

Sự vận động trong cấu trúc lợi thế so sánh của Việt Nam: Phân cụm RSCA ở cấp độ HS2 giai đoạn 2000- 2023

An Như Hưng

Học viện Ngân hàng, Việt Nam

Ngày nhận: 23/12/2025

Ngày nhận bản sửa: 25/03/2026

Ngày duyệt đăng: 27/03/2026

Tóm tắt: Bài viết phân tích sự vận động trong cấu trúc lợi thế so sánh của Việt Nam ở cấp độ ngành HS 2 chữ số giai đoạn 2000- 2023. Sử dụng phân cụm K-means trên véc-tơ đặc trưng của sự vận động về mức lợi thế, độ biến động, độ bền, độ bất ổn và xu hướng, các ngành HS 2 chữ số được phân thành bốn cụm có mẫu hình vận động tương đối đồng nhất. Trên nền phân cụm, bài viết mô tả vai trò của từng cụm trong cơ cấu xuất khẩu và đóng góp vào tăng/giảm xuất khẩu qua thời gian, đồng thời đo lường mức tập trung nội bộ cụm. Ma trận chuyển trạng thái Markov được sử dụng sau đó để lượng hóa xác suất duy trì, hình thành và đánh mất lợi thế theo các giai đoạn. Kết quả cho thấy động lực tăng trưởng xuất khẩu dịch chuyển theo thời gian giữa các cụm, trong khi mức tập trung nội bộ khác biệt đáng kể. Từ đó, nghiên cứu hàm ý cách tiếp

Dynamics of Viet Nam's comparative advantage structure: RSCA Clustering at the HS2 Level, 2000- 2023

Abstract: This paper examines the dynamics of Viet Nam's comparative advantage structure at the two-digit HS (HS2) level over 2000-2023. Applying K-means clustering on a feature vector that summarizes each HS2 industry's dynamics (advantage strength, volatility, persistence, instability, and trend), the HS2 structure are classified into four clusters with relatively homogeneous patterns. Building on this classification, the paper characterizes each cluster's role in the export structure and its contribution to export growth over time, and it further measures within-cluster concentration. A Markov transition framework is then employed to quantify the probabilities of maintaining, gaining, and losing comparative advantage across sub-periods. The results indicate that the drivers of export growth have shifted across clusters over time and that within-cluster concentration differs markedly, implying a risk-growth portfolio-management approach to policy design, in which policy packages are tailored to clustered-specific characteristics so as to harness growth drivers while reducing concentration and dependency risks.

Keywords: Comparative advantage, K-means, Markov, Export structure, Portfolio management, Viet Nam

Link Doi: <https://doi.org/10.59276/JELB.2026.04.3139>

An Nhu Hung

Email: hungan@hvn.edu.vn

Organization: Banking Academy of Vietnam

cận thiết kế chính sách theo hướng quản trị danh mục cân bằng rủi ro-tăng trưởng, trong đó gói chính sách cần được điều chỉnh theo đặc thù từng cụm để vừa khai thác động lực tăng trưởng, vừa giảm rủi ro tập trung và phụ thuộc.

Từ khóa: Lợi thế so sánh, K-means, Markov, Cơ cấu xuất khẩu, Quản trị danh mục, Việt Nam

Trích dẫn: An Như Hưng. (2026). Sự vận động trong cấu trúc lợi thế so sánh của Việt Nam: Phân cụm RSCA ở cấp độ HS2 giai đoạn 2000- 2023. *Tạp chí Kinh tế - Luật và Ngân hàng*, 28(4), 74-89. <https://doi.org/10.59276/JELB.2026.04.3139>

1. Giới thiệu

Trong hai thập niên qua, Việt Nam đạt tốc độ tăng trưởng xuất khẩu nhanh và hội nhập sâu vào thương mại quốc tế. Tuy nhiên, đánh giá thành công xuất khẩu không nên dừng ở quy mô kim ngạch, mà cần chuyển sang một câu hỏi có ý nghĩa chính sách hơn: chất lượng của tăng trưởng xuất khẩu. Cụ thể, tăng trưởng đó chủ yếu dựa trên những lợi thế so sánh (LTSS) có tính bền vững và được củng cố theo thời gian, hay dựa nhiều vào những lợi thế dao động mạnh và dễ đảo chiều khi điều kiện thị trường, tiêu chuẩn, chu kỳ công nghệ hoặc các cú sốc bên ngoài thay đổi.

Trong bối cảnh Việt Nam, nhận thức phổ biến thường nhấn mạnh sự nổi lên của một số trụ cột xuất khẩu (chẳng hạn các nhóm hàng chế biến chế tạo gắn với khu vực FDI) song hành với các nhóm hàng truyền thống. Vấn đề là, các mô tả trực giác như vậy, dù có sức thuyết phục, thường khó chuyển hóa thành những hàm ý chính sách thiết thực nếu thiếu một cách đo lường có hệ thống về (i) mức độ ổn định của LTSS theo thời gian, (ii) mức độ bền vững của LTSS mới, và (iii) vai trò thực tế của các cụm LTSS khác nhau trong việc tạo ra tăng trưởng xuất khẩu. Nói cách khác, cần đặt song song hai lát cắt: lát cắt tĩnh (cấu trúc LTSS ở một thời điểm) và lát cắt động (cách LTSS vận động, duy trì hay biến mất

theo thời gian).

Bài viết đề xuất một cách tiếp cận tổng hợp để trả lời trực diện các vấn đề này ở cấp độ ngành HS 2 chữ số (HS2) cho Việt Nam giai đoạn 2000- 2023. LTSS được đo lường bằng chỉ số LTSS hiển thị (Revealed Comparative Advantage- RCA) theo Balassa (1965), đồng thời sử dụng biến đổi đối xứng Symmetric RCA (RSCA) để hỗ trợ cho mô tả phân phối và so sánh động theo thời gian (Laursen, 1998). Trên nền chuỗi trạng thái có/không có LTSS, bài viết ước lượng ma trận chuyển trạng thái (Transition Probability Matrix- TPM) để lượng hóa xác suất duy trì, gia nhập, và rời khỏi LTSS, cách tiếp cận đã được sử dụng trong một số nghiên cứu về sự vận động trong cấu trúc LTSS (Doanh & Heo, 2011; Widodo, 2009). Trên cơ sở đó, đặc trưng vận động của RSCA từng ngành được tóm tắt và phân cụm bằng phương pháp K-means để quy chiếu danh mục HS2 lớn thành một số cụm điển hình (MacQueen, 1967). Trên nền phân cụm, bài viết nối kết ba lớp thông tin: (i) độ bền/mong manh của LTSS theo cụm; (ii) cơ cấu và đóng góp của từng cụm vào thay đổi kim ngạch xuất khẩu theo thời gian; và (iii) mức tập trung nội bộ trong từng cụm như một chỉ báo rủi ro phụ thuộc.

Từ thiết kế trên, bài viết tập trung trả lời ba câu hỏi nghiên cứu: (1) Xác suất duy trì/đánh mất và hình thành LTSS của các

ngành HS2 biến thiên như thế nào giữa các giai đoạn? (2) Khi quy chiếu HS2 thành các cụm, cơ cấu và động lực tăng trưởng xuất khẩu chủ yếu thuộc về các cụm nào theo thời gian? (3) Mức độ tập trung nội bộ trong từng cụm thay đổi ra sao, và điều đó gợi ý gì về rủi ro phụ thuộc vào một/một vài trụ cột trong cấu trúc xuất khẩu?

Đóng góp của bài viết không ở việc liệt kê ngành có LTSS mà theo hướng hỗ trợ thảo luận chính sách, ở đó phân biệt LTSS bền vững với LTSS mong manh, chỉ ra cụm LTSS đang dẫn dắt tăng trưởng xuất khẩu, và nhận diện rủi ro cấu trúc thông qua mức tập trung nội bộ theo cụm. Kết quả nghiên cứu cung cấp hàm ý thiết thực cho cơ quan điều hành theo hướng quản trị danh mục, vừa khai thác các trụ cột tăng trưởng vừa giảm rủi ro phụ thuộc.

Phần còn lại của bài viết được tổ chức như sau. Mục 2 tổng quan nghiên cứu và định vị khoảng trống. Mục 3 trình bày dữ liệu, cách đo lường RCA/RSCA và thiết kế thực nghiệm. Mục 4 báo cáo kết quả. Mục 5 thảo luận hàm ý chính sách và các hạn chế. Mục 6 kết luận.

2. Tổng quan nghiên cứu và khoảng trống

2.1. Cơ cấu xuất khẩu và chất lượng tăng trưởng: Từ quy mô sang cấu trúc và khả năng chống chịu

Một nhánh nghiên cứu cho rằng cơ cấu xuất khẩu không chỉ là kết quả của tăng trưởng, mà còn phản ánh năng lực sản xuất/công nghệ và gắn với triển vọng tăng trưởng dài hạn. Theo hướng này, những gì quốc gia xuất khẩu hàm chứa thông tin về mức độ tinh vi và hàm lượng năng lực (capabilities) của nền kinh tế, từ đó ảnh hưởng đến khả năng nâng cấp và đuổi kịp (Hausmann và cộng sự, 2007; Hidalgo & Hausmann, 2009). Theo nghĩa đó, đánh giá

thành công xuất khẩu về mặt chính sách không thể dừng ở quy mô, mà cần xét tới chất lượng chuyển dịch cơ cấu, nền kinh tế đang dịch chuyển về phía danh mục xuất khẩu có khả năng tạo năng suất, lan tỏa, và nâng cấp hay không.

Một nhánh khác nhấn mạnh đa dạng hóa và tập trung hóa như hai mặt của tiến trình phát triển. Bằng chứng thực nghiệm cho thấy nhiều quốc gia có xu hướng mở rộng đa dạng hóa xuất khẩu khi còn ở mức thu nhập thấp và trung bình, song tái tập trung ở mức phát triển cao hơn (Imbs & Wacziarg, 2003; Cadot và cộng sự, 2011). Điều này gợi ý chuyển dịch cơ cấu không nên được hiểu đơn giản là “đa dạng hóa càng nhiều càng tốt”, mà là quá trình cấu trúc danh mục vận động theo giai đoạn, theo năng lực và theo ràng buộc thị trường. Nhánh thứ ba chú trọng vào rủi ro cấu trúc gắn với tập trung danh mục. Khi danh mục xuất khẩu quá phụ thuộc vào số ít nhóm hàng/chuỗi giá trị, một cú sốc nhỏ trong nhóm hàng/chuỗi giá trị đó cũng có thể khuếch đại thành biến động vĩ mô lớn hơn. Bằng chứng thực nghiệm cho thấy các nền kinh tế chuyên môn hóa thường dễ biến động hơn (Koren & Tenreyro, 2007). Nghiên cứu gần đây cũng chỉ ra rằng sự phụ thuộc trong phạm vi hẹp các nhà cung ứng làm tăng tính dễ tổn thương và lan truyền cú sốc (Berthou và cộng sự, 2024). Từ góc nhìn điều hành, điều này hàm ý rằng phân tích cơ cấu xuất khẩu cần đi kèm một cách nhìn về khả năng chống chịu (resilience), thay vì chỉ là “đang xuất cái gì và bao nhiêu”.

Trong bài viết này, chất lượng chuyển dịch cơ cấu được hiểu theo nghĩa hẹp và có thể đo lường được, gồm độ bền của LTSS, vai trò của từng cụm trong động lực tăng trưởng xuất khẩu, và mức độ tập trung nội bộ như một chỉ báo rủi ro phụ thuộc.

2.2. Lợi thế so sánh trong nghiên cứu thực nghiệm

Nghiên cứu thực nghiệm về LTSS có thể được phân biệt theo hai hướng lớn. Thứ nhất là tiếp cận tĩnh, thường tập trung vào việc xác định nhóm hàng có LTSS tại một thời điểm hoặc theo trung bình giai đoạn, từ đó mô tả cấu trúc chuyên môn hóa và so sánh giữa các nhóm ngành/quốc gia. Cách tiếp cận này có ưu điểm đơn giản, song bỏ qua nhiều vấn đề như có LTSS nhưng đang suy giảm, có LTSS nhưng biến động mạnh, hoặc vừa đạt LTSS nhưng mong manh. Thứ hai là tiếp cận động, nhấn mạnh việc LTSS được hình thành, duy trì hay mất đi như thế nào theo thời gian. Hướng nghiên cứu này sử dụng mô tả về độ bền (persistence), khuynh hướng hội tụ/phân tán, và đặc biệt là TPM trong khuôn khổ chuỗi Markov để lượng hóa xác suất ở lại/gia nhập/rời khỏi trạng thái LTSS qua thời gian (Doanh & Heo, 2011; Widodo, 2009). Điểm mạnh của tiếp cận động là tạo ra căn cứ định lượng cho mức độ bền vững hay mong manh của LTSS, nội dung có giá trị chính sách cao hơn so với việc chỉ liệt kê nhóm hàng có LTSS.

Tuy vậy, ngay cả khi dùng tiếp cận động, một khó khăn thường gặp là kết quả dễ bị phân mảnh ở cấp mã hàng. Khi danh mục bao gồm nhiều nhóm ngành, việc phân tích các chuỗi động học riêng lẻ thường dẫn đến mô tả dài và phân tán, gây khó khăn trong việc tổng hợp thành các kết luận mang tính cấu trúc, đặc biệt khi chuyển hóa thành hàm ý chính sách.

2.3. Từ động học sang cấu trúc

Nếu mục tiêu là đánh giá chất lượng chuyển dịch cơ cấu xuất khẩu, cần một cách chuyển hóa thông tin động học của từng nhóm hàng sang một câu chuyện cấu trúc. Câu

chuyện này thường xoay quanh ba lớp nội dung: (i) LTSS nào bền, lợi thế nào mới nổi nhưng có xu hướng trưởng thành, và lợi thế nào mang tính mong manh/dễ đảo chiều. (ii) Vai trò của từng cụm trong cơ cấu và động lực tăng trưởng xuất khẩu. (iii) Mức độ tập trung nội bộ của từng cụm như một chỉ báo rủi ro phụ thuộc. Ngay cả khi một cụm đóng vai trò trụ cột tăng trưởng, cấu trúc bên trong cụm có thể hẹp (phụ thuộc vào ít nhóm hàng) hoặc rộng (chia sẻ bởi nhiều nhóm hàng). Khía cạnh này kết nối trực tiếp với tranh luận về chuyên môn hóa- đa dạng hóa và khả năng chống chịu (Koren & Tenreyro, 2007; Berthou và cộng sự, 2024).

Tóm lại, các mô tả động học hiện vẫn thiếu một bước tổng hợp có hệ thống để quy chiếu danh mục lớn thành một số cụm điển hình, từ đó liên hệ sang cơ cấu xuất khẩu, động lực tăng trưởng và rủi ro cấu trúc.

2.4. Khoảng trống thực nghiệm

Tổng hợp lại, hai nhận định được rút ra để định vị bài viết. Thứ nhất, các nghiên cứu dựa trên RCA/RSCA đã cung cấp nền tảng vững chắc về đo lường LTSS (Balassa, 1965; Laursen, 2015) và một số nghiên cứu đã đi theo hướng động học với Markov/TPM để lượng hóa sự dịch chuyển LTSS (Doanh & Heo, 2011; Widodo, 2009). Thứ hai, và quan trọng hơn với mục tiêu thảo luận chính sách, thiếu một khung thực nghiệm có thể (i) tổng hợp và phân loại mẫu hình vận động LTSS; (ii) gắn các mẫu hình vận động đó với cơ cấu và động lực tăng trưởng xuất khẩu theo giai đoạn; (iii) bổ sung thước đo trực quan về rủi ro tập trung nội bộ như một thành phần của chất lượng chuyển dịch cơ cấu.

Theo hướng này, bài viết được định vị như một nỗ lực tái tổ chức câu chuyện LTSS theo hướng gắn kết động học, cơ cấu và

rủi ro phụ thuộc. Khái niệm “quản trị danh mục” ở đây chỉ được hiểu theo trục giác đánh đổi giữa lợi ích kỳ vọng và rủi ro, nhấn mạnh vai trò của đa dạng hóa, chứ không hàm ý áp dụng Lý thuyết Danh mục đầu tư Hiện đại (Markowitz, 1952).

3. Phương pháp nghiên cứu

3.1. Dữ liệu và tiền xử lý

Dữ liệu xuất khẩu được lấy từ UNComtrade cho 95 ngành HS2 (không bao gồm ngành 93- Đạn dược và vũ khí) theo từng năm, giai đoạn 2000- 2023. Đơn vị phân tích là cặp (HS2, năm) của Việt Nam với thị trường đối tác là *World*. Giai đoạn nghiên cứu được chia thành bốn giai đoạn nhỏ 2000- 2007, 2008- 2013, 2014-2019, và 2020- 2023. Cách chia này nhằm phản ánh các bối cảnh vĩ mô khác nhau, đồng thời đảm bảo độ dài đủ để ước lượng TPM và các thước đo theo cụm một cách ổn định.

3.2. Đo lường lợi thế so sánh

Ký hiệu $x_{i,t}$ là kim ngạch xuất khẩu của Việt Nam cho ngành HS2 i và năm t , $X_t = \sum_i x_{i,t}$ là tổng xuất khẩu của Việt Nam trong năm t ; tương tự $x_{i,t}^{\{W\}}$ và $X_t^{\{W\}}$ là kim ngạch và tổng xuất khẩu của thế giới. Chỉ số (Balassa, 1965) cho ngành i trong năm t được tính như sau:

$$RCA_{i,t} = (x_{i,t}/X_t) / (x_{i,t}^{\{W\}}/X_t^{\{W\}})$$

Để thuận tiện cho phân tích sự vận động và phân cụm, chỉ số được biến đổi sang chỉ số cân xứng RSCA (Laursen, 2015):

$$RSCA_{i,t} = ((RCA_{i,t} - 1) / (RCA_{i,t} + 1)) \in [-1, 1]$$

Khi $RSCA_{i,t} > 0$ tương ứng $RCA_{i,t} > 1$, tức Việt Nam có LTSS đối với nhóm hàng i tại năm t .

3.3. Xây dựng véc-tơ đặc trưng cho sự vận động trong lợi thế so sánh

Nghiên cứu sử dụng cách tiếp cận dựa trên đặc trưng (feature-based, hay characteristic-based) để tóm tắt mỗi chuỗi RSCA thành một véc-tơ đặc trưng ngắn nhưng giàu tính diễn giải. Cách tiếp cận này phổ biến trong phân cụm chuỗi thời gian khi mục tiêu là vừa giảm chiều dữ liệu vừa tăng khả năng diễn giải kinh tế (Wang và cộng sự, 2006; Fulcher & Jones, 2014). Trên tinh thần đó, bộ đặc trưng được lựa chọn bám sát các chiều kinh tế then chốt của LTSS: mức LTSS trung bình, độ biến động, độ bền, độ bất ổn, và xu thế dài hạn. Cụ thể, với mỗi ngành, các chỉ báo sau được tính toán (ký hiệu là số năm trong khung thời gian):

Mức LTSS trung bình phản ánh độ mạnh trung bình của LTSS (hoặc không có LTSS) trong toàn kỳ, tách khỏi dao động ngắn hạn.

$$\mu_i = 1/T \sum_t RSCA_{i,t}$$

Độ biến động đo mức độ dao động của LTSS theo thời gian.

$$\sigma_i = \sqrt{(1/(T-1)) \sum_t (RSCA_{i,t} - \mu_i)^2}$$

Độ bền của LTSS đo thời gian (năm) một ngành nằm trong vùng có LTSS trong toàn kỳ quan sát.

$$Dur_i = \sum_t I(RSCA_{i,t} > 0)$$

Độ bất ổn phản ánh sự đổi trạng thái và rủi ro mong manh của LTSS. Chỉ số s_i ; giá trị càng cao cho thấy LTSS của ngành có xu hướng đảo trạng thái thường xuyên quanh ngưỡng.

$$s_i = 1/(T-1) \sum_t I(\text{sign}(RSCA_{i,t}) \neq \text{sign}(RSCA_{i,t-1}))$$

Xu thế phản ánh hướng dịch chuyển dài hạn của LTSS (củng cố hay xói mòn), hữu ích để phân biệt LTSS mới nổi nhưng dần củng cố với LTSS từng mạnh nhưng suy giảm. Ước lượng hệ số β_i trong hồi quy

tuyến tính đơn:

$$RSCA_{i,t} = \alpha_i + \beta_i \cdot t + \varepsilon_{i,t}$$

Hệ số β_i phản ánh ngành có xu hướng tăng cường hay suy giảm lợi thế theo thời gian. Như vậy, mỗi ngành được đại diện bởi vector đặc trưng:

$$X_i = (\mu_p, \sigma_p, p_p, s_p, \beta_i)$$

Chuẩn hoá biến: Do các chỉ báo có thang đo khác nhau, trước khi phân cụm, từng thành phần của được chuẩn hoá theo z-score trên toàn bộ HS2 nhằm tránh đặc trưng có thang đo/độ phân tán lớn chi phối khoảng cách trong k-means (MacQueen, 1967).

$$Z_{i,j} = (X_{i,j} - \bar{X}_j) / (sd(X_j))$$

Tập vector Z_i là đầu vào trực tiếp cho bước phân cụm. Trong đó, j là chỉ số của từng thành phần đặc trưng trong véc-tơ (X_i).

3.4. Phân cụm K-means và lựa chọn số cụm K

3.4.1. Thuật toán K-means

Nghiên cứu áp dụng phân cụm K-means trên véc-tơ đặc trưng chuẩn hoá. Thuật toán này tìm cách tối thiểu hoá tổng bình phương khoảng cách trong cụm (within-cluster sum of squares, WSS):

$$\min_{\{C_1, \dots, C_k\}} \sum_{k=1}^K \sum_{i \in C_k} \|Z_i - \mu_k\|^2$$

trong đó μ_k là trung bình cụm k , và $\|\cdot\|$ là khoảng cách Euclid. Để giảm rủi ro hội tụ vào nghiệm cục bộ, K-means được chạy với nhiều khởi tạo ngẫu nhiên (multiple random starts); nghiệm có WSS nhỏ nhất được chọn làm phân cụm cuối cùng. Trong nghiên cứu này, phân cụm được thực hiện trên giai đoạn đủ dài 2000- 2019 để phản ánh tính cấu trúc của LTSS theo HS2, đồng thời tránh việc đặc trưng cụm bị chi phối bởi các biến động ngắn hạn của giai đoạn sau năm 2019 với cú sốc Covid-19 và các bất định hậu Covid. Việc phân cụm trên giai đoạn 2000- 2019 vì vậy được sử dụng để nhận diện cấu trúc dài hạn của các mẫu

hình RSCA. Theo nghĩa đó, tên cụm trong bài viết nên được hiểu là một phân loại mang tính cấu trúc để hỗ trợ diễn giải, chứ không hàm ý rằng bản chất động học của từng ngành là bất biến tuyệt đối qua mọi giai đoạn.

3.4.2. Chọn K

Nghiên cứu lựa chọn giá trị dựa trên các tiêu chí chuẩn trong phân cụm: (i) Elbow của WSS: quan sát mức giảm WSS khi tăng, chọn tại điểm gãy (marginal gain giảm mạnh) (MacQueen, 1967); (ii) Silhouette trung bình: ưu tiên cho silhouette cao hơn (cụm tách biệt tốt hơn) (Rousseeuw, 1987); (iii) Tính diễn giải và kích thước cụm: tránh cụm quá nhỏ/khó diễn giải; ưu tiên cho phép liên hệ rõ với các kiểu vận động trong RSCA (ổn định cao, mới nổi, suy giảm, bất ổn...).

3.4.3. Nguyên tắc đặt tên cụm và cách diễn giải

Do K-means chỉ tạo ra nhãn cụm mang tính kỹ thuật (cụm 1, cụm 2,...) và nhãn này có thể hoán đổi giữa các lần chạy hoặc giữa các lựa chọn K, nghiên cứu không diễn giải dựa trên số thứ tự cụm. Thay vào đó, sau khi phân cụm các ngành HS2 theo véc-tơ đặc trưng chuẩn hoá, chúng tôi lập hồ sơ cụm (cluster profile) để mô tả nhất quán đặc trưng của từng cụm.

Hồ sơ cụm được xây dựng từ các thống kê tóm tắt trên từng đặc trưng. Để hỗ trợ đối chiếu với cấu trúc xuất khẩu, hồ sơ cụm đồng thời ghi nhận quy mô cụm (số ngành HS2) và các chỉ báo xuất khẩu (kim ngạch bình quân và tỷ trọng theo giai đoạn).

Trên cơ sở hồ sơ cụm, các cụm được gắn nhãn diễn giải theo các quy tắc nhất quán dựa trên tổ hợp đặc trưng (ví dụ: cụm có độ bền cao và bất ổn thấp; cụm có xu hướng cải thiện; cụm có xu hướng suy giảm; cụm dao động mạnh). Các nhãn diễn giải được hiểu

theo nghĩa vận hành gắn với các chỉ báo thực nghiệm. Cụ thể, “bền vững” chỉ các cụm có độ bền cao, xác suất duy trì LTSS cao và độ bất ổn thấp; “mong manh” chỉ các cụm có LTSS hoặc cận LTSS nhưng tần suất đổi trạng thái cao hay xác suất rời LTSS lớn; còn “động lực tăng trưởng” được dùng để chỉ các cụm có đóng góp lớn vào phân gia tăng kim ngạch xuất khẩu giữa các giai đoạn. Cũng cần lưu ý rằng các nhân diện giải này phản ánh đặc trưng chi phối của ngành trong giai đoạn dùng để phân cụm; trong thực tế, một ngành có thể thay đổi bản chất động học theo thời gian và dịch chuyển gần hơn tới một kiểu cụm khác.

3.5. Chuỗi Markov chuyển trạng thái RSCA (2×2)

Gọi $S_{i,t} \in \{0,1\}$ là trạng thái LTSS của ngành i tại năm t , trong đó $S_{i,t} = 1$ nếu $RSCA_{i,t} \geq 0$ (có LTSS) và $S_{i,t} = 0$ nếu $RSCA_{i,t} < 0$ (không có LTSS). Trên cơ sở đó, nghiên cứu sử dụng khung Markov bậc nhất 2×2 ở hai cấp độ. Thứ nhất, một ma trận chuyển trạng thái tổng hợp được tính cho toàn bộ danh mục HS2 nhằm cung cấp bức tranh nền về sự vận động trong cơ LTSS của cơ cấu xuất khẩu. Thứ hai, các ma trận chuyển trạng thái được tính theo cụm và theo giai đoạn để nhận diện sự khác biệt trong sự vận động giữa các nhóm ngành sau khi phân cụm. Với mỗi cụm và mỗi giai đoạn, nghiên cứu xây dựng ma trận chuyển trạng thái bậc nhất 2×2 :

trong đó $p_{a \rightarrow b} = \Pr(S_{i,t+1} = b | S_{i,t} = a)$ là tỷ lệ số lần quan sát chuyển từ trạng thái a ở năm t sang trạng thái b ở năm $t+1$ trên tổng số lần quan sát ở trạng thái a tại năm t ; $a, b \in \{0,1\}$. Ví dụ, $p_{0 \rightarrow 0}$ là xác suất ngành bất kỳ không có LTSS ở năm t và tiếp tục

$$P = \begin{pmatrix} | & \begin{matrix} P_{0 \rightarrow 0} & P_{0 \rightarrow 1} \end{matrix} & | \\ \hline \begin{matrix} P_{1 \rightarrow 0} & P_{1 \rightarrow 1} \end{matrix} & | & \hline \end{pmatrix}$$

không có LTSS ở năm $t+1$. Khung Markov bậc nhất được lựa chọn nhằm cung cấp một mô tả gọn, trực quan và dễ diễn giải về động học LTSS. Theo cách tiếp cận này, xác suất chuyển trạng thái được điều kiện hóa theo trạng thái ở năm liền trước. Các ma trận theo cụm được tính cho các giai đoạn 2000- 2007, 2008- 2013, 2014- 2019 và 2020- 2023; ngoài ra, một ma trận tổng hợp cho toàn bộ HS2 được báo cáo như bức tranh nền chung.

3.6. Chỉ báo cơ cấu, đóng góp và tập trung nội bộ theo cụm

Ba nhóm chỉ báo được báo cáo gồm (i) Cơ cấu xuất khẩu: kim ngạch xuất khẩu bình quân năm của từng cụm; (ii) Đóng góp vào tăng trưởng: phân rã chênh lệch xuất khẩu bình quân năm giữa hai giai đoạn liên tiếp thành đóng góp tuyệt đối và tỷ trọng đóng góp của từng cụm; (iii) Tập trung nội bộ: trong mỗi cụm, tính tỷ trọng xuất khẩu trong cụm của 5 ngành xuất khẩu nhiều nhất mỗi cụm; tỷ trọng càng cao phản ánh mức phụ thuộc càng lớn vào một số ít ngành. Tất cả các tính toán và hình vẽ minh họa được thực hiện với phần mềm R .

4. Kết quả nghiên cứu

4.1. Bức tranh Markov tổng hợp

Bảng 1 trình bày ma trận chuyển trạng thái Markov 2×2 chung cho tất cả các ngành HS2. Trước khi đi vào phân tích theo cụm, ma trận chuyển trạng thái tổng hợp cho toàn bộ các ngành HS2 được trình bày. Ở cấp độ tổng thể, cách tiếp cận này cho biết LTSS nhìn chung có xu hướng được duy trì, hình thành hay biến mất theo thời gian. Kết quả cho thấy quán tính cao của LTSS ở cấp ngành. Trong giai đoạn 2000- 2023, xác suất tiếp tục không có LTSS vào năm

kế tiếp là 0,97, còn xác suất tiếp tục duy trì LTSS là 0,91. Ngược lại, xác suất gia nhập LTSS chỉ khoảng 0,03 mỗi năm, thấp

Bảng 1. Ma trận chuyển trạng thái Markov 2x2 tổng hợp (HS2)

	0 (RSCA < 0)	1 (RSCA ≥ 0)
0 (RSCA < 0)	0,97	0,03
1 (RSCA ≥ 0)	0,09	0,91

Nguồn: Tính toán của tác giả

hơn đáng kể so với xác suất rời khỏi LTSS là 0,09. Điều này cho thấy việc tạo lập LTSS mới diễn ra tương đối chậm, trong khi LTSS hiện hữu có thể bị suy giảm với tần suất cao hơn. Hình 1 củng cố nhận định trên khi báo cáo xác suất gia nhập hay rời khỏi theo từng giai đoạn.

Như vậy, phân tích Markov tổng hợp cho thấy một bức tranh nền về quán tính LTSS, nhưng chưa đủ để trả lời câu hỏi cấu trúc, rằng quán tính đó nằm ở đâu, thuộc về nhóm ngành nào, và đi kèm với kiểu động học nào. Do đó, bước tiếp theo của bài viết là quy chiếu danh mục HS2 thành các cụm có mẫu hình vận động tương đối đồng nhất, từ đó bóc tách các khác biệt bị che khuất trong kết quả tổng hợp.

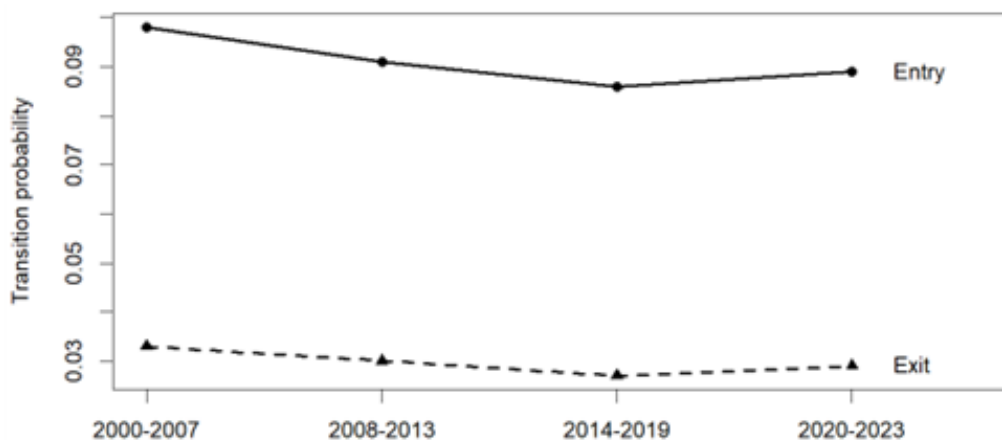
4.2. Lựa chọn số cụm K và hồ sơ cụm

4.2.1. Bằng chứng Elbow/WSS và Silhouette

Kết quả tính toán WSS theo các giá trị của K cho thấy mức giảm WSS là đáng kể khi tăng K ở các mức thấp, nhưng sau ngưỡng K = 4, lợi ích biên của việc tăng thêm cụm giảm rõ rệt (thể hiện ở điểm gãy của đường elbow) (Hình 1). Silhouette trung bình đạt mức cao nhất cũng ở K = 4: khi K < 4, các cụm gộp thô khiến độ tách biệt thấp; khi K > 4, cụm bị phân mảnh và silhouette giảm do khoảng cách giữa các cụm không còn cải thiện tương xứng (Hình 2). Do đó, nghiên cứu lựa chọn K = 4 làm cơ sở phân cụm để trình bày các kết quả.

4.2.2. Hồ sơ cụm

Bảng 2 cho thấy sự phân hoá rõ về sự vận động trong bốn cụm. Cụm 4 là nhóm nổi bật nhất: RSCA trung bình dương (0,46), độ bền rất cao (18,5 năm) và độ bất ổn thấp (0,02, xem thêm đối chiếu Markov theo cụm ở mục 4.3). Mặc dù chỉ gồm 24 mã HS2, cụm này chiếm khoảng 49,3% kim ngạch toàn kỳ 2000- 2019, phản ánh vai trò trụ cột trong cơ cấu xuất khẩu. Ở chiều ngược lại, cụm 3 có RSCA trung bình âm

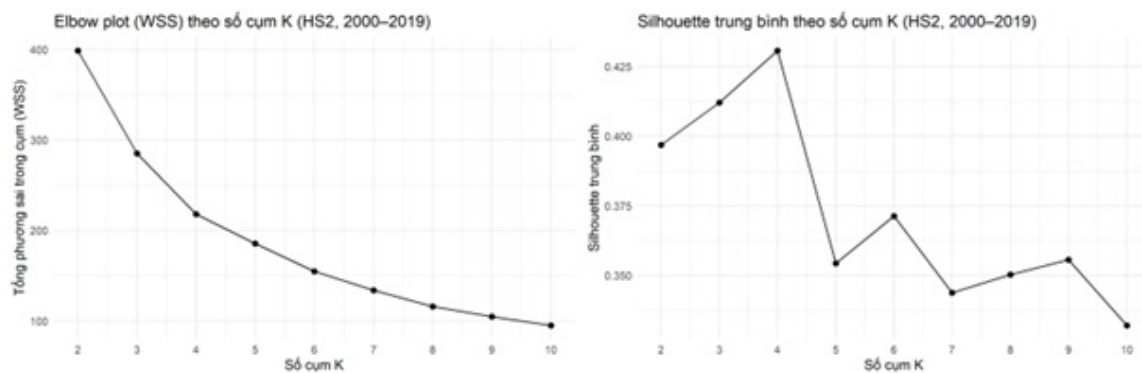


Nguồn: Tính toán của tác giả

Hình 1. Tỷ lệ gia nhập (entry) và rời khỏi (exit) theo thời gian

Sự vận động trong cấu trúc lợi thế so sánh của Việt Nam:

Phân cụm RSCA ở cấp độ HS2 giai đoạn 2000- 2023



Nguồn: Tính toán của tác giả

Hình 2. Elbow/WSS và Silhouette trung bình theo số cụm

sâu (-0,67), độ bền gần như bằng không (0,03) và độ bất ổn rất thấp (0,00); cụm này gồm nhiều mã HS2 (39) nhưng chỉ đóng góp khoảng 12,1% kim ngạch toàn kỳ, gợi ý một nhóm ngành bất lợi tương đối dai dẳng và ít có khả năng bứt phá qua ngưỡng RSCA = 0.

Hai cụm còn lại (cụm 1 và cụm 2) có RSCA trung bình âm ở mức vừa phải (khoảng -0,26) nhưng khác biệt đáng kể về mức độ dao động và quỹ đạo. Cụm 2 có độ bất ổn cao nhất (0,13) và xu thế âm (-0,03), trong khi cụm 1 thể hiện biến động cao (0,36) nhưng xu thế dương (0,05), hàm ý khả năng cải thiện LTSS trong dài hạn dù còn dao động quanh ngưỡng.

4.2.3. Thành phần hàng hoá tiêu biểu trong từng cụm

Nghiên cứu xác định 05 ngành có kim ngạch

xuất khẩu cao nhất trong từng cụm ở giai đoạn gần nhất 2014- 2019 nhằm định danh và diễn giải nội dung kinh tế của các cụm trên cơ sở giai đoạn gần nhất trước cú sốc Covid-19, qua đó tránh để việc định danh cụm bị chi phối quá mức bởi biến động ngắn hạn của giai đoạn 2020- 2023. Bảng 3 báo cáo tỷ trọng của từng ngành (trong 05 ngành lựa chọn) trong cụm và trong tổng xuất khẩu, qua đó cho phép nhận diện mức độ quan trọng của từng ngành dẫn dắt trong mỗi cụm.

Cụm 1 gắn với các nhóm điện- điện tử và hàng liên quan gồm HS85, HS90, HS72, HS52, và HS25, trong đó HS85 đồng thời là một trụ cột rất lớn của tổng xuất khẩu giai đoạn 2014- 2019. Cụm 2 gắn với nhóm nhiên liệu-khoáng sản (HS27 đứng đầu) và một số nhóm hàng giá trị cao như HS71... Cụm 3 phản ánh nhóm máy móc- thiết bị

Bảng 2. Hồ sơ cụm theo véc-tơ đặc trưng và tỷ trọng trong tổng xuất khẩu (2000- 2019)

Cụm (k)	Số ngành HS2	RSCA bình quân	Độ biến động	Độ bền	Độ bất ổn	Xu thế	Tỷ trọng (2000-2019)
1	13	-0,27	0,36	5,93	0,07	0,05	23,1%
2	18	-0,26	0,27	4,61	0,13	-0,03	15,4%
3	40	-0,67	0,11	0,03	0,00	0,01	12,1%
4	24	0,46	0,13	18,50	0,02	-0,01	49,3%

Ghi chú: Các đặc trưng như xây dựng ở Mục 3.3. Phân cụm và các đặc trưng cho sự vận động trong RSCA được xây dựng trên giai đoạn 2000- 2019 như đã thảo luận ở Mục 3.4.1.

Nguồn: Tính toán của tác giả

Bảng 3. Top 5 HS2 theo từng cụm giai đoạn 2014- 2019

Cụm	Xếp hạng trong cụm	Mã HS2	Diễn giải	Kim ngạch xuất khẩu (tỷ USD)	Tỷ trọng trong cụm (%)	Tỷ trọng trong tổng xuất khẩu (%)
1	1	85	Thiết bị điện, điện tử, linh kiện	83,89	76,48	26,97
1	2	90	Thiết bị quang học, đo lường y tế (dụng cụ chính xác)	6,54	5,97	2,10
1	3	72	Sắt và thép	4,71	4,29	1,51
1	4	52	Bông (nguyên liệu bông)	3,26	2,97	1,05
1	5	25	Khoáng phi kim/vật liệu xây dựng (muối, đất đá, vôi, xi măng)	2,12	1,93	0,68
2	1	27	Nhiên liệu khoáng, dầu khoáng và sản phẩm chưng cất	14,24	64,40	4,58
2	2	71	Đá quý, kim loại quý, trang sức (kể cả tiền xu)	1,41	6,40	0,45
2	3	20	Chế phẩm từ rau quả/hạt (đồ hộp, mứt, nước quả...)	1,07	4,83	0,34
2	4	21	Chế phẩm thực phẩm khác (thực phẩm chế biến)	0,84	3,80	0,27
2	5	34	Xà phòng, chất tẩy rửa, chất hoạt động bề mặt (đánh bóng, bôi trơn...)	0,72	3,25	0,23
3	1	84	Máy móc thiết bị cơ khí và bộ phận (nồi hơi, máy móc...)	18,98	45,10	6,10
3	2	39	Nhựa và sản phẩm bằng nhựa	5,36	12,73	1,72
3	3	87	Xe cộ và phụ tùng (trừ đường sắt)	3,82	9,09	1,23
3	4	73	Sản phẩm bằng sắt/thép	3,50	8,32	1,13
3	5	95	Đồ chơi, trò chơi, dụng cụ thể thao	1,44	3,42	0,46
4	1	64	Giày dép và bộ phận	23,13	16,86	7,44
4	2	62	May mặc không dệt kim (dệt thoi)	21,84	15,92	7,02
4	3	61	May mặc dệt kim/móc	19,29	14,06	6,20
4	4	3	Thủy sản (cá, giáp xác, thân mềm...)	10,60	7,73	3,41
4	5	94	Nội thất, chăn đệm, đèn chiếu sáng, nhà lắp ghép	10,50	7,65	3,38

Ghi chú: Danh sách đầy đủ mã HS2 theo từng cụm được trình bày tại Phụ lục A dưới dạng ma trận 'nhóm ngành × cụm', nhằm minh bạch thành phần cụm và hỗ trợ việc gán nhãn cụm; Bảng 3 chỉ báo cáo các nhóm hàng dẫn dắt theo kim ngạch để phục vụ định danh nhanh.

Nguồn: Tính toán của tác giả

và công nghiệp chế tạo (HS84 dẫn dắt, cùng HS39, HS87, HS73, HS95). Cụm 4 thiên về nhóm hàng tiêu dùng thâm dụng lao động (giày dép, dệt may) và một số nhóm nông sản như thủy sản, nội thất.

4.3. Đối chiếu Markov theo cụm

Mục này đóng vai trò đối chiếu/kiểm chứng theo cụm: sau khi các cụm được xác định từ các véc-tơ đặc trưng (mục 4.2), nghiên cứu sử dụng chuỗi Markov 2×2 quanh ngưỡng $RSCA = 0$ để xem các cụm có khác biệt về *duy trì* và *gia nhập lợi thế* hay không. Kết quả cho thấy sự phân hoá rất rõ. Cụm 4 có

Sự vận động trong cấu trúc lợi thế so sánh của Việt Nam:

Phân cụm RSCA ở cấp độ HS2 giai đoạn 2000- 2023

Bảng 4. Đối chiếu Markov theo cụm: xác suất chuyển trạng thái RSCA (2000- 2019)

Cụm	N	Duy trì LTSS $P(1 \rightarrow 1)$	Gia nhập LTSS $P(0 \rightarrow 1)$	Rời khỏi LTSS $P(1 \rightarrow 0)$	Duy trì không có LTSS $P(0 \rightarrow 0)$	Độ bất ổn (tỷ lệ đổi trạng thái)
1	14	0,92	0,06	0,08	0,94	0,07
2	18	0,74	0,07	0,26	0,93	0,13
3	39	-	0,00	-	0,99	0,00
4	24	0,99	0,16	0,01	0,84	0,02

Ghi chú: $P(1 \rightarrow 1)$, $P(0 \rightarrow 1)$, $P(0 \rightarrow 0)$, được tính theo cách gộp các chuyển trạng thái trong từng cụm trên giai đoạn 2000-2019. Dấu “-” thể hiện xác suất không xác định do trạng thái $S_{i,t} = 1$ không xuất hiện trong cụm tương ứng (mẫu số bằng 0).

Nguồn: Tính toán của tác giả

xác suất duy trì lợi thế cao nhất $P(1 \rightarrow 1) = 0,99$ và xác suất rời khỏi LTSS rất thấp $P(1 \rightarrow 0) = 0,01$, đồng thời có khả năng gia nhập tương đối cao $P(0 \rightarrow 1) = 0,16$ và mức chuyển trạng thái thấp (0,02), phản ánh đặc trưng ổn định và bền vững. Cụm 1 cũng tương đối ổn định với $P(1 \rightarrow 1) = 0,92$, $P(1 \rightarrow 0) = 0,08$ và tỷ lệ đổi trạng thái 0,07. Trái lại, cụm 2 là nhóm biến động nhất: $P(1 \rightarrow 1) = 0,74$ thấp hơn đáng kể, $P(1 \rightarrow 0) = 0,26$ cao, và tỷ lệ đổi trạng thái 0,13, phù hợp với trực giác về một cụm có xác suất đảo trạng thái quanh ngưỡng cao hơn. Cuối cùng, cụm 3 gần như luôn ở trạng thái bất lợi: $P(0 \rightarrow 1) = 0,00$ và $P(0 \rightarrow 0) = 0,99$; do trạng thái có LTSS không xuất hiện trong giai đoạn quan sát nên $P(1 \rightarrow 1)$ và $P(1 \rightarrow 0)$ không xác định. Nhìn chung, Bảng 4 củng cố diễn giải kinh tế các cụm ở mục 4.2 theo một thước đo động học độc lập (chuỗi Markov), mà không thay đổi logic phân cụm K-means.

4.4. Tỷ trọng xuất khẩu theo cụm

Hình 3 cho thấy tỷ trọng xuất khẩu theo cụm biến đổi rõ rệt theo thời gian. Trong giai đoạn 2000- 2007, Cụm 4 giữ vai trò chi phối (chiếm trên một nửa tổng kim ngạch); sang 2008- 2013, tỷ trọng của Cụm 1 tăng nhanh và đến giai đoạn 2020- 2023 đã trở thành cụm lớn nhất. Ngược lại lại,

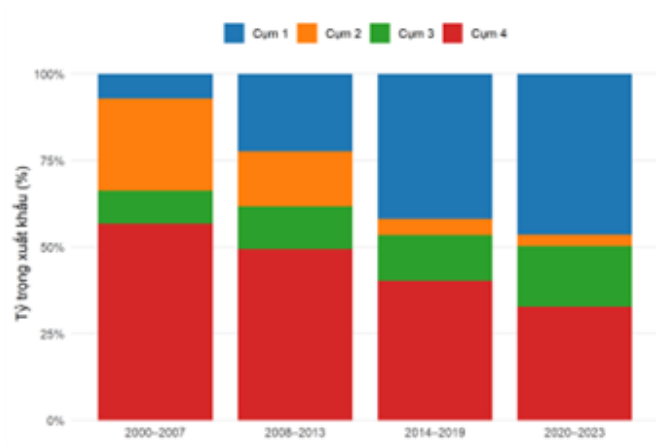
Cụm 2 thu hẹp dần, còn Cụm 3 tăng nhưng chủ yếu đóng vai trò hỗ trợ. Nhìn tổng thể, kết quả này hàm ý một sự tái cơ cấu theo cụm trong đó trọng tâm kim ngạch xuất khẩu dịch chuyển khỏi cụm chi phối ban đầu (Cụm 4) sang một cụm dẫn dắt mới (Cụm 1).

4.5. Đóng góp của các cụm vào tăng/giảm xuất khẩu

Hình 4. G1 cho thấy quy mô kim ngạch xuất khẩu bình quân năm tăng mạnh qua các giai đoạn, đồng thời cấu phần theo cụm thay đổi rõ rệt. Cụm 4 giữ vai trò trụ cột trong giai đoạn đầu, nhưng từ 2014- 2019 trở đi, Cụm 1 mở rộng nhanh và trở thành động lực chính của tổng xuất khẩu; Cụm 3 cũng gia tăng vai trò ở giai đoạn cuối, còn Cụm 2 duy trì quy mô nhỏ.

Hình 4. G2 cho thấy động lực tăng trưởng xuất khẩu dịch chuyển mạnh về phía Cụm 1. Từ 2008- 2013 sang 2014- 2019 và tiếp tục sang 2020- 2023, hơn một nửa phần tăng xuất khẩu đến từ cụm này; trong khi đó, Cụm 2 có giai đoạn đóng góp âm, còn Cụm 3 tăng dần vai trò hỗ trợ. Kết quả này củng cố nhận định rằng cơ cấu xuất khẩu của Việt Nam không chỉ thay đổi về tỷ trọng, mà còn thay đổi về nơi phát sinh động lực tăng trưởng.

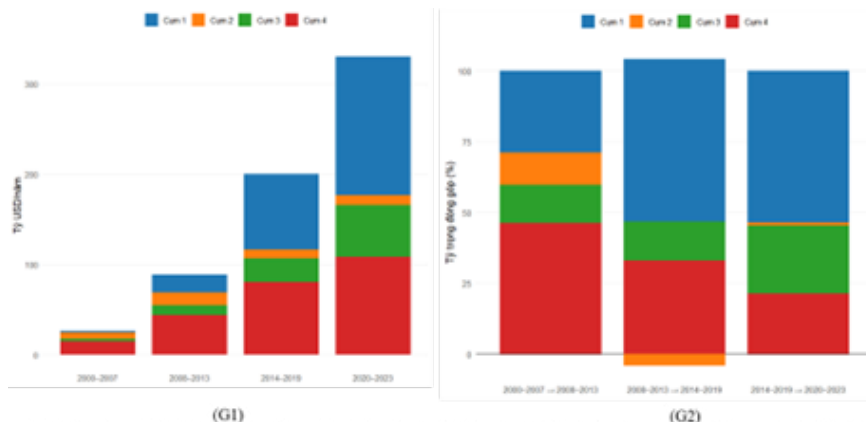
4.6. Tập trung nội bộ trong từng cụm



Ghi chú: Tỷ trọng 2020-2023 được tính bằng cách giữ cố định nhân cụm và tổng hợp kim ngạch HS2 theo cùng phân loại cụm.

Nguồn: Tính toán của tác giả

Hình 3. Tỷ trọng kim ngạch xuất khẩu theo cụm qua các giai đoạn



Ghi chú: Kim ngạch được tính cho xuất khẩu của Việt Nam sang Thế giới, theo mã HS2; đơn vị tỷ USD/năm. Tỷ trọng đóng góp được tính trên mức thay đổi kim ngạch xuất khẩu bình quân năm giữa hai giai đoạn liên tiếp. Giá trị âm phản ánh cụm có đóng góp theo hướng làm giảm tăng trưởng ròng của xuất khẩu trong giai đoạn chuyển tiếp (khi một số cụm suy giảm, phần đóng góp dương của các cụm còn lại có thể vượt 100% để bù trừ). Các cụm được xác định từ phân cụm k-means trên giai đoạn 2000-2019 và được giữ cố định khi tổng hợp cho 2020-2023

Nguồn: Tính toán của tác giả

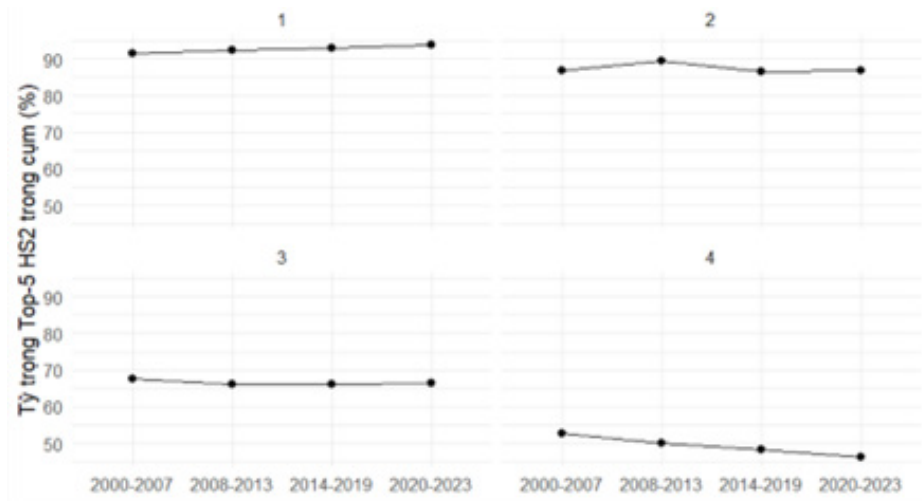
Hình 4. Kim ngạch xuất khẩu bình quân (G1) và tỷ trọng đóng góp vào tăng trưởng xuất khẩu (G2) theo cụm và giai đoạn

Hình 5 phản ánh mức độ tập trung nội bộ trong từng cụm theo thời gian thông qua tỷ trọng của 05 mã có kim ngạch xuất khẩu lớn nhất (tỷ trọng cao hàm ý kim ngạch phụ thuộc vào một số ít mã trụ cột). Kết quả cho thấy Cụm 1 có cấu trúc tập trung rất cao và ổn định (92- 95%), cho thấy đây là cụm mang tính chuyên môn hoá mạnh, biến động kim ngạch chủ yếu gắn với một

vài nhóm hàng dẫn dắt. Cụm 2 cũng tập trung cao (87- 90%) và chỉ dao động nhẹ, hàm ý cấu trúc tương đối ổn định nhưng vẫn có độ nhạy nhất định theo chu kỳ. Cụm 3 ở mức trung bình (66-68%), gợi ý khả năng có nhiều mã đa trụ cột hơn so với hai cụm đầu. Trái lại, Cụm 4 có mức tập trung thấp nhất và giảm dần (53% xuống 47%), phản ánh xu hướng phân tán/đa dạng hoá

Sự vận động trong cấu trúc lợi thế so sánh của Việt Nam:

Phân cụm RSCA ở cấp độ HS2 giai đoạn 2000- 2023



Chú thích: Tỷ trọng Top-5 HS2 được xác định theo kim ngạch xuất khẩu trong từng cụm-giai đoạn (%); tỷ trọng càng cao hàm ý mức chuyên môn hoá nội cụm càng lớn.

Nguồn: Tính toán của tác giả

Hình 5. Mức độ tập trung nội bộ trong từng cụm theo giai đoạn

nội bộ theo thời gian, tức tăng trưởng được tạo bởi một nhóm hàng rộng hơn thay vì phụ thuộc vào một vài ngành chủ lực.

Tuy nhiên, cần thận trọng khi diễn giải ý nghĩa phát triển của các cụm tăng trưởng nhanh, đặc biệt các cụm gắn mạnh với điện tử/chế biến chế tạo, bởi dữ liệu hiện tại không cho phép phân tách đóng góp của khu vực FDI và khu vực doanh nghiệp trong nước. Do đó, từ kết quả hiện có, bài viết chỉ có thể khẳng định vai trò ngày càng lớn của các cụm này trong tăng trưởng xuất khẩu và rủi ro tập trung nội bộ, chưa thể kết luận trực tiếp về mức độ lan tỏa của động lực đó sang năng lực của doanh nghiệp nội địa.

5. Hàm ý chính sách và hạn chế của nghiên cứu

5.1. Hàm ý chính sách

Từ các kết quả trên, trọng tâm chính sách không nằm ở việc chọn đúng ngành thắng cuộc (*picking winners*) theo quan điểm truyền thống, mà ở quản trị danh mục xuất khẩu theo hướng cân bằng rủi ro-tăng

trưởng. Cụ thể, mỗi cụm phản ánh một tổ hợp khác nhau giữa động lực tăng trưởng, độ bền LTSS và mức tập trung nội bộ; vì vậy, gói chính sách phù hợp không nên đồng nhất giữa các cụm. Các hàm ý dưới đây được rút ra ở cấp độ HS2 và vì thế mang tính định hướng ở cấp nhóm ngành rộng; việc thiết kế công cụ can thiệp cụ thể cần được chi tiết hóa tiếp ở cấp HS4/HS6 và, khi có thể, ở cấp doanh nghiệp.

(i) Cụm 1: Trụ cột tăng trưởng nhưng rủi ro tập trung cao. Khi đóng góp vào tăng trưởng xuất khẩu dựa mạnh vào một cụm có cấu trúc nội bộ rất tập trung, tồn tại rủi ro “một trụ cột” và phụ thuộc vào một số chuỗi giá trị then chốt. Với cụm này, hàm ý chính sách trọng tâm là giảm rủi ro phụ thuộc vào một số chuỗi giá trị then chốt, đồng thời tăng năng lực hấp thụ trong nước. Ưu tiên phù hợp là phát triển công nghiệp hỗ trợ, tăng liên kết giữa khu vực FDI và doanh nghiệp nội địa, và cải thiện các điều kiện nền như kỹ năng, logistics và thể chế chất lượng. Về dài hạn, cần mở rộng chiều sâu của cụm thay vì chỉ tiếp tục mở rộng kim ngạch.

(ii) Cụm 2: Nhóm dễ biến động và có giai đoạn kéo lùi. Đóng góp âm cho thấy rủi ro chu kỳ/giá cả và tính dễ tổn thương trước biến động bên ngoài. Hàm ý chính sách là hạn chế kỳ vọng tăng trưởng dựa vào cụm này; thay vào đó là (a) công cụ quản trị biến động và rủi ro giá; (b) định hướng chế biến sâu và hạ nguồn để giảm phụ thuộc xuất thô; (c) lồng ghép mục tiêu chuyển dịch xanh nhằm tránh đóng khung vào các phân khúc chịu sức ép carbon/tiêu chuẩn mới.

(iii) Cụm 3: Đa trụ cột mức trung bình. Mức tập trung trung bình và ổn định gợi ý đây là không gian phù hợp để đẩy mạnh nâng cấp công nghiệp theo chiều sâu. Với cụm này, hàm ý chính sách là hỗ trợ những ngành có khả năng cải thiện LTSS nhưng chưa bứt hẳn lên. Điều này đòi hỏi nhấn mạnh vào năng lực công nghệ, tiêu chuẩn, đổi mới quy trình và các công cụ hỗ trợ đầu tư có chọn lọc để giúp doanh nghiệp vượt ngưỡng cạnh tranh.

(iv) Cụm 4: Danh mục rộng và vai trò đệm. Cụm 4 có mức tập trung thấp và xu hướng phân tán tăng dần, phản ánh một cấu trúc có tiềm năng trở thành bộ đệm trước những cú sốc nhưng biên lợi nhuận dễ thu hẹp khi yêu cầu tiêu chuẩn của thị trường ngày càng khắt khe. Trọng tâm chính sách vì vậy có thể theo hướng giảm chi phí giao dịch và hỗ trợ doanh nghiệp nâng cấp năng lực tuân thủ theo chuẩn thị trường: (a) giảm chi phí tuân thủ và logistics; (b) chuẩn hóa truy xuất nguồn gốc, tiêu chuẩn lao động- môi trường và đáp ứng các yêu cầu EU-US, đặc biệt trong bối cảnh tiêu chuẩn bền vững ngày càng ràng buộc; (c) hỗ trợ doanh nghiệp giảm dần phụ thuộc vào gia công thuần túy, chuyển sang các khâu có giá trị gia tăng cao hơn như thiết kế, phát triển sản phẩm, xây dựng thương hiệu và mở rộng kênh phân phối; (d) đa dạng hóa thị trường/đối tác để tăng khả năng hấp thụ cú sốc. Với Cụm 4, chính sách hiệu quả không

nhằm đẩy một trụ cột lên thật cao mà nhằm nâng mặt bằng năng lực cạnh tranh trên diện rộng.

Ở cấp điều hành, kết quả phân cụm có thể được sử dụng như một công cụ theo dõi đồng thời đóng góp tăng trưởng theo cụm và mức tập trung nội bộ để cảnh báo sớm rủi ro phụ thuộc. Tuy nhiên, đây chỉ nên được hiểu như một công cụ nhận diện tín hiệu cấu trúc ở cấp HS2, không phải căn cứ đủ để thiết kế trực tiếp biện pháp can thiệp chi tiết cho từng phân ngành.

5.2. Hạn chế và hướng nghiên cứu tiếp theo

Nghiên cứu tồn tại một số hạn chế sau. Thứ nhất, việc phân tích ở cấp độ HS2 giúp nhận diện bức tranh cơ cấu với nhóm ngành rộng, nhưng đồng thời che khuất khác biệt đáng kể bên trong từng nhóm. Điều này đặc biệt quan trọng đối với các nhóm ngành lớn như thiết bị điện, điện tử, nơi các sản phẩm có thể rất khác nhau về công nghệ, giá trị gia tăng và mức độ tham gia của doanh nghiệp nội địa. Vì vậy, các kết luận theo cụm trong bài viết cần được hiểu là các nhận định ở cấp nhóm ngành rộng, không thay thế cho phân tích sâu hơn ở cấp HS4/HS6. Thứ hai, giai đoạn dài giúp ổn định xu hướng nhưng làm giảm khả năng nhận diện cú sốc ngắn hạn (ví dụ đứt gãy chuỗi cung ứng) và các hiệu ứng chính sách theo năm. Thứ ba, kết quả phân cụm phụ thuộc vào lựa chọn thước đo và thuật toán; dù $K = 4$ có cơ sở kỹ thuật và diễn giải, vẫn cần nhìn nhận phân cụm như một công cụ tóm tắt cấu trúc hơn là một ranh giới cứng nhắc giữa các ngành. Theo đó, nhãn cụm trong nghiên cứu này phản ánh đặc trưng chi phối của ngành trong giai đoạn dùng để phân cụm, chứ không hàm ý rằng mỗi ngành thuộc về một cụm bất biến theo thời gian. Thứ tư, khung Markov bậc nhất giúp mô tả động học một cách

ngắn gọn và trực quan, nhưng chưa phản ánh đầy đủ hiệu ứng bộ nhớ hay phụ thuộc theo thời lượng duy trì LTSS. Cuối cùng, phân tích hiện tại chưa phân rõ được đóng góp giữa khu vực FDI và khu vực doanh nghiệp trong nước, một hạn chế đặc biệt quan trọng trong bối cảnh Việt Nam. Hạn chế này ảnh hưởng trực tiếp đến độ sắc của các hàm ý chính sách, nhất là đối với các cụm tăng trưởng nhanh gắn với điện tử/chế biến chế tạo.

Các hướng mở rộng tiếp theo vì vậy chủ yếu nằm ở việc tăng độ sâu dữ liệu, cho phép phân rõ tốt hơn theo sở hữu/doanh nghiệp, và kiểm tra độ ổn định của cấu trúc cụm trong các khung động học linh hoạt hơn.

6. Kết luận

Bài viết phân tích sự vận động trong cấu trúc lợi thế so sánh của Việt Nam sử dụng phương pháp phân cụm K-means và phân

tích chuỗi Markov. Kết quả cho thấy cơ cấu xuất khẩu Việt Nam không chỉ thay đổi về nhóm ngành có lợi thế so sánh, mà còn thay đổi về cách thức tăng trưởng xuất khẩu được dẫn dắt, với những cụm đóng vai trò trụ cột xuất khẩu mang tính tập trung cao trong khi những cụm khác đóng vai trò bộ đệm nhờ danh mục phân tán. Điểm cốt lõi vì vậy không phải là tìm ra một ngành thắng cuộc, mà là quản trị danh mục xuất khẩu vừa khai thác trụ cột vừa giảm rủi ro phụ thuộc. Theo tinh thần chính sách công nghiệp dựa trên năng lực và quản trị rủi ro, hàm ý quan trọng là thiết kế gói công cụ mang tính cấu trúc cho từng cụm. Nếu được vận hành như một bảng điều khiển theo dõi cấu trúc, khung RSCA-phân cụm-Markov có thể hỗ trợ điều hành xuất khẩu theo hướng vừa tăng trưởng vừa chống chịu trong bối cảnh cú sốc và tiêu chuẩn thị trường ngày càng bất định hiện nay. ■

Tài liệu tham khảo

- Balassa, B. (1965). Trade liberalisation and “revealed” comparative advantage. *The Manchester School*, 33(2), 99–123. <https://doi.org/10.1111/j.1467-9957.1965.tb00050.x>
- Berthou, A., Haramboure, A., & Samek, L. (2024). *Mapping and testing product-level vulnerabilities in granular production networks* (OECD Science, Technology and Industry Working Papers, No. 2024/02). OECD Publishing. <https://doi.org/10.1787/9bcde495-en>
- Cadot, O., Carrère, C., & Strauss-Kahn, V. (2011). Export diversification: What’s behind the hump? *The Review of Economics and Statistics*, 93(2), 590–605. https://doi.org/10.1162/REST_a_00078
- Doanh, N. K., & Heo, Y. (2011). Dynamic patterns in Vietnam’s comparative advantage. *Journal of International Logistics and Trade*, 9(1), 89–114. <https://doi.org/10.24006/jilt.2011.9.1.89>
- Fulcher, B. D., & Jones, N. S. (2014). Highly comparative feature-based time-series classification. *IEEE Transactions on Knowledge and Data Engineering*, 26(12), 3026–3037. <https://doi.org/10.1109/TKDE.2014.2316504>
- Hausmann, R., Hwang, J., & Rodrik, D. (2007). What you export matters. *Journal of Economic Growth*, 12(1), 1–25. <https://doi.org/10.1007/s10887-006-9009-4>
- Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26), 10570–10575. <https://doi.org/10.1073/pnas.0900943106>
- Imbs, J., & Wacziarg, R. (2003). Stages of diversification. *American Economic Review*, 93(1), 63–86. <https://doi.org/10.1257/00028280321455160>
- Koren, M., & Tenreyro, S. (2007). Volatility and development. *The Quarterly Journal of Economics*, 122(1), 243–287. <https://doi.org/10.1162/qjec.122.1.243>
- Laursen, K. (1998). *Revealed comparative advantage and the alternatives as measures of international specialisation* (DRUID Working Paper No. 98-30). DRUID, Copenhagen Business School. <https://wp.druid.dk/wp/19980030.pdf>
- Laursen, K. (2015). Revealed comparative advantage and the alternatives as measures of international specialization. *Eurasian Business Review*, 5(1), 99–115. <https://doi.org/10.1007/s40821-015-0017-1>
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 5. 1, pp. 281–297). University of California Press. <https://projecteuclid.org/ebooks/berkeley-symposium-on-mathematical->

- [statistics-and-probability/Proceedings-of-the-Fifth-Berkeley-Symposium-on-Mathematical-Statistics-and/chapter/Some-methods-for-classification-and-analysis-of-multivariate-observations/bmsmp/1200512992](https://doi.org/10.1111/j.1540-6261.1952.tb01525.x)
- Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77–91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Wang, X. C., Smith, K. A., & Hyndman, R. J. (2006). Characteristic-based clustering for time series data. *Data Mining and Knowledge Discovery*, 13(3), 335–364. <https://doi.org/10.1007/s10618-005-0039-x>
- Widodo, T. (2009). Dynamic comparative advantages in the ASEAN+3. *Journal of Economic Integration*, 24(3), 505–529. <https://doi.org/10.11130/jei.2009.24.3.505>

Phụ lục A. Danh sách mã HS2 theo cụm (K = 4)

Nhóm ngành	Cụm 1 (n=13)	Cụm 2 (n=18)	Cụm 3 (n=40)	Cụm 4 (n=24)
Nông–thủy sản & thực phẩm (HS01–24)	-	HS04, HS05, HS12, HS13, HS15, HS17, HS20 HS21, HS24	HS01, HS02, HS06, HS18, HS22, HS23	HS03, HS07, HS08, HS09, HS10, HS11, HS14 HS16, HS19
Khoáng sản & nhiên liệu (HS25–27)	HS25	HS26, HS27	-	-
Hóa chất (HS28–38)	HS28	HS31, HS34	HS29, HS30, HS32, HS33, HS35, HS36, HS37 HS38	-
Nhựa–cao su (HS39–40)	-	-	HS39	HS40
Da, gỗ, giấy... (HS41–49)	HS41, HS43	-	HS45, HS47, HS48, HS49	HS42, HS44, HS46
Dệt may (HS50–63)	HS52, HS59, HS60	HS53, HS56	HS51, HS57, HS58	HS50, HS54, HS55, HS61, HS62, HS63
Giày dép/mũ/ô... (HS64–67)	-	HS67	HS66	HS64, HS65
Đá–gốm–thủy tinh (HS68–70)	HS70	-	HS68	HS69
Kim loại (HS71–83)	HS72, HS78	HS71, HS80	HS73, HS74, HS75, HS76, HS79, HS81, HS82 HS83	-
Máy móc & thiết bị điện (HS84–85)	HS85	-	HS84	-
Phương tiện vận tải (HS86–89)	HS89	-	HS86, HS87, HS88	-
Thiết bị chính xác (HS90–92)	HS90	-	HS91, HS92	-
Khác (HS93–97)	-	-	HS95, HS97	HS94, HS96

Ghi chú: Tổng cộng 95 mã HS2 trong mẫu. Nhóm ngành được phân theo mã HS2; dấu “-” biểu thị không có mã thuộc nhóm đó.

Nguồn: Tính toán và tổng hợp của tác giả từ dữ liệu xuất khẩu của UNComtrade