PHÂN TÍCH PHÂN CỤM CÁC DOANH NGHIỆP NGÀNH XÂY DỰNG Ở VIỆT NAM: ỨNG DỤNG PHƯƠNG PHÁP PHÂN TÍCH DỮ LIỆU ĐA HỢP

PGS.TS. Đào Thị Thanh Bình* - ThS. Nguyễn Anh Tuấn* - ThS. Lại Hoài Phương*

Nghiên cứu này sử dụng các chỉ số tài chính để phân cụm doanh nghiệp bằng phương pháp phân tích dữ liệu đa hợp. Từ bộ dữ liệu báo cáo tài chính của 162 công ty xây dựng tại Việt Nam, nghiên cứu nhận diện các nhóm doanh nghiệp và đặc điểm nổi bật của từng cụm trong giai đoạn 2012–2023. Kết quả cho thấy dữ liệu được chia thành ba cụm riêng biệt, phản ánh sự khác biệt về hiệu quả kinh doanh và mức độ rủi ro tài chính.

• Từ khóa: *phân tích dữ liệu đa hợp; phân tích phân cụm; chỉ số tài chính; ngành xây dựng.*

This study employs financial ratios to cluster firms using a compositional data analysis approach. Drawing on financial statement data from 162 construction companies in Vietnam, the research identifies distinct groups of firms and characterizes the defining features of each cluster over the period 2012-2023. The findings reveal three clearly differentiated clusters, reflecting substantial variation in operational performance and financial risk.

• Key words: *compositional data analysis; cluster analysis; financial ratios; construction industry.*

JEL codes: C58; L74

Ngày nhận bài: 14/6/2025
Ngày gửi phản biện: 23/8/2025
Ngày duyệt đăng: 16/12/2025

DOI: https://doi.org/10.71374/jfarv.v25.i303.14

1. Giới thiệu

Phân nhóm doanh nghiệp là quá trình chia doanh nghiệp thành các nhóm dựa trên đặc điểm như quy mô, loại hình sở hữu, lĩnh vực hoạt động, mức độ hội nhập, đổi mới sáng tạo và trách nhiệm xã hội. Tỷ số tài chính là công cụ quan trọng để đánh giá hiệu quả hoạt động, đóng vai trò biến đầu vào cho các phương pháp định lượng nhằm đo lường sức khỏe tài chính và hiệu suất vận hành doanh nghiệp (Gajdosikova et al., 2024). Chúng cũng hỗ trợ so sánh hiệu quả giữa các doanh nghiệp cùng ngành và theo dõi biến động hiệu suất theo thời gian.

Giai đoạn 2012-2023, ngành xây dựng Việt Nam duy trì tăng trưởng với CAGR 11% nhờ quá trình công nghiệp hóa và đô thị hóa. Năm 2023, ngành đạt hơn 640,72 nghìn tỷ đồng, tương đương 6,27% GDP; trong 6 tháng đầu năm 2025, ngành xây dựng ghi nhận mức tăng trưởng cao nhất trong khu vực công nghiệp & xây dựng: 9,62% so với cùng kỳ. Tuy nhiên, ngành cũng chịu sức ép từ đại dịch và khủng hoảng bất động sản.

Do đó, nghiên cứu này áp dụng phân tích phân cụm để phân loại các doanh nghiệp xây dựng niêm yết giai đoạn 2012-2023. Thay vì dùng trực tiếp các tỷ số tài chính, nghiên cứu lựa chọn dữ liệu đa hợp để phục vụ phân loại.

2. Tổng quan nghiên cứu

Phương pháp phân tích dữ liệu đa hợp được sử dụng nhiều trong các nghiên cứu thực nghiệm trong những năm gần đây, đặc biệt từ những năm 2000 khi các chuyển đổi số liệu đa hợp xuất hiện (Greenacre, 2018). Các ứng dụng của CoDa trong phân tích tài chính bao gồm các báo cáo kế toán của các công ty ngành rượu vang tại Tây Ban Nha. Đặc biệt, thông qua hệ thống tỷ lệ tài chính của phân tích Dupont, nhóm tác giả cho thấy phương pháp CoDa là hữu ích khi thực hiện các phân nhóm doanh nghiệp (Jofre-Campuzano & Coenders, 2022). Trong các nghiên cứu này, dữ liệu đa hợp trong tài chính của các công ty là véctơ đa hợp 6 thành phần (tài sản cố định, tài sản lưu động, nợ dài hạn, nợ ngắn hạn, doanh thu tổng và chi phí hoạt động) (Creixans-Tenas và cộng sự, 2019). Phương pháp CoDa giúp xử lý các tỷ lệ tài chính một cách đối xứng sao cho kết quả không phụ thuộc vào tử số và mẫu số. Phân tích dữ liệu đa hợp cũng giúp làm giảm thiểu các vấn đề liên quan đến giá trị ngoại biên, việc sử dụng quá nhiều tỷ số tài chính, mối quan hệ phi tuyến tính giữa các tỷ số và dữ liệu không phân phối chuẩn (Creixans-Tenas và cộng sự, 2019; Linares-Mustarós và cộng sự, 2018). Trong đó, một vài nghiên cứu đã ứng dụng phân tích dữ liệu đa hợp trong phân nhóm hồ sơ doanh nghiệp, ví dụ như Linares-Mustarós và cộng sự (2018). Phân tích dữ liệu đa hợp, khi so sánh với việc sử dụng các

** Trường Đại học Hà Nội; email: binhdtt@hanu.edu.vn - tuanna@hanu.edu.vn -lhphuong@hanu.edu.vn*

tỷ lệ tài chính căn bản, đem lại các kết quả hoàn toàn khác biệt (Linares-Mustarós và cộng sự, 2018). Tại Việt Nam, Đào Thị Thanh Bình và cộng sự (2022) đã sử dụng phương pháp phân tích số liệu đa hợp trong phân tích mối quan hệ giữa hiệu suất hoạt động và cấu trúc tài chính của doanh nghiệp ngành thực phẩm ở Việt Nam, thông qua véctơ đa hợp 3 thành phần (tỉ lệ nợ ngắn hạn, tỉ lệ nợ dài hạn và tỉ lệ vốn chủ sở hữu).

3. Phương pháp nghiên cứu

3.1. Dữ liệu

Mẫu dữ liệu bao gồm 162 doanh nghiệp niêm yết liên tục trên HOSE, HNX hoặc UpCom trong giai đoạn 12 năm từ 2012 đến 2023, hoạt động trong ngành xây dựng, được thu thập thông qua cơ sở dữ liệu của Refinitiv.

3.2. Dữ liệu đa hợp trong phân tích báo cáo tài chính

Theo nguyên tắc độ đo Aitchison (1986) trong phương pháp phân tích dữ liệu đa hợp, dữ liệu được biểu diễn dưới dạng một vec-tơ đa hợp gồm D dữ liệu đa hợp bao gồm một tập hợp số dương, có giá trị quan trọng đối với các nhà nghiên cứu.

$(x_1, x_2, ..., x_D)$ trong đó $x_j > 0, j = 1,2, ...,D$

Các số liệu kế toán có được từ báo cáo tài chính là các D dữ liệu đa hợp. Do đó, để biểu diễn các số liệu kế toán dưới dạng một vec-tơ đa hợp, các giá trị âm phải được loại trừ (Creixans-Tenas và cộng sự, 2019). Mặc dù một số tỷ số tài chính có thể bao gồm các số liệu kế toán âm, nhưng các nghiên cứu thực nghiệm trong lĩnh vực tài chính không khuyến khích sử dụng các chỉ số này vì có thể gây ra một số vấn đề tiềm ẩn như gián đoạn, giá trị ngoại lai hoặc diễn giải không chính xác trong trường hợp xuất hiện số liệu âm trong mẫu số (Lev & Sunder, 1979).

Nghiên cứu này sử dụng sáu chỉ số tài chính dương: x_1: Tài sản dài hạn, x_2: Tài sản ngắn hạn, x_3: Nợ dài hạn, x_4: Nợ ngắn hạn, x_5: Doanh thu, x_6: Chi phí

Giải thích các biến

- Khả năng thanh toán

• Khả năng thanh toán hiện hành: $\frac{x_2}{x_4}$

• Tỷ số nợ dài hạn trên tổng nợ: $\frac{x_3}{x_3+x_4}$

• Cấu trúc vốn (Tổng nợ trên tổng tài sản): $\frac{x_3+x_4}{x_1+x_2}$

- Phân tích Dupont

• Tỷ lệ vòng quay tổng tài sản: $\frac{x_5}{x_1+x_2}$

• Biên lợi nhuận ròng: $\frac{x_5-x_6}{x_5}$

• Tỷ suất lợi nhuận trên tổng tài sản (ROA): $\frac{x_5-x_6}{x_1+x_2}$

• Hệ số đòn bẩy tài chính: $\frac{x_1+x_2}{(x_1+x_2)-(x_3+x_4)}$

• Tỷ suất lợi nhuận trên vốn chủ sở hữu (ROE): $\frac{x_5-x_6}{(x_1+x_2)-(x_3+x_4)}$

3.3. Phép biến đổi logarit trung tâm có trọng số

CoDa chuyển đổi dữ liệu đa hợp vào không gian Euclid trước khi áp dụng các kỹ thuật thống kê truyền thống, qua đó cho phép sử dụng các phương pháp phân cụm cổ điển. Theo Pawlowsky-Glahn và cộng sự (2015), phép biến đổi logarit trung tâm (CLR) lưu giữ đầy đủ thông tin về ý nghĩa tương đối của các số liệu kế toán. Một đặc điểm quan trọng của CLR là khả năng bảo toàn khoảng cách: khoảng cách Aitchison giữa hai dữ liệu đa hợp bằng đúng khoảng cách Euclid giữa các vectơ sau biến đổi.

Phép biến đổi CLR được định nghĩa như sau:

$$CLR_j = log\left(\frac{x_j}{\sqrt[D]{x_1^{\frac{1}{D}} x_2^{\frac{1}{D}} \cdots x_D^{\frac{1}{D}}}}\right) = log\left(\frac{x_j}{\sqrt[D]{x_1 \cdots x_D}}\right), trong\ đó\ j = 1, \dots, D$$

Mỗi CLR so sánh một số liệu kế toán với giá trị trung bình của toàn bộ các số liệu của doanh nghiệp. Tuy nhiên, một hạn chế của CoDa là các khoản mục có giá trị nhỏ có thể tạo ra phương sai CLR lớn, làm lệch kết quả phân cụm. Vấn đề này có thể được khắc phục bằng cách áp dụng phép biến đổi CLR có trọng số (Jofre-Campuzano & Coenders, 2022).

$$wCLR_j = \sqrt{w_j}\ log\ log\left(\frac{x_j}{x_1^{w1} x_2^{w2} \cdots x_D^{wD}}\right), trong\ đó\ j = 1, \dots, D$$

Đối với quá trình phân cụm, Greenacre (2018) chỉ ra rằng CLR có trọng số được tính toán bằng hàm CLR từ mã easyCODA trong gói R và được xử lý như dữ liệu đầu vào trong phân cụm Ward(1963) bằng hàm WARD. Phương pháp này thường được sử dụng cho dữ liệu báo cáo tài chính trong số các kỹ thuật phân cụm dựa trên khoảng cách Euclid (Linares-Mustarós và cộng sự, 2018). Để xác định số lượng cụm, các biểu đồ scree của tổng bình phương trong cụm đã được kiểm tra (Dolnicar và cộng sự, 2018).

3.4. Các bước phân tích cụm dữ liệu đa hợp có trọng số

Quy trình chúng tôi sử dụng để phân tích cụm có trọng số của báo cáo tài chính dựa trên phương pháp được phác thảo bởi Jofre-Campuzano và Coenders (2022) và Đào Thị Thanh Bình và cộng sự (2024). Trước khi tính toán CLR có trọng số, cần phải quy ước bất kỳ giá trị bằng không nào (Martín-Fernández và cộng sự, 2011). Theo các giả định khác nhau, phương pháp tiếp cận dữ liệu đa hợp cung cấp một kỹ thuật hiệu quả để gán giá trị bằng không. Do đó, chúng tôi đã sử dụng hàm lrEM từ gói R zCompositions (Palarea-Albaladejo và Martín-Fernández, 2008).

Quy trình phân tích gồm 6 bước:

- Làm sạch số liệu

- Thống kê mô tả

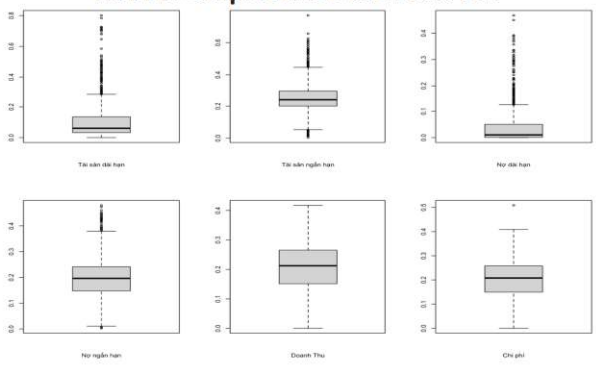
- Tính CLR có trọng số
- Phân tích cụm Ward
- Tính toán chi số tài chính cho mỗi cụm và gán tên
- Phân tích cụm với các biến liên quan

4. Kết quả nghiên cứu và thảo luận

Kết quả cho thấy, các tỷ số tài chính nhìn chung duy trì sự ổn định tương đối trong suốt giai đoạn 2012-2023, không ghi nhận những biến động đột ngột hay xu hướng thay đổi mạnh theo thời gian; đặc biệt các tỷ số có biến động cùng chiều một cách hợp lý, có nghĩa là trong những năm hiệu quả hoạt động tốt thì doanh nghiệp cũng tăng sử dụng đòn bẩy tài chính một cách hợp lý và hiệu quả để đạt được lợi nhuận trên vốn chủ sở hữu ở một mức tốt hơn. Tuy nhiên, một điểm đáng chú ý là chỉ số Khả năng thanh toán hiện hành có sự biến động đáng kể vào năm 2019 - thời điểm đại dịch Covid-19 bắt đầu ảnh hưởng đến hoạt động kinh doanh tại Việt Nam. Diễn biến này cho thấy các doanh nghiệp đã có xu hướng chủ động giảm nợ ngắn hạn nhằm bảo đảm khả năng thanh toán trong bối cảnh bất định, khi tỷ lệ giữa tài sản ngắn hạn và nợ ngắn hạn được điều chỉnh theo hướng an toàn hơn.

Hình 1 minh họa phân bố của sáu biến tài chính cơ bản dưới dạng biểu đồ box plot. Các biến được chuẩn hóa theo tổng tài sản để so sánh giữa các doanh nghiệp có quy mô khác nhau. Biểu đồ boxplot cung cấp thông tin trực quan về trung vị, khoảng tứ phân vị, cũng như các giá trị ngoại biên, giúp đánh giá mức độ phân tán và phân phối dữ liệu của từng biến.

Hình 1: Boxplot cho 6 biến đầu vào



Nguồn: Tác giả

Tài sản dài hạn: Phân bố có trung vị thấp (xấp xỉ 0,03), với phần lớn quan sát nằm trong khoảng từ 0,02 đến 0,10. Tuy nhiên, phần trên của hộp khá rộng, cho thấy sự phân tán đáng kể về tỷ trọng tài sản dài hạn giữa các doanh nghiệp. Tài sản ngắn hạn: Biến này có trung vị khoảng 0,25, và phân bố đối xứng với khoảng biến thiên tập trung hơn so với các biến khác. Đây là yếu tố chính tạo nên tính thanh khoản ngắn hạn của doanh nghiệp. Nợ dài hạn: Trung vị rất thấp, gần 0.005, cho thấy đa phần các doanh nghiệp có tỷ trọng nợ dài hạn trên tổng tài sản thấp. Khoảng tứ phân vị rộng phía trên cho thấy một số doanh nghiệp vẫn sử dụng đòn bẩy dài hạn ở mức cao hơn. Nợ ngắn hạn: Biến này có phân bố hẹp hơn, trung vị khoảng 0.19, phản ánh mức độ phổ biến hơn trong việc sử dụng nợ ngắn hạn thay vì nợ dài hạn trong cấu trúc vốn của doanh nghiệp Việt Nam. Doanh thu: Trung vị rơi vào khoảng 0.21, với dữ liệu tương đối cân bằng; thể hiện sự cân bằng tương đối trong hiệu quả tạo doanh thu trên tài sản sẵn có. Chi phí: Biến chi phí có trung vị gần 0.21 và phân bố khá đồng đều, với khoảng tứ phân vị không quá rộng. Dữ liệu cho thấy sự đồng nhất tương đối trong cấu trúc chi phí của các doanh nghiệp trong mẫu.

Tổng thể, biểu đồ cho thấy có sự khác biệt rõ rệt về mức độ phân tán và cấu trúc phân phối giữa các biến. Các biến liên quan đến tài sản và nợ có xu hướng phân phối tập trung hơn, trong khi doanh thu và chi phí thể hiện mức độ phân tán lớn hơn. Điều này hàm ý rằng các doanh nghiệp trong mẫu có sự đa dạng rõ rệt về cấu trúc tài chính và hiệu quả hoạt động, phản ánh đặc trưng của ngành với quy mô khác nhau.

Phân tích phân cụm phân cấp cho thấy cấu trúc dữ liệu được chia thành ba cụm rõ rệt. Sử dụng biểu đồ Screeplot minh họa điểm gãy tại cụm thứ ba, củng cố cơ sở cho việc lựa chọn ba nhóm làm nền tảng cho các phân tích chuyên sâu tiếp theo.

Bảng 1: Đặc điểm các nhóm đã được phân tách

Các chỉ số tài chính	Nhóm 1	Nhóm 2	Nhóm 3
	276	415	874
Khả năng thanh toán hiện hành	2.39	1.44	1.56
Tỷ số nợ dài hạn trên tổng nợ	-	0.36	0.07
Hệ số đòn bẩy tài chính	7.02	5.05	5.75
Tỷ lệ vòng quay tổng tài sản	0.87	0.40	0.82
Biên lợi nhuận ròng	2.46%	3.59%	1.03%
Tỷ suất lợi nhuận trên tổng tài sản (ROA)	2.01%	2.01%	1.70%
Tỷ suất lợi nhuận trên vốn chủ sở hữu (ROE)	25.89%	7.78%	-2.71%

Nguồn: Tác giả

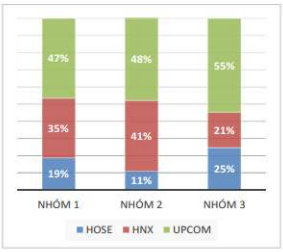
Kết quả phân nhóm cho thấy ba nhóm doanh nghiệp khác biệt, với 276 quan sát ở nhóm 1, 415 ở nhóm 2 và 874 ở nhóm 3. Nhóm 1 đạt ROE cao nhất, tiếp đến là nhóm 2, trong khi nhóm 3 có kết quả kém nhất. Sự phân hóa này phản ánh khác biệt về chiến lược tài chính, hiệu quả hoạt động và khả năng quản trị rủi ro.

Nhóm 1 không chỉ vượt trội về ROE mà còn có ROA cao, cho thấy kết hợp hiệu quả giữa biên lợi nhuận và khả năng quay vòng tài sản. Nhóm này cũng sử dụng đòn bẩy tài chính ở mức cao nhất nhưng theo hướng hợp lý, nhờ nền tảng hoạt động vững vàng, giúp khuếch đại lợi nhuận cho cổ đông. Đồng thời,

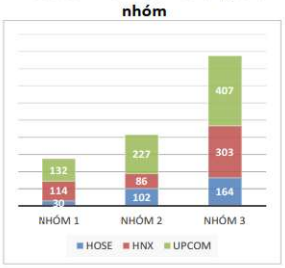
khả năng thanh toán hiện hành cao và ổn định phản ánh quản trị vốn lưu động an toàn.

Ngược lại, nhóm 3 có ROE âm, cho thấy nhiều hạn chế trong hiệu quả tài chính. Dù ROA chỉ thấp hơn nhóm 1 một phần, biên lợi nhuận rất thấp và đòn bẩy cao khiến các doanh nghiệp nhóm này không chuyển hóa được tài sản và vốn vay thành lợi nhuận. Khi hiệu quả hoạt động yếu, việc sử dụng nợ nhiều làm ROE giảm mạnh, minh họa rủi ro tài chính khi quản trị kém: đòn bẩy trở thành gánh nặng thay vì tạo giá trị.

Hình 2: Tỉ lệ niêm yết theo sàn trong 3 nhóm



Hình 3: Số lượng quan sát phân loại theo sàn niêm yết và trong các nhóm



Nguồn: Tác giả

Từ kết quả trên, có thể rút ra kết luận sơ bộ rằng các bộ chỉ số tài chính được sử dụng đều phản ánh hợp lý chất lượng hoạt động của doanh nghiệp, đồng thời khẳng định tính hiệu quả của phương pháp phân nhóm. Một doanh nghiệp muốn tối ưu hóa lợi nhuận cho cổ đông phải xây dựng nền tảng kinh doanh bền vững (thể hiện qua ROA), đồng thời kết hợp với chiến lược quản lý rủi ro tài chính phù hợp. Nếu hai yếu tố này được thực thi thành công, lợi ích mang lại sẽ rất vượt trội, như minh chứng ở nhóm 1 với ROE trung bình đạt 25,89%, so với mức âm ở nhóm 3.

Phân tích cơ cấu quan sát cho thấy nhóm 1 - nhóm hoạt động tốt nhất - có quy mô nhỏ nhất với 276 quan sát, trong khi nhóm 3 có 874 quan sát, gần gấp ba lần. Tỷ lệ doanh nghiệp UPCOM duy trì ổn định quanh mức 50% ở cả ba nhóm. Khác biệt chủ yếu nằm ở sàn HOSE: nhóm 3 có 25% quan sát từ HOSE, trong khi nhóm 1 chỉ 19% (30 quan sát). Điều này phần nào do tỷ trọng doanh nghiệp bất động sản trên HOSE trong toàn bộ mẫu chỉ khoảng 20% trên 1.566 quan sát, khiến số quan sát tốt thuộc HOSE bị hạn chế. Nhìn chung, không có sự phân hóa đáng kể về chất lượng giữa các sàn, khi doanh nghiệp tốt phân bố khá đồng đều trên HOSE, HNX và UPCOM; chỉ nhóm 2 có tỷ lệ doanh nghiệp HOSE thấp hơn đáng kể.

5. Kết luận và khuyến nghị

Nghiên cứu khẳng định tính hữu ích của phương pháp phân tích dữ liệu đa hợp (CoDa) kết hợp phân cụm trong nhận diện cấu trúc tài chính của các doanh nghiệp xây dựng niêm yết tại Việt Nam giai đoạn 2012-2023. Việc sử dụng dữ liệu đa hợp giúp khắc phục các hạn chế của tỷ số tài chính truyền thống, cho phép đánh giá toàn diện hơn mối quan hệ giữa tài sản, nợ, doanh thu và chi phí. Kết quả phân tách thành ba cụm đặc trưng, phản ánh sự khác biệt rõ rệt về hiệu quả hoạt động và mức độ rủi ro. Nhóm 1 là nhóm có khả năng sinh lời vượt trội, với ROE và ROA cao, tỷ lệ đòn bẩy được sử dụng hợp lý và khả năng thanh khoản tốt. Ngược lại, nhóm 3 hiệu quả thấp, biên lợi nhuận thấp và mức độ phụ thuộc vào nợ cao, dẫn đến ROE âm. Những kết quả này cho thấy sự phân hóa bắt nguồn từ nội lực doanh nghiệp, chứ không đến từ sự khác biệt giữa các sàn giao dịch, khi chất lượng doanh nghiệp phân bố tương đối đồng đều ở HOSE, HNX và UPCOM.

Nghiên cứu đề xuất một số khuyến nghị nhằm nâng cao hiệu quả quản trị và giảm thiểu rủi ro cho doanh nghiệp. Các doanh nghiệp thuộc nhóm hiệu quả thấp cần ưu tiên cải thiện biên lợi nhuận thông qua tái cấu trúc chi phí, tối ưu hóa sử dụng tài sản và lựa chọn dự án có khả năng sinh lời cao hơn. Doanh nghiệp cần quản lý đòn bẩy tài chính một cách thận trọng, nợ chỉ tạo giá trị khi doanh nghiệp duy trì nền tảng hoạt động vững chắc; tăng nợ trong điều kiện lợi nhuận yếu có thể làm suy giảm mạnh ROE. Quản trị thanh khoản cần được chú trọng hơn, đặc biệt sau các biến động như giai đoạn 2019 khi doanh nghiệp buộc phải điều chỉnh cấu trúc nợ ngắn hạn để ứng phó rủi ro. Doanh nghiệp và cơ quan quản lý có thể ứng dụng phân cụm CoDa để xây dựng hệ thống cảnh báo sớm, theo dõi sự dịch chuyển của doanh nghiệp giữa các cụm nhằm nhận diện xu hướng suy giảm hoặc cải thiện tài chính. Nhà đầu tư có thể sử dụng kết quả phân cụm như một công cụ hỗ trợ ra quyết định, dựa trên hồ sơ rủi ro - lợi nhuận đặc trưng của từng nhóm doanh nghiệp.

Lời cảm ơn: *Nghiên cứu này được tài trợ bởi Bộ Giáo dục và Đào tạo Việt Nam, trong khuôn khổ đề tài nghiên cứu khoa học cấp Bộ, Mã số B2024-NHF-04, theo Quyết định số 1393/QĐ-BGDĐT ngày 16 tháng 5 năm 2023.*

Tài liệu tham khảo:

Aitchison, J. (1986). The Statistical Analysis of Compositional Data (Monographs on Statistics and Applied Probability) (1st ed.). Springer.

Creixans-Tenas, J., Coenders, G., & Arimany-Serrat, N. (2019). Corporate social responsibility and financial profile of Spanish private hospitals. Heliyon, 5(10), e02623. https://doi.org/10.1016/j.heliyon.2019.e02623

Dao, B. T. T., Coenders, G., Lai, P. H., Dam, T. T. T., & Trinh, H. T. (2024). An empirical examination of financial performance and distress profiles during COVID-19: The case of fishery and food production firms in Vietnam. Journal of Financial Reporting and Accounting.

Dolnicar, S., Grün, B. and Leisch, F. (2018), Market Segmentation Analysis: Understanding It, Doing It, and Making It Useful, Springer Nature, Singapore.

Gajdosikova, D., & Vojtekova, S. (2024). Comparative Analysis of Business Environment Dynamics in Central and Eastern Europe: A Multi-Criteria Approach. Economies, 12(12), 320.

Greenacre, M. (2018). Compositional Data Analysis in Practice (1st ed.). New York: Chapman and Hall/CRC. https://doi.org/10.1201/9780429455537

Jofre-Campuzano, P., & Coenders, G. (2022). Compositional classification of financial statement profiles: The weighted case. Journal of Risk and Financial Management, 15(12), 546. https://doi.org/10.3390/JRFM15120546

Lev, B., & Sunder, S. (1979). Methodoiogical issues in the use of financial ratios. Journal of Accounting and Economics, 1(3), 187-210. https://doi.org/10.1016/0165-4101(79)90007-7

Linares-Mustarós, S., Coenders, G., & Vives-Mestres, M. (2018). Financial performance and distress profiles. From classification according to financial ratios to compositional classification. Advances in Accounting, 40, 1-10. https://doi.org/10.1016/J.ADIAC.2017.10.003

Martín-Fernández, J.A., Palarea-Albaladejo, J. and Olea, R.A. (2011), "Dealing with zeros", in Pawlowsky-Glahn, V. and Buccianti, A. (Eds.), Compositional Data Analysis: Theory and Applications, John Wiley & Sons, Chichester, pp. 43-58.

Palarea-Albaladejo, J., & Martín-Fernández, J. A. (2008). A modified EM alr-algorithm for replacing rounded zeros in compositional data sets. Computers & Geosciences, 34(8), 902-917. https://doi.org/10.1016/J.CAGEO.2007.09.015

Pawlowsky-Glahn, V., Egozcue, J.J. and Tolosana-Delgado, R. (2015), Modeling and Analysis of Compositional Data, John Wiley & Sons, Chichester.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. Journal of the American statistical association, 58(301), 236-244.