

ỨNG DỤNG CÁCH TIẾP CẬN TRUNG BÌNH HÓA MÔ HÌNH KIỂU BAYES (BMA) TRONG VIỆC XÂY DỰNG MÔ HÌNH CHẤM ĐIỂM RỦI RO TÍN DỤNG CHO KHÁCH HÀNG SME TẠI VIỆT NAM

Vương Minh Giang

Ngân hàng Vietcombank
Email: giangvm@gmail.com

Nguyễn Thanh Thiên

Ngân hàng Vietcombank

Ngày nhận: 7/11/2016

Ngày nhận bản sửa: 30/12/2016

Ngày duyệt đăng: 25/7/2017

Tóm tắt:

Bài viết giới thiệu tổng quan về phương pháp trung bình hóa mô hình kiểu Bayes (Bayesian Model Averaging - BMA), và ứng dụng cách tiếp cận này vào việc đưa ra một chiến lược mô hình hóa gồm các bước ứng dụng đảm bảo khả năng xây dựng và lựa chọn mô hình chấm điểm rủi ro tín dụng một cách toàn diện và thấu đáo. Kết quả áp dụng cho phân khúc doanh nghiệp nhỏ và vừa (SME) trong ngành Công nghiệp và Xây dựng, dựa trên nguồn dữ liệu của Trung tâm Thông tin tín dụng Quốc gia Việt Nam (CIC), cho thấy các chỉ tiêu thanh khoản Vốn lưu động/Tổng tài sản, Tài sản ngắn hạn/Nợ ròng, Tài sản ngắn hạn/(Tổng tài sản-Tài sản vô hình), và (Nợ phải trả - Tiền - Các khoản tương đương tiền)/Doanh thu thuần là nhóm nhân tố tài chính quyết định tới rủi ro không trả được nợ của các khách hàng này.

Từ khóa: Trung bình hóa mô hình kiểu Bayes (BMA); Rủi ro tín dụng, khách hàng SME.

Application of BMA approach to develop credit scoring models for Vietnam SMEs

Abstract:

This paper reviews the Bayesian Model Averaging, and based on this approach proposes a modeling strategy that ensures the development and selection of thorough and exhaustive credit scoring model. As a result of its real application, a scoring model for the segment of SMEs in Industry and Construction is built based on a data set sourced from the National Credit Information Center of Vietnam. The application of the model in the SME segment gives evidence that the liquidity ratios of Working Capital/Total Assets, Total Current Assets/Net Borrowing, Total Current Assets/(Total Assets- Intangible), and (Total Liabilities- Cash- Cash Equivalents)/Net Sales are the key financial drivers of this firm segment's default risk.

Keywords: Bayesian Model Averaging (BMA); Credit risk, SME customers

1. Giới thiệu

Vấn đề xây dựng và lựa chọn mô hình tốt nhất luôn là một thách thức trong mô hình hóa rủi ro tín dụng. Trong nghiên cứu và thực hành, những kỹ thuật như cây quyết định (decision trees), hồi quy từng bước (stepwise regression),... thường được sử dụng phổ biến, tuy nhiên không đảm bảo được

khả năng xem xét, lựa chọn mô hình một cách toàn diện và thấu đáo bởi thông thường chỉ đưa ra được một mô hình dựa trên một tập điều kiện xác định trước. Bài viết này giới thiệu tổng quan về cách tiếp cận trung bình hóa mô hình kiểu Bayes (Bayesian Model Averaging - BMA) và đưa ra một chiến lược mô hình hóa gồm các bước ứng dụng đảm bảo khả năng xây dựng và lựa chọn mô hình chấm điểm rủi

ro tín dụng tốt nhất, với kết quả minh họa cho phân khúc khách hàng doanh nghiệp nhỏ và vừa (SME) trong ngành Công nghiệp và xây dựng, dựa trên nguồn dữ liệu của Trung tâm Thông tin tín dụng Quốc gia Việt Nam (CIC).

2. Tổng quan về BMA

Giả sử rằng có nhiều hơn một mô hình với khả năng giải thích tương đương nhau, nếu chỉ lựa chọn một mô hình thì có thể sẽ dẫn đến rủi ro từ việc mô hình không ổn định trong quá trình ứng dụng. Ý tưởng về sự kết hợp các mô hình được đưa ra đầu tiên bởi Barnard (1963), tiếp đó Roberts (1965) đề xuất về một mô hình trung bình hóa của hai mô hình dựa trên xác suất hậu nghiệm. Leamer (1978) kế thừa và đưa ra cơ sở phương pháp luận cho BMA. Tuy nhiên, sau đó kỹ thuật này cũng ít nhận được sự quan tâm của giới nghiên cứu do những khó khăn trong thực hành tính toán. Chỉ đến những năm 1990, nhờ những tiến bộ của khoa học máy tính, BMA dần dần được quan tâm phát triển nhiều hơn cả về lý thuyết và thực hành, điển hình là các nghiên cứu của Draper (1995), Chatfield (1995), Kass & Raftery (1995) và George (1999). Cho đến nay, kỹ thuật này đã được ứng dụng trong nhiều lĩnh vực khác nhau.

Có thể kể đến, trong y học, Fried & cộng sự (1991) sử dụng BMA xây dựng mô hình dự báo bệnh tim mạch tại nước Mỹ; Greenland (1993) ứng dụng BMA trong bài toán dự báo tỷ lệ tử vong ở trẻ sơ sinh. Trong xã hội học, Raftery (1995, 1999) nghiên cứu về tiêu chuẩn thông tin Bayes (BIC - chỉ số được sử dụng trong kỹ thuật BMA) và ứng dụng kỹ thuật này trong các bài toán dự đoán tỷ lệ tội phạm, hội chứng xã hội,... Trong kinh tế học, Ley & Steel (2007), Eicher & cộng sự (2011) sử dụng BMA trong bài toán xác định các yếu tố tác động đến tăng trưởng kinh tế. Jon & cộng sự (2012) xây dựng mô hình dự báo kinh tế theo thời gian thực với cách tiếp cận BMA, dựa trên một số lượng lớn các chỉ số tài chính.

Trong lĩnh vực rủi ro tín dụng, Silvia (2012) chỉ ra rằng mô hình hóa theo BMA đem lại khả năng phân biệt khách hàng tốt hơn phương pháp hồi quy logistic cổ điển, với bộ dữ liệu doanh nghiệp SME của nước Đức. Bằng cách tiếp cận này, Carlos & Enrique (2013) xác định được các nhân tố tài chính ảnh hưởng đến rủi ro vỡ nợ của doanh nghiệp nước Mỹ là tỷ lệ Vốn lưu động/Tổng tài sản, tỷ lệ Lợi nhuận giữ lại/Tổng tài sản và tỷ lệ Tổng nợ phải trả/Tổng tài sản. Trong một nghiên cứu khác, Hayden

& cộng sự (2014) cũng kết luận rằng mô hình xây dựng theo BMA có khả năng phân biệt cao hơn so với phương pháp hồi quy từng bước. Maltritz & cộng sự (2014) sử dụng BMA để xác định những yếu tố ảnh hưởng đến rủi ro tín dụng của một quốc gia. Kết quả nghiên cứu chỉ ra rằng mức tổng sản phẩm quốc nội (GDP) và tỷ lệ Nợ nước ngoài/GDP có ảnh hưởng cao đến rủi ro vỡ nợ của cả nhóm nền kinh tế mới nổi và nhóm nền kinh tế phát triển. Ngoài ra, đối với những nền kinh tế phát triển, các yếu tố lạm phát và tăng trưởng nhập khẩu cũng có tác động quan trọng; trong khi đối với những nền kinh tế đang phát triển, các yếu tố ảnh hưởng là tỷ lệ nợ, tỷ lệ dự trữ ngoại hối và lịch sử các sự kiện vỡ nợ gần đây. Ngoài ra, Matthias & Johannes (2016) cũng ứng dụng BMA trong phân tích các yếu tố nhằm phục vụ định giá các hợp đồng hoán đổi rủi ro tín dụng (credit default swaps).

Về kỹ thuật, khi đưa ra kết luận về một đại lượng Δ chưa biết với k mô hình thống kê M_1, M_2, \dots, M_k dựa trên tập dữ liệu D , xác suất hậu nghiệm của Δ được xác định bằng công thức sau:

$$P(\Delta|D) = \sum_{i=1}^k P(\Delta|M_i, D) P(M_i|D) \quad (1)$$

trong đó $P(\Delta|M_i, D)$ là xác suất hậu nghiệm của Δ khi cho biết mô hình M_i và $P(M_i|D)$ là xác suất hậu nghiệm của mô hình M_i khi cho trước dữ liệu D . Lưu ý rằng, k có thể rất lớn, ví dụ trong mô hình dự báo khả năng không trả được nợ của khách hàng: nếu có p nhân tố tài chính được đưa vào phân tích hồi quy logistic, thì k có thể bằng 2^p .

Công thức trên cho thấy rằng xác suất hậu nghiệm của Δ là trung bình có trọng số của xác suất hậu nghiệm của Δ trong mỗi mô hình.

Xác suất hậu nghiệm của M_i cho bởi công thức:

$$P(M_i|D) = \frac{P(D|M_i) P(M_i)}{\sum_{j=1}^k P(D|M_j) P(M_j)} \quad (2)$$

trong đó $P(M_i)$ là xác suất tiên nghiệm của mô hình M_i và $P(D|M_i)$ được xác định bởi công thức sau:

$$P(D|M_i) = \int P(D|\mu_i, M_i) P(\mu_i|M_i) d\mu_i \quad (3)$$

trong đó μ_i là hệ số của mô hình M_i và $P(D|\mu_i, M_i)$ là hàm hợp lý (*likelihood*) của hệ số μ_i trong mô hình M_i với D .

Trung bình hậu nghiệm và phương sai hậu nghiệm được suy ra từ công thức (1) sau:

$$E(\Delta|D) = \sum_{i=1}^k \hat{\Delta}_i P(M_i|D)$$

$$Var(\Delta|D) = \sum_{i=1}^k Var(\Delta|D, M_i) + \widehat{\Delta}_i^2 P(M_i|D) - E(\Delta|D)^2$$

trong đó: $\Delta_i = E(\widehat{\Delta}|D, M_i)$

Trong thực tiễn, quá trình triển khai tính toán theo kỹ thuật BMA nảy sinh một số vấn đề sau:

- Hàm tổng trong công thức (1) có thể chứa rất nhiều số hạng, gây khó khăn cho việc tính toán;
- Tích phân trong công thức (3) khó tính toán trong nhiều trường hợp, đặc biệt khi có nhiều chiều.

Gói ứng dụng BMA phiên bản 3.18.6 (năm 2015) được Raftery, Hoeting, Volinsky, Painter và Yeung phát triển trên ngôn ngữ thống kê R đã giải quyết được các vấn đề trên. Cụ thể, đối với vấn đề thứ nhất, hàm tổng ở công thức (1) sẽ được rút gọn dựa trên phương pháp *Occam's window* của Madigan & Raftery (1994). Đầu tiên, những mô hình có xác suất hậu nghiệm rất nhỏ so với mô hình có xác suất hậu nghiệm lớn nhất sẽ bị loại bỏ, còn lại các mô hình trong tập hợp sau:

$$\mathcal{A} = \{M_j, \text{Max}_i [P(M_i|D)] \leq C \cdot P(M_j|D)\}$$

trong đó C là hằng số được tùy chọn theo mục đích của người phân tích.

Tiếp đó, loại bỏ những mô hình có nhiều biến nhưng xác suất hậu nghiệm lại nhỏ hơn mô hình ít biến hơn:

$$\mathcal{B} = \{M_l, \exists M_j \in \mathcal{A}, M_j \subset M_l, P(M_j|D) > P(M_l|D)\}$$

Đặt $\mathcal{R} = \mathcal{A} \setminus \mathcal{B}$, thì hàm tổng trong (1) được rút gọn lại như sau:

$$P(\Delta|D) = \sum_{i \in \mathcal{R}} P(\Delta|M_i, D) P(M_i|D) \quad (4)$$

Hàm *bic.glm* trong gói ứng dụng BMA nói trên đã sử dụng thuật toán của Furnival và Wilson (1974) để có thể tìm kiếm nhanh chóng các mô hình trong tập \mathcal{R} nói trên.

Đối với vấn đề thứ hai, theo Raftery (1995), tích phân trong công thức (3) có thể được xấp xỉ bởi chỉ số BIC (Bayesian Information Criterion):

$$P(D|M_i) \approx e^{-BIC_i/2}$$

trong đó BIC_i là chỉ số đo khả năng giải thích của mô hình M_i đối với dữ liệu D .

Về tổng quan, hàm *bic.glm* sẽ tìm kiếm những mô hình tốt nhất (với nghĩa có xác suất hậu nghiệm cao nhất) dựa trên phương pháp *Occam's window*, sau đó lấy “trung bình hóa” của các mô hình này theo công thức (4) để ra mô hình cuối cùng.

3. Ứng dụng BMA trong xây dựng mô hình chấm điểm rủi ro tín dụng

3.1. Phương pháp xây dựng

Phương pháp xây dựng và lựa chọn mô hình chấm điểm rủi ro tín dụng trong bài viết gồm các bước sau:

- Bước 1: trên cơ sở danh sách dài (long list) các biến tài chính, tiến hành phân tích đơn biến để (i) loại bớt các biến có tương quan cặp cao và có ý nghĩa thống kê thấp hoặc ngược với kì vọng về ý nghĩa kinh tế trong việc phân biệt khách hàng, và (ii) biến đổi biến nhằm tăng khả năng phân biệt của từng nhân tố.

- Bước 2: trên cơ sở danh sách ngắn (short list) các biến tài chính có chất lượng tốt thu được từ bước 1, sử dụng gói ứng dụng BMA để tìm ra các mô hình trong tập \mathcal{A} và tập \mathcal{R} .

- Bước 3: thay vì lấy kết quả mô hình trung bình hóa, từng mô hình trong tập \mathcal{R} được xem xét và so sánh để lựa chọn ra mô hình cuối cùng; trong một số trường hợp, có thể mở rộng xem xét trên tập \mathcal{A} bởi trong số các mô hình có xác suất hậu nghiệm lớn có những mô hình rất ít biến. Các tiêu chí phổ biến được kết hợp xem xét trong quá trình so sánh, lựa chọn ra mô hình tốt nhất gồm: độ lớn và độ ổn định của giá trị thống kê AR (Accuracy Ratio - đo hiệu năng phân biệt khách hàng tốt/xấu; tham khảo BCBS, 2005), số lượng biến, số nhóm biến, dấu ý nghĩa kinh tế của từng biến, ... trong mô hình.

3.2. Dữ liệu

Theo thông lệ, cấu trúc dữ liệu phục vụ bài toán chấm điểm rủi ro tín dụng bao gồm dữ liệu tài chính và dữ liệu phi tài chính. Cụ thể, dữ liệu tài chính là các chỉ tiêu, tỷ số tài chính doanh nghiệp, được tính toán từ các báo cáo tài chính của công ty (bao gồm Báo cáo cân đối kế toán, Báo cáo kết quả hoạt động kinh doanh, và Báo cáo lưu chuyển tiền tệ). Ngoài ra, ở các thị trường phát triển, dữ liệu giao dịch cổ phiếu của công ty (thị giá, khối lượng giao dịch, lợi suất, ...) cũng được xem xét dưới vai trò các biến độc lập cho việc xây dựng mô hình chấm điểm rủi ro tín dụng ở khía cạnh tài chính. Dữ liệu phi tài chính thường bao gồm các nội dung như trình độ quản lý và môi trường nội bộ doanh nghiệp, vị thế cạnh tranh của công ty, triển vọng ngành hoạt động của công ty, các hỗ trợ tài chính và kỹ thuật từ Chính phủ hoặc công ty mẹ, ... Trong thực tế, khác với dữ liệu tài chính là dễ thu thập và chuẩn hóa cấu trúc, việc thu thập dữ liệu phi tài chính tốn nhiều chi phí

Bảng 1: Thống kê tập dữ liệu CIC theo cách phân khúc khách hàng

Phân khúc	Siêu nhỏ			Nhỏ và vừa			Lớn		
	Tổng số quan sát	Số quan sát <i>bad</i>	Tỷ lệ <i>bad</i>	Tổng số quan sát	Số quan sát <i>bad</i>	Tỷ lệ <i>bad</i>	Tổng số quan sát	Số quan sát <i>bad</i>	Tỷ lệ <i>bad</i>
Nông, lâm nghiệp và thủy sản	102	0	0%	286	3	1,05%	218	6	2,75%
Công nghiệp và xây dựng	7.218	44	0,61%	15.907	134	0,84%	7.351	122	1,66%
Thương mại và dịch vụ	9.448	51	0,54%	7.232	71	0,98%	1.729	39	2,26%

Nguồn: Tác giả thống kê tổng hợp.

và thường chỉ được tiến hành bởi các định chế tài chính trong quá trình đánh giá rủi ro tín dụng khách hàng, đối tác.

Bộ dữ liệu thô được sử dụng trong nghiên cứu này là thông tin báo cáo tài chính và tình trạng nợ xấu tại giai đoạn 2009-2010 của 25.000 doanh nghiệp Việt Nam trong cơ sở dữ liệu của Trung tâm Thông tin tín dụng quốc gia (CIC). Cấu trúc về các phân khúc khách hàng của tập dữ liệu theo cách phân loại doanh nghiệp của Chính phủ (2009) được trình bày trong Bảng 1.

Trong đó, tình trạng *bad* được hiểu là doanh nghiệp có nợ xấu trong vòng 12 tháng sau khi công bố báo cáo tài chính.

Trên cơ sở bộ dữ liệu này, mẫu dữ liệu về các

doanh nghiệp nhỏ và vừa (SME) trong ngành Công nghiệp và Xây dựng, với 134 quan sát *bad* trong tổng số 15.907 quan sát, được sử dụng cho việc xây dựng mô hình chấm điểm rủi ro tín dụng cho phân khúc khách hàng này. Bảng 4 tại Phụ lục cung cấp danh sách các chỉ tiêu, tỷ số phân tích tài chính doanh nghiệp (long list), được tính toán từ dữ liệu báo cáo tài chính, phục vụ như các biến số đầu vào cho quá trình phân tích.

3.3. Kết quả ước lượng mô hình

Kết quả phân tích đơn biến, được trình bày trong Bảng 5 tại Phụ lục, là danh sách ngắn (short list) gồm 22 biến tài chính có chất lượng tốt cả về mặt thống kê lẫn ý nghĩa kinh tế.

Thuật toán của hàm *bic.glm* trong gói ứng dụng

Bảng 2: Kết quả mô hình

Biến*	Hệ số ước lượng	Độ lệch chuẩn	P-value	Mô tả
Intercept	-0.7609	0.4018	0.0583	Hệ số chặn
X18	-0.0229	0.0116	0.0497	Vốn lưu động/Tổng tài sản
X22	-0.0168	0.00715	0.0186	Tài sản ngắn hạn/Nợ rỗng
X23	-0.0134	0.00785	0.0881	Tài sản ngắn hạn/(Tổng tài sản-Tài sản vô hình)
X28	-0.0222	0.0114	0.0523	(Nợ phải trả - Tiền - Các khoản tương đương tiền)/Doanh thu thuần
X52	-0.0106	0.00539	0.0503	Lợi nhuận thuần từ hoạt động sản xuất kinh doanh/ Chi phí lãi vay
Accuracy Ratio: 40% Posterior probability: 1.44E-05 BIC=1515.912				

Ghi chú: * các giá trị biến đã trải qua bước biến đổi biến (variable transformation, trong đó có Z-score standardization).

Nguồn: Tác giả tính toán.

Bảng 3: Ma trận hệ số tương quan

Correlation	X18	X22	X23	X28	X52
X18	1	0.02	0.17	0.7	0.09
X22	0.02	1	0.45	0.32	0.25
X23	0.17	0.45	1	0.19	0.05
X28	0.7	0.32	0.19	1	0.2
X52	0.09	0.25	0.05	0.2	1

Nguồn: Tác giả tính toán.

BMA, với lựa chọn sử dụng hồi quy logistic – kỹ thuật phổ biến theo thông lệ thực hành xây dựng mô hình chấm điểm rủi ro tín dụng cho các khách hàng SME, chia tập hợp tất cả các mô hình thành nhiều nhóm nhỏ, với mỗi nhóm là một tập hợp các mô hình có i biến ($i=1, 2, \dots, 22$), sau đó lấy ra 500 mô hình (quy định qua tham số $nbest$) có xác suất hậu nghiệm cao nhất trong từng nhóm. Trên cơ sở tập hợp tất cả các kết quả 500 mô hình của từng nhóm, phương pháp *Occam's window* (quy định hằng số qua hai tham số $Or=50$ và $Or:fix=5$ – viết tắt của *Occam's razor*) lược bỏ bớt số lượng các mô hình để cho ra tập A gồm hơn 2.000 mô hình và tập R gồm

22 mô hình. (Bảng 6 tại Phụ lục cung cấp danh sách 20 mô hình có AR lớn nhất trong tập hợp A). Trong số các mô hình này, dựa trên việc xem xét kết hợp các tiêu chí được trình bày ở trên, mô hình tốt nhất được lựa chọn có kết quả được trình bày tại Bảng 2.

Kết quả ma trận tương quan, được trình bày tại Bảng 3, đảm bảo rằng không tồn tại sự tương quan cặp quá cao giữa các biến trong mô hình.

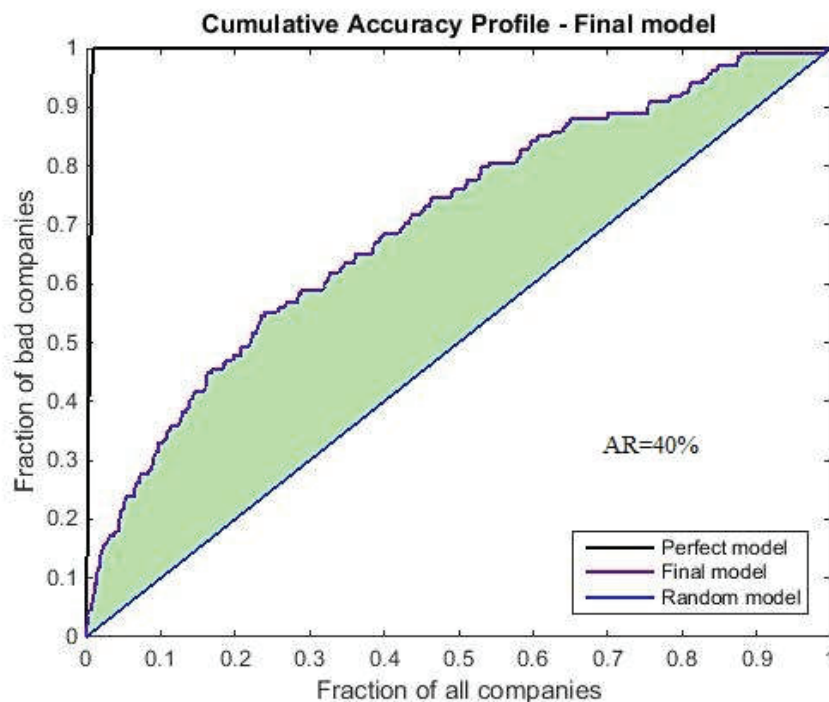
Khả năng phân biệt khách hàng của mô hình (tham khảo BCBS, 2005) có thể kết luận là tốt, được thể hiện một cách kỹ thuật qua hình dạng đường CAP và đường phân bố tích lũy được cung cấp tại Hình 1 và Hình 2.

Như vậy, điểm rủi ro tín dụng của từng doanh nghiệp SME có thể được xác định như sau:

$$\text{Score} = -0.7609 - 0.0229 * X18 - 0.0168 * X22 - 0.0134 * X23 - 0.0222 * X28 - 0.0106 * X52$$

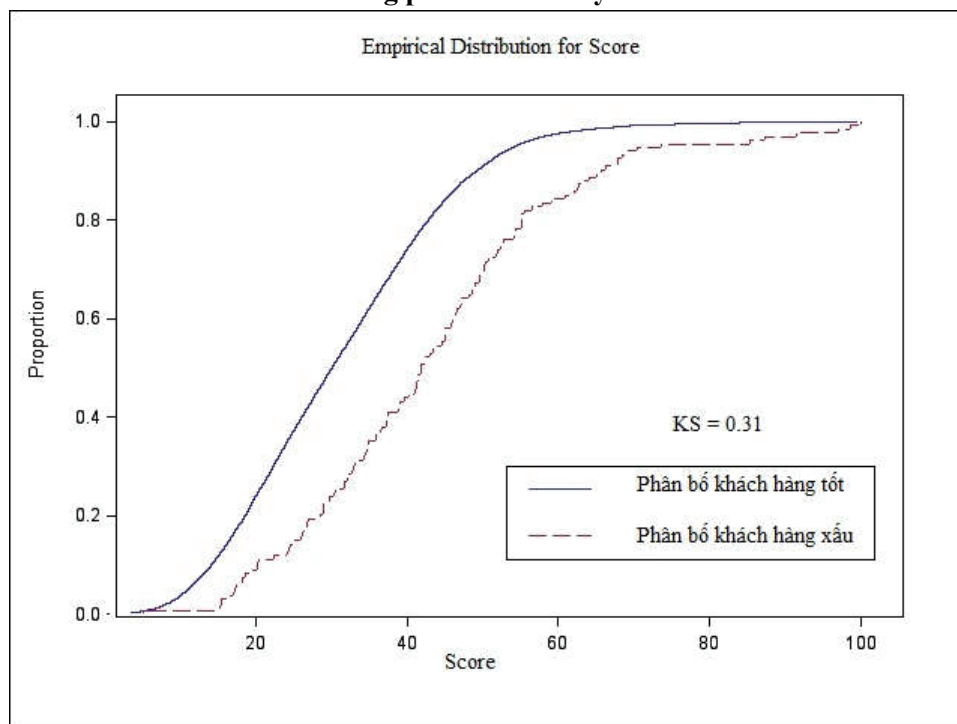
Nhằm thuận tiện cho việc ứng dụng, điểm rủi ro thường được quy về khoảng giá trị $[0, 100]$ trong khi vẫn giữ nguyên thứ tự (rank order) qua phép biến đổi max/min standardization (xem Baesens, 2014) như sau:

$$X_{\text{new}} = \frac{X_{\text{old}} - \min(X_{\text{old}})}{\max(X_{\text{old}}) - \min(X_{\text{old}})} * (\text{newmax} - \text{newmin}) + \text{newmin}$$

Hình 1: Đường CAP của mô hình

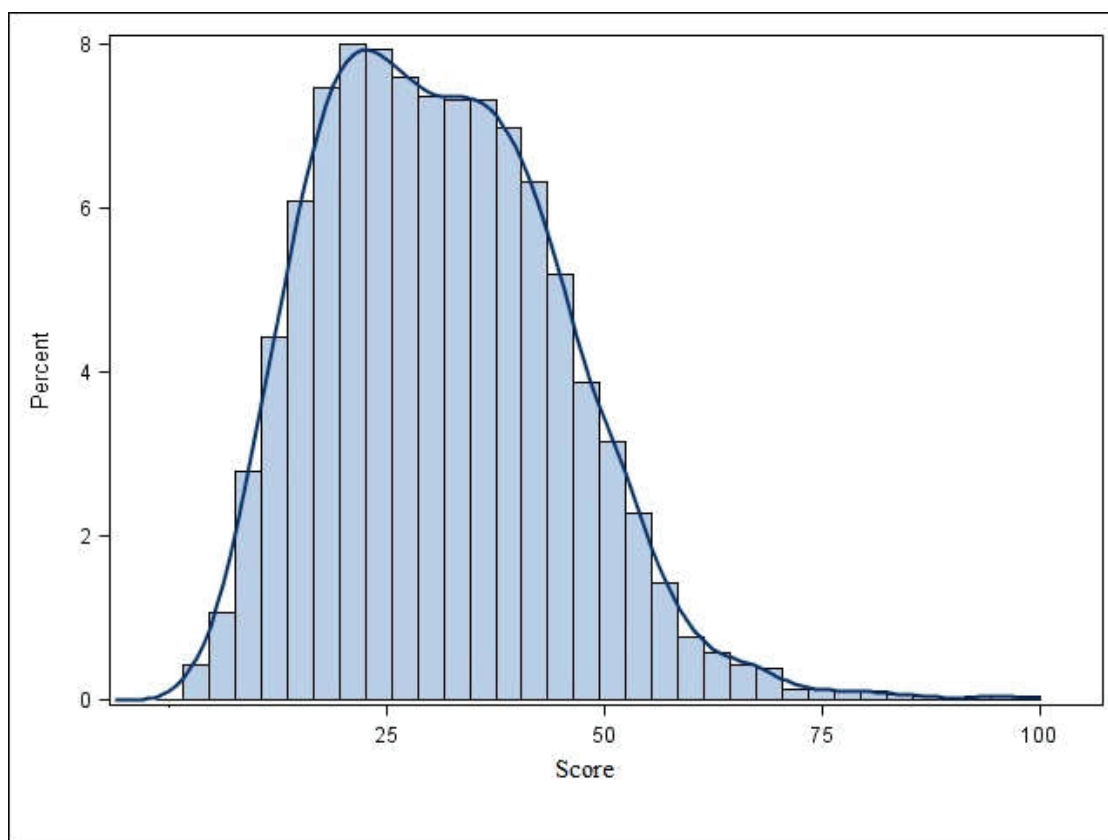
Nguồn: Tác giả tính toán và minh họa.

Hình 2: Đường phân bố tích lũy của mô hình



Nguồn: Tác giả tính toán và minh họa.

Hình 3: Phân bố điểm rủi ro tín dụng



Nguồn: Tác giả tính toán và minh họa.

Bảng 3: Các nhân tố tài chính có vai trò quyết định tới rủi ro tín dụng của khách hàng

Tên biến	Nhóm	Công thức	Dấu kì vọng*	AR
X18	Thanh khoản	Vốn lưu động/Tổng tài sản	+	32.40%
X22	Thanh khoản	Tài sản ngắn hạn/Nợ rỗng	+	26.61%
X23	Thanh khoản	Tài sản ngắn hạn/(Tổng tài sản-Tài sản vô hình)	+	25.27%
X28	Thanh khoản	(Nợ phải trả - Tiền - Các khoản tương đương tiền)/Doanh thu thuần	-	35.82%
X40	Quy mô	Nợ rỗng	-	26.34%

Ghi chú: * Dấu kì vọng trước bước biến đổi biến: (+) giá trị càng lớn, rủi ro càng thấp; (-) giá trị càng lớn, rủi ro càng cao.

Nguồn: Tác giả tính toán.

trong đó $newmax$ và $newmin$ là giá trị cực đại và cực tiểu của khoảng giá trị mới, ở trường hợp này là 0 và 100.

Kết quả phân bố điểm rủi ro tín dụng của mẫu dữ liệu doanh nghiệp SME trong ngành Công nghiệp và Xây dựng được trình bày tại Hình 3.

Ngoài ra, dựa trên thống kê về tần suất xuất hiện trong các mô hình từ kết quả phân tích BMA (xem Hình 5 tại Phụ lục) và kết quả phân tích đơn biến, những nhân tố tài chính có vai trò quyết định tới rủi ro không trả được nợ của khách hàng có thể được xác định rõ tại Bảng 3, cùng với các thông tin chi tiết về đường CAP của từng nhân tố tại Hình 4.

4. Kết luận

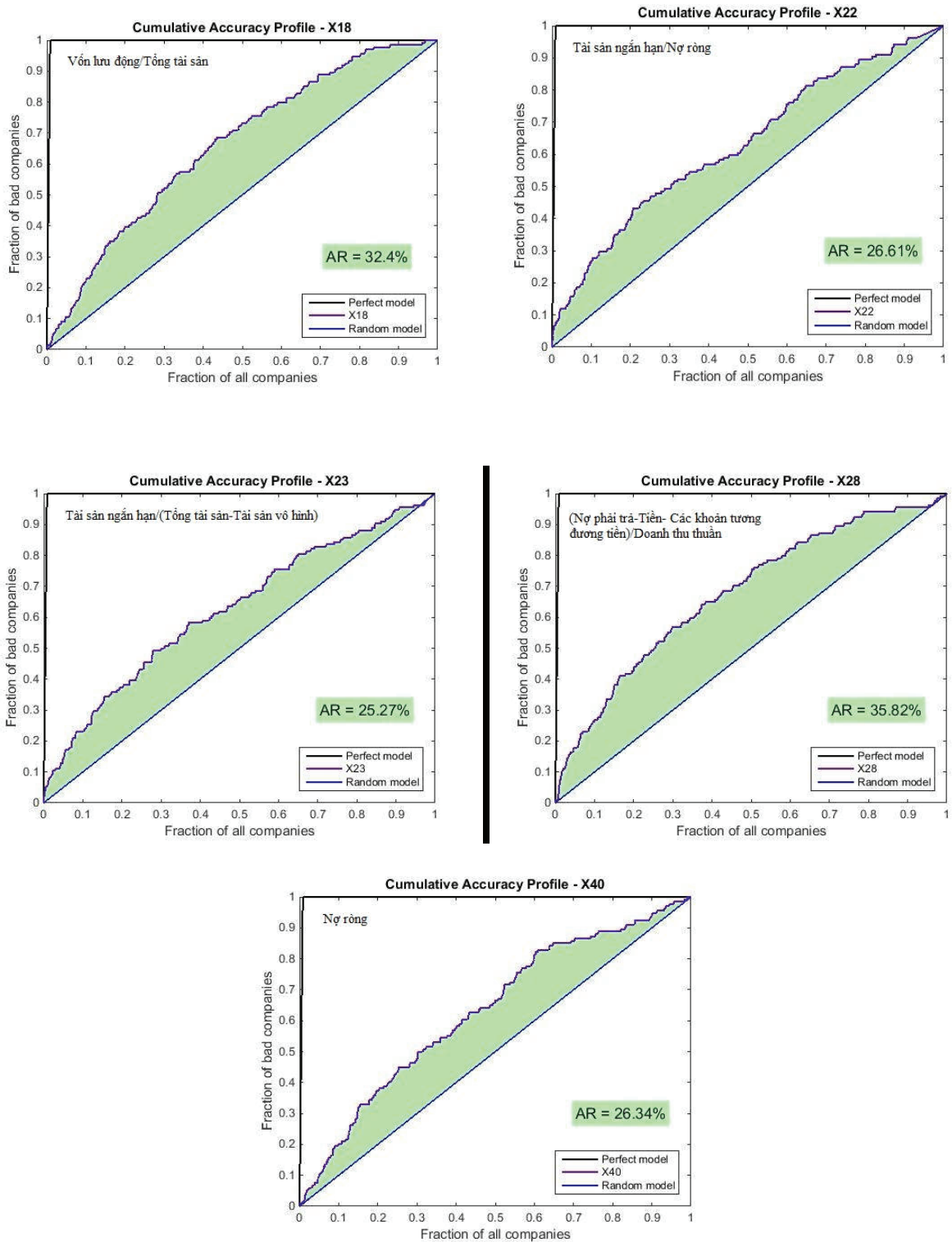
So với những kỹ thuật xây dựng mô hình truyền thống, cách tiếp cận BMA thể hiện hiệu quả vượt trội khi cho phép xem xét, so sánh và lựa chọn giữa nhiều mô hình tốt hơn. Bên cạnh việc đưa ra một chiến lược mô hình hóa gồm các bước ứng dụng đảm bảo khả năng xây dựng và lựa chọn mô hình chấm điểm rủi ro tín dụng một cách toàn diện và thấu đáo, kết quả nghiên cứu thực nghiệm trong bài viết, dựa trên mẫu dữ liệu từ nguồn Trung tâm Thông tin tín dụng Quốc gia Việt Nam (CIC), gợi ý rằng các chỉ tiêu thanh khoản: Vốn lưu động/Tổng tài sản, Tài sản ngắn hạn/Nợ rỗng, Tài sản ngắn hạn/(Tổng tài sản-Tài sản vô hình), và (Nợ phải trả - Tiền - Các khoản tương đương tiền)/Doanh thu thuần là những nhân tố rủi ro tài chính có nhiều giá trị thông tin về khả năng không trả được nợ của khách hàng doanh nghiệp Nhỏ và vừa (SME) trong ngành Công nghiệp và xây dựng.

Kết quả này hàm ý rằng, việc quản lý khả năng thanh toán trong các doanh nghiệp này là công tác rất quan trọng; đối với các tổ chức cho vay nói chung và ngân hàng thương mại nói riêng, việc theo dõi sát sao các chỉ tiêu thanh khoản của doanh nghiệp sẽ có tác dụng dự báo tốt cho việc khách hàng có đáp ứng được nghĩa vụ tín dụng trong tương lai hay không, từ đó giúp bên cho vay có các biện pháp hỗ trợ hoặc giảm thiểu rủi ro tín dụng kịp thời như điều chỉnh kì hạn trả nợ, giảm hạn mức tín dụng, trích lập dự phòng rủi ro,...

Về mặt ứng dụng, các ngân hàng có thể kế thừa sử dụng kết quả xây dựng mô hình trong nghiên cứu này như là cấu phần lõi về chấm điểm rủi ro tài chính, đồng thời phát triển thêm cấu phần về chấm điểm rủi ro phi tài chính, dựa trên các dữ liệu phi tài chính có thể thu thập, và các cấu phần điều chỉnh định tính khác (Overlays) để kết hợp, nhằm tăng cường hơn nữa hiệu quả của công tác quản trị rủi ro tín dụng – một nội dung đặc biệt quan trọng trong hành trình nâng cao năng lực quản trị rủi ro của các ngân hàng thương mại Việt Nam theo bộ tiêu chuẩn Basel II.

Tuy nhiên, là nghiên cứu ứng dụng BMA trong xây dựng mô hình chấm điểm rủi ro tín dụng đầu tiên tại Việt Nam, cũng cần lưu ý rằng, do hạn chế về số lượng quan sát *bad* và chiều dài thời gian của tập dữ liệu hiện tại, mô hình chấm điểm tín dụng cần được kiểm định ngoại mẫu (out-of-sample validation), kiểm định qua thời gian (out-of-time validation) khi có thêm dữ liệu, và ước lượng cập nhật các tham số nếu cần thiết. Khi dữ liệu để xây dựng mô hình đủ lớn và có tính đại diện cao, cách tiếp cận BMA sẽ

Hình 4: Đường CAP của các nhân tố tài chính quyết định



Nguồn: Tác giả tính toán và minh họa.

ngày càng cho hiệu quả lớn hơn.

Ngoài ra, phương pháp tiếp cận này có thể tiếp tục được ứng dụng mở rộng cho các phân khúc khách

hàng có lượng dữ liệu đủ lớn khác như khách hàng SME trong từng ngành/tiểu ngành kinh tế, khách hàng cá nhân, khách hàng hộ gia đình,...

PHỤ LỤC

Bảng 5: Các chỉ tiêu, tỷ số phân tích tài chính doanh nghiệp (long list)

Tên biến	Nhóm	Công thức	Dấu kỳ vọng
X01	Profit	Lợi nhuận gộp từ bán hàng và cung cấp dịch vụ/Doanh thu thuần	1
X02	Profit	Lợi nhuận từ hoạt động kinh doanh (không bao gồm hoạt động tài chính)/Doanh thu thuần	1
X03	Profit	Lợi nhuận sau thuế/Doanh thu thuần	1
X04	Profit	Lợi nhuận thuần từ hoạt động sản xuất kinh doanh/Giá vốn hàng bán	1
X05	Profit	Lợi nhuận thuần từ hoạt động sản xuất kinh doanh/Doanh thu thuần	1
X06	Profit	Lợi nhuận gộp/ Chi phí hoạt động	1
X07	Profit	Tổng lợi nhuận kế toán trước thuế/Giá vốn hàng bán	1
X08	Profit	Tổng lợi nhuận kế toán trước thuế/Doanh thu thuần	1
X09	Profit	Lợi nhuận sau thuế thu nhập doanh nghiệp/Giá vốn hàng bán	1
X10	Profit	Lợi nhuận sau thuế thu nhập doanh nghiệp/Doanh thu thuần	1
X11	Profit	Lợi nhuận sau thuế thu nhập doanh nghiệp/Chi phí hoạt động	1
X12	Profit	Lợi nhuận gộp/ Giá vốn hàng bán	1
X13	Profit	Lợi nhuận gộp/ Doanh thu thuần	1
X14	Efficiency	Chi phí hoạt động/ Doanh thu thuần	-1
X15	Liquidity	Tài sản ngắn hạn/Nợ ngắn hạn	1
X16	Liquidity	(Tài sản ngắn hạn – Hàng tồn kho)/ Nợ ngắn hạn	1
X17	Liquidity	(Tiền và các khoản tương đương tiền)/Nợ ngắn hạn	1
X18	Liquidity	Vốn lưu động / Tổng tài sản	1
X19	Liquidity	Tài sản ngắn hạn / Tổng tài sản	1
X20	Liquidity	(Tài sản ngắn hạn – Nợ ngắn hạn)/ Nợ phải trả	1
X21	Liquidity	Tài sản ngắn hạn/ Nợ phải trả	1
X22	Liquidity	Tài sản ngắn hạn/Nợ rỗng	1
X23	Liquidity	Tài sản ngắn hạn / (Tổng tài sản-Tài sản vô hình)	1
X24	Liquidity	Tài sản ngắn hạn / Nợ dài hạn	1
X25	Liquidity	(Tài sản ngắn hạn – Nợ ngắn hạn)/ Doanh thu thuần	1
X26	Liquidity	(Tiền+Các khoản tương đương tiền+Phải thu khách hàng)/Nợ ngắn hạn	1
X27	Liquidity	(Tiền+Các khoản tương đương tiền+Phải thu khách hàng+Hàng tồn kho)/Nợ ngắn hạn	1
X28	Liquidity	(Nợ phải trả - Tiền - Các khoản tương đương tiền)/Doanh thu thuần	-1
X29	Liquidity	(Khoản phải thu khách hàng + Tiền + Các khoản tương đương tiền - Phải trả khách hàng)/ (Giá vốn hàng bán + Chi phí lãi vay)	1
X30	Cap_usage	Tài sản cố định/ Tổng tài sản	-1
X31	Cap_usage	Tài sản cố định/ (Tổng tài sản - Nợ ngắn hạn)	-1
X32	Cap_usage	Tài sản cố định/ (Tổng tài sản-Tài sản vô hình)	-1
X33	Cap_usage	Tài sản ngắn hạn/ (Tổng tài sản - Nợ ngắn hạn)	1
X34	Cap_usage	Tài sản hữu hình ròng / Tài sản cố định	1
X35	Size	Lợi nhuận gộp	1
X36	Size	Tổng lợi nhuận kế toán trước thuế	1
X37	Size	Lợi nhuận thuần từ hoạt động sản xuất kinh doanh	1
X38	Size	Tài sản ngắn hạn	1
X39	Size	Tổng tài sản	1
X40	Size	Nợ rỗng	-1
X41	Size	Tài sản cố định	1
X42	Size	Doanh thu thuần	1
X43	Size	Nợ phải trả	-1
X44	Solvency	Nợ phải trả/Tổng tài sản	-1
X45	Solvency	Nợ dài hạn/Vốn chủ sở hữu	-1
X46	Solvency	(Lợi nhuận trước thuế + Chi phí lãi vay)/Chi phí lãi vay	1
X47	Solvency	Tổng tài sản/Vốn chủ sở hữu	-1
X48	Solvency	Nợ ngắn hạn/ (Tổng tài sản - Nợ ngắn hạn)	-1
X49	Solvency	Nợ ngắn hạn/ Tổng tài sản	-1
X50	Solvency	Nợ ngắn hạn/ Nợ phải trả	-1
X51	Solvency	Tổng tài sản/ Nợ dài hạn	1
X52	Solvency	Lợi nhuận thuần từ hoạt động sản xuất kinh doanh/ Chi phí lãi vay	1
X53	Solvency	Lợi nhuận sau thuế/ Chi phí lãi vay	1
X54	Solvency	Lợi nhuận gộp/ Chi phí lãi vay	1
X55	Solvency	Chi phí lãi vay/ Doanh thu thuần	-1

Nguồn: Tác giả tổng hợp.

Bảng 6: Danh sách các biến tài chính có chất lượng tốt (short list)

Tên biến	Nhóm	Công thức	Dấu kì vọng	AR
X15	Liquidity	Tài sản ngắn hạn/Nợ ngắn hạn	1	17.29%
X16	Liquidity	(Tài sản ngắn hạn – Hàng tồn kho)/ Nợ ngắn hạn	1	16.10%
X17	Liquidity	(Tiền và các khoản tương đương tiền)/Nợ ngắn hạn	1	23.03%
X18	Liquidity	Vốn lưu động / Tổng tài sản	1	32.40%
X21	Liquidity	Tài sản ngắn hạn/ Nợ phải trả	1	23.20%
X22	Liquidity	Tài sản ngắn hạn/Nợ ròng	1	26.61%
X23	Liquidity	Tài sản ngắn hạn / (Tổng tài sản-Tài sản vô hình)	1	25.27%
X24	Liquidity	Tài sản ngắn hạn / Nợ dài hạn	1	19.39%
X25	Liquidity	(Tài sản ngắn hạn – Nợ ngắn hạn)/ Doanh thu thuần	1	8.09%
X28	Liquidity	(Nợ phải trả - Tiền - Các khoản tương đương tiền)/Doanh thu thuần	-1	35.82%
X31	Cap_usage	Tài sản cố định/ (Tổng tài sản - Nợ ngắn hạn)	-1	20.88%
X33	Cap_usage	Tài sản ngắn hạn/ (Tổng tài sản - Nợ ngắn hạn)	1	17.18%
X40	Size	Nợ ròng	-1	26.34%
X42	Size	Doanh thu thuần	1	8.73%
X43	Size	Nợ phải trả	-1	17.91%
X44	Solvency	Nợ phải trả/Tổng tài sản	-1	5.12%
X45	Solvency	Nợ dài hạn/Vốn chủ sở hữu	-1	18.47%
X46	Solvency	(Lợi nhuận trước thuế + Chi phí lãi vay)/Chi phí lãi vay	1	11.92%
X51	Solvency	Tổng tài sản/ Nợ dài hạn	1	18.31%
X52	Solvency	Lợi nhuận thuần từ hoạt động sản xuất kinh doanh/ Chi phí lãi vay	1	14.17%
X54	Solvency	Lợi nhuận gộp/ Chi phí lãi vay	1	11.07%
X55	Solvency	Chi phí lãi vay/ Doanh thu thuần	-1	12.79%

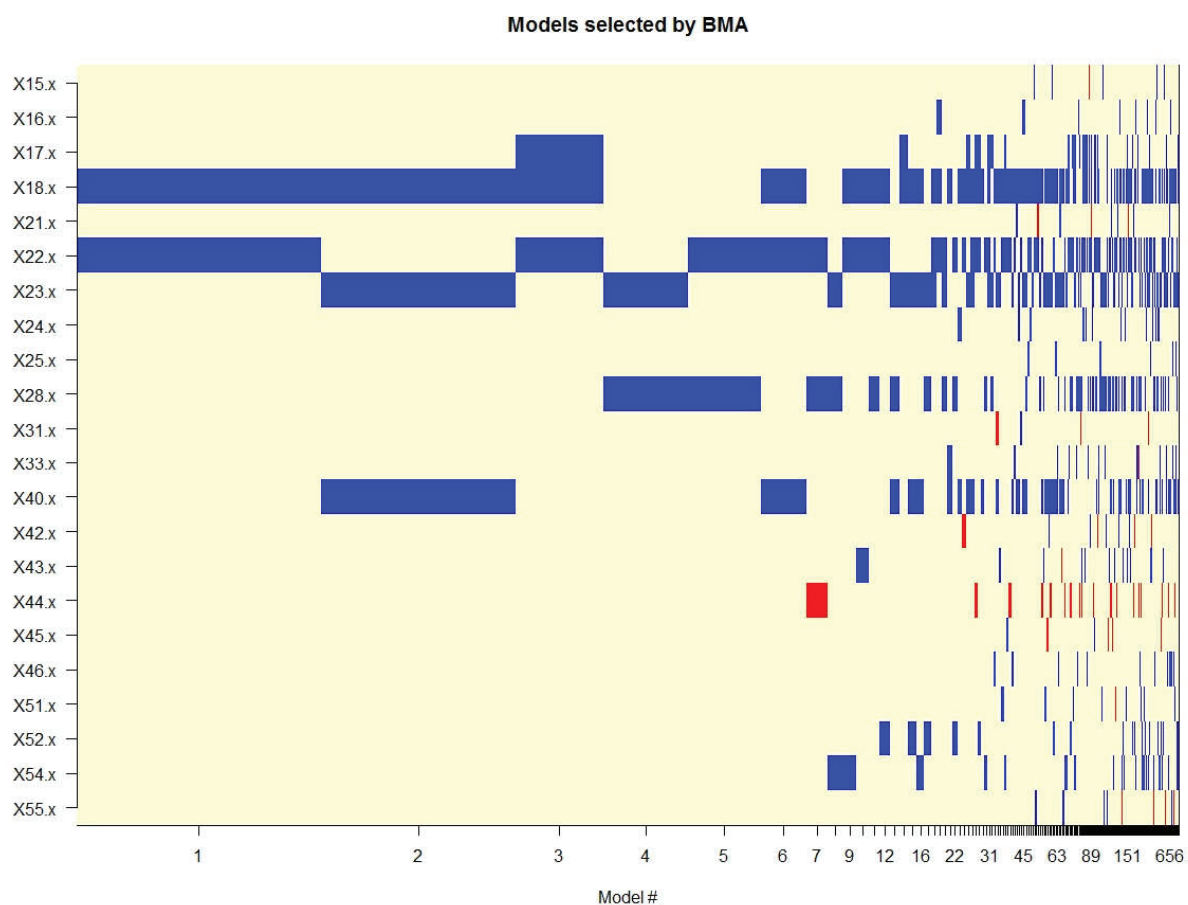
Nguồn: Tác giả tính toán.

Bảng 7: Các mô hình có AR lớn trong tập hợp A

STT	Mô hình	Số nhóm	Số biến	AR
1	X18 X22 X23 X28 X52	2	5	0.400095
2	X18 X22 X23 X54	2	4	0.410137
3	X18 X22 X28 X52	2	4	0.407387
4	X18 X22 X23 X28	1	4	0.389673
5	X17 X22 X28 X33	2	4	0.388153
6	X15 X18 X24 X54	2	4	0.384106
7	X22 X23 X28 X52	2	4	0.382391
8	X16 X18 X23 X52	2	4	0.3783
9	X16 X18 X23 X24	1	4	0.37675
10	X18 X23 X51 X52	2	4	0.374209
11	X18 X23 X24 X52	2	4	0.373087
12	X15 X18 X23 X51	2	4	0.36889
13	X18 X23 X54	2	3	0.407644
14	X18 X22 X43	2	3	0.406362
15	X18 X22 X28	1	3	0.406136
16	X18 X22 X54	2	3	0.402137
17	X28 X33 X54	3	3	0.397558
18	X23 X28 X54	2	3	0.397349
19	X28 X40 X54	3	3	0.395775
20	X17 X18 X24	1	3	0.393521

Nguồn: Tác giả tính toán.

Hình 5: Tần suất xuất hiện biến trong các mô hình từ kết quả phân tích BMA



Nguồn: Tác giả tính toán và minh họa.

Tài liệu tham khảo

- Baesens, B. (2014), *Analytics in a big data world: The essential guide to data science and its application*, Wiley.
- Barnard, G.A. (1963), 'New methods of quality control', *Journal of the Royal Statistical Society, Series A (General)*, 126, 255-258.
- BCBS (2005), 'Studies on the validation of internal rating systems', *Working paper No. 14*, Basel Committee on Banking Supervision.
- Carlos, G-A & Enrique, M-B. (2013). 'Determinants of corporate default: a BMA approach' *Applied Economics Letters*, 20(6), 511-514.
- Chatfield, C. (1995), 'Model uncertainty, data mining, and statistical inference (with discussion)', *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 158, 419-466.
- Chính phủ (2009), Nghị định Số 56/2009/NĐ-CP, về trợ giúp phát triển doanh nghiệp nhỏ và vừa, ban hành ngày 30 tháng 06 năm 2009.
- Draper, D. (1995), 'Assessment and propagation of model uncertainty', *Journal of the Royal Statistical Society, Series B*, 57, 45-97.
- Eicher, T., Papageorgiou, C. & Raftery, A. (2011), 'Default priors and predictive performance in Bayesian model averaging, with application to growth determinants', *Journal of Applied Econometrics*, 26, 30-55.

- Fried, L.P., Borhani, N.O., Enright, P., Furberg, C.D., Gardin, J.M., Kronmal, R.A., Kuller, L. H., Manolio, T.A., Mittelmark, M.B., Newman, A., O’Leary, D.H., Psaty, B., Rautaharju, P., Tracy, R.P. & Weiler, P.G. (1991), ‘The cardiovascular health study: design and rationale’, *Annals of Epidemiology*, 1, 263–276.
- Furnival, G.M. & Wilson, R.W. (1974), ‘Regression by leaps and bounds’, *Technometrics*, 16, 99–511.
- Greenland, S. (1993), ‘Methods for epidemiologic analyses of multiple exposures — A review and comparative study of maximum-likelihood, preliminary testing, and empirical Bayes regression’, *Statistics in Medicine*, 12, 717–736.
- George, E.I. (1999), ‘Bayesian model selection’, *Encyclopedia of Statistical Sciences*, Update 3, Wiley, New York, 39-46.
- Hayden, E., Stomper, A. & Westerkamp, A. (2014), ‘Selection versus averaging of logistic credit risk models’, *Journal of Risk*, 16(5), 39-52.
- Jon, F., Simon, G., Jonathan H.W. & Egon, Z. (2012), ‘Credit Spreads as Predictors of Real-Time Economic Activity: A Bayesian Model-Averaging Approach’, *Finance and Economics Discussion Series: 2012-77*, Federal Reserve Board.
- Kass, R.E. & Raftery, A.E. (1995), ‘Bayes factors’, *Journal of the American Statistical Association*, 90, 773–795.
- Leamer, E.E. (1978), *Specification Searches*, 1st edition, Wiley, New York.
- Ley, E. & Steel, M. (2007), ‘Jointness in Bayesian variable selection with applications to growth regressions’, *Journal of Macroeconomics*, 29, 476-493.
- Madigan, D. & Raftery, A.E. (1994), ‘Model selection and accounting for model uncertainty in graphical models using Occam’s window’, *Journal of the American Statistical Association*, 89, 1535–1546.
- Maltritz, D. & Molchanov, A. (2014), ‘Country credit risk determinants with model uncertainty’, *International Review of Economics & Finance*, 29(C), 224-234.
- Matthias, Pelster & Johannes, V. (2016), ‘The Determinants of CDS Spreads: Evidence from the model space’, *Deutsche Bundesbank Discussion Paper No. 43/2016*, Frankfurt.
- Raftery, A.E. (1995), ‘Bayesian model selection in social research’, *Sociological Methodology*, 25, 111–163.
- Raftery, A.E. (1999), ‘Bayes factors and BIC: Comment on “A critique of the Bayesian information criterion for model selection”’, *Sociological Methods and Research*, 27, 411–427.
- Roberts, H.V. (1965), ‘Probabilistic prediction’, *Journal of the American Statistical Association*, 60, 50–62.
- Silvia, F. (2012), ‘Bayesian model averaging for financial credit risk measurement’, *Working paper*, University of Pavia, Italy.