาผลของขนาดตัวอย่าง สัดส่วนของคุณค่าที่ถูกตัดปลายทางขวา โดยปัจจัยที่ได้ศึกษาเพิ่มเติมจากงานวิจัยก่อนหน้านี้ ได้แก่ อัตราส่วนความแปรปรวนของตัวแปรอิสระแต่ละตัว และอัตราส่วนความแปรปรวนรวมของตัวแปรอิสระต่อความแปรปรวนของความคลาดเคลื่อน เนื่องจากในการวิเคราะห์การถดถอย ตัวแปรอิสระและความคลาดเคลื่อนต่างมีผลต่อการประมาณค่าตัวแปรตาม และจากการทบทวนงานวิจัยพบว่า

พีรวัฒน์ เสรีวัฒนกุล (2555) ได้ทำการศึกษาเปรียบเทียบวิธีการประมาณสำหรับตัวแบบการถดถอยเชิงเส้นพหุที่ค่าความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ เมื่อตัวแปรอิสระ 2 ตัวและความคลาดเคลื่อนมีค่าเฉลี่ยเท่ากับศูนย์ และแจกแจงปกติ วิธีประมาณที่ใช้ในการศึกษา ได้แก่ วิธีการกำลังสองน้อยที่สุด (Ordinary Least Square; OLS), วิธีการแปลงของ Box และ Cox (Box-Cox Transformation) และวิธี Iteratively Reweighted Least Square (IRWLS) ส่วนเกณฑ์การเปรียบเทียบ คือ ค่าเฉลี่ยของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Average Mean Square Error;

AMSE) โดยงานวิจัยครั้งนี้ได้ศึกษาปัจจัยด้านอัตราส่วนความแปรปรวนตัวแปรอิสระตัวที่ 1 ต่อตัวแปรอิสระตัวที่ 2 ที่มี 3 แบบ ได้แก่ 1:2, 1:1 และ 2:1 ซึ่งกำหนดให้ความแปรปรวนรวมของตัวแปรอิสระทั้ง 3 แบบมีค่าเท่ากัน คือ 900 จากการศึกษาพบว่า อัตราส่วนความแปรปรวนตัวแปรอิสระระหว่างตัวแปรอิสระที่ 1 และตัวแปรอิสระที่ 2 ไม่ส่งผลต่อค่า AMSE นอกจากนั้นได้ศึกษาในกรณีที่ความแปรปรวนของความคลาดเคลื่อนมีค่าคงที่ โดยที่กำหนดให้อัตราส่วนความแปรปรวนรวมของตัวแปรอิสระต่อความแปรปรวนของความคลาดเคลื่อน คือ 1:2, 1:1, 2:1 และจากการศึกษาพบว่า เมื่อความแปรปรวนของความคลาดเคลื่อนเพิ่มขึ้นจะทำให้ค่า AMSE เพิ่มขึ้น

อุษณีย์ วงศ์อำมาตย์ (2555) ศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายเมื่อตัวแปรตามมีการสูญหายแบบนอนอินฟอร์เมเบิลในการวิเคราะห์การถดถอยเชิงเส้นพหุ โดยที่ตัวแปรอิสระ 3 ตัวและความคลาดเคลื่อนมีค่าเฉลี่ยเท่ากับศูนย์และแจกแจงปกติ ข้อมูลที่ใช้ในการศึกษามีสัดส่วนของการสูญหาย 3 ระดับ คือ 10%, 20%, 30% และมีระดับของการสูญหายแบบนอนอินฟอร์เมเบิล 3 ระดับ คือ ไม่มี ปานกลาง สูง ขนาดตัวอย่างเท่ากับ 50, 100, 200 วิธีที่ใช้ในการประมาณค่าสูญหาย คือ วิธี EM Algorithm (EM) วิธี K-Nearest Neighbor Imputation (KNN) และวิธี Predictive Mean Matching Imputation (PMM) โดยได้ศึกษาปัจจัยด้านความแปรปรวนของความคลาดเคลื่อนที่มี 3 ระดับ ได้แก่ 100, 900 และ 8,100 เกณฑ์การเปรียบเทียบ คือ AMSE จากการศึกษาพบว่า เมื่อความแปรปรวนของความคลาดเคลื่อนมีค่ามากขึ้น แล้วค่า AMSE ของวิธีประมาณค่าสูญหายทั้ง 3 วิธีจะสูงขึ้น และนอกจากนั้นศึกษาปัจจัยด้านอัตราส่วนความแปรปรวนตัวแปรอิสระตัวที่ 1 ต่อตัวแปรอิสระตัวที่ 2 ที่มี 2 แบบ ได้แก่ 1:1:1 และ 1:3:5 โดยที่ความแปรปรวนรวมของตัวแปรอิสระทั้ง 2 แบบมีค่าเท่ากัน คือ 900 จากการศึกษา พบว่า อัตราส่วนของความแปรปรวนตัวแปรอิสระระหว่างตัวแปรอิสระที่ 1 และตัวแปรอิสระที่ 2 ที่แตกต่างกัน จะไม่ส่งผลต่อค่า AMSE เนื่องจากในงานวิจัยนี้ได้กำหนดให้ตัวแปรตามที่สร้างจากตัวแปรอิสระต่างๆชุดกันมีค่าเฉลี่ยและความแปรปรวนรวมของตัวแปรอิสระเท่ากัน

นอกจากนั้นเกณฑ์ที่ใช้ในการเปรียบเทียบวิธีที่มีประสิทธิภาพมากที่สุดในการประมาณค่าสำหรับการวิเคราะห์การถดถอยในงานวิจัยครั้งนี้ก็มีความแตกต่างจากงานวิจัยที่ผ่านมา กล่าวคือ ค่า AMSE ในงานวิจัยครั้งนี้คำนวณจากการเปรียบเทียบระหว่างค่าประมาณของตัวแปรตามเทียบกับค่าที่แท้จริงของตัวแปรตามของข้อมูลทั้งหมดทั้งในส่วนที่เป็นข้อมูลที่ถูกต้องและไม่ได้ถูกตัดปลายทางขวาประเภทที่ 1

## 1.2 วัตถุประสงค์การวิจัย

เพื่อเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์จากตัวแบบการถดถอย เมื่อตัวแปรตามมีความสัมพันธ์กับตัวแปรอิสระและความคลาดเคลื่อนที่มีการแจกแจงล็อกนอร์มอล และบางค่าของตัวแปรตามเป็นข้อมูลที่ถูกต้องตัดปลายทางขวาแบบที่ 1 ด้วยวิธีกำลังสองต่ำสุด (Ordinary Least Squares Method; OLS) วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation; MLE) วิธีของแชตเตอร์จีและแมคลีช (Chatterjee and McLeish Method; CM) และวิธีภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็ม (Maximum Likelihood Estimation using the EM Algorithm; MLE\_EM)

## 2. เอกสารและงานวิจัยที่เกี่ยวข้อง

### 2.1 ความหมายของข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 (type-I right-censored data)

Klein และ Moeschberger (2003) อธิบายถึงความหมายข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 คือ ข้อมูลที่มีการกำหนดเวลาในการสังเกตไว้ล่วงหน้า โดยข้อมูลที่เกิดขึ้นระหว่างดำเนินการสังเกตจะเป็นข้อมูลที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 ซึ่งผู้วิจัยจะบันทึกข้อมูลเหล่านี้ให้เท่ากับค่าที่แท้จริงของข้อมูล ส่วนข้อมูลที่เกิดขึ้นหลังจากสิ้นสุดการสังเกตไปแล้วนั้นจะเป็นข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 ซึ่งเป็นข้อมูลที่ไม่สามารถระบุค่าแท้จริงได้ ดังนั้นจึงให้ข้อมูลดังกล่าวมีค่าเท่ากับเวลาที่ใช้ในการสังเกตที่กำหนดไว้ล่วงหน้า เช่น ผู้วิจัยต้องการศึกษาระยะเวลารอดชีวิตของหนูที่ได้รับประทานยาชนิดหนึ่ง จึงได้กำหนดระยะเวลาสังเกต 6 เดือน และเมื่อสิ้นสุดการวิจัยไปแล้ว ผู้วิจัยพบว่า มีหนูบางตัวตายระหว่างการสังเกต แต่หนูบางตัวกลับมีชีวิตรอดแม้จะสิ้นสุดการสังเกตไปแล้ว ดังนั้น ระยะเวลารอดชีวิตของหนูที่ตายระหว่างการสังเกตอยู่นั้น เป็นข้อมูลที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 โดยผู้วิจัยจะบันทึกระยะเวลารอดชีวิตของหนูดังกล่าวให้เท่ากับระยะเวลารอดชีวิตที่แท้จริง ส่วนระยะเวลารอดชีวิตของหนูที่ตายหลังสิ้นสุดการสังเกตไปแล้วนั้น จะเป็นข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 ที่เวลา 6 เดือน เป็นต้น

ให้  $T$  เป็นเวลาในการสังเกตที่กำหนดไว้ล่วงหน้า

$\tau_1, \tau_2, \dots, \tau_n$  เป็นเวลาจริงจนเกิดเหตุการณ์ที่สนใจศึกษา

$t_1, t_2, \dots, t_n$  เป็นเวลาจนเกิดเหตุการณ์ที่สนใจศึกษาที่ถูกบันทึกโดยผู้วิจัย

$$\text{นิยาม } t_i = \min(\tau_i, T) \quad (1)$$

$$\text{กล่าวคือ } t_i = \begin{cases} \tau_i, & \tau_i \leq T \text{ (ถ้าเป็นข้อมูลที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1)} \\ T, & \tau_i > T \text{ (ถ้าเป็นข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1)} \end{cases} \quad (2)$$

เมื่อเวลาจริงจนเกิดเหตุการณ์ที่สนใจศึกษามีค่าน้อยกว่าหรือเท่ากับเวลาในการสังเกตที่กำหนดไว้ล่วงหน้า แล้วเวลาจริงจนเกิดเหตุการณ์ที่สนใจศึกษานั้นจะเป็นข้อมูลที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 และจะถูกบันทึกให้เท่ากับเวลาจริงจนเกิดเหตุการณ์ที่สนใจศึกษา แต่หากเวลาจริงจนเกิดเหตุการณ์ที่สนใจศึกษามีค่ามากกว่าเวลาในการสังเกตที่กำหนดไว้ล่วงหน้า เวลาจริงจนเกิดเหตุการณ์ที่สนใจศึกษานั้นจะเป็นข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 และจะถูกบันทึกให้เท่ากับเวลาในการสังเกตที่กำหนดไว้ล่วงหน้า

Jöreskog (2002) กล่าวถึง ฟังก์ชันภาวะน่าจะเป็นของข้อมูลที่บางส่วนของข้อมูลนั้นได้ถูกตัดปลายทางขวาประเภทที่ 1 ไว้ดังนี้

$$L(t_i) = \begin{cases} f(\tau_i), & \tau_i \leq T \text{ (ถ้าเป็นข้อมูลที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 มีจำนวน } m \text{ ค่า คือ } i=1, 2, \dots, m) \\ P(\tau_i > T) = S(T), & \tau_i > T \text{ (ถ้าเป็นข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 มีจำนวน } n-m \text{ ค่า คือ } i=m+1, \dots, n) \end{cases} \quad (3)$$

โดยที่  $f(\tau_i)$  คือ ฟังก์ชันความหนาแน่นจะเป็นของ  $\tau_i$

$P(\tau_i > T) = S(T)$  คือ ฟังก์ชันการอยู่รอดของ  $\tau_i$  หรือเป็นความน่าจะเป็นที่  $\tau_i$  มีค่ามากกว่า  $T$

$$\text{ดังนั้น ฟังก์ชันภาวะน่าจะเป็นรวม คือ } L = \prod_{i=1}^m f(\tau_i) \times \prod_{i=m+1}^n S(T) \quad (4)$$

### 2.2 วิธีการประมาณที่ใช้ในการศึกษา

ก่อนจะเริ่มกล่าวถึงวิธีการประมาณที่ใช้ในการศึกษา มีข้อตกลงดังนี้

|     |                 |                                                                                                    |
|-----|-----------------|----------------------------------------------------------------------------------------------------|
| ให้ | $\beta_0$       | คือ ค่าที่แท้จริงของระยะตัดแกน $y$                                                                 |
|     | $\hat{\beta}_0$ | คือ ค่าประมาณของค่าที่แท้จริงของระยะตัดแกน $y$                                                     |
|     | $\beta_j$       | คือ ค่าสัมประสิทธิ์ที่แท้จริงของการวิเคราะห์การถดถอยพหุเชิงเส้นตัวที่ $j$ โดย $j = 1, 2, \dots, p$ |
|     | $\hat{\beta}_j$ | คือ ค่าประมาณของค่าสัมประสิทธิ์ที่แท้จริงของการวิเคราะห์การถดถอยพหุเชิงเส้นตัวที่ $j$              |
|     | $X_{ji}$        | คือ ตัวแปรอิสระตัวที่ $j$ ลำดับที่ $i$                                                             |
|     | $\varepsilon_i$ | คือ ความคลาดเคลื่อนที่แท้จริงลำดับที่ $i$                                                          |
|     | $Y_c$           | คือ เวลาในการสังเกตที่กำหนดไว้ล่วงหน้า                                                             |
|     | $\sigma$        | คือ ส่วนเบี่ยงเบนมาตรฐานที่แท้จริง                                                                 |
|     | $\hat{\sigma}$  | คือ ค่าประมาณของส่วนเบี่ยงเบนมาตรฐาน                                                               |

$$\text{และให้ } Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i \quad (5)$$

คือ ค่าสังเกตของตัวแบบการวิเคราะห์การถดถอยลำดับ  $i$  เมื่อ  $i = 1, 2, \dots, m-1, m, m+1, \dots, n$  และ  $m \leq n$

$$\text{นิยาม } Y_i^* = \min(Y_i, Y_c) \text{ เป็นเวลาที่ผู้วิจัยบันทึกได้} \quad (6)$$

$$\text{กล่าวคือ } Y_i^* = \begin{cases} Y_i, & Y_i \leq Y_c \text{ (ถ้าเป็นตัวอย่างที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 มีจำนวน } m \text{ ค่า คือ } i = 1, 2, \dots, m) \\ Y_c, & Y_i > Y_c \text{ (ถ้าเป็นตัวอย่างที่ถูกตัดปลายทางขวาประเภทที่ 1 มีจำนวน } n-m \text{ ค่า คือ } i = m+1, \dots, n) \end{cases} \quad (7)$$

$$\text{ให้ } \mathbf{X}^{(NC)} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ 1 & X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1m} & X_{2m} & \cdots & X_{pm} \end{bmatrix}_{m \times (p+1)} \quad \text{และ } \mathbf{X}^{(C)} = \begin{bmatrix} 1 & X_{1(m+1)} & X_{2(m+1)} & \cdots & X_{p(m+1)} \\ 1 & X_{1(m+2)} & X_{2(m+2)} & \cdots & X_{p(m+2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{pn} \end{bmatrix}_{(n-m) \times (p+1)}$$

$$\text{ดังนั้น } \mathbf{X} = \left\{ \begin{array}{c} \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1m} & X_{2m} & \cdots & X_{pm} \end{bmatrix} \\ \hline \begin{bmatrix} 1 & X_{1(m+1)} & X_{2(m+1)} & \cdots & X_{p(m+1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{pn} \end{bmatrix} \end{array} \right\} \begin{array}{l} \mathbf{X}^{(NC)} \\ \mathbf{X}^{(C)} \end{array} \quad \text{และให้ } \hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}_{(p+1) \times 1}$$

$$\text{ต่อมาให้ } \mathbf{Y}^{*(NC)} = \begin{bmatrix} Y_1^* \\ Y_2^* \\ \vdots \\ Y_m^* \end{bmatrix}_{m \times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix}_{m \times 1} \quad \text{และ } \mathbf{Y}^{*(C)} = \begin{bmatrix} Y_{m+1}^* \\ Y_{m+2}^* \\ \vdots \\ Y_n^* \end{bmatrix}_{(n-m) \times 1} = \begin{bmatrix} Y_c \\ Y_c \\ \vdots \\ Y_c \end{bmatrix}_{(n-m) \times 1}$$

$$\text{ดังนั้น } \mathbf{Y}^* = \left\{ \begin{array}{c} \begin{bmatrix} Y_1^* \\ \vdots \\ Y_m^* \end{bmatrix} \\ \hline \begin{bmatrix} Y_{m+1}^* \\ \vdots \\ Y_n^* \end{bmatrix} \end{array} \right\} = \left\{ \begin{array}{c} \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix} \\ \hline \begin{bmatrix} Y_c \\ \vdots \\ Y_c \end{bmatrix} \end{array} \right\} \begin{array}{l} \mathbf{Y}^{*(NC)} \\ \mathbf{Y}^{*(C)} \end{array} \quad \text{และเวกเตอร์ของข้อมูลที่มีความสมบูรณ์ คือ } \tilde{\mathbf{Y}} = \begin{bmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \\ \vdots \\ \tilde{Y}_n \end{bmatrix}_{n \times 1}$$

### 2.2.1 วิธีกำลังสองต่ำสุด (Ordinary Least Squares Method)

เป็นวิธีที่ถูกคิดค้นโดย Gauss (1809) โดยมีหลักการคือ ทำให้ผลรวมของกำลังสองของเศษเหลือมีค่าต่ำที่สุด เมื่อนำข้อมูลมาวิเคราะห์จะถือว่าข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 จะเสมือนเป็นข้อมูลที่ไม่ถูกตัดปลายทางขวา

ประเภทที่ 1 และ  $\hat{\beta} = (X^T X)^{-1} X^T Y^*$  (8)

### 2.2.2 วิธีภาวะความน่าจะเป็นสูงสุด (Maximum Likelihood Estimation Method)

เป็นวิธีที่ถูกคิดค้นโดย Gauss (1821 credit in Fisher, 1922) ต่อมา Fisher (1922) ได้ปรับปรุงวิธีการประมาณค่าพารามิเตอร์ และธีระพร วีระถาวร (2536) กล่าวว่า ตัวประมาณภาวะน่าจะเป็นสูงสุด คือ ตัวประมาณที่ทำให้ความน่าจะเป็นในการเกิด  $\theta$  เกิดขึ้นสูงสุดเมื่อสุ่มตัวอย่าง  $X$  ขึ้นมา กล่าวคือ ตัวประมาณภาวะน่าจะเป็นสูงสุด คือ ตัวประมาณที่ทำให้  $L(\theta; X)$  มีค่าสูงสุด

โดย  $L(\theta; X)$  คือ ฟังก์ชันภาวะน่าจะเป็นรวมซึ่งเป็นตัวที่บ่งบอกถึงความน่าจะเป็นที่  $\theta$  จะเกิดขึ้นเมื่อสุ่มตัวอย่าง  $X$  ขึ้นมา

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix}_{p \times 1} \quad \text{คือ เวกเตอร์ของพารามิเตอร์}$$

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}_{n \times 1} \quad \text{คือ เวกเตอร์ของตัวอย่าง}$$

และ  $p(X; \theta)$  คือ ฟังก์ชันความหนาแน่นร่วมของ  $\theta$  เมื่อ  $L(\theta; X) = p(X; \theta)$

เมื่อตัวอย่างของข้อมูลที่สมบูรณ์มีขนาดใหญ่ จะทำให้ตัวประมาณภาวะน่าจะเป็นสูงสุดไม่เอนเอียง (ค่าคาดหวังของตัวประมาณภาวะน่าจะเป็นสูงสุดจะมีค่าเท่ากับค่าพารามิเตอร์) และรูปร่างการกระจายของตัวประมาณจะคล้ายกับการแจกแจงปกติ

### การประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุด

Jöreskog (2002) ได้กล่าวถึงฟังก์ชันภาวะน่าจะเป็นของตัวแปรตาม โดยที่ตัวแปรตามบางส่วนเป็นข้อมูลที่ถูกตัดปลายทางขวาแบบที่ 1 ไว้ดังนี้

$$L(Y_i^*) = \begin{cases} f(Y_i) & , Y_i \leq Y_c \text{ (ตัวแปรตามที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 มีจำนวน } m \text{ ค่า คือ } i = 1, \dots, m) \\ P(Y_i > Y_c) = S(Y_c), Y_i > Y_c \text{ (ตัวแปรตามที่ถูกตัดปลายทางขวาประเภทที่ 1 มีจำนวน } n - m \text{ ค่า คือ } i = m + 1, \dots, n) \end{cases} \quad (9)$$

เมื่อ  $S(Y_c)$  คือ ฟังก์ชันความอยู่รอด หรือความน่าจะเป็นที่ตัวแปรสุ่ม  $Y_i$  จะมีค่ามากกว่า  $Y_c$

ดังนั้น ฟังก์ชันภาวะน่าจะเป็นรวม คือ

$$L(\theta; \mathbf{X}) = \prod_{i=1}^m f(Y_i) \times \prod_{i=m+1}^n S(Y_c)$$

$$= \prod_{i=1}^m \left[ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \left[ \frac{Y_i - (\beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi})}{\sigma} \right]^2} \right] \times \prod_{i=m+1}^n \left[ 1 - \Phi \left( \frac{Y_c - (\beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi})}{\sigma} \right) \right] \quad (10)$$

Tobin (1958) ได้นำเสนอวิธีการที่ทำให้  $L(\theta; \mathbf{X})$  มีค่าสูงสุดโดยไม่ได้ใช้การหาอนุพันธ์อันดับหนึ่งของฟังก์ชัน ภาวะนั้นจะเป็นดังกล่าวโดยตรงเนื่องจากมีความซับซ้อนในการคำนวณ โดยเริ่มต้นให้  $\theta^* = (\beta_0^*, \beta_1^*, \beta_2^*, \dots, \beta_p^*, \sigma^*)$

$$\text{เมื่อ } \beta_j^* = \frac{\beta_j}{\sigma}, \quad \sigma^* = \frac{1}{\sigma} \quad \text{และ } j = 0, 1, 2, \dots, p \quad (11)$$

จากนั้นจึงแทน  $\frac{\beta_0}{\sigma}$  ด้วยค่า  $\beta_0^*$ ,  $\frac{\beta_j}{\sigma}$  ด้วยค่า  $\beta_j^*$  และ  $\frac{1}{\sigma}$  ด้วยค่า  $\sigma^*$  ในสมการ (10)

ดังนั้น ฟังก์ชันภาวะนั้นจะเป็นรวมที่ได้ คือ

$$L(\theta^*; \mathbf{X}) = \prod_{i=1}^m \left[ \frac{1}{\sqrt{2\pi}} \sigma^* e^{-\frac{1}{2} \left[ Y_i \sigma^* - (\beta_0^* + \beta_1^* X_{1i} + \dots + \beta_p^* X_{pi}) \right]^2} \right] \times \prod_{i=m+1}^n \left[ 1 - \Phi \left( Y_c \sigma^* - (\beta_0^* + \beta_1^* X_{1i} + \dots + \beta_p^* X_{pi}) \right) \right] \quad (12)$$

จากนั้นทำให้ฟังก์ชันภาวะนั้นจะเป็นรวมดังกล่าว มีค่าสูงสุดโดยใช้วิธีการ Newton-Raphson ซึ่งเริ่มต้นประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองต่ำที่สุด และใช้กระบวนการวนซ้ำจนได้ค่าประมาณของ  $\theta^* = (\beta_0^*, \beta_1^*, \beta_2^*, \dots, \beta_p^*, \sigma^*)$

$$\text{ต่อมาจึงหาค่าประมาณของ } \theta \text{ ซึ่ง } \hat{\beta}_j = \frac{\hat{\beta}_j^*}{\hat{\sigma}^*}, \quad \hat{\sigma} = \frac{1}{\hat{\sigma}^*} \quad (13)$$

### 2.2.3 วิธีของแชตเตอร์จีและแมคเลิช (Chatterjee and McLeish Method)

Chatterjee และ McLeish (1986) ได้นำหลักการ การแทนค่าที่สูญหายด้วยค่าที่ทำให้ผลรวมกำลังสองของความคลาดเคลื่อนต่ำที่สุด มาประยุกต์กับการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองต่ำสุดเมื่อตัวแปรตามเป็นข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 และใช้วิธีการวนซ้ำ มีขั้นตอนดังนี้

**ขั้นตอน 0**      คำนวณค่า  $\hat{\beta}^{(0)}$  ด้วยวิธีกำลังสองต่ำที่สุด โดยใช้เฉพาะข้อมูลที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1

$$\text{จะได้ } \hat{\beta}^{(0)} = \left( \left( \mathbf{X}^{(NC)} \right)^T \mathbf{X}^{(NC)} \right)^{-1} \left( \mathbf{X}^{(NC)} \right)^T \mathbf{Y}^{(NC)} \quad (14)$$

**ขั้นตอน a**      แทนค่าตัวแปรตามที่ถูกตัดปลายทางขวาประเภทที่ 1 โดย

- ถ้า  $\hat{\beta}_0^{(r)} + \hat{\beta}_1^{(r)} X_{1i} + \dots + \hat{\beta}_p^{(r)} X_{pi} > Y_c$  จะแทนค่าตัวแปรตามที่ถูกตัดปลายทางขวาประเภทที่ 1

ด้วยค่า  $\hat{\beta}_0^{(r)} + \hat{\beta}_1^{(r)} X_{1i} + \dots + \hat{\beta}_p^{(r)} X_{pi}$

- ถ้า  $\hat{\beta}_0^{(r)} + \hat{\beta}_1^{(r)} X_{1i} + \dots + \hat{\beta}_p^{(r)} X_{pi} \leq Y_c$  จะแทนค่าตัวแปรตามที่ถูกตัดปลายทางขวาประเภทที่ 1

ด้วยค่า  $Y_c$  เหมือนเดิม

เมื่อ  $r$  คือ รอบที่ของการประมาณค่าพารามิเตอร์,  $r = 0, 1, 2, 3, \dots$

$r = 0$  คือ รอบที่ 0 ของการประมาณค่าพารามิเตอร์ ซึ่งเป็นการประมาณค่าพารามิเตอร์ในขั้นตอน 0

และ  $i = m+1, m+2, \dots, n$

**ขั้นตอน b** นำตัวแปรตามที่ได้จากการแทนค่าในขั้นตอน a มารวมกับตัวแปรตามที่ไม่ถูกตัดปลายทางขวา

ประเภทที่ 1 เพื่อคำนวณค่า  $\beta^{(r+1)}$  โดยใช้วิธีกำลังสองต่ำสุด

ทำขั้นตอน a และ b วนไปเรื่อยๆ จนกระทั่งได้  $\left| \beta_j^{(r+1)} - \beta_j^{(r)} \right| \leq 0.001$  ทุก  $j = 0, 1, 2, \dots, p$

## 2.2.4 การประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็ม

### (Maximum Likelihood Estimation using the EM algorithm Method)

ขั้นตอนวิธีอีเอ็มถูกคิดค้นโดย Dempster, Laird และ Rubin (1977) และต่อมา Aitkin (1981) ได้นำขั้นตอนวิธีอีเอ็มมาประยุกต์กับการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดโดยใช้วิธีการวนซ้ำ มีขั้นตอนดังนี้

**ขั้นตอน 0** คำนวณค่า  $\beta^{(0)}$  และ  $\sigma^{(0)}$  โดยใช้ข้อมูลทั้งหมด โดยถือว่าข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 เสมือนเป็นข้อมูลที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 ด้วยวิธีภาวะน่าจะเป็นสูงสุด โดยทำให้ฟังก์ชัน

ภาวะน่าจะเป็นรวม  $L = \prod_{i=1}^n f(Y_i^*)$  มีค่าสูงสุด นั่นคือ การคำนวณค่า  $\beta^{(0)}$  และ  $\sigma^{(0)}$  ด้วยวิธีกำลังสองต่ำสุด

(เนื่องจากการคำนวณค่า  $\beta^{(0)}$  และ  $\sigma^{(0)}$  ด้วยวิธีกำลังสองต่ำสุดของข้อมูลที่สมบูรณ์และมีการแจกแจงปกติ

จะเท่ากับการคำนวณค่า  $\beta^{(0)}$  และ  $\sigma^{(0)}$  ด้วยวิธีภาวะน่าจะเป็นสูงสุด) ดังนั้น  $\beta^{(0)} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}^*$  (15)

**ขั้นตอน E (Expectation Step : E Step)** การแทนค่าตัวแปรตามที่ถูกตัดปลายทางขวาประเภทที่ 1 ด้วยค่าคาดหวัง

แบบมีเงื่อนไข  $E(Y_i^* | Y_i > Y_c, \beta, \sigma) = \mu_i^{(r)} + \left[ \sigma^{(r)} \times h(z_i^{(r)}) \right]$  (16)

โดย  $\mu_i^{(r)} = \beta_0^{(r)} + \beta_1^{(r)} X_{1i} + \dots + \beta_p^{(r)} X_{pi}$  (17)

$$z_i^{(r)} = \frac{Y_c - \mu_i^{(r)}}{\sigma^{(r)}} \quad (18)$$

$$f(z_i^{(r)}) = \left( \frac{1}{\sqrt{2\pi}} \right) e^{-\left( \frac{(z_i^{(r)})^2}{2} \right)} \quad (19)$$

$$S(z_i^{(r)}) = 1 - F(z_i^{(r)}) = \int_{z_i^{(r)}}^{\infty} f(t) dt \quad (20)$$

$$h(z_i^{(r)}) = \frac{f(z_i^{(r)})}{1 - F(z_i^{(r)})} \quad (21)$$

เมื่อ  $r$  คือ รอบที่ของการประมาณค่าพารามิเตอร์,  $r = 0, 1, 2, 3, \dots$

$r = 0$  คือ รอบที่ 0 ของการประมาณค่าพารามิเตอร์ ซึ่งเป็นการประมาณค่าพารามิเตอร์ในขั้นตอน 0  
และ  $i = m+1, m+2, \dots, n$

### ขั้นตอน M (Maximization Step : M Step)

นำตัวแปรตามที่ได้จากการแทนค่าใน **ขั้นตอน E** มารวมกับตัวแปรตามที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 จะได้ข้อมูลที่มีความสมบูรณ์  $\left( \tilde{Y}_i \right)$ ,  $i = 1, 2, \dots, n$  โดย

$$\tilde{Y}_i = \begin{cases} Y_i^* = Y_i & , Y_i \leq Y_c \text{ (ถ้าเป็นตัวแปรตามที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 มีจำนวน } m \text{ ค่า)} \\ \mu_i^{(r)} + \left[ \sigma^{(r)} \times h(z_i^{(r)}) \right] & , Y_i > Y_c \text{ (ถ้าเป็นตัวแปรตามที่ถูกตัดปลายทางขวาประเภทที่ 1 มีจำนวน } n - m \text{ ค่า)} \end{cases} \quad (22)$$

จากนั้นนำข้อมูลสมบูรณ์ดังกล่าวมาประมาณค่า  $\beta^{(r+1)}$  และ  $\sigma^{(r+1)}$  เพื่อให้ฟังก์ชันภาวะน่าจะเป็นรวม

$$L = \prod_{i=1}^n f(\tilde{Y}_i) \text{ มีค่าสูงสุด นั่นคือ หาค่า } \beta^{(r+1)} \text{ ด้วยวิธีกำลังสองต่ำที่สุด จะได้ } \beta^{(r+1)} = (X^T X)^{-1} X^T \tilde{Y} \quad (23)$$

และหา  $\sigma^{(r+1)}$  จากสูตร  $\sigma^{(r+1)} = \sqrt{\frac{\sum_{i=1}^m \left( y_i - \mu_i^{(r)} \right)^2 + \left[ \sigma^{(r)} \right]^2 \sum_{m+1}^n \left[ 1 + \left( z_i^{(r)} h(z_i^{(r)}) \right) \right]}{n}}$  (24)

เมื่อ  $\sigma^{(0)}$  คือ ค่าประมาณของส่วนเบี่ยงเบนมาตรฐานในขั้นตอน 0

ทำ**ขั้นตอน E** และ **M** วนไปเรื่อยๆ จนกระทั่งได้  $\left| \beta_j^{(r+1)} - \beta_j^{(r)} \right| \leq 0.001$  ทุก  $j = 0, 1, 2, \dots, p$

## 3.วิธีการศึกษา

### 3.1 ขอบเขตของการวิจัย

#### 3.1.1 ตัวแบบของค่าสังเกตของตัวแปรตามที่ใช้ในการศึกษาครั้งนี้ คือ

$$Y_i = e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i} \quad (25)$$

หรือ  $\log(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \quad (26)$

เมื่อ  $i = 1, 2, \dots, n$  และตัวแปรอิสระแต่ละตัวไม่มีความสัมพันธ์กัน

โดยที่  $Y_i$  คือ ค่าสังเกตของตัวแปรตาม ลำดับที่  $i$

$X_{1i}$  คือ ตัวแปรอิสระตัวที่ 1 ลำดับที่  $i$  เมื่อ  $X_{1i} \stackrel{iid}{\sim} N(0, \sigma_{x_1}^2)$

$X_{2i}$  คือ ตัวแปรอิสระตัวที่ 2 ลำดับที่  $i$  เมื่อ  $X_{2i} \stackrel{iid}{\sim} N(0, \sigma_{x_2}^2)$

$\varepsilon_i$  คือ ความคลาดเคลื่อน ลำดับที่  $i$  เมื่อ  $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$

$\beta_j$  คือ ค่าสัมประสิทธิ์ที่แท้จริงตัวที่  $j$  โดย  $j = 1, 2$

$\beta_0$  คือ ค่าที่แท้จริงของระยะตัดแกน  $y$

$\sigma_{x_1}^2$  คือ ความแปรปรวนของตัวแปรอิสระตัวที่ 1

$\sigma_{x_2}^2$  คือ ความแปรปรวนของตัวแปรอิสระตัวที่ 2

$\sigma_{x_1 + x_2}^2$  คือ ความแปรปรวนรวมของตัวแปรอิสระตัวที่ 1 และตัวที่ 2

$\sigma_{\varepsilon}^2$  คือ ความแปรปรวนของความคลาดเคลื่อน

$n$  คือ ขนาดตัวอย่าง

เนื่องจาก  $\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$  นั้น มีการแจกแจงปกติ ดังนั้น  $Y_i$  จะมีการแจกแจงล็อกนอร์มอล โดยที่มีความสัมพันธ์กับความคลาดเคลื่อน  $\varepsilon_i$  ที่มีการแจกแจงล็อกนอร์มอลและตัวแปรอิสระอีก 2 ตัว ได้แก่  $X_{1i}$  และ  $X_{2i}$  ที่มีการแจกแจงปกติ โดยจะทำการประมาณค่าพารามิเตอร์จากสมการ (26)

### 3.1.2 เงื่อนไขที่ทำการศึกษา

1) ขนาดตัวอย่าง ( $n$ ) คือ 30, 50, 100

2) เปอร์เซนต์ของตัวแปรตามที่ถูกตัดปลายทางขวา ( $r$ ) คือ 10%, 20%, 30%

3)  $\beta_0 = 0.3, \beta_1 = 1, \beta_2 = 1$

4) อัตราส่วนความแปรปรวนของตัวแปรอิสระตัวที่ 1 ต่อตัวแปรอิสระตัวที่ 2 ได้แก่

- 1:1 ( $\sigma_{x_1}^2 = 0.045, \sigma_{x_2}^2 = 0.045$ )

- 1:2 ( $\sigma_{x_1}^2 = 0.03, \sigma_{x_2}^2 = 0.06$ )

- 1:5 ( $\sigma_{x_1}^2 = 0.015, \sigma_{x_2}^2 = 0.075$ )

การกำหนดอัตราส่วนความแปรปรวนของตัวแปรอิสระตัวที่ 1 ต่อตัวแปรอิสระตัวที่ 2 เป็นอัตราส่วนดังกล่าว เนื่องจากต้องการศึกษากรณีที่ความแปรปรวนของตัวแปรอิสระทั้งสองเท่ากัน แตกต่างกันน้อย และแตกต่างกันมาก

5) อัตราส่วนความแปรปรวนรวมของตัวแปรอิสระต่อความแปรปรวนของความคลาดเคลื่อน ได้แก่

- 2:1 ( $\sigma_{x_1+x_2}^2 = 0.09, \sigma_{\varepsilon}^2 = 0.045$ )

- 1:1 ( $\sigma_{x_1+x_2}^2 = 0.09, \sigma_{\varepsilon}^2 = 0.09$ )

- 1:2 ( $\sigma_{x_1+x_2}^2 = 0.09, \sigma_{\varepsilon}^2 = 0.18$ )

การกำหนดอัตราส่วนความแปรปรวนรวมของตัวแปรอิสระต่อความแปรปรวนของความคลาดเคลื่อนเป็นอัตราส่วนดังกล่าวเนื่องจากต้องการศึกษาในกรณีที่ความคลาดเคลื่อนมีความแปรปรวนเท่ากับ มากกว่า และน้อยกว่าตัวแปรอิสระ

เขียนเป็นตารางสรุปกรณีที่ต้องการศึกษา ดังนี้

ตารางที่ 1  $\sigma_{x_1}^2$ ,  $\sigma_{x_2}^2$  และ  $\sigma_\varepsilon^2$  ที่ต้องการศึกษา

| กรณี | $\sigma_{x_1}^2$ | $\sigma_{x_2}^2$ | $\sigma_\varepsilon^2$ |
|------|------------------|------------------|------------------------|
| 1    | 0.015            | 0.075            | 0.045                  |
| 2    | 0.030            | 0.060            | 0.045                  |
| 3    | 0.045            | 0.045            | 0.045                  |
| 4    | 0.015            | 0.075            | 0.090                  |
| 5    | 0.030            | 0.060            | 0.090                  |
| 6    | 0.045            | 0.045            | 0.090                  |
| 7    | 0.015            | 0.075            | 0.180                  |
| 8    | 0.030            | 0.060            | 0.180                  |
| 9    | 0.045            | 0.045            | 0.180                  |

สำหรับการศึกษาค้างนี้ใช้โปรแกรม R เวอร์ชัน 3.4.4 ในการจำลองข้อมูลเพื่อเปรียบเทียบวิธีการประมาณค่าจากตัวแบบการถดถอยจำนวน 4 วิธี วิธีละ 81 สถานการณ์ และสถานการณ์ละ 10,000 รอบ

### 3.2 วิธีการดำเนินการวิจัย

#### 3.2.1 การจำลองข้อมูล

**ขั้นที่ 1** สร้าง  $X_{1i} \stackrel{iid}{\sim} N(0, \sigma_{x_1}^2)$ ,  $X_{2i} \stackrel{iid}{\sim} N(0, \sigma_{x_2}^2)$  โดยที่ตัวแปรอิสระ 2 ตัว ไม่มีความสัมพันธ์กัน และ  $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$  เมื่อ  $i=1,2,3,\dots, n$  ซึ่งความแปรปรวนของตัวแปรอิสระทั้ง 2 ตัวและความแปรปรวนของความคลาดเคลื่อนใช้เงื่อนไขตามตารางสรุปกรณีที่ต้องการศึกษาจำนวน 9 กรณี และขนาดตัวอย่าง ( $n$ ) มี 3 ขนาด ได้แก่ 30, 50, 100

**ขั้นที่ 2** สร้างค่าจริงของตัวแปรตามที่มีความสัมพันธ์กับตัวแปรอิสระ จากสมการ  $Y_i' = e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}}$  (27) ที่มีการแจกแจงลอการิธึมเมื่อ  $\beta_0 = 0.3$ ,  $\beta_1 = 1$ ,  $\beta_2 = 1$

ในงานวิจัยครั้งนี้เลือกกำหนดให้  $\beta_1 = 1$  และ  $\beta_2 = 1$  เพราะต้องการควบคุมปัจจัยในด้านความแปรปรวนที่เกิดจากตัวแปรอิสระ และ  $\beta_0 = 0.3$  เนื่องจากต้องการศึกษากรณีที่สัมประสิทธิ์การแปรผกผันซึ่งเป็นอัตราส่วนระหว่างส่วนเบี่ยงเบนมาตรฐานต่อค่าเฉลี่ยของข้อมูล มีค่ามากกว่า น้อยกว่า หรือเท่ากับ 1

**ขั้นที่ 3** สร้างค่าสังเกตของตัวแปรตามให้มีการแจกแจงลอการิธึมจากสมการ (25)

#### 3.2.2 การสร้างตัวแปรตามที่ประกอบไปด้วยข้อมูลที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 และข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1

**ขั้นที่ 1** หาค่า  $Y_c$  ซึ่งเป็นค่าเปอร์เซ็นต์ไทล์ที่  $100-r$  ของตัวแปรสุ่มที่มีการแจกแจงเหมือนกับ  $Y_i$

**ขั้นที่ 2** แบ่ง  $Y_i$  ออกเป็น 2 ส่วน ได้แก่  $Y_i$  ที่เป็นข้อมูลที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 ( $Y_i \leq Y_c$ ) และ  $Y_i$  ที่เป็นข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 ( $Y_i > Y_c$ ) จากนั้นสร้างชุดข้อมูลใหม่ของตัวแปรตาม ( $Y_i^*$ ) ที่ประกอบไปด้วยตัวแปรตามที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 และตัวแปรตามที่ถูกตัดปลายทางขวาประเภทที่ 1 ซึ่งมีเกณฑ์ในการสร้าง ดังนี้

- ถ้า  $Y_i$  เป็นข้อมูลที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 ให้  $Y_i^*$  ยังมีค่าเท่ากับ  $Y_i$  เหมือนเดิม
- ถ้า  $Y_i$  เป็นข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 ให้  $Y_i^*$  มีค่าเท่ากับ  $Y_c$

ตัวอย่างการสร้าง  $Y_i^*$

- 1) กำหนด  $n = 10$ ,  $r = 20\%$  และ  $\beta_0 = 0.3$ ,  $\beta_1 = 1$ ,  $\beta_2 = 1$
- 2) สร้าง  $X_{1i} \stackrel{iid}{\sim} N(0, 0.015)$ ,  $X_{2i} \stackrel{iid}{\sim} N(0, 0.075)$ ,  $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.045)$
- 3) สร้าง  $Y_i' = e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}}$  และ  $Y_i = e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i}$
- 4) หาค่า  $Y_c$  ซึ่งเป็นค่าเปอร์เซ็นต์ไทล์ที่  $100-r = 100-20 = 80$  ของตัวแปรสุ่มที่แจกแจงเหมือนกับ  $Y_i$  นั่นคือ ตัวแปรสุ่มที่มีการแจกแจง  $\text{LOGNOR}(\mu = 0.3, \sigma^2 = 0.135)$  ดังนั้น  $Y_c = 1.839$
- 5) แบ่ง  $Y_i$  ออกเป็น 2 ส่วน ได้แก่
  - $Y_i$  ที่เป็นข้อมูลที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 ( $Y_i \leq Y_c$ ) ซึ่งได้แก่  $Y_1, Y_2, Y_4, Y_5, Y_6, Y_9, Y_{10}$
  - $Y_i$  เป็นข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 ( $Y_i > Y_c$ ) ซึ่งได้แก่  $Y_3, Y_7, Y_8$
- 6) สร้างชุดข้อมูลใหม่ของตัวแปรตาม ( $Y_i^*$ ) ที่ประกอบไปด้วยตัวแปรตามที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 และตัวแปรตามที่ถูกตัดปลายทางขวาประเภทที่ 1 ซึ่งมีเกณฑ์การสร้างดังนี้
  - ถ้า  $Y_i$  เป็นข้อมูลที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 ซึ่งได้แก่  $Y_1, Y_2, Y_4, Y_5, Y_6, Y_9, Y_{10}$  จะกำหนดให้  $Y_i^*$  เมื่อ  $i = 1, 2, 4, 5, 6, 9, 10$  ยังมีค่าเท่ากับ  $Y_i$  ตัวเดิม หรือกล่าวได้ว่า  $Y_1^* = Y_1, Y_2^* = Y_2, Y_4^* = Y_4, Y_5^* = Y_5, Y_6^* = Y_6, Y_9^* = Y_9, Y_{10}^* = Y_{10}$
  - ถ้า  $Y_i$  เป็นข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 ซึ่งได้แก่  $Y_3, Y_7, Y_8$  จะกำหนดให้ค่า  $Y_i^*$  เมื่อ  $i = 3, 7, 8$  มีค่าเท่ากับ  $Y_c = 1.839$  หรือกล่าวได้ว่า  $Y_3^* = Y_7^* = Y_8^* = 1.839$

ตารางที่ 2 ตัวอย่างการสร้าง  $Y_i^*$

| i  | $X_{1i} \stackrel{iid}{\sim} N(0, 0.015)$ | $X_{2i} \stackrel{iid}{\sim} N(0, 0.075)$ | $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.045)$ | $Y_i'$ | $Y_i$ | $Y_i^*$ |
|----|-------------------------------------------|-------------------------------------------|--------------------------------------------------|--------|-------|---------|
| 1  | -0.082                                    | -0.416                                    | -0.338                                           | 0.820  | 0.585 | 0.585   |
| 2  | 0.051                                     | -0.427                                    | -0.095                                           | 0.927  | 0.843 | 0.843   |
| 3  | -0.245                                    | 0.191                                     | 0.434                                            | 1.279  | 1.974 | 1.839   |
| 4  | -0.263                                    | 0.115                                     | -0.141                                           | 1.164  | 1.011 | 1.011   |
| 5  | -0.049                                    | 0.118                                     | -0.053                                           | 1.447  | 1.372 | 1.372   |
| 6  | -0.206                                    | 0.129                                     | 0.050                                            | 1.250  | 1.314 | 1.314   |
| 7  | 0.045                                     | 0.664                                     | 0.376                                            | 2.743  | 3.995 | 1.839   |
| 8  | 0.063                                     | 0.358                                     | 0.010                                            | 2.057  | 2.077 | 1.839   |
| 9  | 0.019                                     | 0.239                                     | -0.001                                           | 1.747  | 1.745 | 1.745   |
| 10 | -0.013                                    | -0.332                                    | -0.193                                           | 0.956  | 0.788 | 0.788   |

### 3.2.3 การหาค่าประมาณของตัวแปรตาม

หาค่า  $\log(Y_i^*)$  ก่อนจากนั้นจึงประมาณค่าของ  $\beta_0, \beta_1, \beta_2$  โดยใช้วิธี OLS, CM, MLE และ MLE\_EM

ต่อมาหาค่าประมาณของตัวแปรตามของทั้ง 4 วิธี จากสมการ  $\hat{Y}_i = e^{\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}}$  (28)

### 3.2.4 การเปรียบเทียบวิธีการประมาณค่าจากตัวแบบการถดถอย

ในการประยุกต์ จะให้ความสนใจในการประมาณค่าตัวแปรตามมากกว่า ดังนั้นจึงสรุปผลการวิจัย

จากค่า  $AMSE(\hat{Y}_a)$  และ  $RE(\hat{Y}_a)$  เมื่อ

1)  $AMSE(\hat{Y}_a)$  คือ ค่าเฉลี่ยของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนของตัวแปรตามของวิธี a

$$\text{จากการจำลองสถานการณ์ 10,000 รอบ โดยที่ } AMSE(\hat{Y}_a) = \frac{\sum_{j=1}^{10000} MSE(\hat{Y}_a^{(j)})}{10,000} \quad (29)$$

เมื่อ  $MSE(\hat{Y}_a^{(j)})$  คือ ค่าเฉลี่ยกำลังสองของค่าความคลาดเคลื่อนของตัวแปรตามจากการจำลอง

$$\text{สถานการณ์รอบที่ } j \text{ ของวิธี } a \text{ โดยที่ } MSE(\hat{Y}_a^{(j)}) = \frac{\sum_{i=1}^n (\hat{Y}_{i,a} - Y_i')^2}{n} \quad (30)$$

$\hat{Y}_{i,a}$  คือ ค่าประมาณของตัวแปรตามลำดับที่ i โดยวิธีการ a

$Y_i'$  คือ ค่าที่แท้จริงของตัวแปรตามลำดับที่ i

j คือ รอบที่ของการจำลองสถานการณ์ โดย  $j = 1, 2, \dots, 10000$

a คือ วิธีการประมาณค่าจากตัวแบบการถดถอย ซึ่ง  $a = \text{OLS, CM, MLE, MLE\_EM}$

2)  $RE(\hat{Y}_a)$  คือ อัตราส่วนระหว่าง  $AMSE(\hat{Y}_{MLE})$  กับ  $AMSE(\hat{Y}_a)$  หรือเป็นอัตราส่วน

$$\text{ระหว่าง } AMSE(\hat{Y}) \text{ ของวิธี MLE กับ } AMSE(\hat{Y}) \text{ ของวิธี } a \text{ ซึ่ง}$$

$$RE(\hat{Y}_a) = \frac{AMSE(\hat{Y}_{MLE})}{AMSE(\hat{Y}_a)} \quad (31)$$

สำหรับการศึกษานี้เลือกวิธี MLE เป็นเกณฑ์ในการเปรียบเทียบเพราะว่า วิธี MLE เป็นวิธีการประมาณค่าพื้นฐานที่มีประสิทธิภาพสูง และเป็นวิธีที่ใช้ข้อมูลได้ครบถ้วนทั้งส่วนที่เป็นข้อมูลที่ถูกต้องปลายทางขอบประเภทที่ 1 และข้อมูลที่ไม่วัดปลายทางขอบประเภทที่ 1 รวมถึงเป็นตัวประมาณที่ยืนยงภายใต้การเปลี่ยนแปลงที่เหมาะสม

### เกณฑ์การเปรียบเทียบ

กำหนดให้ a และ b คือ วิธี OLS, MLE, CM, MLE\_EM โดยที่  $a \neq b$

1) วิธีที่มีค่า  $AMSE(\hat{Y}_a)$  ต่ำสุดจะเป็นวิธีที่มีประสิทธิภาพสูงสุดในการประมาณค่าจากตัวแบบการถดถอย

2) วิธีการที่มีค่า  $RE(\hat{Y}_a)$  มากกว่า 1 แสดงว่าวิธีการ a มีประสิทธิภาพมากกว่าวิธี MLE

3) วิธีการที่มีค่า  $RE(\hat{Y}_a)$  เท่ากับ 1 แสดงว่าวิธีการ a มีประสิทธิภาพเท่ากับวิธี MLE

4) วิธีการที่มีค่า  $RE(\hat{Y}_a)$  น้อยกว่า 1 แสดงว่าวิธีการ a มีประสิทธิภาพน้อยกว่าวิธี MLE

5) วิธีที่มีประสิทธิภาพสูงสุดในการประมาณค่าจากตัวแบบการถดถอย คือ วิธีที่มีค่า  $RE(\hat{Y}_a)$  มากที่สุด

หรือวิธีที่มีค่า  $RE(\hat{Y}_b)$  น้อยกว่าค่า  $RE(\hat{Y}_a)$  ที่มีค่ามากที่สุดสถานการณ์เดียวกันไม่เกิน 0.01

#### 4. ผลการวิจัย

ในการพิจารณาวิธีที่มีประสิทธิภาพสูงสุด จะพิจารณาจากค่า  $RE(\hat{Y}_a)$  เป็นหลัก โดยถ้าพิจารณาจากค่า  $AMSE(\hat{Y}_a)$  ก็จะได้ผลสรุปที่เหมือนกัน

ตารางที่ 3 ค่า  $RE(\hat{Y}_a)$  เมื่อขนาดตัวอย่างเท่ากับ 30 และ a คือ วิธี OLS, MLE, CM, MLE\_EM

| n  | $\sigma_{x_1+x_2}^2 : \sigma_{\varepsilon}^2$ | r   | $\sigma_{x_1}^2 : \sigma_{x_2}^2$ | $RE(\hat{Y}_a)$ |             |             |             |
|----|-----------------------------------------------|-----|-----------------------------------|-----------------|-------------|-------------|-------------|
|    |                                               |     |                                   | a = OLS         | MLE         | CM          | MLE_EM      |
| 30 | 2:1                                           | 10% | 1:5                               | 1.01            | 1.00        | <b>1.10</b> | 1.01        |
|    |                                               |     | 1:2                               | 0.97            | 1.00        | <b>1.09</b> | 1.01        |
|    |                                               |     | 1:1                               | 0.97            | 1.00        | <b>1.08</b> | 1.01        |
|    |                                               | 20% | 1:5                               | 0.61            | <b>1.00</b> | 0.95        | <b>1.01</b> |
|    |                                               |     | 1:2                               | 0.61            | <b>1.00</b> | 0.95        | <b>1.00</b> |
|    |                                               |     | 1:1                               | 0.59            | <b>1.00</b> | 0.93        | <b>1.00</b> |
|    |                                               | 30% | 1:5                               | 0.41            | <b>1.00</b> | 0.78        | <b>1.01</b> |
|    |                                               |     | 1:2                               | 0.42            | <b>1.00</b> | 0.80        | <b>1.01</b> |
|    |                                               |     | 1:1                               | 0.40            | <b>1.00</b> | 0.76        | <b>1.01</b> |
| 30 | 1:1                                           | 10% | 1:5                               | <b>1.30</b>     | 1.00        | 1.25        | 1.01        |
|    |                                               |     | 1:2                               | <b>1.28</b>     | 1.00        | 1.23        | 1.01        |
|    |                                               |     | 1:1                               | <b>1.28</b>     | 1.00        | 1.23        | 1.01        |
|    |                                               | 20% | 1:5                               | 0.91            | 1.00        | <b>1.05</b> | 1.00        |
|    |                                               |     | 1:2                               | 0.95            | 1.00        | <b>1.09</b> | 1.00        |
|    |                                               |     | 1:1                               | 0.91            | 1.00        | <b>1.06</b> | 1.00        |
|    |                                               | 30% | 1:5                               | 0.67            | <b>1.00</b> | 0.88        | <b>1.01</b> |
|    |                                               |     | 1:2                               | 0.70            | <b>1.00</b> | 0.92        | <b>1.01</b> |
|    |                                               |     | 1:1                               | 0.68            | <b>1.00</b> | 0.90        | <b>1.01</b> |
|    | 1:2                                           | 10% | 1:5                               | <b>1.46</b>     | 1.00        | 1.38        | 1.00        |
|    |                                               |     | 1:2                               | <b>1.52</b>     | 1.00        | 1.41        | 1.00        |
|    |                                               |     | 1:1                               | <b>1.56</b>     | 1.00        | 1.46        | 1.00        |
|    |                                               | 20% | 1:5                               | <b>1.33</b>     | 1.00        | <b>1.32</b> | 1.00        |
|    |                                               |     | 1:2                               | <b>1.34</b>     | 1.00        | <b>1.33</b> | 1.00        |
|    |                                               |     | 1:1                               | <b>1.36</b>     | 1.00        | <b>1.35</b> | 1.00        |
|    |                                               | 30% | 1:5                               | 1.05            | 1.00        | <b>1.14</b> | 1.00        |
|    |                                               |     | 1:2                               | 1.02            | 1.00        | <b>1.11</b> | 1.00        |
|    |                                               |     | 1:1                               | 1.06            | 1.00        | <b>1.17</b> | 1.00        |

ตารางที่ 4 ค่า  $RE(\hat{Y}_a)$  เมื่อขนาดตัวอย่างเท่ากับ 50 และ a คือ วิธี OLS, MLE, CM, MLE\_EM

| n  | $\sigma_{x_1+x_2}^2 : \sigma_{\varepsilon}^2$ | r   | $\sigma_{x_1}^2 : \sigma_{x_2}^2$ | $RE(\hat{Y}_a)$ |      |      |        |
|----|-----------------------------------------------|-----|-----------------------------------|-----------------|------|------|--------|
|    |                                               |     |                                   | a = OLS         | MLE  | CM   | MLE_EM |
| 50 | 2:1                                           | 10% | 1:5                               | 0.78            | 1.00 | 0.97 | 1.00   |
|    |                                               |     | 1:2                               | 0.78            | 1.00 | 0.97 | 1.00   |
|    |                                               |     | 1:1                               | 0.77            | 1.00 | 0.96 | 1.00   |
|    |                                               | 20% | 1:5                               | 0.39            | 1.00 | 0.69 | 1.00   |
|    |                                               |     | 1:2                               | 0.38            | 1.00 | 0.68 | 1.00   |
|    |                                               |     | 1:1                               | 0.38            | 1.00 | 0.69 | 1.00   |
|    |                                               | 30% | 1:5                               | 0.25            | 1.00 | 0.51 | 1.01   |
|    |                                               |     | 1:2                               | 0.24            | 1.00 | 0.50 | 1.01   |
|    |                                               |     | 1:1                               | 0.24            | 1.00 | 0.50 | 1.01   |
| 50 | 1:1                                           | 10% | 1:5                               | 1.06            | 1.00 | 1.09 | 1.00   |
|    |                                               |     | 1:2                               | 1.05            | 1.00 | 1.08 | 1.00   |
|    |                                               |     | 1:1                               | 1.06            | 1.00 | 1.09 | 1.00   |
|    |                                               | 20% | 1:5                               | 0.61            | 1.00 | 0.77 | 1.00   |
|    |                                               |     | 1:2                               | 0.62            | 1.00 | 0.77 | 1.00   |
|    |                                               |     | 1:1                               | 0.62            | 1.00 | 0.77 | 1.00   |
|    |                                               | 30% | 1:5                               | 0.38            | 1.00 | 0.53 | 1.00   |
|    |                                               |     | 1:2                               | 0.38            | 1.00 | 0.54 | 1.00   |
|    |                                               |     | 1:1                               | 0.38            | 1.00 | 0.53 | 1.00   |
|    | 1:2                                           | 10% | 1:5                               | 1.26            | 1.00 | 1.23 | 1.00   |
|    |                                               |     | 1:2                               | 1.27            | 1.00 | 1.24 | 1.00   |
|    |                                               |     | 1:1                               | 1.24            | 1.00 | 1.21 | 1.00   |
|    |                                               | 20% | 1:5                               | 0.88            | 1.00 | 0.93 | 1.00   |
|    |                                               |     | 1:2                               | 0.90            | 1.00 | 0.95 | 1.00   |
|    |                                               |     | 1:1                               | 0.90            | 1.00 | 0.95 | 1.00   |
|    |                                               | 30% | 1:5                               | 0.60            | 1.00 | 0.67 | 1.00   |
|    |                                               |     | 1:2                               | 0.59            | 1.00 | 0.66 | 1.00   |
|    |                                               |     | 1:1                               | 0.61            | 1.00 | 0.69 | 1.00   |

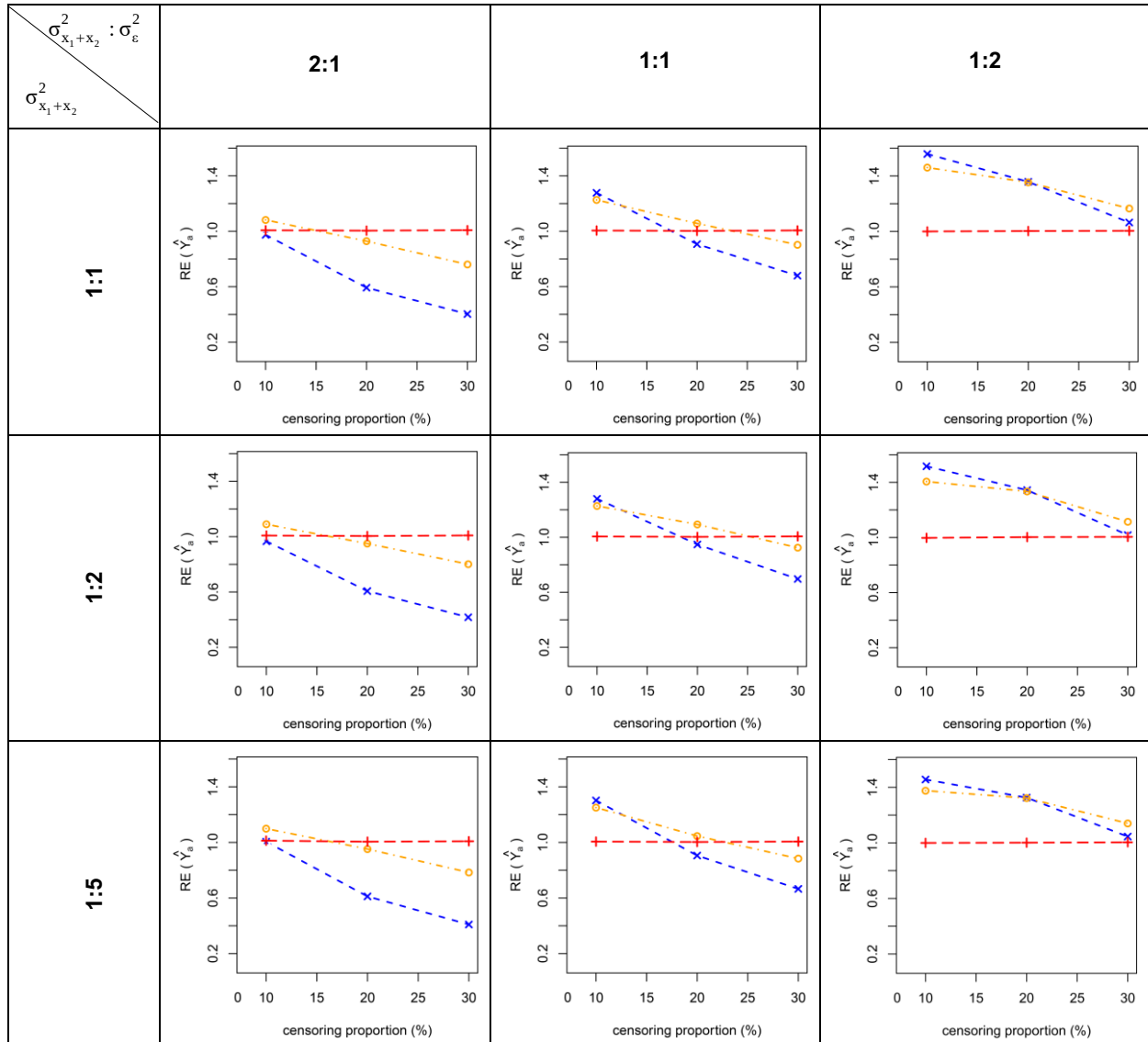
ตารางที่ 5 ค่า  $RE(\hat{Y}_a)$  เมื่อขนาดตัวอย่างเท่ากับ 100 และ a คือ วิธี OLS, MLE, CM, MLE\_EM

| n   | $\sigma_{x_1+x_2}^2 : \sigma_{\varepsilon}^2$ | r   | $\sigma_{x_1}^2 : \sigma_{x_2}^2$ | $RE(\hat{Y}_a)$ |      |      |        |
|-----|-----------------------------------------------|-----|-----------------------------------|-----------------|------|------|--------|
|     |                                               |     |                                   | a = OLS         | MLE  | CM   | MLE_EM |
| 100 | 2:1                                           | 10% | 1:5                               | 0.52            | 1.00 | 0.76 | 1.00   |
|     |                                               |     | 1:2                               | 0.53            | 1.00 | 0.76 | 1.00   |
|     |                                               |     | 1:1                               | 0.53            | 1.00 | 0.76 | 1.00   |
|     |                                               | 20% | 1:5                               | 0.21            | 1.00 | 0.41 | 1.00   |
|     |                                               |     | 1:2                               | 0.21            | 1.00 | 0.42 | 1.00   |
|     |                                               |     | 1:1                               | 0.21            | 1.00 | 0.42 | 1.00   |
|     |                                               | 30% | 1:5                               | 0.12            | 1.00 | 0.27 | 1.01   |
|     |                                               |     | 1:2                               | 0.12            | 1.00 | 0.26 | 1.01   |
|     |                                               |     | 1:1                               | 0.12            | 1.00 | 0.26 | 1.01   |
| 100 | 1:1                                           | 10% | 1:5                               | 0.76            | 1.00 | 0.84 | 1.00   |
|     |                                               |     | 1:2                               | 0.75            | 1.00 | 0.83 | 1.00   |
|     |                                               |     | 1:1                               | 0.78            | 1.00 | 0.85 | 1.00   |
|     |                                               | 20% | 1:5                               | 0.34            | 1.00 | 0.45 | 1.00   |
|     |                                               |     | 1:2                               | 0.34            | 1.00 | 0.45 | 1.00   |
|     |                                               |     | 1:1                               | 0.35            | 1.00 | 0.46 | 1.00   |
|     |                                               | 30% | 1:5                               | 0.19            | 1.00 | 0.27 | 1.00   |
|     |                                               |     | 1:2                               | 0.19            | 1.00 | 0.27 | 1.00   |
|     |                                               |     | 1:1                               | 0.19            | 1.00 | 0.27 | 1.00   |
|     | 1:2                                           | 10% | 1:5                               | 0.98            | 1.00 | 0.99 | 1.00   |
|     |                                               |     | 1:2                               | 0.97            | 1.00 | 0.98 | 1.00   |
|     |                                               |     | 1:1                               | 0.98            | 1.00 | 0.98 | 1.00   |
|     |                                               | 20% | 1:5                               | 0.52            | 1.00 | 0.57 | 1.00   |
|     |                                               |     | 1:2                               | 0.54            | 1.00 | 0.58 | 1.00   |
|     |                                               |     | 1:1                               | 0.52            | 1.00 | 0.57 | 1.00   |
|     |                                               | 30% | 1:5                               | 0.30            | 1.00 | 0.34 | 1.00   |
|     |                                               |     | 1:2                               | 0.30            | 1.00 | 0.34 | 1.00   |
|     |                                               |     | 1:1                               | 0.29            | 1.00 | 0.33 | 1.00   |

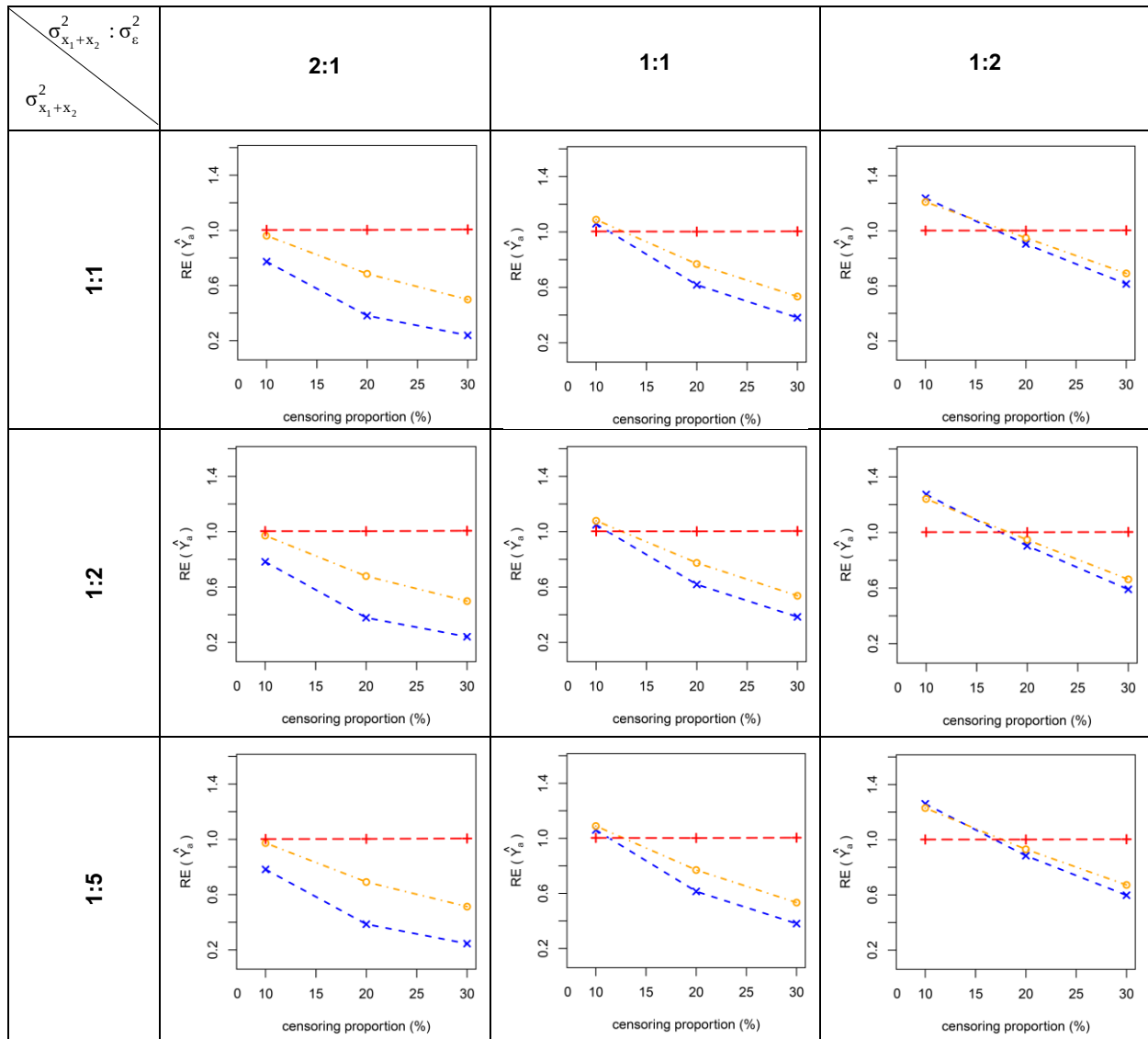
ส่วนต่อไปเป็นส่วนของกราฟแสดงค่า  $RE(\hat{Y}_a)$  จึงกำหนดเส้นกราฟแทนวิธีการประมาณค่าที่ใช้ในงานวิจัย ดังนี้



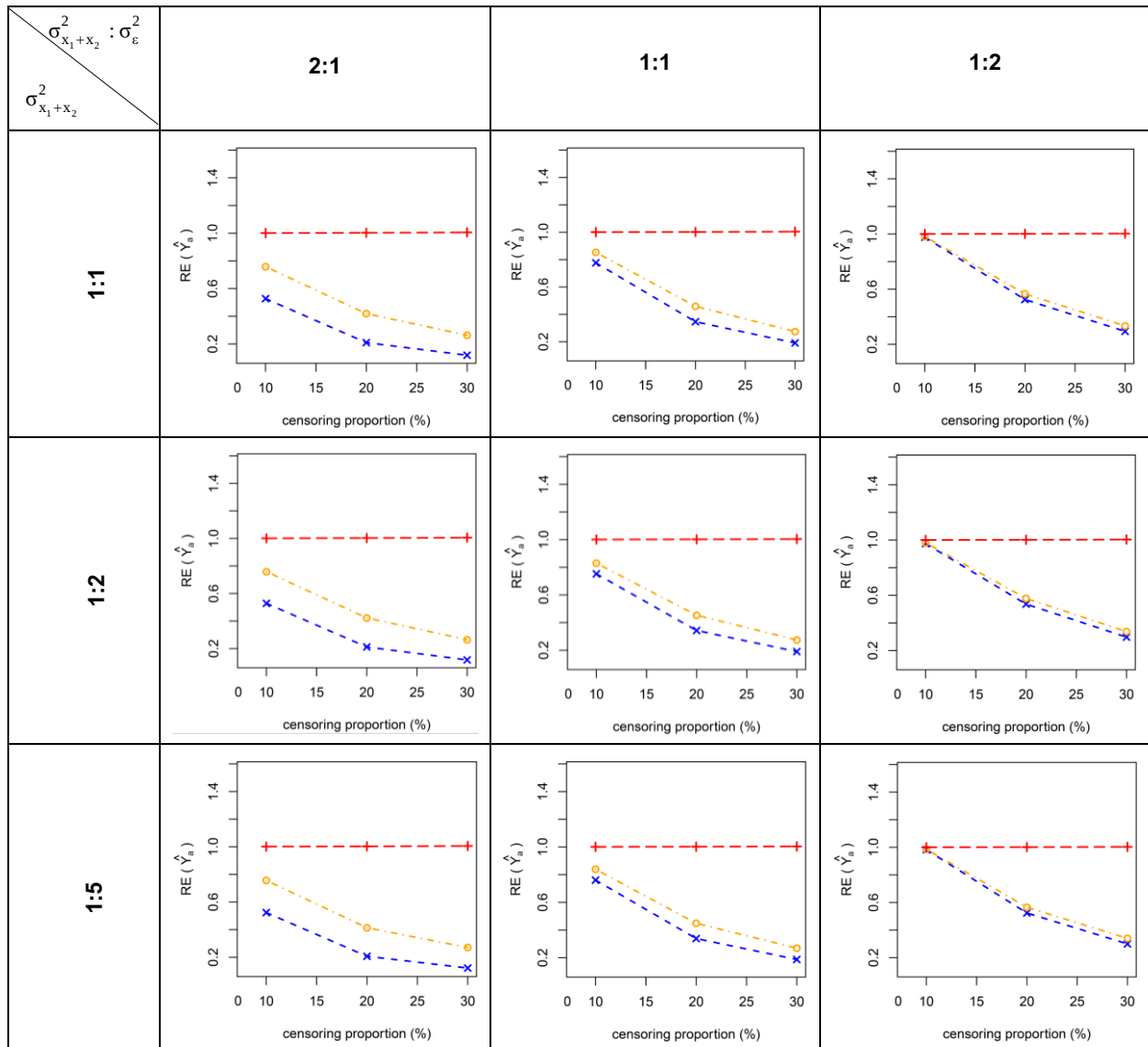
รูปภาพที่ 1 ค่า  $RE(\hat{Y}_a)$  เมื่อขนาดตัวอย่างเท่ากับ 30 และ  $a$  คือ วิธี OLS, CM, MLE\_EM



รูปภาพที่ 2 ค่า  $RE(\hat{Y}_a)$  เมื่อขนาดตัวอย่างเท่ากับ 50 และ  $a$  คือ วิธี OLS, CM, MLE\_EM



รูปภาพที่ 3 ค่า  $RE(\hat{Y}_a)$  เมื่อขนาดตัวอย่างเท่ากับ 100 และ  $\sigma^2_{x_1+x_2} : \sigma^2_{\epsilon}$  คือ วิธี OLS, CM, MLE\_EM



ตารางที่ 6 วิธีที่มีประสิทธิภาพสูงสุดในการประมาณค่าพารามิเตอร์จากตัวแบบการถดถอย

| $r$ | 10% |     |     | 20% |     |     | 30% |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $n$ |     |     |     |     |     |     |     |     |     |
| 30  | 1:2 | 1:1 | 2:1 | 1:2 | 1:1 | 2:1 | 1:2 | 1:1 | 2:1 |
| 50  | 1:2 | 1:1 | 2:1 | 1:2 | 1:1 | 2:1 | 1:2 | 1:1 | 2:1 |
| 100 | 1:2 | 1:1 | 2:1 | 1:2 | 1:1 | 2:1 | 1:2 | 1:1 | 2:1 |

\*\*\*หมายเหตุ  $\sigma^2_{x_1+x_2} : \sigma^2_{\epsilon} = 1:2, 1:1, 2:1$

|  |     |  |    |  |                |  |            |
|--|-----|--|----|--|----------------|--|------------|
|  | OLS |  | CM |  | MLE และ MLE_EM |  | OLS และ CM |
|--|-----|--|----|--|----------------|--|------------|

จากตารางที่ 3-5 และรูปภาพที่ 1 - 3 พบว่า เมื่อตัวอย่างมีขนาดใหญ่ ( $n=100$ ) วิธี MLE กับ วิธี MLE\_EM มีประสิทธิภาพสูงสุดในการประมาณค่าพารามิเตอร์จากตัวแบบการถดถอย

เมื่อตัวอย่างขนาดปานกลาง ( $n = 50$ ) และตัวแปรตามถูกตัดปลายทางขวาปานกลางถึงมาก ( $r = 20\%, 30\%$ ) พบว่า วิธี MLE และวิธี MLE\_EM ยังคงเป็นวิธีที่มีประสิทธิภาพสูงสุด แต่เมื่อตัวอย่างมีขนาดปานกลาง ( $n = 50$ ) และตัวแปรตามถูกตัดปลายทางขวาน้อย ( $r = 10\%$ ) พบว่า อัตราส่วนการกระจายตัวของความคลาดเคลื่อนต่อการกระจายตัวของตัวแปรอิสระมีผลต่อการประมาณค่าพารามิเตอร์จากตัวแบบการถดถอย โดยที่วิธี MLE และวิธี MLE\_EM นั้นมีประสิทธิภาพสูงสุด เมื่อความคลาดเคลื่อนกระจายตัวน้อยกว่าตัวแปรอิสระ ส่วนวิธี CM มีประสิทธิภาพสูงสุด เมื่อความคลาดเคลื่อนกระจายตัวเท่ากับตัวแปรอิสระ และวิธี OLS มีประสิทธิภาพสูงสุด เมื่อความคลาดเคลื่อนกระจายตัวมากกว่าตัวแปรอิสระ

เมื่อตัวอย่างมีขนาดเล็ก ( $n = 30$ ) พบว่า ทั้งอัตราส่วนการกระจายตัวของความคลาดเคลื่อนต่อการกระจายตัวของตัวแปรอิสระ และร้อยละการถูกตัดปลายของตัวแปรตามล้วนมีผลต่อการประมาณค่าพารามิเตอร์จากตัวแบบการถดถอย โดยที่วิธี MLE และวิธี MLE\_EM มีประสิทธิภาพสูงสุด ในกรณีที่ความคลาดเคลื่อนตัวน้อยกว่าตัวแปรอิสระและตัวแปรตามถูกตัดปลายทางขวาปานกลางถึงมาก ( $r = 20\%, 30\%$ ) และในกรณีที่ความคลาดเคลื่อนนั้นกระจายตัวเท่ากับตัวแปรอิสระและตัวแปรตามถูกตัดปลายทางขวามาก ( $r = 30\%$ ) ส่วนวิธี OLS นั้นมีประสิทธิภาพสูงสุด ในกรณีที่ตัวแปรตามถูกตัดปลายทางขวาน้อย ( $r = 10\%$ ) และความคลาดเคลื่อนมีการกระจายตัวมากกว่าหรือเท่ากับตัวแปรอิสระ ส่วนวิธี CM มีประสิทธิภาพสูงสุดใน 3 กรณี ได้แก่ กรณีที่ตัวแปรตามถูกตัดปลายทางขวาน้อย ( $r = 10\%$ ) และความคลาดเคลื่อนกระจายตัวน้อยกว่าตัวแปรอิสระ และกรณีที่ตัวแปรตามถูกตัดปลายทางขวาปานกลาง ( $r = 20\%$ ) และความคลาดเคลื่อนกระจายตัวเท่ากับตัวแปรอิสระ รวมถึงกรณีที่ตัวแปรตามถูกตัดปลายทางขวามาก ( $r = 30\%$ ) และความคลาดเคลื่อนกระจายตัวมากกว่าตัวแปรอิสระ รวมถึงพบว่าทั้งวิธี OLS และวิธี CM เป็นวิธีที่มีประสิทธิภาพสูงสุดเมื่อตัวแปรตามถูกตัดปลายทางขวาปานกลาง ( $r = 20\%$ ) และความคลาดเคลื่อนกระจายตัวมากกว่าตัวแปรอิสระ

นอกจากนั้นพบว่า เมื่อตัวอย่างใหญ่ขึ้น ทุกวิธีต่างมีประสิทธิภาพในการประมาณค่าพารามิเตอร์จากตัวแบบการถดถอยมากขึ้น โดยเมื่อตัวอย่างขนาดใหญ่ วิธี OLS และวิธี CM มีประสิทธิภาพต่ำกว่าวิธี MLE และวิธี MLE\_EM แต่เมื่อความคลาดเคลื่อนกระจายตัวมากกว่าตัวแปรอิสระ และตัวแปรตามถูกตัดปลายทางขวาน้อย ( $r = 10\%$ ) ทั้งวิธี OLS และวิธี CM จะมีประสิทธิภาพใกล้เคียงกับวิธี MLE และวิธี MLE\_EM มากขึ้น และพบว่าอัตราส่วนการกระจายตัวของตัวแปรอิสระต่อการกระจายของความคลาดเคลื่อนนั้นมีผลต่อการประมาณค่าพารามิเตอร์จากตัวแบบการถดถอย โดยเฉพาะอย่างยิ่งเมื่อความคลาดเคลื่อนมีการกระจายตัวมากกว่าตัวแปรอิสระมากขึ้น จะทำให้ทุกวิธีมีประสิทธิภาพลดลง แต่วิธี CM และวิธี OLS จะมีประสิทธิภาพดีขึ้นเมื่อเทียบกับวิธี MLE และวิธี MLE\_EM รวมถึงวิธี OLS จะมีประสิทธิภาพใกล้เคียงวิธี CM มากยิ่งขึ้น ส่วนเมื่อตัวแปรตามถูกตัดปลายทางขวาน้อยลงจะทำให้ทุกวิธีมีประสิทธิภาพได้ดีขึ้น โดยที่วิธี OLS กับวิธี CM จะมีประสิทธิภาพดีขึ้นเมื่อเทียบกับวิธี MLE และวิธี MLE\_EM ทั้งนี้ อัตราส่วนการกระจายของตัวแปรอิสระตัวที่ 1 ต่อตัวแปรอิสระตัวที่ 2 ไม่มีผลต่อการประมาณค่าพารามิเตอร์จากตัวแบบการถดถอย

## 5. การสรุปและอภิปรายผลการวิจัย

วิธี OLS เป็นวิธีที่คำนวณได้ง่าย แต่วิธี OLS ไม่ได้ให้ความสำคัญกับข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 เพราะในการคำนวณถือว่าตัวแปรตามที่ถูกตัดปลายทางขวาประเภทที่ 1 เหมือนเป็นตัวแปรตามที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 ดังนั้น ค่าประมาณของตัวแปรตามที่ได้จะต่ำกว่าค่าแท้จริงของตัวแปรตาม ส่วนวิธี CM เป็นวิธีการที่นำวิธี OLS มาประยุกต์ใช้และมีการคำนวณที่ซับซ้อนมากกว่าวิธี OLS และให้ความสำคัญกับตัวแปรตามที่ถูกตัดปลายทางขวาประเภทที่ 1 มากกว่าวิธี OLS เพราะมีการจัดการกับตัวแปรตามที่ถูกตัดปลายทางขวาประเภทที่ 1 ก่อนนำไปประมาณค่าพารามิเตอร์ แต่การจัดการกับตัวแปรตามไม่ได้มีความซับซ้อนในการคำนวณเท่ากับวิธี MLE และวิธี MLE\_EM ดังนั้นถือว่า CM ใช้ข้อมูลในส่วนของตัวแปรตามที่ถูกตัดปลายทางขวาประเภทที่ 1 แต่ยังไม่ครอบคลุมสมบูรณ์ ส่วนวิธี MLE และ วิธี MLE\_EM นี้ได้ให้ความสำคัญและใช้ข้อมูลทั้งสองส่วน ได้แก่ ข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 และข้อมูลที่ไม่ถูกตัดปลายทางขวาประเภทที่ 1 ได้ครบสมบูรณ์ ดังนั้นส่วนใหญ่ทั้งสองวิธีนี้จึงมีประสิทธิภาพในการประมาณค่าจากตัวแบบการถดถอยเท่ากัน โดยวิธี MLE เป็นวิธีที่มีประสิทธิภาพสูงแต่มีความซับซ้อนในการคำนวณ ส่วนวิธี MLE\_EM เป็นวิธีที่ได้ประยุกต์ใช้วิธี MLE และมีการคำนวณง่ายกว่าวิธี MLE

และจากผลการวิจัย สรุปได้ว่า วิธี MLE และวิธี MLE\_EM นั้นมีประสิทธิภาพสูงสุดเมื่อตัวอย่างมีขนาดใหญ่ ( $n = 100$ ) หรือตัวแปรตามถูกตัดปลายทางขวามาก ( $r = 30\%$ ) ในทางกลับกัน วิธี OLS เป็นวิธีที่มีประสิทธิภาพสูงสุดเมื่อตัวอย่างมีขนาดเล็ก ( $n = 30$ ) หรือตัวแปรตามถูกตัดปลายทางขวาน้อย ( $r = 10\%$ ) และวิธี CM จะมีประสิทธิภาพสูงที่สุดในสถานการณ์ที่เหลือ กล่าวคือ เมื่อตัวอย่างมีขนาดปานกลาง ( $n = 50$ ) หรือตัวแปรตามถูกตัดปลายทางขวาน้อยลง ( $r = 20\%$ ) นอกจากนี้พบว่า ทุกวิธีมีประสิทธิภาพมากขึ้น เมื่อตัวอย่างมีขนาดใหญ่ขึ้นหรือตัวแปรตามถูกตัดปลายทางขวาน้อยลงหรือความคลาดเคลื่อนกระจายตัวน้อยกว่าตัวแปรอิสระ

## 6. ข้อเสนอแนะ

- 6.1 สามารถนำผลการวิจัยครั้งนี้ใช้เป็นแนวทางในการเลือกตัวประมาณที่เหมาะสมกับข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 เพื่อให้ค่าประมาณของตัวแปรตามใกล้เคียงกับค่าที่แท้จริงของตัวแปรตาม
- 6.2 ตัวแปรตามที่ได้จากการประมาณสามารถนำไปวิเคราะห์ทางสถิติต่อไปได้ เช่น การหาค่าเฉลี่ย เป็นต้น
- 6.3 สามารถนำผลการวิจัยครั้งนี้ใช้เป็นแนวทางการศึกษากรณีอื่นๆ ดังนี้
  - 1) ตัวแปรตามมีการแจกแจงอื่น เช่น แกมมา เอ็กซ์โปเนนเชียล เป็นต้น
  - 2) ตัวแปรตามเป็นข้อมูลแบบอื่น เช่น ข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 2 ข้อมูลที่ถูกตัดปลายทางแบบสุ่ม ข้อมูลที่ถูกตัดปลายทางซ้าย เป็นต้น
  - 3) ตัวแปรอิสระมีมากกว่า 2 ตัว หรือตัวแปรอิสระมีความสัมพันธ์กันในระดับต่างๆ
  - 4) เพิ่มระดับอัตราส่วนของความแปรปรวนระหว่างตัวแปรอิสระกับค่าความคลาดเคลื่อน

## 7. เอกสารอ้างอิง

- จำเนียง จันทร์รักษ์. (2539). การพยากรณ์ในสมการถดถอยเชิงเส้นพหุเมื่อค่าตัวแปรตามถูกตัดปลายทางขวา. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ สาขาวิชาการประกันภัย บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- ธีระพร ธีระถาวร. (2536). การอนุมานเชิงสถิติขั้นกลาง : โครงสร้างและความหมาย. (พิมพ์ครั้งที่ 2). กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- บังอร กุมพล. (2539). การวิเคราะห์การถดถอยเชิงเส้นพหุ เมื่อตัวแปรตามมีค่าถูกตัดทั้งทางขวากรณีค่าตัดทั้งประเภทที่ 1. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ สาขาวิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- พีรวัฒน์ เสรีวัฒนกุล. (2555). การเปรียบเทียบวิธีการประมาณสำหรับตัวแบบการถดถอยเชิงเส้นพหุที่ค่าความคลาดเคลื่อนมีความแปรปรวนไม่คงที่. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย.
- อุษณีย์ วงศ์อำมรต์. (2555). การเปรียบเทียบวิธีการประมาณค่าสูญหายแบบหอนอกนอร์เรเบิลในการวิเคราะห์การถดถอยเชิงเส้นพหุ. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย.
- Aitkin, M. (1981, May). A Note on the Regression Analysis of Censored Data. **Technometrics**, 23(2), 161-163.
- Chatterjee, S., and McLeish, D.L. (1986, February). Fitting Linear Regression to Censored Data by Least Squares and Maximum Likelihood Methods. **Communications in Statistics - Theory and Methods**, 15(1), 3227-3243.
- Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977, January). Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm. **Journal of the Royal Statistical Society Series B (Methodological)**, 39(1), 1-38.
- Fisher, R.A. (1922, January). On the Mathematical Foundations of Theoretical Statistic. **Philosophical Transactions of the Royal Society of London Series A**, 222(1), 309-368.
- Gauss, C.F. (1809). **Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium**. Retrieved on 11 November 2017 from [https://archive.org/details/bub\\_gb\\_ORUOAAAAQAAJ](https://archive.org/details/bub_gb_ORUOAAAAQAAJ)
- Jöreskog, K.G. (2002). **Censored Variables and Censored Regression**. Retrieved on 11 November 2017 from <http://www.ssicentral.com/lisrel/techdocs/censor.pdf>
- Klein, J.P., and Moeschberger, M.L. (2003). **Survival Analysis Techniques for Censored and Truncated Data**. (Second Edition). New York: Springer-Verlag New York, Inc.
- Smith, B.E., and Merceret, F.J., (2000, September). The Lognormal Distribution. **The College Mathematics Journal**, 31(4), 259-261.
- Tobin, J. (1958, January). Estimation of Relationships for Limited Dependent Variables. **Econometrica**, 26(1), 24-36.