การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ด้วยวิธีเบย์เซียน วิธีกำลังสองน้อยที่สุด และวิธีบูตสแตรปแบบใช้พารามิเตอร์ สำหรับตัวแบบถดถอยเชิงเส้นอย่างง่าย

A Comparison of Parameter Estimations of Bayesian, Ordinary Least Square and Parametric Bootstrap Methods for Simple Linear Regression Model

ภูวษา แซ่อุ้ย*, ธิดาพร ศุภภากร และ ประสิทธิ์ พยัคฆพงษ์

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

แขวงลาดยาว เขตจตุจักร กรุงเทพมหานคร 10900

Puwasa Sae-ui*, Thidaporn Supapakorn and Prasit Payakkapong

Department of Statistics, Faculty of Science, Kasetsart University,

Ladyao, Chatuchak, Bangkok 10900

บทคัดย่อ

งานวิจัยนี้มีจุดประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์แบบจุดด้วยวิธีเบย์เซียน (Bayesian method) วิธีกำลังสองน้อยที่สุด (ordinary least square method) และวิธีบูตสแตรปแบบใช้พารามิเตอร์ (parametric bootstrap method) ในการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย เกณฑ์ในการพิจารณาประสิทธิภาพของวิธีการประมาณค่า คือ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) ขนาดตัวอย่างที่ใช้ในการศึกษา คือ 15, 30, 75 และ 100 การแจกแจงของความคลาดเคลื่อนสุ่ม คือ การแจกแจงแบบปกติมีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.15, 0.5 และ 1 ตัวแปรอิสระเป็นค่าคงที่ซึ่งสุ่มมาจากการแจกแจงปกติด้วยค่าเฉลี่ยเท่ากับ 1 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.1 ข้อมูลที่ใช้ในการวิจัยได้จากการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลกระทำซ้ำ 1000 ครั้งในแต่ละสถานการณ์ ซึ่งผลการวิจัยสรุปได้ดังนี้ วิธีเบย์เซียนมีประสิทธิภาพดีที่สุดสำหรับทุก ๆ สถานการณ์ที่กำหนด รองลงมาคือวิธีกำลังสองน้อยที่สุดและวิธีบูตสแตรปแบบใช้พารามิเตอร์มีประสิทธิภาพใกล้เคียงกัน

คำสำคัญ : การแจกแจงแบบปกติ; ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ย; การประมาณค่าพารามิเตอร์แบบจุด

***ผู้รับผิดชอบบทความ :** puwasa_3@hotmail.co.th

Abstract

The objective of this research is to compare three methods of point estimation: Bayesian method, ordinary least square method and parametric bootstrap method for the simple linear regression model. The criterion is the average mean squared error (AMSE) and the sample sizes are 15, 30, 75 and 100. The distribution of error term is normal with mean of 0 and standard deviation of 0.15, 0.5 and 1 and the independent variable is selected at random from normal distribution with mean of 1 and standard deviation of 0.1. The data is generated by using the Monte Carlo technique with the repetition of 1000 times. It is found that the Bayesian method is the most efficient, followed by the ordinary least square method and the parametric bootstrap method, respectively.

Keywords: normal distribution; average mean squared error; point estimation

1. บทนำ

งานวิจัยนี้สนใจศึกษาตัวแบบถดถอยเชิงเส้นอย่างง่าย $y = \beta_0 + \beta_1 x + \varepsilon$ โดยตัวแปรที่สนใจศึกษาเป็นตัวแปรต่อเนื่องประกอบด้วยตัวแปรอิสระ (independent variable) หรือตัวแปร x เป็นตัวแปรอิสระไม่ขึ้นอยู่กับตัวแปรอื่น ๆ สามารถเปลี่ยนแปลงค่าได้ยาก และตัวแปรตาม (dependent variable) หรือตัวแปร y เป็นตัวแปรที่แปรผันตามตัวแปรอิสระ แล้วนำมาศึกษาการประมาณค่าพารามิเตอร์ โดยที่การประมาณค่าพารามิเตอร์เป็นการอนุมานทางสถิติอาศัยหลักการที่ว่านำเอาข้อมูลที่ได้จากตัวอย่างไปคาดคะเนคุณลักษณะประชากร โดยทั่วไปการประมาณค่าพารามิเตอร์มี 2 แบบ [1] คือ การประมาณค่าแบบจุด (point estimation) เป็นการประมาณค่าพารา มิเตอร์ที่สนใจด้วยค่าสถิติเพียงค่าเดียว ซึ่งการประมาณค่าแบบนี้อาจมีค่าเท่ากับค่าพารามิเตอร์หรืออาจมีโอกาสที่จะได้ค่าคาดเคลื่อนไปจากพารามิเตอร์ทั้งนี้ขึ้นอยู่กับหน่วยตัวอย่างที่นำมาวิเคราะห์ และการประมาณค่าแบบช่วง (interval estimation) เป็นการประมาณค่าพารามิเตอร์ว่าจะอยู่ช่วงใดช่วงหนึ่งด้วยความเชื่อมั่นที่กำหนด ในงานวิจัยนี้สนใจศึกษาการประมาณค่าแบบจุด ซึ่งเป็นวิธีในการหาตัวประมาณค่าแบบจุดที่มีสมบัติต่าง ๆ ที่ดีเพียงตัวเดียว ซึ่งมีอยู่หลายวิธี ในที่นี้สนใจศึกษาเฉพาะ

วิธีเบย์เซียน (Bayesian method) [2] ซึ่งเป็นวิธีการประมาณค่าพารามิเตอร์บนพื้นฐานความน่าจะเป็นแบบมีเงื่อนไข ซึ่งวิธีนี้อาศัยหลักการที่ว่าข้อมูลในอนาคตหรือเรียกว่าการแจกแจงภายหลัง แปรผันกับผลคูณของข้อมูลในปัจจุบันหรือเรียกว่าฟังก์ชันควรจะเป็น และข้อมูลในอดีตหรือเรียกว่าการแจกแจงก่อน

วิธีกำลังสองน้อยที่สุด (ordinary least square method, OLS) [3] เป็นวิธีที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนมีค่าต่ำที่สุดที่ ภายใต้เงื่อนไขที่ว่ามีความเป็นเส้นตรง กล่าวคือ เป็นฟังก์ชันเชิงเส้นตรงของตัวแปรตามและตัวแปรอิสระ และเป็นตัวประมาณที่ไม่เอนเอียง กล่าวคือ ค่าคาดหมายของค่าประมาณพารามิเตอร์จะเท่ากับค่าพารามิเตอร์ และมีความแปรปรวนต่ำสุดในบรรดาตัวประมาณค่าที่ไม่เอนเอียงอื่น ๆ

วิธีบูตสแตรปแบบใช้พารามิเตอร์ (parametric bootstrap method, PB) [4] วิธีนี้แตกต่างจากวิธีข้างต้นคือเป็นการประมาณค่าพารามิเตอร์โดยใช้การ

สุ่มตัวอย่างจากประชากรแบบใส่คืน ภายใต้เงื่อนไขของการใช้สถิติมีพารามิเตอร์ [5] ที่ว่าตัวแปรที่ศึกษาอยู่ในมาตราวัดระดับช่วงขึ้นไป ข้อมูลที่ได้จากกลุ่มตัวอย่างมีการแจกแจงแบบปกติและประชากรแต่ละกลุ่ม ที่นำศึกษามีความแปรปรวนเท่ากัน

จากการตรวจเอกสาร พบว่า ในปี ค.ศ. 2007 Sahinler และ Topuz [6] ได้ศึกษาวิธีบูตสแตรปและวิธีแจ็คไนฟ์สำหรับการประมาณค่าพารามิเตอร์ในสมการถดถอย โดยจะประมาณค่าความเอนเอียง ค่าความแปรปรวน และค่าพารามิเตอร์ค่าเฉลี่ยแบบจุดและสร้างช่วงความเชื่อมั่นสำหรับพารามิเตอร์ ถัดมาปี ค.ศ. 2008 Amiri และคณะ [7] ได้ศึกษาเปรียบเทียบวิธีบูตสแตรปแบบใช้พารามิเตอร์และวิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์สำหรับการแจกแจงแบบปกติ โดยจะประมาณพารามิเตอร์ค่าเฉลี่ย ความแปรปรวน และความโด่ง นอกจากนี้ ปี ค.ศ. 2011 Larsson [8] ได้ศึกษาวิธีเบย์เซียน โดยการแจกแจงก่อนที่พิจารณามี 2 แบบ คือ การแจกแจงก่อนที่ให้ข้อมูลและการแจกแจงก่อนที่ไม่ให้ข้อมูล โดยการประมาณค่าพารามิเตอร์ หลังจากนั้นปี ค.ศ. 2014 Yahya และคณะ [9] ได้ศึกษาการประมาณค่าพารามิเตอร์ด้วยวิธีเบย์เซียนโดยใช้การแจกแจงก่อนคู่สังยุคเป็นการแจกแจงแบบปกติกับวิธีกำลังสองน้อยที่สุด โดยจะประมาณค่าพารามิเตอร์และค่าความคลาดเคลื่อน

ดังนั้นผู้วิจัยจึงสนใจที่จะศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์แบบจุดทั้ง 3 วิธี ได้แก่ วิธีเบย์เซียน โดยที่การแจกแจงก่อนที่พิจาณาเป็นการแจกแจงก่อนที่ให้ข้อมูล คือ เป็นการแจกแจงที่ให้ข้อมูลเกี่ยวกับพารามิเตอร์อย่างชัดเจน ในงานวิจัยนี้เลือกการแจกแจงก่อนที่ให้ข้อมูลเป็นการแจกแจงก่อนคู่สังยุค (conjugate prior) วิธีกำลังสองน้อยที่สุด และวิธีบูตสแตรปแบบใช้พารามิเตอร์ โดยพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยเป็นเกณฑ์ในการเปรียบเทียบ เพื่อหาวิธีที่เหมาะสมที่สุดเพื่อเป็นแนวทางในการประมาณค่าพารามิเตอร์สำหรับตัวแบบอื่น ๆ

2. วิธีการวิจัย

2.1 ตัวแบบที่ใช้ศึกษาในการวิจัย คือ ตัวแบบถดถอยเชิงเส้นอย่างง่าย $y = \beta_0 + \beta_1 x + \varepsilon$ หรือสามารถเขียนในรูปเวกเตอร์และเมทริกซ์ได้ดังนี้ $\underset{\sim}{y} = \boldsymbol{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ [10] และสร้างข้อมูลของตัวแปรตาม $\left(\underset{\sim}{y}\right)$ ที่มีค่าสังเกต n ค่าจะได้ $\underset{\sim}{y} = \left(y_1, y_2, ..., y_n\right)'$ ที่มีรูปแบบความสัมพันธ์เชิงเส้นอย่างง่ายโดยมี $\boldsymbol{X}$ เป็นเมทริกซ์เขียนได้เป็น $\boldsymbol{X}' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix}_{2\times n}'$ ที่มีแถวแรกเป็น 1 และแถวที่ 2 เป็นตัวแปรอิสระ ซึ่งตัวแปรอิสระ x สุ่มจากประชากรที่มีการแจกแจงปกติ ค่าเฉลี่ยเท่ากับ 1 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.1 และค่าความคลาดเคลื่อนสุ่ม $\left(\underset{\sim}{\varepsilon}\right)$ มีการแจกแจงแบบปกติด้วยค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.15, 0.5 และ 1 กำหนดขนาดตัวอย่างที่ใช้ในการศึกษา $\left(n\right)$ มีค่าเท่ากับ 15, 30, 75 และ 100 กำหนดค่าสัมประสิทธิ์ถดถอย $\underset{\sim}{\beta} = \left(\beta_0, \beta_1\right)'$ เท่ากับ 1 และ 2 ตามลำดับแล้วจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล (Monte Carlo) แล้วนำมาประมาณค่าพารามิเตอร์ในแต่ละวิธี โดยใช้โปรแกรม R เวอร์ชั่น 3.2.2 ทดลองซ้ำ 1000 ครั้ง

2.2 นำข้อมูล $\underset{\sim}{y}$ และ $\underset{\sim}{x}$ ที่ได้จากการจำลองในข้อ 2.1 มาประมาณค่าพารามิเตอร์แบบจุดวิธีการประมาณค่าที่ศึกษามีดังนี้

2.2.1 วิธีเบย์เซียน

Koop [2] กล่าวว่าวิธีเบย์เซียนเกิดจากการวิเคราะห์บนพื้นฐานความน่าจะเป็นแบบมีเงื่อนไข สมการของเบย์เซียนสามารถเขียนได้ดังนี้

$$p\left(\underset{\sim}{\beta},\sigma^2\middle|\underset{\sim}{y}\right)=\frac{p\left(\underset{\sim}{y}\middle|\underset{\sim}{\beta},\sigma^2\right)p\left(\underset{\sim}{\beta},\sigma^2\right)}{p\left(\underset{\sim}{y}\right)}\propto p\left(\underset{\sim}{y}\middle|\underset{\sim}{\beta},\sigma^2\right)p\left(\underset{\sim}{\beta},\sigma^2\right)$$

โดยที่ $p\left(\underset{\sim}{y}\right)$ เรียกว่าการแจกแจงของข้อมูล

$p\left(\underset{\sim}{\beta},\sigma^2\right)$ เรียกว่า การแจกแจงก่อนร่วมสำหรับ $\left(\underset{\sim}{\beta},\sigma^2\right)$

$p(\underset{\sim}{\beta},\sigma^2|\underset{\sim}{y})$ เรียกว่าการแจกแจงภายหลังร่วมสำหรับ $\left(\underset{\sim}{\beta},\sigma^2\right)$

$p(\underset{\sim}{y}|\underset{\sim}{\beta},\sigma^2)$ เรียกว่าฟังก์ชันภาวะน่าจะเป็นร่วม (likelihood function) เมื่อทราบข้อมูล $\underset{\sim}{\beta}$ และ σ^2 วิธีเบย์เซียนมีขั้นตอนดังนี้

(1) เลือกการแจกแจงก่อนร่วม (joint prior distribution) สำหรับ $\left(\underset{\sim}{\beta},\sigma^2\right)$ ในงานวิจัยนี้เลือกใช้การแจกแจงก่อนที่ให้ข้อมูล เป็นการแจกแจงก่อนคู่สังยุค (conjugate prior) สำหรับ $\underset{\sim}{\beta}$ และกำหนดให้ทราบ σ^2 มีรูปแบบสมการดังนี้

$$p(\underset{\sim}{\beta},\sigma^2)\propto\exp\left\{-\frac{1}{2}(\underset{\sim}{\beta}-\bar{\underset{\sim}{\beta}})'\bar{\underset{\sim}{\Sigma}}_{\underset{\sim}{\beta}}^{-1}(\underset{\sim}{\beta}-\bar{\underset{\sim}{\beta}})\right\}$$

$\bar{\underset{\sim}{\beta}}$ คือ เวกเตอร์ค่าเฉลี่ยก่อน (prior mean) ของ $\underset{\sim}{\beta}$

$\bar{\underset{\sim}{\Sigma}}_{\underset{\sim}{\beta}}$ คือ เมทริกซ์ความแปรปรวนร่วมก่อน (prior covariance) ของ $\bar{\underset{\sim}{\beta}}$

(2) คำนวณฟังก์ชันภาวะน่าจะเป็น (likelihood function) เมื่อทราบข้อมูล $\bar{\underset{\sim}{\beta}}$ และ σ^2 จากประชากรที่มีการแจกแจงแบบปกติ (normal distribution) มีรูปแบบสมการดังนี้

$$p(\underset{\sim}{y}|\underset{\sim}{\beta},\sigma^2)\propto\exp\left\{-\frac{1}{2\sigma^2}(\underset{\sim}{y}-\boldsymbol{X}\underset{\sim}{\beta})'(\underset{\sim}{y}-\boldsymbol{X}\underset{\sim}{\beta})\right\}$$

(3) คำนวณการแจกแจงภายหลังร่วม (joint posterior distribution) สำหรับ $\left(\underset{\sim}{\beta},\sigma^2\right)$ จากสมการของเบย์เซียนมีรูปแบบสมการดังนี้

$$p(\underset{\sim}{\beta},\sigma^2|\underset{\sim}{y})\propto\exp\left\{\frac{-1}{2}\left(\underset{\sim}{\beta}-\bar{\bar{\underset{\sim}{\beta}}}\right)'\bar{\bar{\underset{\sim}{\Sigma}}}_{\underset{\sim}{\beta}}^{-1}\left(\underset{\sim}{\beta}-\bar{\bar{\underset{\sim}{\beta}}}\right)\right\}$$

โดยที่ $\underset{\sim}{\beta}\sim N\left(\bar{\bar{\underset{\sim}{\beta}}},\bar{\bar{\underset{\sim}{\Sigma}}}_{\underset{\sim}{\beta}}\right)$ และ $\bar{\bar{\underset{\sim}{\beta}}}$ คือ เวกเตอร์ค่าเฉลี่ยภายหลัง (posterior mean) ของ $\underset{\sim}{\beta}$

$\bar{\bar{\underset{\sim}{\Sigma}}}_{\underset{\sim}{\beta}}$ คือ เมทริกซ์ความแปรปรวนร่วมภายหลัง (posterior covariance) ของ $\underset{\sim}{\beta}$

ดังนั้นจะได้ค่าประมาณพารามิเตอร์คือ $\bar{\bar{\underset{\sim}{\beta}}}=\left(\bar{\bar{\underset{\sim}{\Sigma}}}_{\underset{\sim}{\beta}}^{-1}\right)\left(\bar{\underset{\sim}{\Sigma}}_{\underset{\sim}{\beta}}^{-1}\bar{\underset{\sim}{\beta}}+\frac{\boldsymbol{X}'\boldsymbol{X}}{\sigma^2}\underset{\sim}{b}\right)$ โดยที่ $\bar{\bar{\underset{\sim}{\Sigma}}}_{\underset{\sim}{\beta}}^{-1}=\bar{\underset{\sim}{\Sigma}}_{\underset{\sim}{\beta}}^{-1}+\frac{\boldsymbol{X}'\boldsymbol{X}}{\sigma^2}$

2.2.2 วิธีกำลังสองน้อยที่สุด

วิธีการหาตัวประมาณพารามิเตอร์วิธีนี้ [3] มีหลักเกณฑ์อยู่ว่าหาตัวประมาณพารามิเตอร์ที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อน (sum square error) มีค่าต่ำที่สุด

โดยที่ผลบวกกำลังสองของความคลาดเคลื่อนเขียนได้ดังนี้

$$\hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}}=\left(\underset{\sim}{y}-\boldsymbol{X}\hat{\underset{\sim}{\beta}}\right)'\left(\underset{\sim}{y}-\boldsymbol{X}\hat{\underset{\sim}{\beta}}\right)$$
$$=\underset{\sim}{y}'\underset{\sim}{y}-2\hat{\underset{\sim}{\beta}}'\boldsymbol{X}'\underset{\sim}{y}+\hat{\underset{\sim}{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\underset{\sim}{\beta}}$$

หาอนุพันธ์ของ $\hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}}$ เทียบกับ $\hat{\underset{\sim}{\beta}}$ และกำหนดให้เท่ากับ 0 นั่นคือ

$$\frac{\partial}{\partial\hat{\underset{\sim}{\beta}}}\hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}}=\frac{\partial}{\partial\hat{\underset{\sim}{\beta}}}\left(\underset{\sim}{y}-\boldsymbol{X}\hat{\underset{\sim}{\beta}}\right)'\left(\underset{\sim}{y}-\boldsymbol{X}\hat{\underset{\sim}{\beta}}\right)=0$$
$$-2\boldsymbol{X}'\underset{\sim}{y}+2\left(\boldsymbol{X}'\boldsymbol{X}\right)\hat{\underset{\sim}{\beta}}=0$$
$$\left(\boldsymbol{X}'\boldsymbol{X}\right)\hat{\underset{\sim}{\beta}}=\boldsymbol{X}'\underset{\sim}{y}$$

ดังนั้นจะได้ค่าประมาณพารามิเตอร์คือ $\hat{\underset{\sim}{\beta}}=\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\underset{\sim}{y}\right)$

2.2.3 วิธีบูตสแตรปแบบใช้พารามิเตอร์

Efron [4] กล่าวว่า วิธีบูตสแตรปเป็นวิธีการประมาณค่าโดยใช้การสุ่มตัวอย่างจากประชากรแบบใส่คืน (replacement) นั่นคือเป็นการสุ่มตัวอย่างที่ยอมให้หน่วยตัวอย่างซ้ำกันได้โดยที่แต่ละหน่วยตัวอย่างมีโอกาสในการถูกสุ่มเท่ากัน ทำการสุ่มตัวอย่าง

จำนวนด้วยจำนวนครั้งที่มากพอ เพื่อสร้างการแจกแจงของตัวสถิติตัวอย่างแล้วนำมาใช้ในการประมาณค่าพารามิเตอร์ที่สนใจ ในงานวิจัยได้ศึกษาวิธีบูตสแตรปแบบใช้พารามิเตอร์ มีขั้นตอนดังต่อไปนี้

(1) ทำการสุ่มตัวอย่างซ้ำแบบคืนที่จำนวน n ตัว จะได้ตัวอย่างสุ่มชุดใหม่ คือ X_1^*, X_2^*, ..., X_n^* ซึ่งเรียกว่าตัวอย่างบูตสแตรป (bootstrap sample)

(2) คำนวณค่าประมาณพารามิเตอร์ของ $\underset{\sim}{\beta}^*$ จากวิธีกำลังสองน้อยที่สุด จะได้ค่าประมาณ คือ $\underset{\sim}{\hat{\beta}}_{ols}$

(3) สุ่ม $\underset{\sim}{\varepsilon}^*$ จากการแจกแจงความคลาดเคลื่อนภายใต้เงื่อนไขที่ว่า $E(\varepsilon)=\underset{\sim}{0}$

(4) แทนค่า $\underset{\sim}{\varepsilon}^*$ ในสมการ $\underset{\sim}{y}^* = \boldsymbol{X}^* \underset{\sim}{\hat{\beta}}_{ols} + \underset{\sim}{\varepsilon}^*$

(5) นำ $\underset{\sim}{y}^*$ และ $\boldsymbol{X}^*$ ที่ได้จากข้อ (4) มาประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีบูตสแตรป

(6) ทำตามขั้นตอนข้อที่ (3), (4) และ (5) ตามลำดับ ด้วยจำนวนซ้ำของการทำซ้ำแบบบูตสแตรป (B) 1000 ครั้ง

(7) จะได้ค่าประมาณพารามิเตอร์ คือ $\hat{\beta}_B = \dfrac{\sum_{k=1}^{B} \hat{\beta}_k^*}{B}$

2.3 เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์แบบจุดทั้ง 3 วิธี โดยพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) [11] เป็นค่าวัดความถูกต้องที่วัดจากขนาดของความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์ที่ได้จากกำลังสองของค่าความคลาดเคลื่อนค่านี้จะมีหน่วยวัดเป็นกำลังสองของหน่วยวัดของค่าสังเกต และมีสูตรการคำนวณ ดังนี้

$$\mathrm{AMSE} = \frac{\sum_{i=1}^{p} \mathrm{MSE}_i}{p}, \quad \mathrm{MSE}_i = \frac{\sum_{j=1}^{1000} \left(\beta_i - \hat{\beta}_{ij}\right)^2}{j}$$

เมื่อ β_i คือ สัมประสิทธิ์การถดถอยตัวที่ i, $i = 0, 1$

$\hat{\beta}_{ij}$ คือ ตัวประมาณสัมประสิทธิ์การถดถอยตัวที่ i จากการประมาณรอบที่ j โดยที่ $j = 1, 2, ..., 1000$

p คือ จำนวนพารามิเตอร์ในตัวแบบการถดถอย ซึ่งเท่ากับ 2

n คือ ขนาดตัวอย่าง

MSE_i คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณตัวที่ i

3. ผลการวิจัย

งานวิจัยนี้ใช้เทคนิคการจำลองแบบมอนติ-คาร์โล โดยจำลองข้อมูลตัวอย่างที่มีการแจกแจงแบบปกติเพื่อเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์แบบจุดทั้ง 3 วิธี คือ วิธีเบย์เซียน เมื่อใช้การแจกแจงก่อนที่ให้ข้อมูลวิธีกำลังสองน้อยที่สุด และวิธีบูตสแตรปแบบใช้พารามิเตอร์ จำลองข้อมูลในแต่ละสถานการณ์ที่กำหนด 1000 ครั้ง โดยพิจารณาเปรียบเทียบจากค่าเฉลี่ยความคลาด เคลื่อนกำลังสองเฉลี่ย ผลการศึกษาสรุปได้ดังตารางที่ 1

4. อภิปรายผล

เมื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์แบบจุดทั้ง 3 วิธี คือ วิธีเบย์เซียน วิธีกำลังสองน้อยที่สุด และวิธีบูตสแตรปแบบใช้พารามิเตอร์ โดยพิจารณาเปรียบเทียบจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) พบว่าค่า MSE และ AMSE จะแปรผันตามค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มและแปรผกผันกับขนาดตัวอย่าง กล่าวคือ เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มมีค่าสูงพบว่าค่า MSE และ AMSE มีแนวโน้มเพิ่มขึ้น และเมื่อขนาดตัวอย่าง

เพิ่มขึ้นพบว่าค่า MSE และ AMSE มีแนวโน้มลดลง ซึ่งสอดคล้องกับงานวิจัยของ วีรพา [12] ยกเว้นวิธีเบย์เซียน ไม่มีแนวโน้มค่า MSE และ AMSE เพิ่มขึ้นหรือลดลงอย่างชัดเจน ซึ่งสอดคล้องกับงานวิจัยของปรางค์ทิพย์ [13] และจากการเปรียบเทียบประสิทธิภาพของค่า MSE และ AMSE จากทั้ง 3 วิธี พบว่าวิธีเบย์เซียนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ทุก ๆ สถานการณ์ที่ศึกษา รองลงมาเป็นวิธีกำลังสอง และน้อยที่สุดเป็นวิธีบูตสแตรปแบบใช้พารามิเตอร์เพียงเล็กน้อย สอดคล้องกับงาน วิจัยของ ณัฐภาภรณ์ [14]

ตารางที่ 1 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของ β_0 และ β_1 และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยที่คำนวณได้จากการประมาณค่าพารามิเตอร์การถดถอยเชิงเส้นอย่างง่ายทั้ง 3 วิธี

สถานการณ์		วิธีการประมาณค่าพารามิเตอร์								
		Bayesian			OLS			PB		
n	$\sigma(\varepsilon_i)$	MSE		AMSE	MSE		AMSE	MSE		AMSE
		β_0	β_1		β_0	β_1		β_0	β_1	
15	0.15	0.1294	0.1276	**0.1285**	0.1892	0.1864	0.1878	0.1891	0.1864	0.1878
	0.5	0.0548	0.0543	**0.0546**	2.0875	2.0649	2.0762	2.0878	2.0658	2.0768
	1	0.0224	0.0219	**0.0221**	8.3509	8.2605	8.3057	8.3691	8.2765	8.3228
30	0.15	0.0783	0.0774	**0.0779**	0.0876	0.0866	0.0871	0.0878	0.0868	0.0873
	0.5	0.1259	0.1243	**0.1251**	0.8908	0.8773	0.8841	0.8936	0.8800	0.8868
	1	0.0261	0.0247	**0.0254**	3.8256	3.7670	3.7963	3.8278	3.7701	3.7990
75	0.15	0.0303	0.0299	**0.0301**	0.0310	0.0306	0.0308	0.0311	0.0307	0.0309
	0.5	0.1646	0.1630	**0.1638**	0.3259	0.3230	0.3244	0.3263	0.3234	0.3249
	1	0.0821	0.0804	**0.0812**	1.3828	1.3660	1.3744	1.3833	1.3659	1.3746
100	0.15	0.0239	0.0238	**0.0238**	0.0238	0.0237	**0.0238**	0.0240	0.0238	0.0239
	0.5	0.1691	0.1689	**0.1690**	0.2649	0.2639	0.2644	0.2656	0.2645	0.2651
	1	0.1171	0.1182	**0.1176**	1.0599	1.0558	1.0578	1.0607	1.0569	1.0588

หมายเหตุ : ตัวหนาหมายถึง MSE และ AMSE ต่ำที่สุด

5. สรุป

การพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยพบว่าวิธีเบย์เซียนมีประสิทธิภาพดีที่สุด [9] สำหรับทุก ๆ สถานการณ์ที่กำหนด ส่วนวิธีกำลังสองน้อยที่สุดและวิธีบูตสแตรปแบบใช้พารามิเตอร์มีประสิทธิภาพใกล้เคียงกันโดยที่วิธีกำลังสองน้อยที่สุดจะมีประสิทธิภาพดีกว่าวิธีบูตสแตรปแบบใช้พารามิเตอร์เพียงเล็กน้อย

6. ข้อเสนอแนะ

6.1 งานวิจัยนี้ผู้วิจัยได้ศึกษาเกณฑ์เฉพาะค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ย ผู้ที่สนใจศึกษา

เพิ่มเติมควรศึกษาเกณฑ์นอกเหนือจากนี้

6.2 ผู้ที่สนใจศึกษาเพิ่มเติมอาจศึกษาเปรียบเทียบโดยใช้วิธีอื่น ๆ ได้แก่ วิธีกำลังสองน้อยที่สุดสามขั้น วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก เป็นต้น

7. รายการอ้างอิง

[1] ธีระพงษ์ กระการดี, การประมาณค่า**,** แหล่งที่มา : http://www.stvc.ac.th/elearning/stat/csu5.html, 6 ตุลาคม2558.

[2] Koop, G. 2003**,** Bayesian Econometrics, John Wiley, England.

[3] ทรงศิริ แย้มสมบัติ, 2548, การวิเคราะห์การถดถอย, พิมพ์ครั้งที่ 3, สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพฯ, 294 น.

[4] Efron, B., 1979, Bootstrap methods another look at the Jackknife, Ann. Stat. 7: 1-26.

[5] ฉัตรศิริ ปิยะพิมลสิทธิ์, สถิติ : ความหมายและประเภท, แหล่งที่มา : http://www.watpon.com/Elearning/stat1.htm, 24 พฤษศจิกายน 2558.

[6] Sahinler, S. and Topuz, D., 2007, Boot strapand Jackknife resampling algorithms for estimation of regression parameters, J. Appl. Quant. Methods 2: 189-199.

[7] Amiri, S., von Rosen, D. and Zwanzig, S., 2008, On the comparison of parametric and nonparametric bootstrap, J. Disease Month 15: 1-15.

[8] Larsson, R., 2011, How informative is non-informative prior, Working Paper 2: 1-10.

[9] Yahya, W.B., Olaniran, O.R. and Ige, S.O., 2014, On Bayesian conjugate normal linear regression and ordinary least square regression methods: A Monte Carlo study, J. Sci. 1: 216-227.

[10] Kutner, M.H., Nachtsheim, C.J. and Neter, J. 2008, Applied Linear Regression Models, 4th Ed., McGraw-Hill, New York.

[11] Sosa, J.C. and Diaz, L.G. 2012, Random time-varyring cofficient method estimation through radial basis functions, J. Revista Colombiana de Estadistica 35: 167-184.

[12] วีรพา ฐานะปรัชญ์, 2542, การวิเคราะห์เชิงเบย์สำหรับตัวแบบการถดถอยเชิงเส้นเชิงเดี่ยว, วิทยานิพนธ์ปริญญาโท, จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ, 222 น.

[13] ปรางค์ทิพย์ รัชตะปีติ, 2550, การเปรียบเทียบการประมาณค่าพารามิเตอร์ด้วยวิธีเบย์เซียนกับวิธีวิเคราะห์ความถดถอยพหุคูณแบบกำลังสองน้อยสุดสองขั้น, วิทยานิพนธ์ปริญญาโท, จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ, 315 น.

[14] ณัฐภาภรณ์ รอดรัตษะ, 2553, การศึกษาเปรียบเทียบการประมาณตัวแบบความถดถอยเชิงเส้นด้วยวิธีบูตสแตรปแบบใช้พารามิเตอร์และไม่ใช้พารามิเตอร์, วิทยานิพนธ์ปริญญาโท, จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพฯ, 148 น.