
ƯỚC TÍNH TÁC ĐỘNG CỦA CÁC YẾU TỐ LÊN THỜI GIAN SỐNG SÓT CỦA KHOẢN VAY CỦA KHÁCH HÀNG CÁ NHÂN TẠI NGÂN HÀNG THƯƠNG MẠI BẰNG MÔ HÌNH LAPLACE

Đoàn Trọng Tuyên

Trường Đại học Kinh tế Quốc dân
Email: doantrongtuyen@neu.edu.vn

Nguyễn Thị Minh

Trường Đại học Kinh tế Quốc dân
Email: minhntkt@neu.edu.vn

Bùi Quốc Hoàn

Trường Đại học Kinh tế Quốc dân
Email: buiquochoan@neu.edu.vn

Ngày nhận: 01/5/2021

Ngày nhận bản sửa: 13/5/2021

Ngày duyệt đăng: 15/5/2021

Tóm tắt:

Việc ước lượng và dự báo thời điểm mà khoản vay bị vỡ nợ là bài toán quan trọng trong việc quản trị rủi ro của ngân hàng. Người ta thường sử dụng các mô hình Cox PH hay AFT để nghiên cứu bài toán này. Tuy nhiên, các mô hình này dựa trên giả định là tác động của các biến giải thích lên toàn bộ thời gian sống sót của khoản vay là đồng nhất và giả thiết này là không đúng trong nhiều trường hợp. Trong bài viết này, chúng tôi sử dụng mô hình Laplace được đề xuất bởi Bottai & Zhang (2010), là mô hình hồi quy phân vị trong phân tích sống sót. Kết quả hồi quy theo phân vị cho thấy tác động của nhiều biến giải thích như tuổi tác, học vấn, tài sản hay vị trí việc làm lên thời gian sống sót của khoản vay là khác nhau theo thời gian và trên các phân vị. Ngoài ra, chỉ số tài chính được đo bằng tỷ số giữa giá trị của khoản vay trên tổng thu nhập là đồng nhất theo thời gian và trên các phân vị.

Từ khóa: Hồi quy phân vị, hồi quy Laplace, mô hình Cox, mô hình AFT, phân tích sống sót.
Mã JEL: B16, C21, G32

Estimating the impact of factors on the survival time of individual customers' loans at commercial banks with Laplace model

Abstract:

Estimating and forecasting the loan's default time are important for a bank's risk management. Traditionally, the Cox PH and AFT models are often used to study this issue. However, these models are based on the assumption that the impacts of explanatory variables are the same over the lifetime of the loan, which may not be satisfied in many situations. In this paper, we employ the Laplace model, a quantile regression model designed especially for survival analysis, proposed by Bottai & Zhang (2010). The results show that impacts of customer's age, education, asset value, job position on loan's survival time are different over time and over quantiles. And that, financial index, measured as loan's value to income, has the same impact over time and over quantiles.

Keywords: Quantile regression, laplace regression, Cox model, AFT model, survival analysis.
JEL Codes: B16, C21, G32.

1. Giới thiệu

Tính toán và dự báo khả năng vỡ nợ của các khoản vay là một bài toán quan trọng trong quản trị rủi ro tại các ngân hàng thương mại. Việc tính toán hoặc dự báo thiếu chính xác có thể dẫn đến việc sử dụng nguồn vốn một cách kém hiệu quả, như dành quá nhiều tiền cho việc phòng ngừa rủi ro, hoặc nghiêm trọng hơn, có thể dẫn đến phá sản nếu nguồn tiền dành cho việc này là quá thiếu. Vì vậy việc tính toán một cách chính xác khả năng vỡ nợ của các khoản vay là hết sức cần thiết và đã được nhiều nhà nghiên cứu cũng như các nhà hoạch định chính sách quan tâm. Có hai phương pháp tiếp cận chính trong bài toán này: Phương pháp thứ nhất là tính toán khả năng vỡ nợ hay không vỡ nợ của một khoản vay trong một thời kỳ cụ thể nào đó, chẳng hạn một năm hay một quý. Các mô hình được sử dụng cho cách tiếp cận này có thể kể đến mô hình logit – probit, hay các phương pháp học máy như máy vectơ hỗ trợ (SVM) hay mạng nơron nhân tạo, trong khi phương pháp học máy sử dụng các thuật toán và khả năng tính toán của máy tính thì các mô hình logit – probit dựa vào dạng hàm hồi quy để phân loại khách hàng thành nhóm “tốt” và “xấu”. Phương pháp tiếp cận gần đây hơn, còn gọi là phân tích sống sót, trong đó mối quan tâm là thời điểm mà khoản vay bị vỡ nợ. Ưu điểm của phương pháp này là tính được được xác suất để khoản vay có thể tồn tại được đến thời điểm nào đó, từ đó có thể dự đoán được thời điểm mà khoản vay bị vỡ nợ, điều này là khác biệt hoàn toàn với việc đánh giá xác suất khoản vay bị vỡ nợ trong một khoảng thời gian như ở cách tiếp cận đầu tiên.

Tuy nhiên, việc sử dụng các mô hình trong phân tích sống sót cũng gặp phải một số hạn chế, hạn chế lớn nhất là các biến giải thích được giả định thỏa mãn điều kiện thứ tự nguy cơ vỡ nợ của khoản vay không đổi theo thời gian, điều này trong thực tế nhiều khi là không đúng vì với cùng một biến giải thích có thể tác động lên các thời điểm của khoản vay theo những cách khác nhau. Do đó, để có thể dự báo được chính xác khả năng vỡ nợ của khách hàng thì việc đánh giá tác động của các biến giải thích lên khả năng trả nợ khoản vay của khách hàng tại mỗi thời điểm là rất cần thiết. Vì vậy trong bài viết này chúng tôi sẽ sử dụng mô hình Laplace, là mô hình hồi quy phân vị trong phân tích sống sót, để nghiên cứu các tác động khác nhau của các biến giải thích lên thời gian sống sót của khoản vay. Bài viết được sắp xếp như sau, mục 2 tiếp theo là tổng quan mạch nghiên cứu thời gian sống sót của khoản vay, mục 3 là mô hình và kết quả, mục 4 là kết luận và khuyến nghị.

2. Tổng quan nghiên cứu

Trong phân tích sống sót, yếu tố được quan tâm là thời gian sống sót T của đối tượng, hàm sống sót được định nghĩa là xác suất để đối tượng không xảy ra sự kiện đến thời điểm t , nghĩa là $S(t) = Pr(T > t)$, giá trị xác suất này phụ thuộc vào biến giải thích $X = (X_1, X_2, \dots, X_p)$ nên có thể viết là $S(t, X)$. Trong các bài toán xác định khả năng vỡ nợ và thời gian sống sót của khoản vay của khách hàng bằng phân tích sống sót, thời gian sống sót của khoản vay được xác định là thời gian từ khi khoản vay được giải ngân tới khi khoản vay kết thúc hoặc khoản vay bị vỡ nợ. Khoản vay được xác định là vỡ nợ nếu quá hạn 90 ngày mà khách hàng không thể trả được lãi và gốc theo quy định, khi khoản vay bị vỡ nợ thì khoản vay được xác định là xảy ra sự kiện, trong các trường hợp còn lại khoản vay không bị xảy ra sự kiện.

Một trong những mô hình phân tích sống sót được sử dụng phổ biến là mô hình AFT (Accelerated Failure Time Model), được đề xuất bởi Miller (1976). Trong mô hình AFT, hàm $S(t, X)$ được biểu diễn dưới dạng $S(t, X) = S_0(t \cdot e^{-\beta'X})$, trong đó thời gian sống sót T tuân theo một quy luật phân phối nào đó, $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ và $X = (X_1, X_2, \dots, X_p)$ tương ứng là ma trận cột các tham số và ma trận cột các biến giải thích, t là ma trận chuyển vị của t và S_0 là hàm sống sót cơ sở - là hàm sống sót khi tất cả các biến giải thích bị triệt tiêu. Narain (1992) là một trong những người đầu tiên sử dụng mô hình AFT để nghiên cứu bài toán phân tích sống sót cho khách hàng cá nhân. Dữ liệu được tác giả phân tích gồm 1.242 quan sát là các khoản vay trong 24 tháng từ giữa năm 1986 tới giữa năm 1988 của một ngân hàng ở Vương quốc Anh. Tác giả sử dụng mô hình AFT dạng phân phối mũ với 7 biến giải thích là: tình trạng hôn nhân, phần trăm khoản vay trên tổng thu nhập, số điện thoại cố định, tình trạng cư trú, thời gian ở nơi hiện tại, thời gian quan hệ với ngân hàng và thời gian làm việc ở công việc hiện tại. Kết quả là mô hình AFT có khả năng phân loại tốt khách hàng “tốt” và “xấu” trong các khoảng thời gian khác nhau (18, 21, 24 tháng), do đó quyết định cấp tín dụng là tốt hơn nếu điểm tín dụng được hỗ trợ bởi thời gian sống sót của khoản vay. Hạn chế của nghiên cứu này là mẫu nhỏ, điều này có thể dẫn đến giả định về phân phối của thời gian sống sót của khoản vay dễ được

chấp nhận. Các mô hình AFT phức tạp hơn cũng được Banasik & cộng sự (1999) sử dụng để đánh giá xác suất vỡ nợ và xác suất trả nợ sớm của các khoản vay cá nhân trong các khoảng thời gian khác nhau (từ 0 – 12 tháng và từ 12 – 24 tháng). Tuy nhiên kết quả của các mô hình dự báo cho năm thứ 2 cho thấy hiệu suất dự báo còn thấp, nguyên nhân là do giả thiết của mô hình là rủi ro vỡ nợ hoặc trả nợ sớm không thay đổi trong các khoảng thời gian khác nhau, vì vậy với khoảng thời gian dài hơn thì dự báo có thể không còn chính xác.

Một hạn chế của các mô hình AFT là phải có giả thiết thời gian sống sót của khoản vay tuân theo một phân phối nào đó. Mô hình được đề xuất bởi Cox (1972, 1975) là một bổ sung cho những hạn chế của mô hình AFT vì nó không cần giả thiết nào về phân phối của thời gian sống sót. Trong đó, Cox đã mô hình hóa mối quan hệ giữa biến giải thích và nguy cơ xảy ra sự kiện dưới dạng $h(t, X) = h_0(t) \cdot e^{\beta' \cdot X}$, với $h_0(t) = h(t, 0)$ là hàm nguy cơ cơ sở, trong đó hàm nguy cơ cơ sở chỉ phụ thuộc vào thời gian t mà không phụ thuộc vào các biến giải thích X và các biến giải thích X cũng không phụ thuộc thời gian t . Vì các biến giải thích X làm tăng hay giảm nguy cơ theo bội $e^{\beta' \cdot X}$ của hàm nguy cơ cơ sở nên đây còn được gọi là mô hình Cox PH (Cox Proportional Hazard Model).

Stepanova & Thomas (2002) đã sử dụng mô hình Cox PH để giải quyết bài toán tính xác suất vỡ nợ đối với các khách hàng cá nhân. Trong nghiên cứu này, dữ liệu thực nghiệm gồm 50.000 quan sát từ một tổ chức tài chính lớn ở Vương quốc Anh cùng với thông tin trả nợ đến 36 tháng, tác giả sử dụng tới 16 biến giải thích liên quan đến các đặc điểm nhân khẩu học và tình hình tài chính của cá nhân vay vốn như tuổi của khách hàng, lượng tiền vay, ngày mở tài khoản, ngày đóng tài khoản, số năm ở địa chỉ hiện tại, số năm làm việc ở lĩnh vực chuyên môn hiện tại, giới tính của khách hàng, số trẻ em phụ thuộc, tần suất trả nợ, có số điện thoại cố định, phí bảo hiểm, kiểu khoản vay (đơn lẻ hoặc liên kết), tình trạng hôn nhân, thời hạn của khoản vay, sở hữu bất động sản, mục đích của khoản vay. Tác giả đề xuất quy trình xử lý dữ liệu cho phân tích sống sót và xây dựng các mô hình dự báo vỡ nợ và dự báo trả nợ sớm. Kết quả cho thấy hiệu suất dự báo của mô hình là cạnh tranh được với mô hình logistic, tuy nhiên hiệu suất của mô hình dự báo trả nợ sớm và dự báo vỡ nợ trong năm thứ 2 vẫn còn thấp – điều này giống như đối với nghiên cứu trước của Banasik & cộng sự (1999).

Bellotti & Crook (2009) sử dụng mô hình Cox PH để ước tính xác suất vỡ nợ nhưng với mô hình cho phép các biến giải thích phụ thuộc thời gian. Dữ liệu phân tích gồm 100.000 quan sát là các tài khoản thẻ tín dụng cá nhân được mở từ năm 1997 đến giữa năm 2005. Bên cạnh các biến giải thích truyền thống như thu nhập, tuổi, tình trạng nhà ở và công việc cùng với điểm tín dụng tại thời điểm nộp đơn, tác giả còn bổ sung thêm biến giải thích là các biến kinh tế vĩ mô phụ thuộc thời gian như tỷ lệ lãi suất, tỷ lệ thất nghiệp, chỉ số giá nhà... đây là các biến thay đổi theo thời gian và không dễ để đưa vào tập các biến giải thích trong các mô hình logistic và trong mô hình Cox PH truyền thống. Kết quả cho thấy việc bổ sung các biến giải thích kinh tế vĩ mô giúp cải thiện hiệu suất dự báo của các mô hình, các biến kinh tế vĩ mô đều có tác động có ý nghĩa thống kê đến thời gian sống sót của khoản vay như tỷ lệ thất nghiệp, lãi suất,... trong đó, tỷ lệ lãi suất là biến kinh tế vĩ mô quan trọng nhất trong việc ước tính xác suất vỡ nợ. Tuy nhiên, bài viết cũng chỉ ra rằng mặc dù có sự hoàn thiện về mô hình với việc bổ sung thêm biến giải thích và sự cải thiện về hiệu suất dự báo là có ý nghĩa thống kê nhưng cũng cho thấy hiệu quả còn khiêm tốn. Điều này cho thấy, kể cả khi đã bổ sung thêm các biến giải thích là biến kinh tế vĩ mô và sử dụng mô hình Cox PH mở rộng thì khả năng dự báo chính xác thời điểm vỡ nợ vẫn là một bài toán cần lời giải tối ưu hơn nữa.

Như vậy, ta thấy rằng các mô hình trong phân tích sống sót có thể được sử dụng trong các bài toán dự đoán thời điểm vỡ nợ của khách hàng, nhưng hiệu quả chưa cao khi dự đoán ở các khoảng thời gian dài. Điều này có thể là do các mô hình Cox PH và AFT được xây dựng dựa trên giả định là tác động của các biến giải thích lên thời gian sống sót là đồng nhất tại mọi thời điểm của của khoản vay, việc áp đặt giả định này lên thời gian sống sót của khoản vay có thể không đúng trong nhiều trường hợp. Để tách biệt các tác động khác nhau của các biến giải thích lên thời gian của khoản vay thì người ta sử dụng hồi quy phân vị trong phân tích sống sót. Một số thuật toán hồi quy phân vị trong phân tích sống sót đã được phát triển, đầu tiên phải kể đến Powell (1986), trong đó tác giả xét trường hợp thời điểm cuối cùng của các quan sát không xảy ra sự kiện là cố định. Tuy nhiên, trong thực tế thời gian kết thúc các quan sát không xảy ra sự kiện có thể là ngẫu nhiên, Portnoy (2003) đã đề xuất thuật toán để ước lượng các hệ số của mô hình hồi quy phân vị trong trường hợp này. Tương tự, Peng & Huang (2008) đề xuất một thuật toán khác so với Portnoy (2003),

theo Koenker (2008) thì thuật toán của Peng & Huang (2008) đơn giản hơn trong khi hiệu suất của các thuật toán này là tương đồng với nhau. Với các mô hình hồi quy phân vị trong phân tích sống sót hiện có, hồi quy Laplace cho tốc độ tính toán là nhanh nhất và phù hợp nhất với trường hợp các thời điểm cuối cùng của các quan sát không xảy ra sự kiện là ngẫu nhiên.

Mô hình hồi quy Laplace trong phân tích sống sót được đề xuất bởi Bottai & Zhang (2010). Hồi quy Laplace cho phép ước tính phân vị có điều kiện của thời gian sống sót theo các biến giải thích dưới dạng $Q_{\tau}(T|X) = \beta_0(\tau) + \beta_1(\tau).X_1 + \beta_2(\tau).X_2 + \dots + \beta_p(\tau).X_p$, với $\tau \in (0; 1)$ là giá trị xác suất cố định cho trước thể hiện phân vị τ , $Q_{\tau}(T|X) = \inf \{t: Pr(T \leq t|X) \geq \tau\}$ là phân vị τ của thời gian sống sót T với điều kiện X . Giả thiết của mô hình này là thời gian sống sót T tuân theo phân phối Laplace không đối xứng, mặc dù kèm theo điều kiện này nhưng theo Bellavia & cộng sự (2016) thì hiệu suất dự báo của mô hình không bị ảnh hưởng với các dữ liệu có phân phối khác nhau. Mô hình trên có hình thức tương tự như mô hình hồi quy phân vị thông thường được đề xuất bởi Koenker & Baset (1978), nhưng phương pháp ước lượng là hoàn toàn khác biệt do sự khác biệt về bản chất của dữ liệu. Với mô hình Laplace, dữ liệu ghi nhận trạng thái của mỗi quan sát là có xảy ra sự kiện hay không, nếu bỏ qua trạng thái về sự kiện của các quan sát thì các ước lượng trong hồi quy phân vị sẽ không còn chính xác.

Trên thế giới đã có nghiên cứu sử dụng hồi quy phân vị trong lĩnh vực quản trị rủi ro trong ngân hàng, chẳng hạn Miller (2014) khi nghiên cứu các nhân tố ảnh hưởng đến thời gian của khoản vay tín dụng của khách hàng cá nhân của Mỹ, tác giả sử dụng mô hình hồi quy phân vị trong phân tích sống sót, dữ liệu nghiên cứu gồm 17.000 quan sát là khoản vay từ trang web cho vay ngang hàng Prosper.com từ tháng 2/2007 đến tháng 11/2009, trong đó tác giả so sánh mô hình hồi quy phân vị trong phân tích sống sót với mô hình Cox PH truyền thống trong việc đánh giá tác động của các biến giải thích lên thời gian sống sót của khoản vay. Bài viết chỉ ra tác động không đồng nhất của các biến giải thích truyền thống lên thời gian tồn tại của các khoản vay vỡ nợ sớm và muộn. Các yếu tố có tác động không đồng nhất là điểm tín dụng, quá khứ vỡ nợ và số lượng báo cáo thẩm tra tín dụng – những yếu tố này là không thể ghi nhận được bởi mô hình tỷ lệ nguy cơ Cox PH truyền thống. Gần đây, một số nghiên cứu khác đáng chú ý trong lĩnh vực kinh tế và quản trị rủi ro gần đây có sử dụng hồi quy phân vị trong phân tích sống sót là của Di & Hanke (2012), Berlemann & Luik (2014), Loonat (2015), Cajias & cộng sự (2020). Đây là các ứng dụng đa dạng trong nhiều lĩnh vực, từ lĩnh vực ngân hàng tài chính, sự lựa chọn của khách hàng cá nhân, khách hàng doanh nghiệp đến thị trường bất động sản. Các nghiên cứu này đều có kết quả thể hiện ưu điểm vượt trội của hồi quy phân vị trong phân tích sống sót so với các mô hình phân tích sống sót truyền thống khác.

Ở Việt Nam, đã có một số nghiên cứu về thời điểm vỡ nợ của khách hàng cá nhân bằng các mô hình trong phân tích sống sót. Chẳng hạn bài viết của Đoàn Trọng Tuyển & Nguyễn Thị Minh (2020), trong đó các tác giả đã sử dụng mô hình Cox PH và mô hình Cox PH mở rộng để dự báo vỡ nợ của các khoản vay cá nhân trong các khoảng thời gian khác nhau, nghiên cứu được thực hiện trên tập dữ liệu gồm 55.500 quan sát là khách hàng cá nhân có nguồn thu nhập chính từ lương trong khoảng thời gian từ tháng 1/2014 đến tháng 6/2018. Các biến giải thích được sử dụng trong các mô hình là tuổi của khách hàng, thời gian làm việc, vị trí công việc, học vấn, sở hữu bất động sản – nhà ở, hình thức trả lương và tỷ lệ giữa số tiền vay trên tổng thu nhập. Kết quả chính của bài viết cho thấy mô hình Cox PH mở rộng có hiệu suất dự báo tốt hơn mô hình Cox PH thông thường, trong đó các tác động của các biến giải thích lên thời gian sống sót của khoản vay (là thời gian từ khi khoản vay được giải ngân tới khi được xác định là vỡ nợ hay kết thúc khoản vay) đều có ý nghĩa thống kê, bài viết chỉ ra thời điểm khoản vay có khả năng vỡ nợ cao nhất là tháng thứ 16. Tuy nhiên, cũng giống như các công bố khác sử dụng phân tích sống sót bằng mô hình Cox PH hay mô hình AFT, bài viết cũng cho thấy hiệu suất dự báo không cao ở các mô hình với thời gian của khoản vay dài hơn, đặc biệt là trong khoảng 0 – 36 tháng. Vì vậy trong bài viết này, chúng tôi sẽ sử dụng mô hình Laplace, mà theo hiểu biết của chúng tôi thì đây là một phương pháp hầu như chưa được sử dụng ở Việt Nam trong lĩnh vực quản trị rủi ro, để đánh giá các tác động của các biến giải thích cũng như ước tính thời gian sống sót của các khoản vay cá nhân trên các phân vị khác nhau – điều mà mô hình Cox PH hay các mô hình AFT không thể giải quyết được, với dự tính các biến giải thích quan trọng sẽ có tác động khác nhau lên thời gian sống sót của khoản vay, chúng tôi hi vọng sẽ dự đoán chính xác hơn về thời gian sống sót của khoản vay.

Bảng 1. Thống kê mô tả các biến liên tục cùng các phân vị không điều kiện tương ứng

Tổng số khoản vay: 53.981		Số khoản vay bị vỡ nợ: 5.911			
Tỷ lệ số khoản vay bị vỡ nợ trong thời gian nghiên cứu:		10,95%			
Tên biến số	Mô tả biến số	10%	25%	50%	75%
<i>Surv</i>	Thời gian sống sót của khoản vay	7	11	19	28
<i>Loan_per_Income</i>	Tỷ lệ tổng số tiền phải trả trên tổng thu nhập	0,3107	0,4186	0,5046	0,5889

Nguồn: Nghiên cứu của nhóm tác giả.

3. Mô hình và kết quả phân tích

3.1. Số liệu và các biến số

Dữ liệu thứ cấp từ một ngân hàng thương mại cổ phần, với 53.981 quan sát là các khách hàng cá nhân có nguồn thu nhập chính từ lương, dữ liệu được lấy một lần duy nhất khi khách hàng làm thủ tục xin cấp vốn và đã được chấp thuận cho vay. Dữ liệu bao gồm các thông tin về nhân khẩu học, tình hình tài chính và lịch sử trả nợ hàng tháng từ khi khoản vay được giải ngân, khoảng thời gian lấy dữ liệu bắt đầu từ 01/01/2014 đến 01/06/2018, khoảng thời gian nghiên cứu là 54 tháng. Việc phân biến thành các nhóm được thực hiện theo thuật toán được đề xuất bởi Stepanova & Thomas (2002).

Dưới đây là các biến số được sử dụng trong mô hình:

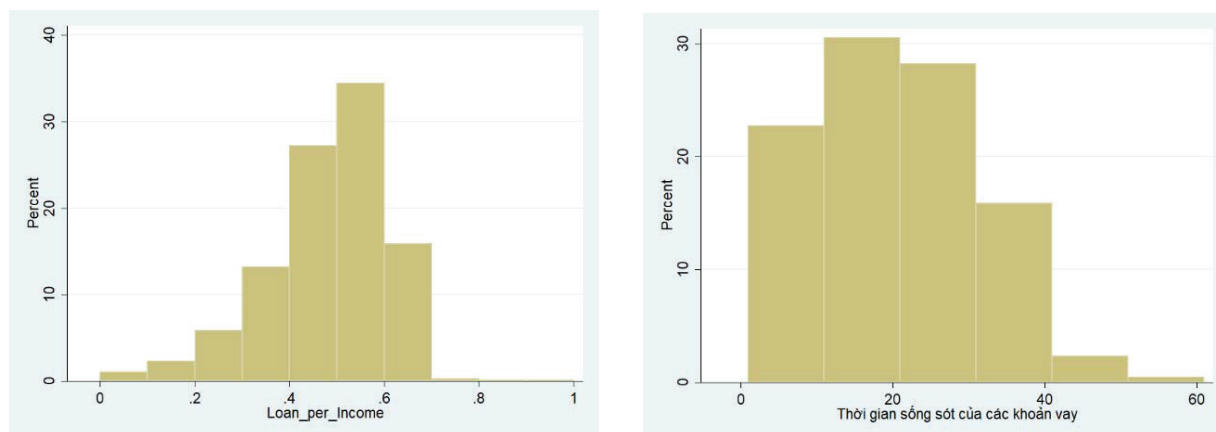
- *Age* là tuổi của khách hàng, đơn vị tính là năm. Tuổi của khách hàng thường được cho là có tương quan với năng lực quản lý tài chính cũng như sở thích rủi ro, nên được cho là có mối liên hệ với khả năng vỡ nợ. Dự kiến về giá trị của hệ số biến tuổi là: Ở mức tuổi càng cao thì giá trị càng lớn do khi tuổi càng cao thì người ta thường có nhiều kinh nghiệm trong công việc, có sự chín chắn trong các quyết định kinh doanh và không có nhiều sự mạo hiểm trong các quyết định, đặc biệt là liên quan tới tài chính.
- *Working time* là thời gian làm việc tại lĩnh vực chuyên môn hiện tại, đơn vị tính là năm. Thông thường, thời gian làm việc càng nhiều thì kinh nghiệm càng được tích lũy nhiều, điều này thường dẫn đến lương được trả cho công việc càng cao, do đó chúng tôi kỳ vọng thời gian làm việc càng lâu thì thời gian sống sót của khoản vay càng nhiều, nghĩa là chúng tôi kỳ vọng nhóm có thời gian làm việc lâu năm sẽ có hệ số lớn hơn so với nhóm có thời gian làm việc ít hơn.
- *Position* là vị trí công việc của khách hàng. Vì lương, thưởng của người lao động là khác nhau tùy thuộc vào vị trí công việc cũng như trách nhiệm của cá nhân đối với công việc, nên kỳ vọng về giá trị hệ số của nhóm lao động được đào tạo nghề hoặc công nhân là thấp nhất, nhóm có vị trí quản lý, chuyên viên có hệ số cao nhất.
- *Educ* là học vấn của khách hàng. Vì học vấn là thước đo trình độ của người lao động, nên chúng tôi kỳ vọng hệ số của các nhóm này tăng dần từ nhóm có trình độ học vấn thấp đến nhóm có trình độ học vấn cao.
- *Land* là tình trạng sở hữu bất động sản, nhà ở. Thông thường, việc sở hữu nhà thứ hai trở lên sẽ cho một nguồn thu nhập ổn định hàng tháng và cũng thể hiện khách hàng có tiềm lực tài chính vững mạnh, do đó hệ số của nhóm này được kỳ vọng có giá trị lớn nhất.
- *Wage* là hình thức trả lương. Do sự tiện lợi cũng như sàng lọc trước đó về khả năng trả nợ mà nhóm khách hàng được trả lương qua chính ngân hàng cho vay được kỳ vọng sẽ có hệ số lớn nhất.
- *Loan_per_Income* là tỷ lệ giữa tổng số tiền phải trả (gốc + lãi) tại tất cả các tổ chức tín dụng trên tổng thu nhập sau thuế, bao gồm cả khoản vay đang xem xét, dữ liệu của biến này nằm trong khoảng , đây là biến liên tục. Giá trị biến số này càng lớn thì khoản vay càng lớn (so với tổng thu nhập sau thuế), nên hệ số của biến này được kỳ vọng là số âm.

Bảng 2. Thống kê mô tả các biến giải thích dạng nhóm

Tên biến số	Mô tả biến số	Số lượng
Age	1 Nhóm tuổi từ [18, 23)	1.446
	2 Nhóm tuổi từ [23, 28)	11.467
	3 Nhóm tuổi từ [28, 33)	14.902
	4 Nhóm tuổi từ [33, 48)	21.830
	5 Nhóm tuổi từ [48, 69)	4.336
Working_time	1 Khoảng [0, 3)	7.419
	2 Khoảng [3, 6)	16.106
	3 Khoảng [6, 9)	10.003
	4 Khoảng [9, 15)	11.201
	5 Khoảng [15, 43)	9.252
Position	1 Lao động được đào tạo nghề hoặc công nhân	11.972
	2 Chuyên viên, nhân viên thất nghiệp hoặc đã nghỉ hưu	32.052
	3 Quản lý, chuyên viên	9.957
Educ	1 Dưới trung cấp	8.607
	2 Trung cấp	12.843
	3 Cao đẳng	8.212
	4 Đại học hoặc trên đại học	24.319
Land	1 Ở nhà thuê	7.892
	2 Ở nhờ	19.927
	3 Có 1 nhà	24.268
	4 Có 2 nhà trở lên	1.894
Wage	1 Được trả lương qua các ngân hàng khác	36.888
	2 Được trả lương bằng tiền mặt	9.465
	3 Được trả lương qua chính ngân hàng cho vay	7.628

Nguồn: Nghiên cứu của nhóm tác giả.

Hình 1. Biểu đồ phân bố Histogram của các biến tỷ lệ giữa tổng số tiền phải trả trên tổng thu nhập sau thuế và thời gian sống sót của các khoản vay (tính theo tháng)



Nguồn: Nghiên cứu của nhóm tác giả.

Bảng 1, Bảng 2 và Hình 1 mô tả cơ bản về các biến trong dữ liệu phân tích và mối liên hệ giữa các biến giải thích với thời gian sống sót của khoản vay.

Mô hình Laplace cho phép ước tính thời gian sống sót của mỗi khoản vay trên các phân vị trong khoảng dưới dạng

$$Q_{\tau}(T|X) = \beta_0(\tau) + \beta_1(\tau).Age + \beta_2(\tau).Working_time + \beta_3(\tau).Position + \beta_4(\tau).Educ + \beta_5(\tau).Land + \beta_6(\tau).Wage + \beta_7(\tau).Loan_per_Income, \text{ với } \tau \in (0;1) \quad (1)$$

Trong đó $Q_{\tau}(T|X)$ là phân vị τ của thời gian sống sót T với điều kiện X , $\beta_i(\tau)$ là hệ số của biến giải thích thứ i trong phân vị τ , thể hiện tác động của biến giải thích thứ i lên phân vị τ của thời gian sống sót của khoản vay.

Bảng 3. Kết quả ước lượng của mô hình hồi quy Laplace

Các hệ số hồi quy	$\tau = 0,1$ [1]	$\tau = 0,25$ [2]	$\tau = 0,5$ [3]	$\tau = 0,75$ [4]
Hệ số chặn $\beta_0(\tau)$	10,10** (0,97)	16,62** (105)	25,99** (0,94)	31,96** (0,89)
<i>Loan_per_Income</i>	-9,56** (1,54)	-11,99** (1,47)	-10,06** (1,30)	-9,39** (1,24)
<i>Age</i>	2	1,87** (0,57)	4,20** (0,73)	4,43** (0,70)
	3	0,94 (0,60)	2,64** (0,77)	3,08** (0,73)
	4	0,25 (0,63)	1,89** (0,81)	2,59** (0,76)
	5	2,08* (1,02)	4,12** (1,14)	4,90** (1,04)
<i>Working_time</i>	2	1,38** (0,41)	2,49** (0,50)	2,36** (0,46)
	3	2,48** (0,50)	5,07** (0,64)	4,60** (0,56)
	4	3,13** (0,58)	4,90** (0,67)	4,68** (0,60)
	5	3,99** (0,74)	5,73** (0,80)	5,39** (0,70)
<i>Position</i>	2	3,64** (0,37)	6,42** (0,51)	6,29** (0,42)
	3	6,20** (0,74)	10,99** (0,76)	11,24** (0,59)
<i>Educ</i>	2	2,43** (0,37)	3,53** (0,51)	3,35** (0,46)
	3	1,87** (0,5)	4,02** (0,66)	3,70** (0,53)
	4	9,52** (0,54)	11,09** (0,59)	10,36** (0,52)
<i>Land</i>	2	1,91** (0,37)	2,50** (0,48)	2,84** (0,44)
	3	4,60** (0,47)	6,22** (0,56)	6,10** (0,49)
	4	8,95** (1,73)	9,64** (1,62)	10,61** (1,68)
<i>Wage</i>	2	4,31** (0,56)	4,45** (0,5)	4,16** (0,46)
	3	13,27** (0,72)	15,68** (0,68)	15,36** (0,66)

Ghi chú: Số trong dấu ngoặc đơn là sai số chuẩn; các ký hiệu * và ** tương ứng biểu thị các mức ý nghĩa 10% và 5%.

Nguồn: Nghiên cứu của nhóm tác giả.

3.2. Kết quả ước lượng

Bảng 3 trình bày kết quả ước lượng mô hình hồi quy Laplace từ phương trình (1) cho các phân vị $\tau = 0,1$, $\tau = 0,25$, $\tau = 0,5$ và $\tau = 0,75$. Các hệ số của các biến đều có ý nghĩa thống kê. Từ định nghĩa của $Q_{\tau}(T|X)$, ta có thể chuyển từ việc tính thời gian sống sót sang nguy cơ vỡ nợ với mức xác suất cụ thể theo phân vị τ , ví dụ nếu $Q_{\tau=0,1}(T|X) = 34$ thì điều này có nghĩa là thời gian sống sót của khoản vay tương ứng với các biến giải thích X là 34 tháng ở phân vị 10%, hay khách hàng này có 10% khả năng bị vỡ nợ ở tháng thứ 34 của khoản vay, hoặc có thể hiểu là xác suất để khoản vay vẫn chưa bị vỡ nợ ở tháng thứ 34 là 90%.

Cột [1], [2], [3] và [4] lần lượt là các hệ số ước lượng tương ứng với các phân vị $\tau = 0,1, \tau = 0,25, \tau = 0,5, \tau = 0,75$. Chẳng hạn giá trị -9,56 trên dòng 3, cột [1] là số âm, điều này có nghĩa là khi giá trị của *Loan_per_Income* càng lớn thì thời gian sống sót của khoản vay càng ngắn lại. Cụ thể hơn, từ phương trình hồi quy

(1) ở trên, nếu tỷ lệ giữa giá trị khoản vay trên tổng thu nhập là α thì thời gian sống sót của khoản vay được trừ bớt đi $9,56\alpha$ vì

$$Q_{\tau}(T|X) = \beta_0(\tau) + \beta_1(\tau).Age + \beta_2(\tau).Working_time + \beta_3(\tau).Position + \beta_4(\tau).Educ + \beta_5(\tau).Land + \beta_6(\tau).Wage - 9,56.\alpha \quad \text{với } \tau = 0,1$$

Từ cột [1] (là hệ số hồi quy cho phân vị) cho thấy: Hệ số của biến Age với nhóm tuổi trong khoảng [48,69) là cao nhất và bằng 2,08. Nghĩa là, thời gian sống sót của các khoản vay của nhóm khách hàng có tuổi trong khoảng [48,69) dài hơn so với nhóm có tuổi trong khoảng [18,23) (nhóm cơ sở) là 2,08 tháng, đồng thời nhóm tuổi trong khoảng [18,23) là nhóm tuổi có thời gian sống sót của khoản vay là ngắn nhất (với hệ số hồi quy bằng 0) so với các nhóm tuổi khác. Hệ số của biến Educ cho biết thời gian sống sót của khoản vay của khách hàng có trình độ từ đại học trở lên cao hơn nhóm có trình độ dưới trung cấp là 9,52 tháng, nhưng điều thú vị ở đây là ở chỗ nhóm có trình độ cao đẳng lại có thời gian sống sót của khoản vay lại ngắn hơn nhóm có trình độ trung cấp.

Các hệ số của các biến giải thích liên quan đến tình hình tài chính và công việc của khách hàng như *Land, Position, Working_time* đều phù hợp với kỳ vọng ban đầu – nghĩa là chúng đều có hệ số hồi quy tăng dần theo khả năng tài chính (thể hiện bằng tình trạng sở hữu bất động sản), vị trí công việc, kinh nghiệm làm việc (thể hiện bằng thời gian làm việc) của các cá nhân vay vốn, những khách hàng được trả lương qua chính ngân hàng đang vay vốn sẽ có thời gian sống sót của khoản vay dài hơn nhóm khách hàng được trả lương qua ngân hàng khác tới 13,27 tháng. Kết quả này cũng đồng nhất với kết quả trong Silva & cộng sự (2020), trong đó tác giả khẳng định nguy cơ vỡ nợ của khoản vay tăng theo giá trị của khoản vay, thời gian vay và tuổi của khách hàng, nhưng giảm nếu khách hàng có nhiều thế tín dụng hay nhận lương từ cùng một ngân hàng cho vay.

So sánh kết quả giữa các phân vị cho thấy: Có sự tương đồng về xu hướng biến thiên của các biến giải thích giữa các phân vị, cụ thể là các biến *Working_time, Position, Land* và *Wage* có các hệ số với xu thế chung là tăng. Tuy nhiên, xu thế của biến Age là không đơn điệu theo nhóm tuổi, nhóm ít tuổi nhất (nhóm cơ sở) có thời gian sống sót của khoản vay thấp nhất, sau đó nhóm tuổi [23,28) có hệ số lại cao hơn so với hai nhóm tiếp theo là nhóm tuổi [28,33) và nhóm tuổi [33,48). Điều này có nghĩa là nhóm tuổi [23,28) có thời gian sống sót của khoản vay dài hơn so với các nhóm [28,33) và [33,48), nhóm tuổi già nhất có hệ số cao nhất – nghĩa là có thời gian sống sót của khoản vay dài nhất, sự biến thiên ở đây giống hình ảnh hàm bậc 3 và đúng cho các phân vị. Với biến *Land* thì không có xu thế chung cho các phân vị, hệ số của biến *Land* trên phân vị của nhóm có trình độ trung cấp lại cao hơn nhóm có trình độ cao đẳng, nhưng ở các phân vị cao hơn thì điều này lại không còn đúng nữa, điều này có nghĩa là đối với các khoản vay vỡ nợ sớm thì tác động lên thời gian sống sót của khoản vay của nhóm có trình độ cao đẳng là thấp hơn so với nhóm có trình độ trung cấp, còn với các khoản vay có thời gian sống sót dài hơn thì điều này là ngược lại.

Kết quả kiểm định (Xem Bảng 4 phần Phụ lục) cũng cho thấy, đa số các biến đều có sự khác biệt giữa các hệ số của các phân vị $\tau = 0,1$ và $\tau=0,5$. Tuy nhiên hệ số của biến *Loan_per_Income* lại không khác biệt giữa các phân vị này.

4. Kết luận và khuyến nghị

Từ kết quả phân tích ở trên, có thể rút ra một số kết luận và khuyến nghị như sau:

Thứ nhất, về tác động của tuổi của khách hàng đến thời gian sống sót của khoản vay. Kết quả cho thấy các nhóm khách hàng rất trẻ ([18; 23)) hoặc trung trung tuổi ([33; 48)) có nguy cơ vỡ nợ cao hơn so với nhóm khách hàng còn lại. Do đó, khi đưa ra các quyết định tín dụng cũng như tính toán phòng ngừa rủi ro, ngân hàng nên tính đến yếu tố này. Chẳng hạn với các khách hàng có nhiều rủi ro này thì cần có thêm các biện pháp để hạn chế rủi ro.

Thứ hai, về học vấn của khách hàng, ở phân vị thấp của thời gian sống sót của khoản vay, tác động của học vấn cao đẳng lại ít hơn học vấn trung cấp, nhưng ở các phân vị cao hơn thì điều này không còn đúng nữa. Điều này ngụ ý rằng, với các khoản vay bị vỡ nợ sớm thì tác động của học vấn trung cấp lại tích cực hơn so với học vấn cao đẳng, có thể có nhiều lao động có trình độ cao đẳng đang làm công việc trái ngành, dẫn đến thu nhập thấp hơn người có trình độ trung cấp.

Thứ ba, nhóm khách hàng được trả lương qua cùng một ngân hàng đang vay có thời gian sống sót của khoản vay cao hơn nhiều so với những nhóm khách hàng được trả lương qua ngân hàng khác hoặc được trả lương bằng tiền mặt. Dựa vào kết quả này, các ngân hàng nên lưu ý các chương trình khuyến mại dành cho nhóm khách hàng thân thiết và tạo ra cộng đồng nhiều khách hàng thân thiết, do nhóm khách hàng này có độ an toàn cao khi vay vốn, điều này đem lại lợi nhuận lớn cho ngân hàng.

Thứ tư, một chỉ số quan trọng đối với sự kiện vỡ nợ là chỉ số tài chính, được đo bằng tỷ lệ giữa số tiền phải trả trên tổng thu nhập sau thuế, là có tác động như nhau đối với các nhóm có thời điểm vỡ nợ khác nhau. Điều này có thể ngụ ý rằng chỉ số tài chính này không thực sự là yếu tố gây nên sự khác nhau về khoảng thời gian sống sót của khoản vay.

Phụ lục: Kiểm định đối với các hệ số hồi quy

Bảng 4. Kiểm định về sự khác nhau của các hệ số hồi quy giữa các phân vị 10%, 50%

$H0: \beta_i(10) = \beta_i(50)$		p-value	$H0: \beta_i(10) = \beta_i(50)$		p-value
<i>Age</i>	2	0,0045	<i>Educ</i>	2	0,1151
	3	0,0233		3	0,0123
	4	0,0178		4	0,2666
	5	0,0534	<i>Land</i>	2	0,1044
<i>Working_time</i>	2	0,1113		3	0,0279
	3	0,0049		4	0,4910
	4	0,0626	<i>Wage</i>	2	0,8411
	5	0,1736		3	0,0335
<i>Position</i>	2	0,0000	<i>Loan_per_Income</i>		0,8068
	3	0,0000			

Nguồn: Nghiên cứu của nhóm tác giả.

Tài liệu tham khảo

- Bellavia, A., Bottai, M. & Orsini, N. (2016), ‘Evaluating Additive Interaction Using Survival Percentiles’, *Epidemiology*, 27(3), 360 – 364.
- Bellotti, T. & Crook, J. (2009), ‘Credit scoring with macroeconomic variables using survival analysis’, *The Journal of the Operational Research Society*, 60(12), 1699–1707.
- Bottai, M. & Zhang, J. (2010), ‘Laplace regression with censored data’, *Biometrical Journal*, 52(4), 487–503.
- Banasik, J., Crook, J. & Thomas, L. (1999), ‘Not if but when will borrowers default’, *The Journal of the Operational Research Society*, 50(12), 1185–1190.
- Berlemann, M. & Luik, M. (2014), ‘Institutional reform and depositors’ portfolio choice: Evidence from censored quantile regressions’, *CESifo Working Paper No. 4782 - Category 7: Monetary Policy and International Finance*.
- Cajias, M., Freudenreich, P. & Freudenreich, A. (2020), ‘Exploring the determinants of real estate liquidity from an alternative perspective: Censored quantile regression in real estate research’, *Journal of Business Economics*, 90, 1057–1086.
- Cox, D.R. (1972), ‘Regression models and life-tables (with discussion)’, *Journal of the Royal Statistical Society, Series B*, 74, 187–220.
- Cox, D.R. (1975), ‘Partial likelihood’, *Biometrika*, 62(2), 269–276.
- Đoàn Trọng Tuyên & Nguyễn Thị Minh (2020), ‘Mô hình Cox và Cox mở rộng trong ước tính xác suất vỡ nợ đối với các khoản vay cá nhân’, *Hội thảo khoa học công cụ toán – thống kê trong nghiên cứu kinh tế - tài chính*, 89–104.

-
- Di, H. & Hanke, S.A. (2012), 'Why do small businesses take on high levels of external loans? A censored quantile regression analysis', *Academy of Accounting and Financial Studies Journal*, 16(Special Issue), 17–25.
- Koenker, R. & Bassett, G.W. (1978), 'Regression quantiles', *Econometrica*, 46(1), 33–50.
- Koenker, R. (2008), 'Censored quantile regression redux', *Journal of Statistical Software*, 27(6), 1–25.
- Miller, R. (1976), 'Least squares regression with censored data', *Biometrika*, 63(3), 449–464.
- Miller, S. (2014), *Risk Factors for Consumer Loan Default: A Censored Quantile Regression Analysis*, Working paper, University of Illinois.
- Narain, B. (1992), 'Survival analysis and the credit granting decision', in Thomas, L.C., Crook, J.N. & Edelman, D.B., *Credit Scoring and Credit Control*, Clarendon Press: Oxford, 109–121.
- Peng, L. & Huang, Y. (2008), 'Survival analysis with quantile regression models', *Journal of the American Statistical Association*, 103, 637–649.
- Portnoy, S. (2003), 'Censored regression quantiles', *Journal of the American Statistical Association*, 98(464), 1001–1012.
- Powell, J. (1986), 'Censored regression quantiles', *Journal of Econometrics*, 32, 143–155.
- Silva, E.C., Lopes, I.C, Correia, A. & Faria, S. (2020), 'A logistic regression model for consumer default risk', *Journal of Applied Statistics*, 47, 2879–2894.
- Stepanova, M. & Thomas, L. (2002). 'Survival analysis methods for personal loan data' *Operations Research Quarterly*, 50(2), 277–289.
- Loonat, Z. (2015), 'Quantile regression methods for censored survival data', Master of Science Dissertation, University of Johannesburg, South Africa.