



Phương pháp ước lượng LASSO: Cơ sở toán học và ứng dụng*

BÙI THỊ THIÊN MỸ**

Trường Đại học Ngân hàng TP. Hồ Chí Minh

THÔNG TIN	TÓM TẮT
Ngày nhận: 26/05/2020 Ngày nhận lại: 01/09/2020 Duyệt đăng: 14/09/2020 Mã phân loại JEL: C1; C6 Từ khóa: Hồi quy; LASSO; Mô hình co gọn. Keywords: Regression; LASSO; Shrinkage model.	Bài báo giới thiệu phương pháp ước lượng LASSO cho phân tích hồi quy. Tương tự phương pháp OLS, LASSO tìm các ước lượng tham số của một mô hình hồi quy bằng cách cực tiểu hóa hàm mục tiêu là tổng bình phương các phần dư. Tuy nhiên, phương pháp ước lượng LASSO áp đặt ràng buộc tổng giá trị tuyệt đối của các tham số không vượt quá một tham số điều chỉnh. Bài viết trình bày tổng quan ý nghĩa, cơ sở toán học của phương pháp ước lượng LASSO, đồng thời giới thiệu một tình huống thực nghiệm áp dụng phương pháp ước lượng LASSO nhằm xác định các yếu tố nội tại có ảnh hưởng đến các tỷ suất sinh lợi trên tổng tài sản, tỷ suất sinh lợi trên vốn chủ sở hữu của các ngân hàng thương mại cổ phần Việt Nam. Kết quả cho thấy LASSO đã loại các yếu tố không có ảnh hưởng đến biến phụ thuộc, cung cấp một mô hình hồi quy đơn giản hơn với sai số MSE nhỏ hơn so với bằng phương pháp ước lượng OLS, tuy vậy, sự cải thiện này chưa đáng kể. Abstract This paper introduces the LASSO estimation method for regression analysis. Similarly OLS estimation method, LASSO finds estimates of the parameters of a regression model by minimizing the objective function as the sum of squared-residuals. However, LASSO imposes a constraint on the sum of absolute values of parameters in the model by a tuning parameter. The paper presents an overview of the meaning and the mathematical background of LASSO. Besides, the paper introduces an empirical case applying LASSO estimation method to identify intrinsic factors affecting the ROA and ROE of Vietnamese joint stock

* Bài viết này thuộc chuyên san Toán Kinh tế do PGS.TS. Lê Xuân Trường (lxuantruong@ueh.edu.vn) phụ trách.

** Tác giả liên hệ: Bùi Thị Thiên Mỹ.

Email: mybtt@buh.edu.vn (Bùi Thị Thiên Mỹ)

Trích dẫn bài viết: Bùi Thị Thiên Mỹ. (2020). Phương pháp ước lượng LASSO: Cơ sở toán học và ứng dụng. *Tạp chí Nghiên cứu Kinh tế và Kinh doanh Châu Á*, 31(4), 05–22.

commercial banks. The results show that LASSO has eliminated factors that do not affect ROA and ROE, providing a simpler regression model with a smaller MSE than OLS method does, although this improvement is not significant.

1. Giới thiệu

Phân tích hồi quy là một công cụ thường được sử dụng trong nghiên cứu kinh tế, tài chính với mục đích dự báo và giải thích ảnh hưởng. Phân tích hồi quy thông thường cung cấp một mô hình toán học với đầu ra là giá trị trung bình của biến phụ thuộc khi biết giá trị đầu vào của các biến độc lập. Xét mô hình hồi quy tổng thể có dạng:

$$Y_i = \beta_o + \beta_1 X_{1i} + \dots + \beta_p X_{pi} + u_i \quad (1)$$

Trong đó,

Y : Biến phụ thuộc (biến được giải thích);

X_j ($j = 1, \dots, p$): Các biến độc lập (biến giải thích);

u : Sai số ngẫu nhiên với kỳ vọng bằng 0.

Nội dung của bài toán ước lượng OLS có thể tóm tắt như sau:

Từ mẫu ngẫu nhiên $\{(Y_i, X_{1i}, \dots, X_{pi})\}_{i=1}^n$, cần tìm ước lượng $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$ sao cho:

$$\min_{\beta \in \mathbb{R}^p} \left\{ \sum_{i=1}^n (Y_i - \beta_o - \beta_1 X_{1i} - \dots - \beta_p X_{pi})^2 \right\} \quad (2)$$

Năm 1822, định lý Gauss – Markov do nhà toán học Gauss và Markov độc lập chứng minh đã khẳng định rằng phương pháp OLS cho kết quả ước lượng mô hình (1) là tuyến tính, không chệch và hiệu quả trong lớp các ước lượng tuyến tính nếu bốn điều kiện sau đây thỏa mãn: (1) Không có tương quan giữa các sai số ngẫu nhiên; (2) kỳ vọng có điều kiện của sai số ngẫu nhiên u bằng 0; (3) phương sai có điều kiện của u là không đổi; và (4) không có đa cộng tuyến hoàn hảo giữa các biến độc lập. Nếu bổ sung thêm giả thiết (5) sai số ngẫu nhiên u tuân theo luật phân phối chuẩn thì ước lượng OLS là hiệu quả trong lớp các ước lượng không chệch, kể cả ước lượng tuyến tính và phi tuyến (Wooldridge, 2012).

Khi sử dụng mô hình hồi quy (1), hai câu hỏi sau đây thường được đặt ra:

- Có phải ít nhất một biến độc lập $X_1, X_2, \dots, X_p$ hữu ích trong việc giải thích hoặc dự báo giá trị của biến phụ thuộc Y không?

- Tất cả các biến độc lập $X_1, X_2, \dots, X_p$ hay chỉ một tập con các biến là cần thiết trong mô hình?

Để trả lời các câu hỏi trên, các kiểm định giả thuyết thống kê về các hệ số hồi quy β_j được thực hiện. Khi đó, các giá trị xác suất (p-value) của các thống kê quan sát được sử dụng để đi đến các

khẳng định về hệ số hồi quy β_j , qua đó giải thích ảnh hưởng của các biến độc lập $X_1, X_2, \dots, X_p$ đến biến phụ thuộc Y . Với những hạn chế, thậm chí sai lầm trong kết luận dựa trên p-value theo các phân tích của Nguyen (2016, 2019), Casella và Berger (2002), Cohen (1994), phương pháp ước lượng OLS có thể gặp bế tắc trong việc tìm câu trả lời cho các câu hỏi (1) và (2). Đặc biệt, trong trường hợp $p > n - 1$, phương pháp ước lượng OLS cho kết quả không duy nhất, không thể xác định xu hướng và mức độ ảnh hưởng của mỗi biến độc lập X_j đến biến phụ thuộc Y . Bên cạnh đó, kết quả ước lượng OLS cho mô hình (1) thường là ước lượng không chệch và phương sai lớn, đặc biệt trong các trường hợp: Các biến độc lập có tương quan tuyến tính chặt chẽ, số biến độc lập trong mô hình khá nhiều (Tibshirani, 1996). Vì phương sai lớn nên tính chính xác của các ước lượng OLS không cao, từ đó, các kết quả suy diễn thống kê không đáng tin cậy.

Vì những bất tiện của phương pháp OLS như vừa chỉ ra, cần tìm một phương pháp ước lượng mô hình hồi quy (1) sao cho chỉ một tập con gồm k biến độc lập ($k < n$) có ảnh hưởng mạnh lên biến phụ thuộc. Vấn đề này được đề cập đến bằng thuật ngữ mô hình co gọn (Shrinkage Model). Việc thu hẹp tập các biến độc lập vừa nhằm mục đích đơn giản hóa trong phân tích ảnh hưởng vừa có thể cải thiện năng lực dự báo của mô hình hồi quy (1). Tibshirani (1996) cho rằng có thể hy sinh tính không chệch đổi lấy sự suy giảm của phương sai của các ước lượng OLS bằng cách thiết lập các ràng buộc về độ lớn cho các tham số của mô hình (1). Phương pháp ước lượng LASSO (The Least Absolute Shrinkage and Selection Operator) lần đầu tiên được Tibshirani (1996) giới thiệu, có nội dung như sau:

Xét mô hình hồi quy (1), từ mẫu ngẫu nhiên $\left\{ \left(Y_i, X_{1i}, \dots, X_{pi} \right) \right\}_{i=1}^n$, cần tìm ước lượng $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$ sao cho:

$$\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^n \left(Y_i - \beta_0 - \beta_1 X_{1i} - \dots - \beta_p X_{pi} \right)^2 \right\} \text{ với điều kiện } \sum_{j=1}^p |\beta_j| \leq t \quad (3)$$

Khi giải bài toán (3), ta nhận được một số ước lượng $\hat{\beta}_j$ (từ đây sẽ gọi là ước lượng LASSO) có giá trị 0, vì thế mô hình (1) thu gọn về ít biến độc lập hơn, từ đó dễ phân tích hơn. Tuy vậy, giải quyết bài toán tối ưu (3) tương đương với giải bài toán tối ưu mà hàm mục tiêu không khả vi tại một số điểm. Nói cách khác, không thể sử dụng phương pháp đạo hàm riêng để giải bài toán tối ưu (3) như đã từng làm đối với bài toán tìm ước lượng OLS. Nghiên cứu này sẽ làm rõ cơ sở toán học, giới thiệu một thuật toán tìm ước lượng LASSO và một tình huống nghiên cứu thực nghiệm ứng dụng phương pháp ước lượng LASSO.

Bài viết gồm 4 phần: Phần 1 giới thiệu ý nghĩa của phương pháp ước lượng LASSO; phần 2 giới thiệu cơ sở lý thuyết, bao gồm: Bài toán tìm ước lượng hệ số LASSO, cơ sở toán học, thuật toán giảm chiều và một số tính chất cơ bản của ước lượng LASSO; phần 3 trình bày ví dụ ứng dụng LASSO; và phần 4 là kết luận.

2. Phương pháp ước lượng LASSO

2.1. Bài toán tìm ước lượng hệ số LASSO

Dùng ngôn ngữ vectơ-ma trận, ta ký hiệu $\sum_{j=1}^p |\beta_j| = \|\beta\|_1$; $\sum_{j=1}^p \beta_j^2 = \|\beta\|_2^2$ với $\beta \in \mathbb{R}^p$ (lưu ý $\|\cdot\|_1, \|\cdot\|_2$ là chuẩn l_1 và chuẩn Euclide trên không gian mêtric $\mathbb{R}^p$), $Y = (Y_1, \dots, Y_n) \in \mathbb{R}^n$ là vectơ biến phụ thuộc, $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^n$, X là ma trận $n \times p$ gồm các vectơ cột là các biến độc lập, $\beta = (\beta_1, \dots, \beta_p)$; bài toán (3) – tìm ước lượng LASSO – được viết lại như sau:

$$\min \left\{ \frac{1}{2n} \|\mathbf{Y} - \beta_0 \cdot \mathbf{1} - \beta \mathbf{X}^T\|_2^2 \right\} \text{ với điều kiện } \|\beta\|_1 \leq t \quad (4)$$

Trong bài toán (3) và (4), giá trị $t > 0$ được gọi là tham số điều chỉnh. Khi giá trị t đủ lớn, các ước lượng $\hat{\beta}_j$ thu được từ bài toán (4) trùng với các ước lượng $\hat{\beta}_j$ thu được từ bài toán (2). Ngược lại, khi giá trị t gần 0, các ước lượng $\hat{\beta}_j$ rất bé và do ràng buộc bởi chuẩn $\|\cdot\|_1$ nên một số ước lượng nhận giá trị bằng 0.

Giả sử rằng các biến độc lập X_j được chuẩn hóa theo công thức sau:

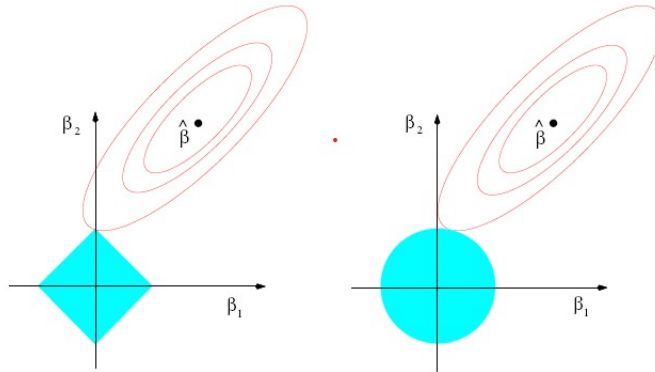
$$X'_j = \frac{X_j - \overline{X_j}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (X_{ji} - \overline{X_j})^2}}, \text{ với } \overline{X_j} = \frac{1}{n} \sum_{i=1}^n X_{ji}$$

Hơn nữa, giả sử biến phụ thuộc Y có trung bình bằng 0. Khi đó, tham số $\beta_0 = 0$ và ta có thể bỏ tham số β_0 ra khỏi mô hình (3), (4). Do đó, trong khuôn khổ của bài viết này, thay vì giải bài toán (3), (4), ta sẽ xét bài toán (5) với giả thiết các biến độc lập và biến phụ thuộc được chuẩn hóa (tức là các biến độc lập có trung bình bằng 0, phương sai bằng 1; biến phụ thuộc có trung bình bằng 0). Khi đó, bài toán (4) trở thành:

$$\min_{\beta} \left\{ \frac{1}{2n} \|\mathbf{Y} - \beta \mathbf{X}^T\|_2^2 \right\} \text{ với điều kiện } \|\beta\|_1 \leq t \quad (5)$$

Ta so sánh bài toán (5) và bài toán ước lượng Ridge như sau:

$$\min_{\beta} \left\{ \frac{1}{2n} \|\mathbf{Y} - \beta \mathbf{X}^T\|_2^2 \right\} \text{ với điều kiện } \|\beta\|_2 \leq t \quad (6)$$



Hình 1. Minh họa hình học hàm mục tiêu và miền giới hạn của LASSO (trái) và Ridge (phải)

Nguồn: Hastie và cộng sự (2015).

Các hàm mục tiêu trong bài toán (5) và (6) giống nhau, tuy nhiên, điều kiện ràng buộc của hai bài toán khác nhau. Hình 1 minh họa kết quả ước lượng từ LASSO và Ridge cho trường hợp $p = 2$. Phần diện tích đậm màu là miền phẳng xác định bởi điều kiện $|\beta_1| + |\beta_2| \leq t$ (trái) và $\beta_1^2 + \beta_2^2 \leq t$ (phải). Cả hai phương pháp đều mong muốn tìm giao điểm đầu tiên của các đường Elip (là đồ thị của hàm mục tiêu) với miền giới hạn. Với miền giới hạn là đa giác có các góc nằm trên các trục tọa độ, nhiều khả năng giao điểm sẽ rơi vào các góc này, khi đó sẽ có một tham số β_j bằng 0. Điều này không thể xảy ra khi miền giới hạn là hình tròn. Do đó, bài toán (6) sẽ cho kết quả các ước lượng khác 0. Khi $p > 2$, miền giới hạn ở phương pháp LASSO sẽ có nhiều góc, cạnh phẳng và mặt, vì thế có nhiều khả năng cho kết quả ước lượng bằng 0 hơn (Hastie và cộng sự, 2009).

2.2. Cơ sở toán học của phương pháp ước lượng LASSO

2.2.1. Bài toán tương đương

Trở lại bài toán ước lượng LASSO (5):

$$\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2n} \|\mathbf{Y} - \beta \mathbf{X}^T\|_2^2 \right\} \text{ với điều kiện } \|\beta\|_1 \leq t$$

Đặt $f(\beta) = \frac{1}{2n} \|\mathbf{Y} - \beta \mathbf{X}^T\|_2^2$ và $g(\beta) = \|\beta\|_1$. Khi đó $f(\beta), g(\beta)$ là các hàm lồi và liên tục trên $\mathbb{R}^p$; hơn nữa tập ràng buộc $\{\beta \in \mathbb{R}^p | g(\beta) - t \leq 0\}$ là tập lồi, đóng và bị chặn trên $\mathbb{R}^p$. Theo các kết quả tối ưu lồi, bài toán (5) tồn tại lời giải được trình bày trong mục 2.3 và có thể đưa về bài toán tương đương như sau:

$$\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2n} \|\mathbf{Y} - \beta \mathbf{X}^T\|_2^2 + \lambda \|\beta\|_1 \right\} \text{ với } \lambda > 0 \quad (7)$$

Trong bài toán (7), tham số λ được gọi là mức phạt của mô hình. Khi λ đủ lớn, tất cả các giá trị ước lượng của β_j từ bài toán (7) bằng 0. Khi λ bé, chỉ một vài giá trị $\hat{\beta}_j$ bằng 0, tương ứng với một số biến độc lập bị loại khỏi mô hình (1). Tham số λ tương ứng một - một với tham số điều chỉnh t

trong bài toán (5). Với mỗi tham số điều chỉnh t , ta có thể xác định duy nhất giá trị λ sao cho $\|\hat{\beta}\|_1 \leq t$. Ngược lại, với mỗi tham số λ , tham số điều chỉnh t được xác định bởi $\widehat{\beta}_{\lambda_1} = t$.

Bài toán (7) là bài toán tối ưu lồi không ràng buộc, với hàm mục tiêu là

$$f(\beta) + \lambda g(\beta) = \frac{1}{2n} \|Y - \beta X^T\|_2^2 + \lambda \|\beta\|_1$$

Tuy nhiên, hàm mục tiêu này không khả vi tại những điểm $\beta_j = 0$. Để giải bài toán (7), ta cần một số kiến thức về hàm lồi, điều kiện cần và đủ cho bài toán tối ưu lồi được trình bày trong mục kế tiếp.

2.2.2. Hàm lồi và cực tiểu của hàm lồi

Các kiến thức về tối ưu lồi sau đây bỏ qua phần chứng minh, độc giả quan tâm vấn đề này có thể tham khảo các tài liệu của Bertsekas (2003), Boyd và Vanderberghe (2004).

- **Tập lồi**

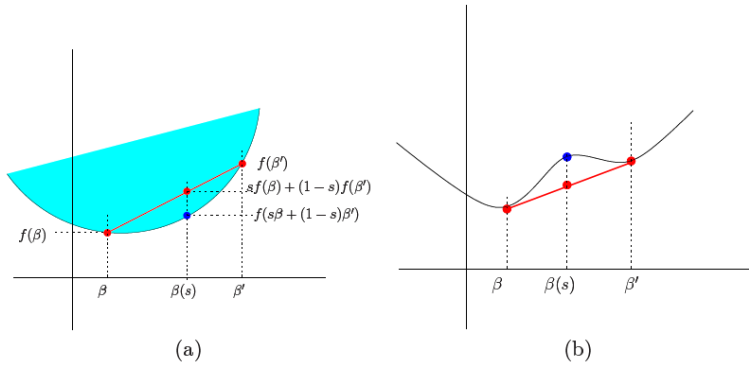
Tập hợp $C \subset \mathbb{R}^p$ được gọi là tập lồi nếu với mọi $\beta, \beta' \in C$ và với mọi số thực $s \in [0, 1]$, ta luôn có $s\beta + (1-s)\beta' \in C$.

- **Hàm lồi**

Hàm số $f: \mathbb{R}^p \rightarrow \mathbb{R}$ được gọi là hàm lồi nếu tập xác định của f là tập lồi và với mọi β, β' thuộc miền xác định của f , với mọi số thực $s \in (0, 1)$, ta luôn có:

$$f(s\beta + (1-s)\beta') \leq sf(\beta) + (1-s)f(\beta').$$

Nếu với mọi $\beta \neq \beta'$, bất đẳng thức trong định nghĩa trên là $<$ thì hàm số f được gọi là lồi chặt.



Hình 2. Đồ thị hàm lồi (a) và không lồi (b)

Nguồn: Hastie và cộng sự (2017).

Hình 2 minh họa đồ thị hàm lồi (a) và không lồi (b). Theo định nghĩa hàm lồi, mọi đoạn thẳng nối hai điểm $(\beta, f(\beta)); (\beta', f(\beta'))$ luôn nằm trên hoặc trùng với phần đồ thị hàm số trên khoảng (β, β') . Hơn nữa, nếu đoạn thẳng nối 2 điểm $(\beta, f(\beta)); (\beta', f(\beta'))$ luôn nằm trên thì hàm số là lồi chặt. Đồ thị (a) là đồ thị của một hàm lồi chặt.

Ví dụ: Trên $\mathbb{R}$, các hàm $f(x) = x^2$, $g(x) = |x|$ là hàm lồi. Trong đó, $f(x) = x^2$ là hàm lồi chặt; $g(x) = |x|$ không là hàm lồi chặt. Các chuẩn trên $\mathbb{R}^p$ là những hàm lồi, trong đó chuẩn Euclidean $\|\cdot\|_2$ là hàm lồi chặt, chuẩn $\|\cdot\|_1$ không là hàm lồi chặt.

- *Gradient* (bậc nhất) của hàm số $f: \mathbb{R}^p \rightarrow \mathbb{R}$ tại $\beta^* \in \mathbb{R}^p$, ký hiệu là $\nabla f(\beta^*)$, là vectơ

$$\nabla f(\beta^*) = \left(\frac{\partial f}{\partial x_1}(\beta^*), \dots, \frac{\partial f}{\partial x_p}(\beta^*) \right) \in \mathbb{R}^p,$$

với $\frac{\partial f}{\partial x_j}(\beta^*)$ là đạo hàm riêng của f theo biến x_j tại β^* .

- *Subgradient* của hàm số $f: \mathbb{R}^p \rightarrow \mathbb{R}$ tại $\beta^* \in \mathbb{R}^p$ là vector z thỏa mãn:

$$f(\beta) \geq f(\beta^*) + (z, \beta - \beta^*), \quad \forall \beta \in \mathbb{R}^p$$

Trong đó, $\langle x, y \rangle$ là ký hiệu tích vô hướng của $x, y \in \mathbb{R}^p$. Tập tất cả các Subgradient của f tại β^* là tập lồi trên $\mathbb{R}^p$ và được ký hiệu là $\partial f(\beta^*)$. Nếu hàm số f khả vi tại β^* thì $\partial f(\beta^*) = \{\nabla f(\beta^*)\}$.

Ví dụ: Subgradient của hàm số $f(\beta) = |\beta|$ là:

$$\partial f(\beta) = \begin{cases} \{1\}, & \beta > 0 \\ \{-1\}, & \beta < 0 \\ [-1; 1], & \beta = 0 \end{cases} =: \text{sign}(\beta)$$

Ta thường dùng kí hiệu $\text{sign}(\beta)$ để chỉ $\partial f(\beta)$ của hàm số $f(\beta) = |\beta|$.

• *Cực tiểu của hàm lồi*

- *Tính chất cực tiểu của hàm lồi*: Giả sử f là một hàm lồi, nếu hàm f đạt cực tiểu địa phương tại x_0 thì x_0 là điểm cực tiểu toàn cục của f . Hơn nữa, nếu f là hàm lồi chặt, có cực tiểu địa phương thì cực tiểu địa phương đó là cực tiểu toàn cục và duy nhất.

- *Điều kiện cần và đủ của cực tiểu toàn cục*: Giả sử $f: \mathbb{R}^p \rightarrow \mathbb{R}$ là hàm lồi và $C \subset \mathbb{R}^p$ là tập lồi. Xét bài toán tối ưu:

$$\min_{\beta \in C} f(\beta) \quad (8)$$

(1) Giả sử f khả vi tại β^* . Khi đó β^* là điểm cực tiểu toàn cục của f trên C khi và chỉ khi $\langle \nabla f(\beta^*), \beta - \beta^* \rangle \geq 0, \forall \beta \in C$.

Trong trường hợp $C = \mathbb{R}^p$, bài toán (8) trở thành tối ưu không có điều kiện ràng buộc. Khi đó, điều kiện cần và đủ để điểm β^* là điểm cực tiểu toàn cục của f là $\nabla f(\beta^*) = 0$.

(2) Giả sử $C = \mathbb{R}^p$ và f không khả vi tại β^* . Khi đó, điều kiện cần và đủ để điểm β^* là điểm cực tiểu toàn cục của f là $0 \in \partial f(\beta^*)$.

Trở lại bài toán (7), điều kiện cần và đủ để β^* là nghiệm là :

$$0 \in \partial f(\beta^*) + \lambda^* \partial g(\beta^*) \text{ hay } 0 = \nabla f(\beta^*) + \lambda^* z^* \quad (9)$$

Trong đó, vector $z^* \in \mathbb{R}^p$ thỏa mãn $z_j^* \in \text{sign}(\beta_j^*)$ với mọi $j = 1, \dots, p$.

2.3. Thuật toán giảm chiều

Trong nhiều thuật toán tìm lời giải cho bài toán (7), giảm chiều (Coordinate Descent – CD) là thuật toán tuy đơn giản nhưng rất hiệu quả.

2.3.1. Trường hợp hồi quy đơn

Bài toán (7) được phát biểu như sau: Từ mẫu $\{(Y_i, X_i)\}_{i=1}^n$, cần tìm $\hat{\beta} \in \mathbb{R}$ sao cho:

$$\min_{\beta \in \mathbb{R}} \left\{ \frac{1}{2n} \sum_{i=1}^n (Y_i - X_i \beta)^2 + \lambda |\beta| \right\} \text{ với } \lambda > 0 \quad (10)$$

Từ điều kiện cần và đủ được cho bởi bài toán (8), nghiệm của bài toán (10) được xác định như sau:

$$\frac{1}{n} \sum_{i=1}^n (-X_i)(Y_i - X_i \hat{\beta}) + \lambda z = 0 \Leftrightarrow \hat{\beta} = \frac{1}{n} \langle Y, X \rangle - \lambda z \text{ (với } z \in \text{sign}(\hat{\beta}))$$

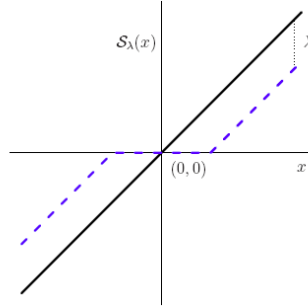
Như vậy:

$$\hat{\beta} = \begin{cases} \frac{1}{n} \langle Y, X \rangle - \lambda, & \frac{1}{n} \langle Y, X \rangle > \lambda \\ 0, & \left| \frac{1}{n} \langle Y, X \rangle \right| \leq \lambda \\ \frac{1}{n} \langle Y, X \rangle + \lambda, & \frac{1}{n} \langle Y, X \rangle < -\lambda \end{cases}$$

Nếu đặt $S_\lambda(x) = \text{sign}(x)(|x| - \lambda)^+$ ($S_\lambda(x)$ được gọi là toán tử biên mềm) thì kết quả ước lượng trên có thể viết như sau:

$$\hat{\beta} = S_\lambda\left(\frac{1}{n} \langle Y, X \rangle\right) \quad (11)$$

Lưu ý rằng $\tilde{\beta} = \frac{1}{n} \langle Y, X \rangle$ là kết quả ước lượng bằng phương pháp OLS.



Hình 3. Đồ thị của toán tử biên mềm

Hình 3 minh họa đồ thị của toán tử biên mềm $S_\lambda(x)$. Khi $\lambda = 0$ thì $S_\lambda(x) = x$. Tuy nhiên, khi $\lambda > 0$ thì $|S_\lambda(x)| \leq |x|$, nói cách khác, toán tử $S_\lambda(x)$ co độ lớn của x về 0. Theo công thức (11), có thể cho rằng trường hợp hồi quy đơn, phương pháp LASSO đã co độ lớn ước lượng OLS về 0.

2.3.2. Trường hợp hồi quy bội

Ý tưởng của thuật toán giảm chiều trong trường hợp hồi quy bội như sau: Lần lượt áp dụng thuật toán giảm chiều, tìm cực tiểu hàm mục tiêu theo một biến độc lập và giữ không đổi các biến độc lập còn lại.

Cụ thể, hàm mục tiêu trong bài toán (7) được viết lại như sau:

$$\frac{1}{2n} \sum_{i=1}^n (Y_i - \sum_{k \neq j} \beta_k X_{ki} - \beta_j X_{ji})^2 + \lambda \sum_{k \neq j} |\beta_k| + \lambda |\beta_j| \quad (12)$$

Gọi vector chỉ phần dư riêng phần thứ j là $r^{(j)}$ với $r_i^{(j)} = Y_i - \sum_{k \neq j} \beta_k X_{ki}$. Tại vòng lặp thứ j , thuật toán giảm chiều thực hiện tối thiểu hóa hàm mục tiêu (12) theo tham số β_j . Hệ số $\hat{\beta}_j$ được cập nhật bằng cách cực tiểu hóa hàm mục tiêu đơn biến khi các hệ số $\hat{\beta}_k, k \neq j$ được giữ không đổi. Theo công thức (11), suy ra tại bước j :

$$\hat{\beta}_j = S_\lambda\left(\frac{1}{n} r^{(j)}, X_j\right) \quad (13)$$

Quá trình cứ thế được lặp lại tuần tự trên các tham số khác cho đến khi hàm mục tiêu đạt cực tiểu toàn cục. Trong thực hành, việc tìm lời giải cho phương pháp ước lượng LASSO thực hiện trên một tập các giá trị của λ thay vì tại một giá trị đơn lẻ. Ban đầu, thuật toán giảm chiều áp dụng với $\lambda = \lambda_{\max}$ đủ lớn (thường chọn $\lambda_{\max} = \max_j \left| \frac{1}{n} \langle Y, X_j \rangle \right|$), kết quả nhận được là $\hat{\beta}_j = 0 \quad \forall j \in \{1, \dots, p\}$. Sau đó, giảm giá trị một lượng nhỏ định trước, tiếp tục áp dụng thuật toán giảm chiều với λ tương ứng; kết quả ước lượng $\hat{\beta}_j$ ở bước trước đó được xem như là giá trị khởi đầu của các tham số $\hat{\beta}_j$ ở bước kế tiếp. Quá trình cứ thế tiếp tục cho đến khi bài toán (7) đạt cực tiểu toàn cục tại $(\hat{\beta}^*; \lambda^*)$. Thuật toán vừa trình bày được gọi là giảm chiều theo quỹ đạo (Pathwise Coordinate Descent). Các thuật toán được tóm tắt ở Bảng 1 và Bảng 2.

Bảng 1.

Thuật toán 1. Thuật toán giảm chiều

Dữ liệu vào	Mẫu $\{(Y_i, X_{i1}, \dots, X_{ip})\}_{i=1}^n$ và λ .
Dữ liệu ra	$\hat{\beta}^* = (\hat{\beta}_1^*, \dots, \hat{\beta}_p^*)$ sao cho thỏa mãn: $\min \left\{ \frac{1}{2n} \sum_{i=1}^n (Y_i - \sum_{k=1}^p \hat{\beta}_k X_{ki})^2 + \lambda \sum_{k=1}^p \hat{\beta}_k \right\}$
Bước 1	Gán $j = 1$ và $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p) = (0, \dots, 0)$
Bước 2	Tính $r_i^{(j)} = Y_i - \sum_{k \neq j} \hat{\beta}_k X_{ki}$
Bước 3	Tính $\hat{\beta}_j = S_{\lambda} \left(\frac{1}{n} \langle r^{(j)}, X_j \rangle \right)$
Bước 4	+ Tính giá trị hàm mục tiêu: $F_j = \frac{1}{2n} \sum_{i=1}^n (Y_i - \sum_{k=1}^p \hat{\beta}_k X_{ki})^2 + \lambda \sum_{k=1}^p \hat{\beta}_k $ + j tăng thêm 1 đơn vị.
Bước 5	Lặp lại từ bước 2 đến bước 4 cho đến khi $j = p + 1$ thì dừng.
Bước 6	Trả về $\hat{\beta}^* = (\hat{\beta}_1^*, \dots, \hat{\beta}_p^*)$ tương ứng với $\min\{F_j\}_{j=1}^p$

Bảng 2.

Thuật toán 2. Thuật toán Giảm chiều theo quỹ đạo

Dữ liệu vào	Mẫu $\{(Y_i, X_{i1}, \dots, X_{ip})\}_{i=1}^n$ và tập hợp các giá trị tăng dần của λ : $\lambda \in \{\lambda_1, \dots, \lambda_N\}$ (với $\lambda_N = \max_j \left \frac{1}{n} \langle Y, X_j \rangle \right $)
Dữ liệu ra	$(\hat{\beta}^*, \lambda^*) = (\hat{\beta}_1^*, \dots, \hat{\beta}_p^*, \lambda^*)$ sao cho thỏa mãn: $\min \left\{ \frac{1}{2n} \sum_{i=1}^n (Y_i - \sum_{k=1}^p \beta_k X_{ki})^2 + \lambda \sum_{k=1}^p \beta_k \right\}$
Bước 1	
Bước 1.1	Gán $i = N$ và $\widehat{\beta} = (\widehat{\beta}_1, \dots, \widehat{\beta}_p) = (0, \dots, 0)$
Bước 1.2	Gán $j = 1$, tính $r_i^{(j)} = Y_i - \sum_{k \neq j} \widehat{\beta}_k X_{ki}$
Bước 1.3	Với $\lambda = \lambda_i$, tính $\widehat{\beta}_j = S_{\lambda_i} \left(\frac{1}{n} \langle r^{(j)}, X_j \rangle \right)$
Bước 1.4	+ Tính giá trị hàm mục tiêu: $F_{ij} = \frac{1}{2n} \sum_{i=1}^n (Y_i - \sum_{k=1}^p \widehat{\beta}_k X_{ki})^2 + \lambda_i \sum_{k=1}^p \widehat{\beta}_k $ + j tăng thêm 1 đơn vị
Bước 1.5	Lặp lại từ Bước 1.2 đến Bước 1.5 cho đến khi $j = p + 1$
Bước 2	+ i giảm một đơn vị + Lặp lại Bước 1 cho đến khi $i = 0$
Bước 3	Trả về $(\hat{\beta}^*; \lambda^*)$ tương ứng với $\min \{F_{ij}\}_{i,j}$.

2.4. Các tính chất của ước lượng LASSO

- Các ước lượng LASSO là ước lượng chệch. Chẳng hạn trường hợp $p = 1$, ước lượng LASSO có công thức tính là $\hat{\beta} = S_\lambda(\tilde{\beta})$ với $\tilde{\beta}$ là ước lượng OLS. Mặt khác, vì giả thiết kỳ vọng có điều kiện của sai số ngẫu nhiên u ở mô hình (1) bằng 0 nên $\tilde{\beta}$ là ước lượng không chệch: $E(\tilde{\beta}) = \beta$. Với $\lambda \neq 0$, ta có:

$$E(\hat{\beta}) = E(S_\lambda(\tilde{\beta})) = E(\text{sign}(\tilde{\beta})(|\tilde{\beta}| - \lambda)^+) = \text{sign}(\tilde{\beta})(E(|\tilde{\beta}|) - \lambda)^+ = \text{sign}(\tilde{\beta})(|\beta| - \lambda)^+ \neq \beta$$

- Ước lượng LASSO không duy nhất. Từ bài toán (6) nhận thấy ước lượng LASSO là điểm cực tiểu toàn cục của một hàm lồi không chặt, do đó có thể không là cực tiểu duy nhất. Một số trường hợp LASSO cho kết quả ước lượng không duy nhất, ví dụ: Các biến độc lập có tương quan tuyến tính hoàn hảo hoặc dữ liệu X có biến định tính, được mã hóa bởi biến giả 0–1. Ngược lại, nếu dữ liệu X được rút ra từ một phân phối xác suất liên tục thì ước lượng LASSO là duy nhất (Hastie và cộng sự, 2015).

- Ước lượng LASSO có thể không duy nhất nhưng kết quả dự báo dựa trên LASSO là duy nhất.

3. Ví dụ áp dụng phương pháp ước lượng LASSO

Trong mục này, phương pháp ước lượng LASSO được sử dụng để tìm các yếu tố nội tại có ảnh hưởng mạnh nhất đến khả năng sinh lợi của các ngân hàng thương mại cổ phần Việt Nam. Tác giả sử dụng suất sinh lợi trên tổng tài sản (Return on Asset – ROA) và suất sinh lợi trên vốn chủ sở hữu (Return on Equity - ROE) để đo lường khả năng sinh lợi của các ngân hàng. Các yếu tố nội tại xét trong mô hình dựa trên các nghiên cứu đã thực hiện tại các nền kinh tế khác nhau. Bảng 3 giới thiệu các biến độc lập của mô hình, thể hiện các yếu tố nội tại của ngân hàng, cách tính toán và cơ sở đề xuất. Dữ liệu các biến được tổng hợp từ các báo cáo tài chính và báo cáo thường niên của 10 ngân hàng thương mại cổ phần Việt Nam trong giai đoạn 2009–2019¹. Các ngân hàng được chọn trong mẫu nghiên cứu bao gồm: Ngân hàng TMCP Ngoại thương Việt Nam (Vietcombank), ngân hàng TMCP Công thương Việt Nam (Vietinbank), ngân hàng TMCP Đầu tư và Phát triển Việt Nam (BIDV), ngân hàng TMCP Á Châu (ACB), ngân hàng TMCP Xuất Nhập khẩu Việt Nam (Eximbank), ngân hàng TMCP Quân đội (MB), ngân hàng TMCP Quốc dân (NCB), ngân hàng TMCP Sài Gòn thương tín (Sacombank), ngân hàng TMCP Sài Gòn - Hà Nội (SHB), và ngân hàng TMCP Việt Nam thịnh vượng (VPBank).

3.1. Mô hình nghiên cứu

Mô hình nghiên cứu:

$$ROA = \beta_0 + \beta_1 SIZE + \beta_2 EQU + \beta_3 CRE + \beta_4 LOAN + \beta_5 LDR + \beta_6 FLE + \beta_7 FOR + \beta_8 STATE + u \quad (14)$$

$$ROE = \beta_0 + \beta_1 SIZE + \beta_2 EQU + \beta_3 CRE + \beta_4 LOAN + \beta_5 LDR + \beta_6 FLE + \beta_7 FOR + \beta_8 STATE + u \quad (15)$$

Trong đó,

β_j ($j = 0, \dots, 8$): Các tham số cần được ước lượng;

β_j ($j > 0$): Thể hiện tác động của biến độc lập đến biến phụ thuộc của mô hình nghiên cứu; và

u : Sai số ngẫu nhiên của mô hình.

Bảng 3.

Các biến độc lập trong mô hình

STT	Yếu tố nội tại	Biến độc lập	Cách tính	Cơ sở đề xuất
1	Quy mô ngân hàng	SIZE	Logarit tự nhiên tổng tài sản ngân hàng	
2		EQU	Tỷ lệ tăng trưởng vốn chủ sở hữu	Isik và cộng sự (2018),
3	Cấu trúc sở hữu	STATE	Tỷ lệ sở hữu cổ phần của nhà nước	Lin và Zhang (2009),
4		FOR	Tỷ lệ sở hữu cổ phần của nước ngoài	

¹ Nhóm tác giả thu thập dữ liệu báo cáo thường niên và báo cáo tài chính trên các website chính của 10 ngân hàng được chọn trong mẫu nghiên cứu.

STT	Yếu tố nội tại	Biến độc lập	Cách tính	Cơ sở đề xuất
5	Rủi ro tín dụng	CRE	Tỷ lệ dự phòng rủi ro tín dụng/dư nợ cho vay khách hàng và cho thuê tài chính	Rahman và Reja (2015), Swai và Mbogela (2014).
6		LDR	Tỷ lệ cho vay khách hàng/tiền gửi khách hàng	
7		LOAN	Tỷ lệ cho vay/tổng tài sản	
8	Đòn bẩy tài chính	FLE	Tỷ lệ nợ/tổng tài sản	

Bảng 4.

Thông kê mô tả các biến

	ROA	ROE	SIZE	EQU	CRE	LOAN	LDR	FLE	FOR	STATE
Số quan sát	110	110	110	110	110	110	110	110	110	110
Giá trị nhỏ nhất	0,01	0,07	16,74	-7,16	-1,01	34,86	58,53	79,60	0,00	0,00
Giá trị lớn nhất	2,54	27,73	21,12	104,27	5,41	74,37	138,01	95,94	30,00	100,00
Khoảng biến thiên	2,53	27,66	4,38	111,43	6,42	39,51	79,48	16,34	30,00	100,00
Trung vị	0,88	13,03	19,23	11,10	0,87	58,92	85,69	92,92	12,28	9,09
Trung bình	0,96	13,05	19,21	17,86	1,08	57,94	88,57	92,36	14,24	29,89
Độ lệch chuẩn	0,59	7,26	1,00	20,22	0,93	10,51	17,11	2,47	11,89	36,78
Đơn vị tính	%	%		%	%	%	%	%	%	%

3.2. Quy trình tính toán

Bài báo sử dụng phần mềm R thực hiện tính toán với quy trình như sau:

- **Bước 1:** Chuẩn hóa dữ liệu các biến độc lập, đối với biến phụ thuộc chỉ thực hiện biến đổi đưa về trung bình bằng 0. Xét tập giá trị của λ gồm 1.000 giá trị cách đều nhau từ 10^{-4} đến 101.

$$A = \{10^{-4}; 1,0116 \times 10^{-4}; 1,0233 \times 10^{-4}; \dots; 9,5490; 9,7719; 10\}$$

- **Bước 2:** Chia ngẫu nhiên tập dữ liệu thành tập huấn luyện và tập kiểm tra theo tỷ lệ 80%–20%.

• **Bước 3:** Vì dữ liệu mẫu là dữ liệu bảng, thực hiện ước lượng các Mô hình (14) và Mô hình (15) bằng kỹ thuật ước lượng hai chiều trong cùng nhóm (Within Estimator for the Two-Way) để kiểm soát các ảnh hưởng của yếu tố thời gian và yếu tố cá biệt của mỗi ngân hàng. Trên cơ sở đó, dùng thuật toán giảm chiều, thủ tục xác thực chéo để tìm λ tốt nhất và thực hiện phương pháp ước lượng LASSO với λ tốt nhất trên tập huấn luyện.

Thủ tục xác thực chéo là một trong những phương pháp được sử dụng để tìm ước lượng tốt nhất cho λ . Trong thủ tục xác thực chéo, tập dữ liệu được chia ngẫu nhiên làm K tập con, ký hiệu $C_1, \dots, C_K$. Với mỗi giá trị λ , thủ tục xác thực chéo được thực hiện như sau: Xem C_i là tập kiểm tra, (K–1) tập còn lại là tập huấn luyện. Áp dụng phương pháp ước lượng LASSO trên tập huấn luyện này thu được mô hình, sau đó sử dụng tập kiểm tra C_i để dự báo giá trị của biến phụ thuộc và ghi nhận lại giá trị MSE_i (Mean Squared Error). Lặp lại thực hiện với $i = 1, 2, \dots, K$. Đặt CVM_λ là trung bình của các MSE_i :

$$CVM_\lambda = \frac{\sum_{i=1}^K MSE_i}{K}$$

Như vậy, mỗi λ tương ứng với một CVM_λ . Khi cho λ biến thiên trên một tập cho trước, giá trị λ tốt nhất là giá trị tương ứng với CVM_λ nhỏ nhất (Hastie và cộng sự, 2015).

- **Bước 4:** Tính MSE của các mô hình tương ứng với phương pháp ước lượng LASSO và OLS trên tập kiểm tra.

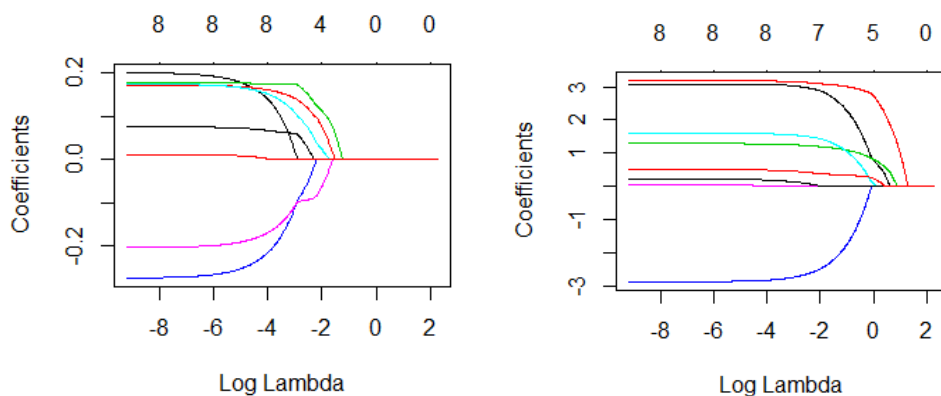
- **Bước 5:** Thực hiện lặp lại 100 lần từ Bước 2 đến Bước 4 để so sánh MSE của các Mô hình 14 và Mô hình 15 tương ứng với phương pháp ước lượng LASSO và OLS trên tập kiểm tra.

3.3. Kết quả tính toán

- **Bước 1:** Chuẩn hóa dữ liệu các biến độc lập, đối với biến phụ thuộc chỉ thực hiện biến đổi đưa về trung bình bằng 0.

- **Bước 2:** Tập huấn luyện và tập kiểm tra được phân chia ngẫu nhiên theo tỷ lệ 80%–20%. Tập kiểm tra gồm 22 quan sát với các quan sát thứ: 1, 2, 5, 7, 9, 17, 22, 25, 32, 35, 42, 51, 53, 56, 60, 75, 76, 80, 83, 91, 101, 102. Tập huấn luyện gồm 88 quan sát còn lại.

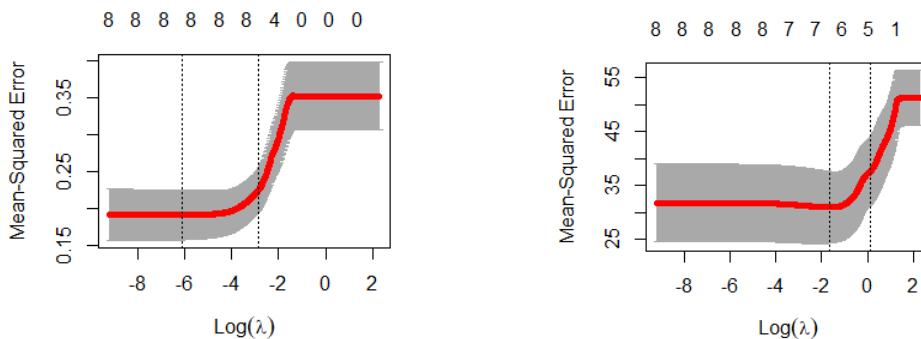
- **Bước 3:** Hình 4 minh họa sự biến thiên của các hệ số ước lượng trong các Mô hình (14) (trái) và Mô hình (15) (phải) khi λ biến thiên trên tập A .



Hình 4. Sự biến thiên của các ước lượng hệ số theo λ

Dùng thủ tục xác thực chéo với $K = 11$ để tìm λ tốt nhất. Hình 5 minh họa sự biến thiên của CVM theo λ đối với Mô hình (14) (trái) và Mô hình (15) (phải). Mức phạt λ tốt nhất lần lượt đối với các Mô hình (14) và Mô hình (15) là $\lambda\text{-best}_{ROA} = 0,007707$; $\lambda\text{-best}_{ROE} = 0,220467$.

Các kết quả ước lượng theo hai phương pháp - LASSO (theo λ tốt nhất) và OLS của các Mô hình (14), Mô hình (15) đồng thời thể hiện ở Bảng 5. Đối với Mô hình (14), biến chỉ thành phần sở hữu nhà nước STATE có ước lượng hệ số hồi quy bằng 0, do đó bị loại ra khỏi mô hình. Đối với Mô hình (15), các biến chỉ số đòn bẩy tài chính (FLE), thành phần sở hữu nhà nước (STATE) và thành phần sở hữu nước ngoài (FOR) bị loại ra khỏi mô hình. Vì các giá trị λ tốt nhất gần giá trị 0 nên kết quả ước lượng LASSO khá gần với kết quả ước lượng OLS. Bên cạnh đó, có thể nhận thấy những hệ số của biến độc lập có ước lượng LASSO bằng 0 thì ước lượng OLS không có ý nghĩa thống kê.



Hình 5. Sự biến thiên của CVM theo λ

Bảng 5.

Kết quả ước lượng bằng phương pháp LASSO với λ tốt nhất

Biến độc lập	Mô hình (14)			Mô hình (15)		
	LASSO ($\lambda=0,007707$)	OLS		LASSO ($\lambda=0,220467$)	OLS	
	Hệ số ước lượng	Hệ số ước lượng	Thống kê t	Hệ số ước lượng	Hệ số ước lượng	Thống kê t
SIZE	0,2208	0,2687*	2,580	3,6685	4,5555***	3,461
EQU	0,2068	0,2119***	4,008	3,6699	3,8580***	5,774
CRE	0,1753	0,1706**	3,333	1,0174	1,0006	1,547
LOAN	-0,2507	-0,2848***	-4,171	-2,4018	-3,2745***	-3,794
LDR	0,1494	0,1654**	2,834	1,2249	1,6554*	2,245
FLE	-0,1905	-0,2105**	-3,220	0,0000	-0,2215	-0,268
FOR	0,0555	0,0495	0,658	0,0000	-0,0615	-0,07
STATE	0,0000	-0,0179	-0,250	0,0000	-0,2282	-0,252

Ghi chú: *, **, *** lần lượt tương ứng với các mức ý nghĩa thống kê 10%, 5% và 1%.

- **Bước 4:** Để so sánh khả năng dự báo của phương pháp LASSO và OLS, ta so sánh MSE trên tập kiểm tra của các ước lượng Mô hình (14) và Mô hình (15) bằng hai phương pháp nói trên. Kết quả ở Bảng 6 cho thấy MSE của các mô hình được ước lượng bằng LASSO nhỏ hơn bằng OLS.

Bảng 6.So sánh MSE từ LASSO với λ tốt nhất và OLS trên tập kiểm tra

Phương pháp Mô hình	LASSO		OLS
	λ tốt nhất	MSE	MSE
Mô hình (14)	0,007707	0,1514108	0,1593538
Mô hình (15)	0,220467	29,97276	33,36601

Như vậy, từ Bước 2 đến Bước 4, ta đã xác định được λ tốt nhất, từ đó thực hiện phương pháp ước lượng LASSO cho các Mô hình (14) và Mô hình (15). Kết quả cho thấy, đối với tỷ suất ROA, tất cả các yếu tố nội tại trong mô hình, trừ thành phần sở hữu nhà nước, đều có ảnh hưởng. Đối với tỷ suất ROE, các yếu tố đòn bẩy tài chính và cấu trúc sở hữu không có ảnh hưởng. Về mức độ ảnh hưởng đến các tỷ suất sinh lợi, ba yếu tố sau đây là mạnh nhất: (1) Tỷ lệ cho vay trên tổng tài sản (LOAN), (2) tổng tài sản (SIZE); và (3) mức tăng trưởng vốn chủ sở hữu (EQU) (thứ tự mức độ ảnh hưởng tùy vào mô hình). Ảnh hưởng mạnh của yếu tố quy mô ngân hàng lên các tỷ suất sinh lợi ngân hàng Việt Nam trùng với kết quả của các nghiên cứu trước của Isik và cộng sự (2018), Lin và Zhang (2009), Rahman và Reja (2015), Swai và Mbogela (2014). Bên cạnh đó, ở cả hai trường hợp Mô hình (14) và Mô hình (15), trên tập kiểm tra, MSE của mô hình được ước lượng từ phương pháp LASSO nhỏ hơn từ OLS. Tuy vậy, sự khác biệt MSE của hai phương pháp ước lượng không quá lớn.

- **Bước 5:** Thực hiện lặp lại 100 lần từ Bước 2 đến Bước 4, thu được MSE của các mô hình từ các phương pháp LASSO và OLS. Hình 6 và Hình 7 mô tả trực quan phân bố mẫu của MSE từ phương pháp ước lượng LASSO và OLS cho các Mô hình (14) và Mô hình (15).



Hình 6. Biểu đồ MSE của Mô hình (14)

Hình 6 cho thấy đối với Mô hình (14), phương pháp LASSO cho MSE nhỏ hơn phương pháp OLS, tuy không đáng kể. Đối với Mô hình (15), MSE từ hai phương pháp là như nhau. Phân bố của MSE từ hai phương pháp tương đối giống nhau ở cả hai mô hình.



Hình 7. Biểu đồ MSE của Mô hình (15)

4. Kết luận

Bài viết đã cung cấp một cái nhìn tổng quan từ góc độ ý nghĩa, cơ sở toán học của thuật toán tìm ước lượng LASSO đến ví dụ ứng dụng phương pháp ước lượng LASSO trong phân tích kinh tế. Có thể kết luận rằng phương pháp ước lượng LASSO là một điều chỉnh của phương pháp OLS với mục đích chính là co gọn mô hình nhằm cung cấp một mô hình dễ giải thích hơn vì xác định rõ các yếu tố có ảnh hưởng mạnh đến biến phụ thuộc. Đây là ưu điểm vượt trội của LASSO so với các phương pháp ước lượng OLS hay phương pháp ước lượng điều chỉnh Ridge. Kết quả thực nghiệm trong bài viết còn cho thấy rằng LASSO có thể cải thiện khả năng dự báo của mô hình. Tuy nhiên, sự khác biệt giữa sai số của mô hình ước lượng bằng LASSO và OLS không đáng kể. Bài viết chưa có được những phân tích chặt chẽ về mặt lý thuyết để khẳng định vấn đề LASSO cải thiện năng lực dự báo của mô hình so với OLS ■

Tài liệu tham khảo

- Bertsekas, D. P., Nedic, A., & Ozdaglar, A. E. (2003). *Convex Analysis and Optimization*. Belmont, MA: Athena Scientific.
- Boyd, S., & Vandenberghe, L. (2004). *Convex Optimization*. Cambridge, UK: Cambridge University Press.

- Casella, G., & Berger, R. (2002). *Statistical Inference* (2nd ed.). Pacific Grove, CA: Thomson Learning.
- Cohen, J. (1994). The earth is around ($p < .05$). *American Psychologist*, 49(12), 997–1003.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (2nd ed.). New York: Springer.
- Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. New York: Chapman and Hall Book.
- Isik, O., Kosaroglu, S., & Demirci, A. (2018). The impact of size and growth decision on Turkish banks' profitability. *International Journal of Economics and Finance Issues*, 8(1), 21–29.
- Lin, X., & Zhang, Y. (2009). Bank ownership reform and bank performance in China. *Journal of Banking and Finance*, 33(1), 20–29.
- Nguyen, H. (2016). Why p-value are banned?. *Thai Statistician*, 14(2), i–iv.
- Nguyen, H. (2019). How to test without p-value?. *Thai Statistician* 17(2), i–x.
- Rahman, A., & Reja, B. (2015). Ownership structure and bank performance. *Journal of Economic, Business and Management*, 3(5), 483–488.
- Swai, J., & Mbogela, C. (2014). Do ownership structure affect banks' performance? An empirical inquiry onto Tanzanian bank industry. *Journal of Finance and Risk Perspectives*, 3(2), 47–66.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B (Methodological)*, 58(1), 267–288.
- Wooldridge, J. M. (2012). *Introductory Econometrics: A Modern Approach* (5th ed). USA: Cengage Learning.