การเปรียบเทียบการประมาณค่าพารามิเตอร์ระหว่างวิธีกำลังสองน้อยสุดเชิงส่วนและวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิดต์ สำหรับตัวแบบความถดถอยพหุเชิงเส้นที่เกิดพหุสัมพันธ์

A COMPARISON OF PARAMETERS ESTIMATION AMONG PARTIAL LEAST SQUARES AND ORDINARY LEAST SQUARES METHODS USING GRAM-SCHMIDT'S DATA TRANSFORMATION FOR MULTIPLE LINEAR REGRESSION MODEL WITH MULTICOLLINEARITY

ผู้วิจัย เด่นนภา จุลเพชร, และรองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา
ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อ

งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อประมาณค่าพารามิเตอร์ที่เหมาะสม โดยการเปรียบเทียบจากค่าความคลาดเคลื่อนมาตรฐาน (SE) และความเอนเอียงของตัวประมาณ สำหรับตัวแบบความถดถอยพหุเชิงเส้นของวิธีกำลังสองน้อยสุดเชิงส่วน (PLS) และวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิดต์ (OLS_G) เมื่อเกิดพหุสัมพันธ์ ซึ่งศึกษาภายใต้ตัวแปรอิสระมีการแจกแจงแบบปกติค่า $\mu=0$, $\sigma_x=2$ และ $\sigma^2=10$ โดยมีจำนวนตัวแปรอิสระ (p) เท่ากับ 2 และ 3 ตัวแปร มีขนาดตัวอย่าง 50, 100 และ 200 มีค่าสัมประสิทธิ์ความถดถอยเบื้องต้น$\beta_0 = \beta_1 = \cdots = \beta_p = 1.$ และระดับความสัมพันธ์ระหว่างตัวแปรอิสระแบ่งเป็น 2 กรณี คือ กรณีตัวแปรอิสระ 2 ตัวแปร มีระดับความสัมพันธ์ (ρ) เป็น 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 และ 0.99 ส่วนกรณีจำนวนตัวแปรอิสระ 3 ตัวแปร มีระดับความสัมพันธ์ (ρ) เป็น 0.91, 0.93, 0.95, 0.97 และ 0.99 ซึ่งในการศึกษาจะทำซ้ำจำนวน 1000 รอบ

ผลการศึกษาพบว่ากรณี p เท่ากับ 2 ตัวแปร จะเกิดปัญหาพหุสัมพันธ์ที่ ρ ตั้งแต่ 0.95 ขึ้นไป และกรณี p เท่ากับ 3 ตัวแปร ขนาดตัวอย่างเท่ากับ 50 และ 100 จะเกิดปัญหาพหุสัมพันธ์ที่ ρ ตั้งแต่ 0.93 ขึ้นไป ส่วนขนาดตัวอย่างเท่ากับ 200 จะเกิดปัญหาพหุสัมพันธ์ที่ ρ ตั้งแต่ 0.95 ขึ้นไป และทั้งสองกรณีมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์สูง ซึ่งส่งผลกระทบต่อการประมาณค่าสัมประสิทธิ์ความถดถอย ดังนั้นกรณี p เท่ากับ 2 ตัวแปร ขนาดตัวอย่างเท่ากับ 50, 100 และ 200 ตามลำดับ และ ρ เท่ากับ 0.9-0.94, 0.9-0.97 และ 0.9-0.98 ตามลำดับ วิธี OLS_G มีประสิทธิภาพดีกว่าวิธี PLS แต่ที่ ρ เท่ากับ 0.95-0.99, 0.98-0.99 และ 0.99 ตามลำดับ วิธี PLS มีประสิทธิภาพดีกว่าวิธี OLS_G และกรณี p เท่ากับ 3 ตัวแปร วิธี PLS มีประสิทธิภาพดีกว่าวิธี OLS_G ทุกกรณีที่ทำการศึกษา โดยปัจจัยที่มีผลต่อค่า SE ของทั้งสองกรณีนั้นพบว่าเมื่อ ρ เพิ่มขึ้นและ p

เท่ากับ 2 ค่า SE ของวิธี PLS จะมีค่าเพิ่มขึ้น ในขณะ p เท่ากับ 3 ค่า SE จะมีค่าคงที่ แต่วิธี OLS_G จะมีค่าเพิ่มขึ้นทั้งสองกรณี และถ้าขนาดตัวอย่างเพิ่มขึ้นค่า SE ของทั้งสองวิธีจะมีค่าลดลง ส่วนปัจจัยที่มีผลต่อค่าความเอนเอียงของทั้งสองกรณีนั้นพบว่าเมื่อ ρ เพิ่มขึ้นค่าความเอนเอียงของวิธี PLS จะมีค่าลดลง แต่วิธี OLS_G จะมีค่าเพิ่มขึ้น และถ้าขนาดเพิ่มขึ้นค่าความเอนเอียงของวิธี PLS จะมีค่าเพิ่มขึ้น แต่วิธี OLS_G จะมีค่าคงที่

คำสำคัญ: วิธีกำลังสองน้อยสุดเชิงส่วน, วิธีกำลังสองน้อยสุด, วิธีกรามชมิดต์, พหุสัมพันธ์ระหว่างตัวแปรอิสระ

Abstract

The objective of this research is to estimate optimal parameters by comparison from the value of the standard error (SE) and the bias of the estimate for multiple linear regression models among Partial Least Square (PLS) and Ordinary Least Square Methods by using Gram-Schmidt's data transformation (OLS_G) to solve multicollinearity problem of the independent variables. This study focuses on Normal distribution of the independent variables with $\mu = 0$, $\sigma_x=2$ and $\sigma^2=10$. We will study on 2 and 3 independent variables (p) under the following condition; the sample size 50, 100 and 200, the initial regression coefficient is $\beta_0 = \beta_1 = \cdots = \beta_p = 1$. For case 2 independent variables, the multicollinearities (ρ) are defined to be 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, and for the other case, 3 independent variables, ρ's are defined to be 0.91, 0.93, 0.95, 0.97 and 0.99. We will generate each case with 1,000 simulation runs.

From the conclusion, we found that problems of multicollinearity occurs when ρ is more than 0.95 for case 2 independent variables, and these problems depend on number of sample size for case 3 independent variables. In particular, for the latter case the problems occurs when ρ is more than 0.93 for sample size 50 and 100, and more than 0.95 for sample size 200. In both cases, the problems of multicollinearity will be occurred in high chance. Therefore, in case of 2 independent variables at the sample size 50,100 and 200 in order and ρ is 0.9-0.94, 0.9-0.97 and 0.9-0.98 in order ,the OLS_G method is more effective than the PLS method. But, if ρ is 0.95-0.99, 0.98-0.99 and 0.99 in order; the PLS method is more effective than the OLS_G method. And for case 3 independent variables the PLS method is more effective than the OLS_G method in every case. The affecting factors to the SE value of both case; when ρ increases and p=2 the SE value of the PLS method will be increase while p=3 the SE value will be constant, but in the OLS_G method the SE value will be increase of both cases. And the sample size is increase the SE value of both cases will be decreases. For the affecting factors to the bias value of both cases; when ρ increases the bias value of the PLS method will be decrease,

but the OLS_G method will be increase. And if the sample size is increases the, PLS method will be increase, but the PLS_G method will be constant.

Key Word (s): Partial Least Squares Method, Ordinary Least Squares Method, Gram-Schmidt Method, Multicollinearity

บทนำ

เนื่องจากในปัจจุบันนี้มีงานวิจัยทางด้านต่างๆ เช่น สังคมศาสตร์ วิทยาศาสตร์ วิศวกรรมศาสตร์ การแพทย์และเภสัชศาสตร์ ได้นำการวิเคราะห์ความถดถอย (Regression Analysis) มาใช้เพื่อศึกษาความสัมพันธ์ของตัวแปรที่สนใจและเพื่อประมาณค่าสัมประสิทธิ์ความถดถอยในการสร้างสมการพยากรณ์อย่างแพร่หลาย แต่หลักการของการวิเคราะห์ความถดถอยมีข้อกำหนดเบื้องต้นหลายประการ อาทิเช่น ความคลาดเคลื่อนมีความเป็นเชิงเส้น ความคลาดเคลื่อนมีความเป็นอิสระจากกัน ความคลาดเคลื่อนมีการแจกแจงแบบปกติมีค่าเฉลี่ยเท่ากับศูนย์และค่าความแปรปรวนคงที่ เป็นต้น และในการสร้างสมการความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระเพื่อสร้างตัวแบบในการพยากรณ์ค่า y นั้นตัวแปรอิสระที่สนใจจะต้องเป็นอิสระซึ่งกันและกัน แต่ในความเป็นจริงตัวแปรอิสระมักเกิดปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระ (Multicollinearity) ซึ่งทำให้ผลลัพธ์ของการวิเคราะห์ความถดถอยไม่มีความแม่นยำ กล่าวคือ ความแปรปรวนของสัมประสิทธิ์ความถดถอยมีค่าสูงเกินไป เครื่องหมายของสัมประสิทธิ์ความถดถอยมีทิศทางตรงข้ามกันและค่าที่ได้อาจมากหรือน้อยกว่าความเป็นจริง และทำให้ค่าสัมประสิทธิ์ความถดถอยเปลี่ยนแปลงไปเมื่อทำการเพิ่มตัวแปรอิสระ เป็นต้น ดังนั้นจึงจำเป็นต้องทำการแก้ไขปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระก่อนทำการวิเคราะห์ความถดถอย ในการศึกษาครั้งนี้จึงสนใจที่จะเปรียบเทียบประสิทธิภาพของตัวประมาณหลังแก้ไขปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระของวิธีกำลังสองน้อยสุดเชิงส่วนและวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิดต์ โดยวิธีกำลังสองน้อยสุดเชิงส่วนเป็นเทคนิคที่เกี่ยวข้องกับวิธีกำลังสองน้อยสุดและวิธีการถดถอยองค์ประกอบหลัก ส่วนวิธีกำลังสองน้อยสุด ในการศึกษาครั้งนี้จะใช้หลักการของพีชคณิตเชิงเส้นด้วยกระบวนการแปลงข้อมูลแบบกรามชมิดต์เข้ามาช่วยแก้ไขปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระ จากนั้นทำการประมาณค่าพารามิเตอร์ความถดถอยด้วยวิธีกำลังสองน้อยสุด ซึ่งค่าประมาณพารามิเตอร์ความถดถอยที่ได้จะเป็นค่าประมาณที่เหมาะสม ทางผู้ศึกษาจึงเห็นว่าในการแก้ปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระในการวิเคราะห์ความถดถอยหลายตัวแปรเป็นสิ่งที่สำคัญในการประมาณค่าพารามิเตอร์ความถดถอยที่เหมาะสมในแต่ละกรณีของจำนวนตัวแปรอิสระ ขนาดตัวอย่าง ระดับความสัมพันธ์ของตัวแปรอิสระ และการแจกแจงของตัวแปรอิสระหลายตัวแปร

ในงานวิจัยของ Garthwaite (1994) [1] ได้เสนอการประมาณค่าสัมประสิทธิ์ความถดถอยวิธีกำลังสองน้อยสุดเชิงส่วนที่มีตัวแปรตามตัวแปรเดียวและตัวแปรอิสระหลายตัวแปร ซึ่งจะใช้เป็นแนวทางในการศึกษาครั้งนี้

วัตถุประสงค์ของการวิจัย

1. เพื่อประมาณค่าพารามิเตอร์ที่เหมาะสม สำหรับตัวแบบความถดถอยพหุเชิงเส้นของวิธีกำลังสองน้อยสุดเชิงส่วนและวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิดต์ เมื่อเกิดพหุสัมพันธ์
2. เพื่อเปรียบเทียบการประมาณค่าพารามิเตอร์ สำหรับตัวแบบความถดถอยพหุเชิงเส้นระหว่างวิธีกำลังสองน้อยสุดเชิงส่วนและวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิดต์ เมื่อเกิดพหุสัมพันธ์

วิธีการวิจัย

ข้อมูลที่ใช้ในการวิจัยได้จากการจำลองค่า โดยใช้โปรแกรม R แต่ละครั้งของการทดสอบจะทำซ้ำจำนวน 1000 รอบ ซึ่งมีขั้นตอนในการดำเนินการวิจัย ดังนี้

1. กำหนดค่าเริ่มต้นในการจำลองข้อมูล ดังนี้

1.1) กำหนดจำนวนตัวแปรอิสระ 2 และ 3 ตัวแปร และจำนวนตัวแปรตาม 1 ตัวแปร

1.2) กำหนดขนาดตัวอย่างเป็น 50, 100 และ 200

1.3) กำหนดค่าสัมประสิทธิ์ความถดถอยเบื้องต้นเป็น $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 1$

1.4) กำหนดระดับความสัมพันธ์ระหว่างตัวแปรอิสระ ซึ่งจำนวนตัวแปรอิสระ 2 ตัวแปร ระดับความสัมพันธ์เป็น 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 และ 0.99 ส่วนจำนวนตัวแปรอิสระ 3 ตัวแปร ระดับความสัมพันธ์เป็น 0.91, 0.93, 0.95, 0.97 และ 0.99

1.5) กำหนดค่าความแปรปรวนของความคลาดเคลื่อนเป็น $\sigma^2 = 10$ และกำหนดค่าความแปรปรวนของตัวแปรอิสระเป็น X_1, X_2 และ X_3 เป็น $\sigma_{x_1}^2 = 4$, $\sigma_{x_2}^2 = 4$ และ $\sigma_{x_3}^2 = 4$ ตามลำดับ

2. จำลองข้อมูลตัวแปรอิสระของการแจกแจงปกติหลายตัวแปร โดยตัวแปรอิสระสร้างจากค่าเฉลี่ยเท่ากับ $(\mu = 0)$ ค่าความแปรปรวนและระดับความสัมพันธ์ที่กำหนด จากนั้นสร้างเวกเตอร์ของข้อมูลตัวแปรตาม y จากตัวแบบความถดถอยพหุเชิงเส้น

---

[1] Paul H. Garthwaite. “An Interpretation of Partial Least Squares.”, Journal of the American Statistical Association, Vol. 89, No. 425 (Mar., 1994): 122-127.

3. คำนวณค่า VIF ซึ่งพิจารณาจากสัมประสิทธิ์สหสัมพันธ์พหุของตัวแปรอิสระ โดยเกณฑ์ในการใช้ตัดสินใจว่าเกิดปัญหา Collinearity สามารถคำนวณตามมาตรวัดทางสถิติได้ ดังนี้

$$\mathrm{VIF_j} = \frac{1}{1-\mathrm{R_j^2}} \qquad ;\ \mathrm{j} = 1,2,\ldots,\mathrm{p}$$

โดยที่ $\mathrm{R_j^2}$ เป็นสัมประสิทธิ์สหสัมพันธ์พหุของตัวแปรอิสระ $\mathrm{P_j}$ กับตัวแปรอิสระอื่นๆ

ถ้าค่า $\mathrm{VIF_j} \leq 10$; $\mathrm{j} = 1,2,\ldots,\mathrm{p}$ แสดงว่าไม่เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

ถ้าค่า $\mathrm{VIF_j} > 10$; $\mathrm{j} = 1,2,\ldots,\mathrm{p}$ แสดงว่าเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

4. นำเวกเตอร์ของตัวแปรตาม **y** และนำเมทริกซ์ของตัวแปรอิสระ **X** ที่ได้มาทำ centering เพื่อทำให้ค่า Intercept หายไปเหลือเพียงค่า X อย่างเดียว

5. สร้างตัวแปรอิสระใหม่ให้ตั้งฉากกัน 2 วิธี คือ

5.1) วิธีกำลังสองน้อยสุดเชิงส่วน เป็นการนำตัวแปรตาม $\boldsymbol{y}^*$ และตัวแปรอิสระเริ่มต้น $\{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \ldots, \boldsymbol{x}_p^*\}$ ที่ได้จากการทำ Centering มาสร้างเป็นมูลฐานเชิงตั้งฉาก $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_p\}$ อย่างเป็นลำดับ แสดงด้วยขั้นตอน ดังนี้

ขั้นที่ 1 ประมาณค่า $\widehat{\widetilde{\boldsymbol{y}}}_{(1)j}^*$ จากสมการความถดถอยด้วยวิธีกำลังสองน้อยสุด ได้จาก

$$\widehat{\widetilde{\boldsymbol{y}}}_{(1)j}^* = b_{(1)j}\widetilde{\boldsymbol{x}}_{(1)j}^*$$

โดยที่ $b_{(1)j} = (\widetilde{\boldsymbol{x}}_{(1)j}^{*}{}'\widetilde{\boldsymbol{x}}_{(1)j}^*)^{-1}\widetilde{\boldsymbol{x}}_{(1)j}^{*}{}'\widetilde{\boldsymbol{y}}_{(1)}^*$ คือ เวกเตอร์ค่าความสัมพันธ์ของเวกเตอร์ของตัวแปรอิสระ $\boldsymbol{x}_1^*$ และเวกเตอร์ของตัวแปรตาม $\boldsymbol{y}^*$

เราจะได้สมการความถดถอยที่ประมาณค่า $\widehat{\widetilde{\boldsymbol{y}}}_{(1)j}^*$ ทั้งหมด p สมการ

ขั้นที่ *2* คำนวณค่าตัวแปรอิสระใหม่ โดยการรวมสมการประมาณค่า $\widehat{\widetilde{\boldsymbol{y}}}_{(1)j}^*$ โดยไม่คำนึงถึงสหสัมพันธ์ระหว่างตัวแปรอิสระ $\boldsymbol{x}_{(1)j}^*$ ได้ดังนี้

$$\mathbf{p}_1 = \sum_{j=1}^{p} w_{(1)j} b_{(1)j}\widetilde{\boldsymbol{x}}_{(1)j}^*$$

$$\mathbf{p}_1 = \sum_{j=1}^{p} w_{(1)j}\widehat{\widetilde{\boldsymbol{y}}}_{(1)j}^*$$

เมื่อ $\sum_{j=1}^{p} w_{(1)j} = 1$ โดยที่ $w_{(1)j} = \frac{1}{p}$

ขั้นที่ 3 ระบุค่าเริ่มต้นในการคำนวณค่าตัวแปรอิสระใหม่ตัวถัดไป โดยการคำนวณค่าส่วนที่เหลือของการถดถอยก่อนหน้า ซึ่งสามารถหา $\widetilde{\boldsymbol{y}}_{(i+1)}^*$ และ $\widetilde{\boldsymbol{x}}_{(i+1)j}^*$ ได้จาก

$$\widetilde{\boldsymbol{y}}_{(i+1)}^* = \widetilde{\boldsymbol{y}}_{(i)}^* - \{(\mathbf{p}_i'\mathbf{p}_i)^{-1}\mathbf{p}_i'\widetilde{\boldsymbol{y}}_{(i)}^*\}\mathbf{p}_i$$

และ

$$\widetilde{\boldsymbol{x}}_{(i+1)j}^* = \widetilde{\boldsymbol{x}}_{(i)j}^* - \{(\mathbf{p}_i'\mathbf{p}_i)^{-1}\mathbf{p}_i'\widetilde{\boldsymbol{x}}_{(i)j}^*\}\mathbf{p}_i$$

ขั้นตอนในการสร้างตัวแปรอิสระใหม่ $\mathbf{p}_2, \mathbf{p}_3, \dots, \mathbf{p}_p$ จะเป็นการทำซ้ำขั้นตอนที่ 1-3

5.2) วิธีแปลงข้อมูลแบบกรามชมิดต์[2] สำหรับ p มิติของผลคูณภายในปริภูมิเวกเตอร์ V เริ่มต้นด้วยการกำหนดค่าพื้นฐาน $\{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \dots, \boldsymbol{x}_p^*\}$ และสร้างเป็นมูลฐานเชิงตั้งฉาก $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_p\}$ แสดงด้วยขั้นตอน ดังนี้

ขั้นที่ 1 ให้ $\mathbf{p}_1 = \boldsymbol{x}_1^*$

ขั้นที่ 2 ให้ $\mathbf{p}_2 = \boldsymbol{x}_\mathbf{2}^* - proj_{\mathbf{p1}}(\boldsymbol{x}_2^*)$

ขั้นที่ 3 ให้ $\mathbf{p}_3 = \boldsymbol{x}_\mathbf{3}^* - proj_{\mathbf{p1}}(\boldsymbol{x}_3^*) - proj_{\mathbf{p2}}(\boldsymbol{x}_3^*)$

$\vdots$

ขั้นที่ p ให้ $\mathbf{p}_p = \boldsymbol{x}_\boldsymbol{p}^* - \sum_{j=1}^{p} proj_{\boldsymbol{p}_j}(\boldsymbol{x}_p^*)$

โดยกำหนดตัวดำเนินการภาพฉาย (Projection Operator) เป็น $proj_{\boldsymbol{p}}(\boldsymbol{x}^*) = \frac{\langle \boldsymbol{x}^*, \boldsymbol{p} \rangle}{\langle \boldsymbol{p}, \boldsymbol{p} \rangle}\boldsymbol{p}$ และเวกเตอร์ $\boldsymbol{e_i}$ เป็นเวกเตอร์เชิงตั้งฉากปกติ คือ $\boldsymbol{e_j} = \frac{\boldsymbol{p_j}}{\|\boldsymbol{p_j}\|}$

6. ประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุด ได้ดังนี้

$$\widetilde{\boldsymbol{b}} = (\mathbf{P}^T\mathbf{P})^{-\mathbf{1}}\mathbf{P}^T\widetilde{\boldsymbol{y}}$$

7. ประมาณค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานและค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอยของทั้ง 2 วิธี ได้ดังนี้

$$\overline{SE}_l = \frac{\sum_{i=1}^{N} \overline{SE}_i}{N}$$

และ

$$\overline{Bias}_l = \frac{\sum_{i=1}^{N} \overline{Bias}_i}{N}$$

โดยที่ l คือ วิธีการประมาณค่าสัมประสิทธิ์ความถดถอย

N คือ จำนวนรอบของการทำซ้ำ

$\overline{SE}_i$ คือ ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยของตัวแปรอิสระแต่ละรอบของการทำซ้ำ คำนวณได้ดังนี้

$$\overline{SE}_i = \frac{\sum_{j=1}^{j+1} \sqrt{Var(\hat{\beta}_j)}}{j+1}$$

เมื่อ $Var(\hat{\beta}_j)$ คือ ค่าประมาณความแปรปรวนของสัมประสิทธิ์ความถดถอยของตัวแปรอิสระแต่ละตัว คำนวณได้ดังนี้

---

[2] Ramesha K et al. “Gram-Schmidt Orthogonalization Based Face Recognition using DWT.”, International Journal of Engineering Science and Technologe (IJEST), Vol. 3 No. 1 Jan 2011, pp. 498.

$$Var(\hat{\beta}_j) = MSE(\mathbf{P}^T\mathbf{P})^{-1}$$

และ $\overline{Bias_i}$ คือ ค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอยของตัวแปรอิสระแต่ละรอบของการทำซ้ำ คำนวณได้ดังนี้

$$\overline{Bias_i} = \frac{\sum_{j=1}^{j+1}(Bias(\hat{\beta}_j))}{j+1}$$

เมื่อ $Bias(\hat{\beta}_j)$ คือ ค่าความเอนเอียงของสัมประสิทธิ์ความถดถอยของตัวแปรอิสระแต่ละตัว คำนวณได้ดังนี้

$$Bias(\hat{\beta}_j) = \tilde{\beta}_j - \tilde{b}_j$$

8. เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของวิธีการประมาณ จะพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานและค่าเฉลี่ยของความเอนเอียงของสัมประสิทธิ์ความถดถอยของทั้ง 2 วิธี

9. สรุปผลการศึกษา

สรุปผลการวิจัย

จากการจำลองข้อมูลและการวิเคราะห์ผลตามขั้นตอนในการวิจัยครั้งนี้ สำหรับข้อมูลของตัวแปรอิสระที่มีพหุสัมพันธ์ตามระดับความสัมพันธ์ต่างๆ เมื่อพิจารณาค่าเฉลี่ยของ VIF ในการทำซ้ำจำนวน 1000 รอบ ของแต่ละกรณีมีผลการวิเคราะห์ดังนี้

ตารางที่ 1 แสดงค่าเฉลี่ยของ VIF กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร

ระดับความสัมพันธ์		0.9	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99
n=50	VIF1	5.5177	6.2075	6.8314	7.8679	9.1504	10.7955	13.3956	17.9124	26.8981	53.5674
	VIF2	5.5177	6.2075	6.8314	7.8679	9.1504	10.7955	13.3956	17.9124	26.8981	53.5674
n=100	VIF1	5.3804	6.0241	6.6770	7.6451	8.8013	10.4928	13.0428	17.3185	26.0257	51.5789
	VIF2	5.3804	6.0241	6.6770	7.6451	8.8013	10.4928	13.0428	17.3185	26.0257	51.5789
n=200	VIF1	5.3649	5.9130	6.6172	7.4615	8.7037	10.3526	12.8869	17.2239	25.6234	50.6278
	VIF2	5.3649	5.9130	6.6172	7.4615	8.7037	10.3526	12.8869	17.2239	25.6234	50.6278

ผลการวิเคราะห์พบว่า เมื่อพิจารณาค่าเฉลี่ยของ VIF ณ ระดับความสัมพันธ์ต่างๆ สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 2 ตัวแปร ที่ระดับความสัมพันธ์เท่ากับ 0.95 จะเป็นจุดเปลี่ยนที่ทำให้เกิดปัญหาของความสัมพันธ์ระหว่างตัวแปรอิสระของทุกขนาดตัวอย่างที่ทำการศึกษา ดังนั้น สรุปได้ว่าจะเกิดปัญหาของความสัมพันธ์ระหว่างตัวแปรอิสระตั้งแต่ระดับความสัมพันธ์ 0.95-0.99 สำหรับในการศึกษาครั้งนี้ และเมื่อ

พิจารณาแนวโน้มของค่าเฉลี่ยของ VIF ที่ได้พบว่าจะแปรผันตามระดับความสัมพันธ์ของตัวแปรอิสระ และจะแปรผกผันกับขนาดตัวอย่าง

ตารางที่ 2 แสดงค่าเฉลี่ยของ VIF กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับความสัมพันธ์สูงสุด		0.91	0.93	0.95	0.97	0.99
n=50	VIF1	8.1552	10.4649	14.6164	24.5235	83.5899
	VIF2	8.1397	10.4536	14.6195	24.5290	245.6718
	VIF3	8.0798	10.3350	15.0099	32.7385	563.5433
n=100	VIF1	7.9043	10.0277	13.9889	23.3483	80.8529
	VIF2	7.8619	9.9975	13.9310	23.4098	237.8624
	VIF3	7.8576	9.9042	14.3633	31.7155	545.7375
n=200	VIF1	7.7085	9.8223	13.7092	22.8013	78.6370
	VIF2	7.7022	9.8117	13.7536	22.8225	232.0279
	VIF3	7.6777	9.7538	14.0353	30.9554	531.5976

ผลการวิเคราะห์พบว่า เมื่อพิจารณาค่าเฉลี่ยของ VIF ณ ระดับความสัมพันธ์ต่างๆ สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 3 ตัวแปร ที่ระดับความสัมพันธ์เท่ากับ 0.93 จะเป็นจุดเปลี่ยนที่ทำให้เกิดปัญหาของความสัมพันธ์ระหว่างตัวแปรอิสระของขนาดตัวอย่างเท่ากับ 50 และ 100 และขนาดตัวอย่างเท่ากับ 200 จะมีจุดเปลี่ยนอยู่ที่ระดับความสัมพันธ์เท่ากับ 0.95 ดังนั้น สรุปได้ว่าที่ขนาดตัวอย่างเท่ากับ 50 และ 100 จะเกิดปัญหาของความสัมพันธ์ระหว่างตัวแปรอิสระตั้งแต่ระดับความสัมพันธ์ 0.93-0.99 และที่ขนาดตัวอย่างเท่ากับ 200 จะเกิดปัญหาของความสัมพันธ์ระหว่างตัวแปรอิสระตั้งแต่ระดับความสัมพันธ์ 0.95-0.99 สำหรับในการศึกษาครั้งนี้ และเมื่อพิจารณาแนวโน้มของค่าเฉลี่ยของ VIF ที่ได้พบว่าจะแปรผันตามระดับความสัมพันธ์ของตัวแปรอิสระ และจะแปรผกผันกับขนาดตัวอย่าง

ตารางที่ 3 แสดงเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระ กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร

ระดับความสัมพันธ์		0.9	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99
n=50	%Multicollinearity	1.4	2.6	5.8	16.4	32.5	55.6	82.7	97.7	100	100
n=100	%Multicollinearity	0	0.1	1.1	6.8	21.8	55.9	88.5	99.6	100	100
n=200	%Multicollinearity	0	0	0.2	1.5	13.4	57.6	96.4	100	100	100

ผลการวิเคราะห์พบว่า เมื่อพิจารณาค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระ ณ ระดับความสัมพันธ์ต่างๆ สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 2 ตัวแปร ที่ระดับความสัมพันธ์ตั้งแต่

0.95-0.99 จะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระสูง และเมื่อพิจารณาแนวโน้มของค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระ พบว่าจะแปรผันตามระดับความสัมพันธ์ของตัวแปรอิสระและขนาดตัวอย่าง

ตารางที่ 4 แสดงเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับความสัมพันธ์สูงสุด		**0.91**	**0.93**	**0.95**	**0.97**	**0.99**
n=50	%Multicollinearity	32.8	70.4	96.4	100	100
n=100	%Multicollinearity	18.7	67.6	98.5	100	100
n=200	%Multicollinearity	5	63	99.9	100	100

ผลการวิเคราะห์พบว่า เมื่อพิจารณาค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระ ณ ระดับความสัมพันธ์ต่างๆ สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 3 ตัวแปร ที่ระดับความสัมพันธ์ตั้งแต่ 0.93-0.99 จะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระสูง และเมื่อพิจารณาแนวโน้มของค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระ พบว่าจะแปรผันตามระดับความสัมพันธ์ของตัวแปรอิสระ ที่ระดับความสัมพันธ์ตั้งแต่ 0.91-0.93 จะแปรผกผันกับขนาดตัวอย่าง และที่ระดับความสัมพันธ์ตั้งแต่ 0.95-0.99 จะแปรผันตามขนาดตัวอย่าง

ตารางที่ 5 แสดงค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร

ระดับความสัมพันธ์	**n=50**		**n=100**		**n=200**	
	PLS	**OLS_G**	**PLS**	**OLS_G**	**PLS**	**OLS_G**
0.9	0.2857	0.4079	0.1924	0.2908	0.1289	0.2064
0.91	0.2864	0.4196	0.1908	0.2980	0.1296	0.2112
0.92	0.2956	0.4309	0.1931	0.3065	0.1301	0.2182
0.93	0.2957	0.4480	0.1956	0.3187	0.1301	0.2265
0.94	0.3064	0.4700	0.1999	0.3346	0.1304	0.2380
0.95	0.3258	0.5019	0.1974	0.3546	0.1320	0.2514
0.96	0.3359	0.5382	0.2080	0.3817	0.1309	0.2700
0.97	0.3708	0.6011	0.2191	0.4260	0.1338	0.3014
0.98	0.3970	0.7126	0.2281	0.5041	0.1371	0.3566
0.99	0.4944	0.9541	0.2513	0.6770	0.1397	0.4793

ผลการวิเคราะห์เมื่อพิจารณาค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย ณ ระดับความสัมพันธ์ต่างๆ สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 2 ตัวแปร พบว่าค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานของวิธีกำลังสองน้อยสุดเชิงส่วนมีค่าน้อยกว่าวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิดต์ ในทุกระดับความสัมพันธ์และทุกขนาดตัวอย่าง และเมื่อพิจารณาแนวโน้มของค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐาน พบว่าที่ระดับความสัมพันธ์ต่างกัน ของแต่ละขนาดตัวอย่างค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานของวิธีกำลังสองน้อยสุดเชิงส่วนที่ขนาดตัวอย่างเท่ากับ 100 และ 200 จะมีแนวโน้มค่อนข้างคงที่ และที่ขนาดตัวอย่างเท่ากับ 50 จะมีแนวโน้มเพิ่มขึ้น ส่วนวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิดต์จะมีแนวโน้มเพิ่มขึ้นตามระดับความสัมพันธ์ และที่ขนาดตัวอย่างต่างกันของแต่ละระดับความสัมพันธ์ค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานของทั้ง 2 วิธี จะมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

ตารางที่ 6 แสดงค่าเฉลี่ยของความเอนเอียงของสัมประสิทธิ์ความถดถอย กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร

ระดับความสัมพันธ์	n=50		n=100		n=200	
	PLS	OLS_G	PLS	OLS_G	PLS	OLS_G
0.9	0.6384	0.3187	0.6533	0.3111	0.6598	0.3082
0.91	0.6525	0.2987	0.6584	0.3019	0.6640	0.3011
0.92	0.6384	0.3041	0.6542	0.3049	0.6613	0.3050
0.93	0.6342	0.3101	0.6479	0.3171	0.6586	0.3110
0.94	0.6586	0.2931	0.6617	0.3096	0.6660	0.3092
0.95	0.6265	0.3199	0.6470	0.3214	0.6573	0.3217
0.96	0.6305	0.2923	0.6491	0.2988	0.6541	0.3108
0.97	0.6183	0.2880	0.6445	0.3081	0.6501	0.3225
0.98	0.5706	0.3462	0.6169	0.3415	0.6407	0.3331
0.99	0.4845	0.3491	0.5803	0.3419	0.6251	0.3383

ผลการวิเคราะห์เมื่อพิจารณาค่าเฉลี่ยของความเอนเอียงของสัมประสิทธิ์ความถดถอย ณ ระดับความสัมพันธ์ต่างๆ สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 2 ตัวแปร พบว่าค่าเฉลี่ยของความเอนเอียงของวิธีกำลังสองน้อยสุดเชิงส่วนมีค่ามากกว่าวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิดต์ ในทุกระดับความสัมพันธ์และทุกขนาดตัวอย่าง และเมื่อพิจารณาแนวโน้มของค่าเฉลี่ยของความเอนเอียง พบว่าที่ระดับความสัมพันธ์ต่างกันของแต่ละขนาดตัวอย่างค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานของวิธีกำลังสอง

น้อยสุดเชิงส่วนจะมีแนวโน้มค่อนข้างคงที่ แต่ที่ระดับความสัมพันธ์เท่ากับ 0.96-0.99 จะมีแนวโน้มลดลงเล็กน้อย ส่วนวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิดต์จะมีแนวโน้มค่อนข้างคงที่ แต่ที่ระดับความสัมพันธ์ 0.97-0.99 จะมีแนวโน้มเพิ่มขึ้นเล็กน้อย และที่ขนาดตัวอย่างต่างกันของแต่ละระดับความสัมพันธ์ค่าเฉลี่ยของความเอนเอียงของวิธีกำลังสองน้อยสุดเชิงส่วนจะมีแนวโน้มเพิ่มขึ้นเล็กน้อย ส่วนวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิดต์จะมีแนวโน้มค่อนข้างคงที่เมื่อขนาดตัวอย่างเพิ่มขึ้น

ตารางที่ 7 แสดงค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับความสัมพันธ์	n=50		n=100		n=200	
	PLS	OLS_G	PLS	OLS_G	PLS	OLS_G
(0.91,0.93,0.95)	0.2884	0.5475	0.1563	0.3896	0.0272	0.2761
(0.91,0.93,0.97)	0.2912	0.6348	0.1554	0.4494	0.0269	0.3183
(0.91,0.93,0.99)	0.3048	0.9802	0.1586	0.6941	0.0234	0.4916
(0.91,0.95,0.97)	0.2952	0.7144	0.1594	0.5062	0.0269	0.3584
(0.91,0.95,0.99)	0.2983	1.5066	0.1464	1.0637	0.0165	0.7526
(0.93,0.95,0.97)	0.2997	0.6657	0.1590	0.4719	0.0290	0.3347
(0.93,0.95,0.99)	0.3138	1.0412	0.1593	0.7359	0.0244	0.5196
(0.93,0.97,0.99)	0.2564	2.5934	0.1373	1.8294	0.0160	1.2956
(0.95,0.97,0.99)	0.3156	1.1488	0.1561	0.8118	0.0252	0.5757
(0.91,0.91,0.91)	0.3078	0.4752	0.1624	0.3373	0.0194	0.2387
(0.93,0.93,0.93)	0.3053	0.5179	0.1573	0.3691	0.0174	0.2612
(0.95,0.95,0.95)	0.3062	0.5967	0.1501	0.4228	0.0174	0.3001
(0.97,0.97,0.97)	0.3031	0.7367	0.1435	0.5236	0.0203	0.3705
(0.99,0.99,0.99)	0.3194	1.2347	0.1536	0.8719	0.0263	0.6176

ตารางที่ 8 แสดงค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย ที่ค่าระดับความสัมพันธ์สูงสุด กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับความสัมพันธ์สูงสุด	n=50		n=100		n=200	
	PLS	OLS_G	PLS	OLS_G	PLS	OLS_G
0.91	0.3078	0.4752	0.1624	0.3373	0.0194	0.2387
0.93	0.3053	0.5179	0.1573	0.3691	0.0174	0.2612
0.95	0.3062	0.5967	0.1563	0.4228	0.0272	0.3001

0.97	0.3031	0.7367	0.1594	0.5236	0.0290	0.3705
0.99	0.3194	2.5934	0.1593	1.8294	0.0263	1.2956

ผลการวิเคราะห์เมื่อพิจารณาค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย ณ ระดับความสัมพันธ์ต่างๆ สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 3 ตัวแปร พบว่าค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานของวิธีกำลังสองน้อยสุดเชิงส่วนมีค่าน้อยกว่าวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิดต์ ในทุกระดับความสัมพันธ์และทุกขนาดตัวอย่าง และเมื่อพิจารณาแนวโน้มของค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐาน พบว่าที่ระดับความสัมพันธ์ต่างกัน ของแต่ละขนาดตัวอย่างค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานของวิธีกำลังสองน้อยสุดเชิงส่วนจะมีแนวโน้มค่อนข้างคงที่ ส่วนวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิดต์จะมีแนวโน้มเพิ่มขึ้นตามระดับความสัมพันธ์ และที่ขนาดตัวอย่างต่างกัน ของแต่ละระดับความสัมพันธ์ค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานของทั้ง 2 วิธี จะมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

ตารางที่ 9 แสดงค่าเฉลี่ยของความเอนเอียงของสัมประสิทธิ์ความถดถอย กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับความสัมพันธ์	n=50		n=100		n=200	
	PLS	OLS_G	PLS	OLS_G	PLS	OLS_G
(0.91,0.93,0.95)	0.7036	0.6216	0.7319	0.6080	0.7430	0.6069
(0.91,0.93,0.97)	0.6894	0.6366	0.7194	0.6394	0.7307	0.6394
(0.91,0.93,0.99)	0.5968	0.6664	0.6812	0.6778	0.7128	0.6795
(0.91,0.95,0.97)	0.6772	0.6076	0.7178	0.6109	0.7316	0.6170
(0.91,0.95,0.99)	0.3964	0.6174	0.5883	0.6380	0.6707	0.6563
(0.93,0.95,0.97)	0.6923	0.6087	0.7208	0.6220	0.7361	0.6228
(0.93,0.95,0.99)	0.5866	0.6568	0.6697	0.6714	0.7070	0.6614
(0.93,0.97,0.99)	0.2871	0.5984	0.2536	0.6327	0.5125	0.6422
(0.95,0.97,0.99)	0.5432	0.6353	0.6493	0.6495	0.6995	0.6585
(0.91,0.91,0.91)	0.7236	0.5751	0.7313	0.5833	0.7360	0.5831
(0.93,0.93,0.93)	0.7128	0.5829	0.7321	0.5856	0.7443	0.5841
(0.95,0.95,0.95)	0.7005	0.5893	0.7321	0.5854	0.7390	0.5922
(0.97,0.97,0.97)	0.6701	0.6094	0.7135	0.6037	0.7342	0.5986
(0.99,0.99,0.99)	0.4623	0.6395	0.6139	0.6278	0.6792	0.6297

ตารางที่ 10 แสดงค่าเฉลี่ยของความเอนเอียงของสัมประสิทธิ์ความถดถอย ที่ค่าระดับความสัมพันธ์สุดสุด กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับความสัมพันธ์สูงสุด	n=50		n=100		n=200	
	PLS	OLS_G	PLS	OLS_G	PLS	OLS_G
0.91	0.7236	0.5751	0.7313	0.5833	0.7360	0.5831
0.93	0.7128	0.5829	0.7321	0.5856	0.7443	0.5841
0.95	0.7036	0.6216	0.7321	0.6080	0.7430	0.6069
0.97	0.6923	0.6366	0.7208	0.6394	0.7361	0.6394
0.99	0.5968	0.6664	0.6812	0.6778	0.7128	0.6795

ผลการวิเคราะห์เมื่อพิจารณาค่าเฉลี่ยของความเอนเอียงของสัมประสิทธิ์ความถดถอย ณ ระดับความสัมพันธ์ต่างๆ สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 3 ตัวแปร พบว่าค่าเฉลี่ยของความเอนเอียงของวิธีกำลังสองน้อยสุดเชิงส่วนมีค่ามากกว่าวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิดต์ ในทุกระดับความสัมพันธ์และขนาดตัวอย่าง ยกเว้นกรณีที่ระดับความสัมพันธ์เท่ากับ 0.99 ขนาดตัวอย่างเท่ากับ 50 และเมื่อพิจารณาแนวโน้มของค่าเฉลี่ยของความเอนเอียง พบว่าที่ระดับความสัมพันธ์ต่างกันของแต่ละขนาดตัวอย่างค่าเฉลี่ยของความเอนเอียงของวิธีกำลังสองน้อยสุดเชิงส่วนและวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิดต์จะมีแนวโน้มค่อนข้างคงที่ และที่ขนาดตัวอย่างต่างกันของแต่ละระดับความสัมพันธ์ค่าเฉลี่ยของความเอนเอียงวิธีกำลังสองน้อยสุดเชิงส่วนจะมีแนวโน้มลดลงเล็กน้อย ส่วนวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิดต์จะมีแนวโน้มเพิ่มขึ้นเล็กน้อยเมื่อขนาดตัวอย่างเพิ่มขึ้น

อภิปรายผล

จากผลการวิจัยสำหรับข้อมูลของตัวแปรอิสระเท่ากับ 2 ตัวแปร ที่ขนาดตัวอย่างเท่ากับ 50 ที่ระดับความสัมพันธ์เท่ากับ 0.9-0.94 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิดต์มีประสิทธิภาพดีกว่าการประมาณค่าด้วยวิธีกำลังสองน้อยสุดเชิงส่วน แต่ที่ระดับความสัมพันธ์เท่ากับ 0.95-0.99 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดเชิงส่วนมี

ประสิทธิภาพดีกว่าการประมาณค่าด้วยวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิดต์ ที่ขนาดตัวอย่างเท่ากับ 100 ที่ระดับความสัมพันธ์เท่ากับ 0.9-0.97 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิดต์มีประสิทธิภาพดีกว่าการประมาณค่าด้วยวิธีกำลังสองน้อยสุดเชิงส่วน แต่ที่ระดับความสัมพันธ์เท่ากับ 0.98-0.99 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดเชิงส่วนมีประสิทธิภาพดีกว่าการประมาณค่าด้วยวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิดต์ และที่ขนาดตัวอย่างเท่ากับ 200 ที่ระดับความสัมพันธ์เท่ากับ 0.9-0.98 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิดต์มีประสิทธิภาพดีกว่าการประมาณค่าด้วยวิธีกำลังสองน้อยสุดเชิงส่วน แต่ที่ระดับความสัมพันธ์เท่ากับ 0.99 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดเชิงส่วนมีประสิทธิภาพดีกว่าการประมาณค่าด้วยวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิดต์ เนื่องจากวิธีกำลังสองน้อยสุดเชิงส่วนเป็นวิธีที่พัฒนามาเพื่อใช้แก้ไขปัญหาการเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ ส่งผลให้ค่าความแปรปรวนและค่าความเอนเอียงของตัวประมาณมีค่าลดลง โดยปัจจัยที่มีผลต่อค่าความคลาเคลื่อนมาตรฐาน (SE) นั้นพบว่าเมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น วิธีกำลังสองน้อยสุดเชิงส่วนจะมีค่า SE เพิ่มขึ้นเล็กน้อย แต่วิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิดต์จะมีค่า SE เพิ่มขึ้น ถ้าขนาดตัวอย่างเพิ่มขึ้นทั้ง 2 วิธีจะมีค่า SE ลดลง ส่วนปัจจัยที่มีผลต่อค่าความเอนเอียง นั้นพบว่า เมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น วิธีกำลังสองน้อยสุดเชิงส่วนจะมีค่าความเอนเอียงลดลงเล็กน้อย แต่วิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิดต์จะมีค่าความเอนเอียงเพิ่มขึ้นเล็กน้อย ถ้าขนาดตัวอย่างเพิ่มขึ้น วิธีกำลังสองน้อยสุดเชิงส่วนจะมีค่าความเอนเอียงเพิ่มขึ้นเล็กน้อย แต่วิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิดต์จะมีค่าความเอนเอียงค่อนข้างคงที่ จะเห็นว่าเมื่อขนาดตัวอย่างเพิ่มจะไม่มีผลกระทบต่อค่าความเอนเอียงสำหรับวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิดต์

และผลการวิจัยสำหรับข้อมูลของตัวแปรอิสระเท่ากับ 3 ตัวแปร การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดเชิงส่วนมีประสิทธิภาพดีกว่าการประมาณค่าด้วยวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิดต์ ทุกกรณีที่ทำการศึกษา เนื่องจากวิธีกำลังสองน้อยสุดเชิงส่วนเป็นวิธีที่พัฒนามาเพื่อใช้แก้ไขปัญหาการเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ ส่งผลให้ค่าความแปรปรวนและค่าความเอนเอียงของตัวประมาณมีค่าลดลง โดยปัจจัยที่มีผลต่อค่าความคลาดเคลื่อนมาตรฐาน (SE) นั้นพบว่า เมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น วิธีกำลังสองน้อยสุดเชิงส่วนจะมีค่า SE ค่อนข้างคงที่ แต่วิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิดต์จะมีค่า SE เพิ่มขึ้น ถ้าขนาดตัวอย่างเพิ่มขึ้นทั้ง 2 วิธีจะมีค่า SE ลดลง ส่วนปัจจัยที่มีผลต่อค่าความเอนเอียง นั้นพบว่า เมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น วิธีกำลังสองน้อยสุดเชิงส่วนจะมีค่าความเอนเอียงลดลงเล็กน้อย แต่วิธีกำลังสองน้อยสุดด้วยการแปลง

ข้อมูลแบบกรามชมิดต์จะมีค่าความเอนเอียงเพิ่มขึ้นเล็กน้อย ถ้าขนาดตัวอย่างเพิ่มขึ้น วิธีกำลังสองน้อยสุดเชิงส่วนจะมีค่าความเอนเอียงเพิ่มขึ้นเล็กน้อย แต่วิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิดต์จะมีค่าความเอนเอียงค่อนข้างคงที่ จะเห็นว่าเมื่อขนาดตัวอย่างเพิ่มจะไม่มีผลกระทบต่อค่าความเอนเอียงสำหรับวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิดต์

ข้อเสนอแนะ

เพื่อเป็นแนวทางในการเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยของตัวแบบความถดถอยพหุเชิงเส้นที่เกิดพหุสัมพันธ์ที่เหมาะสม สำหรับผู้ที่สนใจศึกษาเพิ่มเติมในครั้งต่อไป

1. ในการศึกษาครั้งนี้มีจำนวนตัวแปรอิสระ 2 และ 3 ตัวแปร สำหรับในการศึกษาครั้งต่อไปควรศึกษาเพิ่มเติมในกรณีที่มีจำนวนตัวแปรอิสระมากขึ้น เพื่อทำให้การประมาณค่าสัมประสิทธิ์ความถดถอยเมื่อเกิดปัญหาพหุเชิงเส้นมีประสิทธิภาพมากขึ้น

2. ในการศึกษาครั้งนี้มีจำนวนตัวแปรตามเพียงตัวเดียว สำหรับในการศึกษาครั้งต่อไปควรศึกษาเพิ่มเติมในกรณีที่มีจำนวนตัวแปรตามมากกว่าหนึ่งตัว

3. ในการศึกษาครั้งนี้ทางผู้วิจัยได้ทำการเปรียบเทียบวิธีการประมาณค่า 2 วิธี ดังนั้นในการศึกษาครั้งต่อไปควรเพิ่มวิธีการประมาณค่า เพื่อให้เกิดการตัดสินใจเลือกวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยได้อย่างเหมาะสมสำหรับข้อมูลที่ทำการศึกษา

รายการอ้างอิง

Paul H. Garthwaite. “An Interpretation of Partial Least Squares.”, Journal of the American Statistical Association, Vol. 89, No. 425 (Mar., 1994): 122-127.

Ramesha K et al. “Gram-Schmidt Orthogonalization Based Face Recognition using DWT.”, International Journal of Engineering Science and Technologe (IJEST), Vol. 3 No. 1 Jan 2011, pp. 498.