

Đo lường độ phức tạp trong chuỗi thời gian của các cổ phiếu trong danh mục VN30: Tiếp cận bằng entropy hoán vị

Measuring the complexity of securities' time series in VN30 index: A permutation entropy approach

Trần Thị Tuấn Anh^{1*}

¹Trường Đại học Kinh tế Thành phố Hồ Chí Minh, Việt Nam

*Tác giả liên hệ, Email: anhttt@ueh.edu.vn

THÔNG TIN

TÓM TẮT

DOI:10.46223/HCMCOUJS.
econ.vi.14.1.493.2019

Ngày nhận: 29/08/2018

Ngày nhận lại: 04/11/2018

Duyệt đăng: 14/01/2019

Từ khóa:

chuỗi giá đóng cửa, chuỗi tỷ suất sinh lợi, entropy hoán vị, đo lường độ phức tạp của chuỗi thời gian, entropy hoán vị chuẩn hóa

Keywords:

daily close prices, daily stock returns, measuring the complexity of time series, normalized permutation entropy, permutation entropy

Bài viết áp dụng phương pháp Bandt và Pompe (2002) trên dữ liệu giá đóng cửa và tỷ suất sinh lợi hàng ngày của các cổ phiếu thuộc danh mục VN30 trên thị trường chứng khoán Việt Nam thu thập trong giai đoạn từ tháng 01/2000 đến tháng 08/2018. Kết quả tính toán cho thấy entropy hoán vị chuẩn hóa của chuỗi giá đóng cửa các cổ phiếu không gần 1, nghĩa là biến động của các chuỗi chưa thực sự ngẫu nhiên, còn có tính hình mẫu và có thể dự đoán một phần bằng các hình mẫu hoán vị. Nếu xét theo entropy hoán vị chuẩn hóa của chuỗi tỷ suất sinh lợi thì có thể thấy biến động của chuỗi tỷ suất sinh lợi mang tính ngẫu nhiên nhiều hơn, và vì thế ít có thể dự đoán được hơn so với chuỗi giá. Ngoài ra, entropy hoán vị của chuỗi giá giải thích tốt hơn cho biến động của tỷ suất sinh lợi trung bình, so với dùng entropy của chuỗi tỷ suất sinh lợi và cũng tốt hơn so với khi dùng độ đo rủi ro thông thường là độ lệch chuẩn của tỷ suất sinh lợi.

ABSTRACT

The paper applies the Bandt and Pompe (2002) method to measure the complexity of daily close price and daily return of VN30's stocks during the period from January 2000 to August 2018. The fact that normalized permutation entropy of daily close price series is quite far from 1 implies the stocks' close price is not a pure random walk process. It contains permutation patterns and can be partially predicted. In consideration of normalized permutation entropy of daily returns series, it is noticeable that the fluctuation of daily return series is more random and thus, more unpredictable compared to that of the price series. Additionally, the permutation entropy of the price series better explains the fluctuation of average returns. It is also better compared to using entropy of daily returns series and to using normal risks as returns' standard deviation.

1. Giới thiệu

Đo lường độ phức tạp của chuỗi thời gian là một trong những vấn đề được quan tâm nghiên cứu của chuỗi thời gian, đặc biệt là các chuỗi thời gian tài chính. Bandt và Pompe (2002) đề cập đến nhiều cách tiếp cận đã được sử dụng để đo lường mức độ phức tạp của chuỗi thời gian thông qua việc phân tích tính định kỳ, tính trật tự hoặc phân tích tính ngẫu nhiên của chuỗi. Theo các hướng tiếp cận trên, các đại lượng thường được dùng là các entropy, số chiều fractal (*fractal dimension*), và số mũ Lyapunov (*Lyapunov exponents*).

Riedl, Muller, và Wessel (2013) nhận định rằng việc vận dụng entropy để đo lường độ phức tạp của chuỗi thời gian ngày càng trở nên phổ biến với nhiều cách khác nhau. Tiêu biểu có thể kể đến việc sử dụng các đại lượng như Shannon entropy (Shannon, 1948), Kolmogorov - Sinai (Kantz & Schreiber, 2003), entropy xấp xỉ (*approximate entropy*, Pincus (1991)) ... Tuy nhiên, nhược điểm của các phương pháp trên là bỏ qua tính thứ tự của chuỗi thời gian và đôi khi đòi hỏi độ dài chuỗi thời gian phải rất lớn để kết quả tính toán được chính xác cộng với việc tính toán đôi khi phức tạp.

Bandt và Pompe (2002) đề xuất kết hợp tính toán entropy dựa trên việc biểu tượng hóa các chuỗi thời gian với việc duy trì được tính thứ tự trong chuỗi để tạo ra một đại lượng entropy mới dùng để đo lường độ phức tạp của chuỗi thời gian, gọi là entropy hoán vị. Ngoài ưu điểm về sự kết hợp giữa entropy truyền thống và việc biểu tượng hóa chuỗi ký tự nhưng vẫn quan tâm đến tính thứ tự của chuỗi, entropy hoán vị còn có ưu thế về sự đơn giản trong tính toán và thuyết phục trong lập luận. Sau khi được giới thiệu năm 2002 cho đến nay, entropy hoán vị ngày càng được vận dụng nhiều trong các nghiên cứu có liên quan đến đo lường sự phức tạp, độ biến động, tính ngẫu nhiên cũng như đo lường khả năng dự báo của chuỗi thời gian. Thông qua đó, entropy hoán vị cũng là một công cụ hữu hiệu để đo lường mức độ hiệu quả của thị trường chứng khoán, thị trường vàng, tỷ giá hối đoái hoặc các chuỗi thời gian tài chính khác.

Bài viết này dựa theo phương pháp tính toán entropy hoán vị do Bandt và Pompe (2002) đề xuất để đo lường mức độ phức tạp trong chuỗi tỷ suất sinh lợi và chuỗi giá của các chứng khoán thuộc danh mục VN30 bằng entropy hoán vị và từ đó so sánh độ phức tạp giữa các chuỗi chứng khoán trong danh mục này. Bên cạnh đó, bài viết xem xét mối liên hệ giữa độ phức tạp của chuỗi đo lường bằng entropy hoán vị với tỷ suất sinh lợi hàng ngày trung bình, từ đó cho thấy khả năng dự báo được tỷ suất sinh lợi của chứng khoán dựa trên entropy hoán vị. Với mục tiêu như trên, các phần còn lại của bài viết được tổ chức như sau: Mục 2 giới thiệu một số các nghiên cứu liên quan đến việc ứng dụng entropy trong đo lường độ phức tạp chuỗi thời gian; Mục 3 giới thiệu dữ liệu sử dụng và phương pháp tính toán entropy hoán vị; Mục 4 trình bày và thảo luận kết quả nghiên cứu và Mục 5 nêu kết luận và một số hàm ý rút ra từ kết quả nghiên cứu.

2. Tổng quan lý thuyết

Có ba hướng tiếp cận chính trong việc đo lường độ phức tạp của chuỗi thời gian. Hướng tiếp cận thứ nhất sử dụng độ đo độ phức tạp Lempel và Ziv (1976). Hướng tiếp cận thứ hai sử dụng các độ đo độ phức tạp dựa trên đại lượng entropy thông qua tính ngẫu nhiên của chuỗi; và hướng tiếp cận thứ ba dựa trên số mũ Lyapunov.

Lempel và Ziv (1976) đề xuất một thước đo giúp đánh giá mức độ phức tạp chuỗi thời gian thông qua đánh giá tính ngẫu nhiên của một chuỗi thời gian hữu hạn. Đại lượng này liên quan số bước mà quá trình tự phân định diễn ra ứng với các chuỗi giả định có một độ dài cho trước. Độ đo này được tính toán dựa trên số lần xảy ra và tỷ lệ lặp lại các chuỗi con trong dãy

số thời gian xem xét. Lempel và Ziv (1976) cũng ứng dụng đại lượng này để phân tích tín hiệu y sinh học như nhịp tim, nhịp thở, nồng độ oxy trong máu.

Các độ đo độ phức tạp dựa trên entropy cũng ngày càng trở nên phổ biến cùng với sự phát triển của kinh tế học vật lý (*econophysics*). Pincus (1991) đề xuất sử dụng đại lượng entropy xấp xỉ (*approximate entropy*) trong khi Richman và Moorman (2000) đề xuất sử dụng entropy mẫu (*sample entropy*). Hai đại lượng entropy này đều dựa trên tính quy tắc của một chuỗi thời gian bằng cách đánh giá sự xuất hiện của các mẫu lặp đi lặp lại. Khác với entropy xấp xỉ, entropy mẫu loại trừ lần đầu tiên xuất hiện của một hình mẫu. Entropy mẫu có tốc độ hội tụ ổn định cao hơn và ít phụ thuộc vào độ dài của chuỗi thời gian hơn entropy xấp xỉ.

Cách tiếp cận bằng số mũ của Lyapunov đo lường tính phức tạp của chuỗi thời gian thông qua xác định số mũ cho các mô hình động khác nhau, bao gồm các mô hình thời gian rời rạc, liên tục, xác định và ngẫu nhiên, có thể áp dụng cho cả hai hệ thống đơn giản chỉ với một vài bậc tự do và hệ thống phức tạp.

Trong ba hướng tiếp cận trên, cách tiếp cận đo lường độ phức tạp của chuỗi thời gian bằng entropy ngày càng được mở rộng do ngày càng có nhiều nghiên cứu phát hiện ra sự tương đồng giữa các chuỗi biến động trong vật lý với chuỗi biến động trong kinh tế. Mỗi đại lượng đo lường độ phức tạp dựa trên entropy đều có những ưu nhược điểm riêng. Bài viết này dựa trên entropy hoán vị do Bandt và Pompe (2002) đề xuất để đo lường và so sánh độ phức tạp của các chuỗi giá và tỷ suất sinh lợi của các cổ phiếu thuộc danh mục VN30. Mặc dù khái niệm về entropy hoán vị đã xuất hiện từ năm 2002 nhưng việc sử dụng thử nghiệm entropy hoán vị vào các chuỗi thời gian tài chính ở Việt Nam gần như chưa có một nghiên cứu nào thực hiện và công bố.

3. Dữ liệu và phương pháp nghiên cứu

3.1. Dữ liệu

Bảng 1

Danh sách các mã cổ phiếu thuộc danh mục VN30

STT	Mã cổ phiếu	STT	Mã cổ phiếu	STT	Mã cổ phiếu
1	BMP	11	HSG	21	SAB
2	CII	12	KDC	22	SBT
3	CTD	13	MBB	23	SSI
4	CTG	14	MSN	24	STB
5	DHG	15	MWG	25	VCB
6	DPM	16	NVL	26	VIC
7	FPT	17	PLX	27	VJC
8	GAS	18	PNJ	28	VNM
9	GMD	19	REE	29	VPB
10	HPG	20	ROS	30	VRE

Nguồn: Kết quả phân tích dữ liệu của nhóm nghiên cứu

Bài viết sử dụng dữ liệu chỉ số chứng khoán thu thập theo ngày tại thị trường chứng khoán Việt Nam của các chứng khoán thuộc danh mục VN30, bao gồm các chứng khoán được liệt kê ở Bảng 1 trong giai đoạn từ 02/01/2000 đến ngày 25/8/2018, với khoảng 4300 ngày giao dịch được ghi nhận.

Dữ liệu giá chứng khoán theo ngày, ký hiệu là P_t , được sử dụng để tính tỷ suất sinh lợi theo ngày dạng logarit, ký hiệu là R_t . Cả hai chuỗi chỉ số chứng khoán và tỷ suất sinh lợi đều được sử dụng để thực hiện mục tiêu nghiên cứu của bài viết, nghĩa là được sử dụng để tính toán entropy hoán vị của chuỗi và từ đó kiểm tra mức độ hiệu quả thông tin của thị trường cũng như đo lường mối liên hệ giữa tỷ suất sinh lợi với rủi ro. Các đồ thị thể hiện biến động của chỉ số chứng khoán và tỷ suất sinh lợi được thực hiện bằng Stata, trong khi tính toán entropy hoán vị được lập trình trên phần mềm Python.

3.2. Phương pháp nghiên cứu

Bandt và Pompe (2002) kế thừa và phát triển ý tưởng về việc biểu tượng hóa chuỗi thời gian nhưng không chỉ sử dụng 2 ký hiệu biểu tượng 0 và 1 tương ứng với khi chuỗi tăng hoặc giảm và sử dụng số ký tự biểu tượng tương ứng với độ dài chuỗi được xét Rizzo (2008). Cụ thể, khi xét độ dài D , sẽ có tương ứng D chữ số tự nhiên, bao gồm $1, 2, \dots, D$, được dùng để biểu tượng hóa chuỗi. Với độ trễ τ , Một khung cửa sổ độ dài D với các phần tử $(x_s, x_{s+\tau}, \dots, x_{s+(D-2)\tau}, x_{s+(D-1)\tau})$, trong đó $s=1, \dots, n - (D-1)\tau - 1$; sẽ trượt từ đầu chuỗi đến cuối dãy số thời gian. Tại mỗi vị trí s , tùy theo thứ tự về độ lớn giữa các phần tử nằm trong khung cửa sổ mà bộ các phần tử đó sẽ được xác định ứng với hoán vị nào trong số hoán vị của D số tự nhiên. Sau khi hoàn tất dịch chuyển cửa sổ từ đầu chuỗi đến cuối chuỗi, mỗi hoán vị sẽ được đếm số lần xuất hiện và từ đó tính ra tần suất xuất hiện của hoán vị.

Xét một chuỗi thời gian $\{x_t : t = 1, \dots, n\}$, với độ dài D và độ trễ τ cho trước, các bước để thực hiện tính toán entropy hoán vị theo cách của Bandt và Pompe (2002) như sau:

- (1.) Chọn độ trễ D ; xác định và sắp xếp thứ tự tất cả các hoán vị của các số tự nhiên $1, 2, \dots, D$. Có tất cả $D!$ hoán vị. Ký hiệu các hoán vị tương ứng là π_j với $j = 1, 2, \dots, D!$.
- (2.) Khởi tạo $s = 1$, tương ứng với việc bắt đầu từ vị trí đầu tiên của chuỗi thời gian.
- (3.) Chọn ra một tập con của chuỗi thời gian ban đầu gồm các phần tử ở các vị trí $s, s+\tau, \dots, s+(D-1)\tau$. Ký hiệu tập con đó gồm các phần tử là $(x_s, x_{s+\tau}, \dots, x_{s+(D-2)\tau}, x_{s+(D-1)\tau})$.
- (4.) Tùy theo thứ tự lớn và nhỏ giữa các phần tử trong tập con $(x_s, x_{s+\tau}, \dots, x_{s+(D-2)\tau}, x_{s+(D-1)\tau})$ mà xác định hoán vị π_j tương ứng với tập con này. Từ đó tăng bộ đếm của hoán vị π_j lên một đơn vị.
- (5.) Nếu $s+(D-1)\tau < n$, tiếp tục tăng s lên một đơn vị và lặp lại từ bước 3. Nếu $s+(D-1)\tau \geq n$ thì chuyển sang bước 6
- (6.) Tính toán tần suất xuất hiện p_j của từng hoán vị $\pi_j, j = 1, 2, \dots, D!$.

- (7.) Tính toán entropy hoán vị theo công thức
$$H_p = - \sum_{j=1}^{D!} p(\pi_j) \ln(p(\pi_j)) \quad (1)$$

Entropy theo bước 7 thực chất là Shannon entropy nhưng được tính dựa trên phân phối xác suất của các hoán vị nên đại lượng tính được ở bước 7 được gọi là entropy hoán vị. Nếu chuỗi có độ phức tạp càng cao, các giá trị trong chuỗi càng xuất hiện ngẫu nhiên và khó đoán trước, không thể biết được hoán vị nào sẽ chiếm ưu thế và thường xuyên xảy ra nhiều hơn, và vì vậy xác suất xuất hiện của các hoán vị trở nên gần như bằng nhau. Khi đó entropy hoán vị càng lớn. Ngược lại, nếu biến động của chuỗi đơn giản, sẽ có những hình mẫu biến động thường xuyên xảy ra, do vậy sẽ có những hoán vị xác suất xảy ra cao hơn những hoán vị khác, khi đó

entropy hoán vị tính toán được sẽ càng nhỏ. Trường hợp đơn giản nhất là chuỗi thời gian đơn điệu tăng (hoặc đơn điệu giảm) thì entropy hoán vị sẽ bằng 0 do chỉ có duy nhất một hình mẫu xảy ra và xác suất của hình mẫu đó bằng 1.

Trong một số nghiên cứu, entropy hoán vị còn được chuẩn hóa. Entropy hoán vị chuẩn hóa (normalized Shannon entropy) được tính toán bằng cách chia entropy hoán vị ở bước 7 cho giá trị lớn nhất có thể đạt được của H_p , đó là khi chuỗi thời gian đạt được trạng thái được cho là phức tạp và khó dự đoán nhất khi mà các hoán vị đều có khả năng xuất hiện như nhau. Giá trị lớn nhất chính là $H_{max} = \ln D!$. Khi đó, công thức entropy chuẩn hóa là:

$$H_N = \frac{H_p}{H_{max}} \quad (2)$$

Sau khi chuẩn hóa, $0 \leq H_N \leq 1$. Nếu chuỗi thời gian có độ phức tạp càng lớn, tính ngẫu nhiên trong biến động của chuỗi càng nhiều. Bài viết này tính toán cả hai trường hợp entropy hoán vị và entropy hoán vị chuẩn hóa để người đọc dễ dàng so sánh độ phức tạp của các chuỗi chứng khoán được xem xét. Ngoài ra, bài viết sử dụng entropy hoán vị tính được ở bước 7 để thực hiện hồi quy kiểm định mối liên hệ giữa entropy hoán vị với tỷ suất sinh lợi của cổ phiếu cũng như tìm ra độ dài hoán vị phù hợp nhất để có được entropy hoán vị giúp dự toán tỷ suất sinh lợi.

Bài viết thực hiện tính toán entropy hoán vị lần lượt với các độ dài $D = 4, 5, 6$ và độ trễ $\tau = 1$. Các tính toán về xác suất hoán vị và entropy hoán vị được thực hiện trên phần mềm Python kết hợp với đồ thị chuỗi chứng khoán và thống kê mô tả được thực hiện bằng Stata. Theo định nghĩa, chuỗi thời gian có entropy hoán vị (hoặc entropy hoán vị chuẩn hóa) càng lớn thì độ phức tạp của chuỗi càng tăng và ngược lại. Độ phức tạp của chuỗi càng tăng hàm ý rằng tính ngẫu nhiên của chuỗi càng cao và càng khó dự đoán giá trị tương lai của chuỗi.

4. Kết quả nghiên cứu và thảo luận

4.1. Kết quả nghiên cứu

Bảng 2 mô tả tỷ suất sinh lợi hàng ngày trung bình của các chứng khoán thuộc danh mục VN30 trong thời gian từ tháng 01/2000 đến tháng 08/2018. Theo kết quả mô tả này, cổ phiếu ROS có tỷ suất sinh lợi hàng ngày trung bình cao nhất với con số là 0,24%, tiếp theo là cổ phiếu SAB với tỷ suất sinh lợi trung bình là 0,11%. Có hơn một nửa các cổ phiếu thuộc danh mục VN30 có tỷ suất sinh lợi trung bình mang dấu âm trong thời gian dữ liệu được thu thập.

Bảng 2 cũng cho thấy độ lệch chuẩn tương ứng của chuỗi tỷ suất sinh lợi của các cổ phiếu trong danh sách VN30. Độ lệch chuẩn là một trong những đại lượng truyền thống dùng để đo lường độ biến động của các chuỗi thời gian. Theo kết quả mô tả, cổ phiếu VPB là cổ phiếu có độ lệch chuẩn cao nhất, lên đến 0.25%; kế tiếp là cổ phiếu VRE với độ lệch chuẩn là 0.21% trong khi cổ phiếu có độ biến động đo bằng độ lệch chuẩn thấp nhất là DPM với 0.03%. Thông thường, khi độ biến động của chứng khoán càng lớn, rủi ro khi đầu tư vào cổ phiếu càng cao thì tỷ suất sinh lợi kỳ vọng càng cao. Theo số liệu mô tả ở Bảng 2, hệ số tương quan giữa chuỗi tỷ suất sinh lợi và độ lệch chuẩn mang dấu dương nhưng không có ý nghĩa thống kê. Kết quả này là dấu hiệu ban đầu cho thấy có thể độ lệch chuẩn chưa phải là một công cụ đo lường rủi ro tốt đối với số liệu các cổ phiếu thu thập được trên thị trường chứng khoán Việt Nam.

Bảng 2

Thống kê mô tả tỷ suất sinh lợi

Stt	Chứng khoán	Tỷ suất sinh lợi trung bình	Xếp hạng	Độ lệch chuẩn
1	BMP	-0,0016	15	0,0522
2	CII	-0,0213	20	0,0483
3	CTD	0,0294	11	0,0558
4	CTG	-0,0189	19	0,0453
5	DHG	-0,0419	23	0,0564
6	DPM	-0,0624	27	0,0394
7	FPT	-0,0755	29	0,0464
8	GAS	0,0580	5	0,0559
9	GMD	-0,0084	18	0,0405
10	HPG	-0,0447	24	0,0552
11	HSG	-0,0450	25	0,0618
12	KDC	-0,0214	21	0,0469
13	MBB	0,0314	10	0,0422
14	MSN	0,0338	9	0,0497
15	MWG	0,0384	7	0,1090
16	NVL	0,0195	12	0,0987
17	PLX	0,0981	3	0,1471
18	PNJ	0,0413	6	0,0531
19	REE	0,0189	13	0,0407
20	ROS	0,2432	1	0,1738
21	SAB	0,1121	2	0,1040
22	SBT	-0,0075	17	0,0482
23	SSI	-0,0620	26	0,0548
24	STB	-0,0640	28	0,0477
25	VCB	0,0033	14	0,0459
26	VIC	-0,0043	16	0,0549
27	VJC	0,0913	4	0,1400
28	VNM	0,0357	8	0,0451
29	VPB	-0,1743	30	0,2581
30	VRE	-0,0325	22	0,2106
Hệ số tương quan giữa chuỗi tỷ suất sinh lợi và độ lệch chuẩn				0,0934 (0,6233)

p-value được thể hiện trong dấu ngoặc đơn

Nguồn: tác giả tính toán từ dữ liệu thu thập được

Sau khi mô tả dữ liệu, bài viết tiến hành tính toán entropy hoán vị với các độ dài $D = 4$, 5 và 6 với độ trễ $\tau = 1$. Kết quả entropy hoán vị với chuỗi giá đóng cửa chứng khoán được thể hiện ở Bảng 3 và kết quả hồi quy entropy hoán vị với chuỗi tỷ suất sinh lợi được thể hiện ở Bảng 4.

Bảng 3 thể hiện kết quả entropy hoán vị và entropy hoán vị chuẩn hóa cùng với xếp hạng tăng dần về độ phức tạp của chuỗi giá đóng cửa của 30 cổ phiếu ứng với $D = 4$ (ở cột 2,3, và 4), $D = 5$ (ở cột 5,6, và 7); $D = 6$ (ở cột 8, 9, và 10). Theo định nghĩa, giá trị của entropy

hoán vị càng lớn (hoặc entropy hoán vị chuẩn hóa càng gần 1) độ phức tạp của chuỗi càng cao và chuỗi càng khó dự đoán giá trị tương lai. Với nhận định này, có thể thấy cổ phiếu ROS của Công ty Cổ phần Xây dựng FLC FAROS là chuỗi có entropy hoán vị chuẩn hóa của giá đóng cửa nhỏ nhất ở cả ba độ dài D được xét, nghĩa là có độ phức tạp thấp nhất và dễ dự đoán giá nhất nếu dùng hình mẫu hoán vị. Chuỗi giá cổ phiếu có entropy hoán vị thấp tiếp theo là NVL của Công ty Cổ phần Tập đoàn Đầu tư Địa ốc No Va. Các giá trị entropy chuẩn hóa của giá đóng cửa các cổ phiếu này cách khá xa giá trị 1, cho thấy tính hình mẫu lặp lại của các cổ phiếu này ít mang tính ngẫu nhiên mà có tính quy luật rõ ràng hơn so với các cổ phiếu khác.

Độ phức tạp cao nhất dựa trên entropy hoán vị thuộc về cổ phiếu DPM của Tổng Công ty Phân bón và Hóa chất Dầu khí và cổ phiếu VCB của Ngân hàng TMCP Ngoại Thương Việt Nam. Ở độ trễ $D = 4$ và $D = 6$, cổ phiếu DPM có độ phức tạp cao hơn nhưng ở độ trễ $D = 5$, cổ phiếu VCB có độ phức tạp vượt hơn. Mặc dù hai cổ phiếu này được xem là có độ phức tạp cao nhất trong số 30 cổ phiếu nhưng giá trị entropy chuẩn hóa cũng chỉ ở mức 0,92 ($D = 4$) và 0,86 (ở mức $D = 6$) cho thấy chuỗi giá của các cổ phiếu này vẫn chưa thực sự ngẫu nhiên. Trong số các cổ phiếu có độ phức tạp cao nhất trong danh mục VN30 còn có cổ phiếu PNJ của Công ty Cổ phần Vàng bạc Đá quý Phú Nhuận và cổ phiếu CII của công ty Cổ phần Đầu tư Hạ tầng Kỹ thuật TP.HCM. Một kết quả cũng khá ngạc nhiên là việc cổ phiếu VNM của công ty cổ phần sữa Việt Nam thuộc vào nhóm cổ phiếu có độ phức tạp thấp, nhiều tính hình mẫu và ít có tính ngẫu nhiên so với các cổ phiếu khác.

Bảng 3

Entropy hoán vị với chuỗi giá chứng khoán thuộc danh mục VN30

Mã chứng khoán	D = 4			D = 5			D = 6		
	Entropy hoán vị	Entropy hoán vị chuẩn hóa	Xếp hạng	Entropy hoán vị	Entropy hoán vị chuẩn hóa	Xếp hạng	Entropy hoán vị	Entropy hoán vị chuẩn hóa	Xếp hạng
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
BMP	2,8442	0,8950	12	4,1559	0,8681	15	5,4987	0,8358	20
CII	2,9281	0,9213	28	4,2710	0,8921	27	5,6255	0,8550	27
CTD	2,8610	0,9002	15	4,1910	0,8754	20	5,4990	0,8358	21
CTG	2,9183	0,9183	27	4,2580	0,8894	26	5,5825	0,8485	26
DHG	2,7981	0,8804	3	4,0809	0,8524	8	5,3868	0,8187	10
DPM	2,9540	0,9295	30	4,3156	0,9014	29	5,6987	0,8662	30
FPT	2,8554	0,8985	14	4,1639	0,8697	16	5,4912	0,8346	18
GAS	2,9022	0,9132	25	4,2341	0,8844	25	5,5241	0,8396	23
GMD	2,8448	0,8951	13	4,1394	0,8646	14	5,4575	0,8295	16
HPG	2,8735	0,9042	17	4,1820	0,8735	18	5,4940	0,8350	19
HSG	2,8999	0,9125	24	4,2315	0,8839	24	5,5594	0,8450	24
KDC	2,8276	0,8897	9	4,1187	0,8603	11	5,4246	0,8245	13
MBB	2,8896	0,9092	20	4,1945	0,8761	22	5,4491	0,8282	14

Mã chứng khoán	D = 4			D = 5			D = 6		
	Entropy hoán vị	Entropy hoán vị chuẩn hóa	Xếp hạng	Entropy hoán vị	Entropy hoán vị chuẩn hóa	Xếp hạng	Entropy hoán vị	Entropy hoán vị chuẩn hóa	Xếp hạng
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
MSN	2,8010	0,8814	5	4,0815	0,8525	9	5,3733	0,8167	8
MWG	2,8827	0,9071	18	4,1865	0,8745	19	5,4181	0,8235	11
NVL	2,7152	0,8544	2	3,8372	0,8015	2	4,7580	0,7232	3
PLX	2,8974	0,9117	23	4,1162	0,8598	10	5,0305	0,7646	7
PNJ	2,9157	0,9175	26	4,2747	0,8929	28	5,6404	0,8573	28
REE	2,8384	0,8931	10	4,1267	0,8620	13	5,4499	0,8283	15
ROS	2,2981	0,7231	1	3,1468	0,6573	1	3,8718	0,5885	1
SAB	2,8091	0,8839	6	4,0045	0,8364	5	4,9060	0,7457	5
SBT	2,8946	0,9108	22	4,2268	0,8829	23	5,5667	0,8461	25
SSI	2,8860	0,9081	19	4,1930	0,8758	21	5,5029	0,8364	22
STB	2,8619	0,9005	16	4,1725	0,8715	17	5,4865	0,8339	17
VCB	2,9527	0,9291	29	4,3164	0,9016	30	5,6848	0,8640	29
VIC	2,8256	0,8891	8	4,1200	0,8606	12	5,4186	0,8236	12
VJC	2,8160	0,8861	7	4,0366	0,8432	6	4,9675	0,7550	6
VNM	2,7995	0,8809	4	4,0809	0,8524	7	5,3837	0,8183	9
VPB	2,8395	0,8935	11	3,9616	0,8275	3	4,7199	0,7174	2
VRE	2,8924	0,9101	21	3,9931	0,8341	4	4,7767	0,7260	4

Nguồn: Tác giả tính toán từ dữ liệu thu thập được

Tương tự như Bảng 3, Bảng 4 trình bày kết quả entropy hoán vị, entropy hoán vị chuẩn hóa, và xếp hạng về mức độ phức tạp theo chiều tăng dần của chuỗi tỷ suất sinh lợi các cổ phiếu thuộc danh mục VN30. Khi xét theo entropy hoán vị chuẩn hóa của chuỗi tỷ suất sinh lợi, việc lựa chọn độ trễ D lại có ảnh hưởng rất lớn đến kết quả tính toán entropy hoán vị. Kết quả ứng với các độ trễ khác nhau lại rất khác nhau. Tuy nhiên, với số liệu thể hiện ở cột (3), cột (6), và cột (9), có thể thấy cổ phiếu VJC của công ty Cổ phần Hàng không VietJet có chuỗi tỷ suất sinh lợi theo hình mẫu hoán vị rõ rệt nhất; tiếp theo có thể nói đến cổ phiếu VRE của công ty Cổ phần Vincom Retail.

Các cổ phiếu trong nhóm những cổ phiếu có độ phức tạp cao nhất tính bằng entropy hoán vị cũng có thay đổi rõ rệt. Nhóm ba cổ phiếu có entropy hoán vị cao ứng với ba độ dài D bao gồm cổ phiếu VCB (Ngân hàng TMCP Ngoại Thương Việt Nam); cổ phiếu HGS (Công ty Cổ phần Tập đoàn Hoa Sen) và cổ phiếu CII (Công ty Cổ phần Đầu tư Hạ tầng Kỹ thuật TP.HCM). Giá trị của entropy hoán vị chuẩn hóa của các cổ phiếu này nhìn chung lên đến 0,99; các giá trị này rất gần giá trị lớn nhất là 1. Điều này cho thấy chuỗi tỷ suất sinh lợi có tính ngẫu nhiên cao hơn rõ rệt so với chuỗi giá, khó đoán được diễn biến hơn do đó độ phức tạp đo bằng entropy hoán vị tính được cũng cao hơn rất nhiều. Kết quả này cũng cho thấy rằng việc dự đoán giá chứng khoán

theo hình mẫu hoán vị sẽ dễ dàng hơn so với dự đoán tỷ suất sinh lợi của các cổ phiếu vì tính ngẫu nhiên của chuỗi giá thấp hơn, và vì vậy độ phức tạp của chuỗi giá cũng ít hơn.

Bảng 4

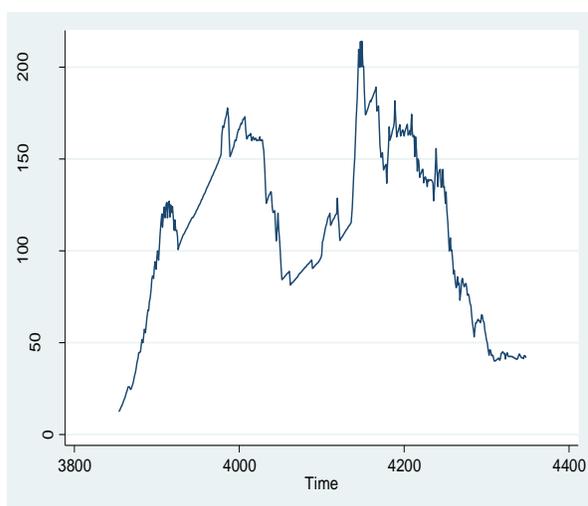
Entropy hoán vị trên chuỗi tỷ suất sinh lợi của các cổ phiếu thuộc danh mục VN30

Mã chứng khoán	D = 4			D = 5			D = 6		
	Entropy hoán vị	Entropy hoán vị chuẩn hóa	Xếp hạng	Entropy hoán vị	Entropy hoán vị chuẩn hóa	Xếp hạng	Entropy hoán vị	Entropy hoán vị chuẩn hóa	Xếp hạng
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
BMP	3,1397	0,9879	9	4,7099	0,9838	13	6,3549	0,9659	17
CII	3,1600	0,9943	25	4,7406	0,9902	26	6,4140	0,9749	30
CTD	3,1589	0,9940	23	4,7224	0,9864	19	6,2844	0,9552	10
CTG	3,1675	0,9967	28	4,7450	0,9911	27	6,3753	0,9690	21
DHG	3,1454	0,9897	13	4,7052	0,9828	12	6,3317	0,9624	14
DPM	3,1540	0,9924	19	4,7316	0,9883	22	6,3840	0,9703	25
FPT	3,1490	0,9909	17	4,7220	0,9863	18	6,3830	0,9702	24
GAS	3,1710	0,9978	30	4,7510	0,9924	29	6,3002	0,9576	12
GMD	3,1588	0,9939	22	4,7315	0,9883	21	6,3970	0,9723	26
HPG	3,1585	0,9939	21	4,7367	0,9894	23	6,3760	0,9691	22
HSG	3,1671	0,9966	27	4,7487	0,9919	28	6,4000	0,9728	27
KDC	3,1479	0,9905	15	4,7239	0,9867	20	6,3718	0,9685	20
MBB	3,1483	0,9906	16	4,6886	0,9794	8	6,2064	0,9433	9
MSN	3,1364	0,9869	7	4,6958	0,9809	10	6,2959	0,9569	11
MWG	3,1473	0,9903	14	4,7036	0,9825	11	6,1507	0,9349	8
NVL	3,1266	0,9838	4	4,5648	0,9535	5	5,6223	0,8546	6
PLX	3,1309	0,9851	5	4,5798	0,9566	6	5,4979	0,8356	4
PNJ	3,1596	0,9942	24	4,7398	0,9900	25	6,3659	0,9676	19
REE	3,1410	0,9883	10	4,7116	0,9842	14	6,3815	0,9699	23
ROS	3,0958	0,9741	2	4,5111	0,9423	3	5,5782	0,8479	5
SAB	3,1648	0,9958	26	4,6468	0,9706	7	5,6613	0,8605	7
SBT	3,1449	0,9896	11	4,7124	0,9843	15	6,3399	0,9636	15
SSI	3,1585	0,9939	20	4,7397	0,9900	24	6,4037	0,9733	29
STB	3,1531	0,9922	18	4,7179	0,9855	17	6,3600	0,9667	18
VCB	3,1705	0,9976	29	4,7566	0,9936	30	6,4006	0,9728	28
VIC	3,1450	0,9896	12	4,7141	0,9847	16	6,3546	0,9659	16
<i>VJC</i>	<i>3,0736</i>	<i>0,9671</i>	<i>1</i>	<i>4,4631</i>	<i>0,9322</i>	<i>2</i>	<i>5,4497</i>	<i>0,8283</i>	<i>3</i>
VNM	3,1359	0,9867	6	4,6912	0,9799	9	6,3102	0,9591	13
VPB	3,1380	0,9874	8	4,5506	0,9505	4	5,2856	0,8034	2

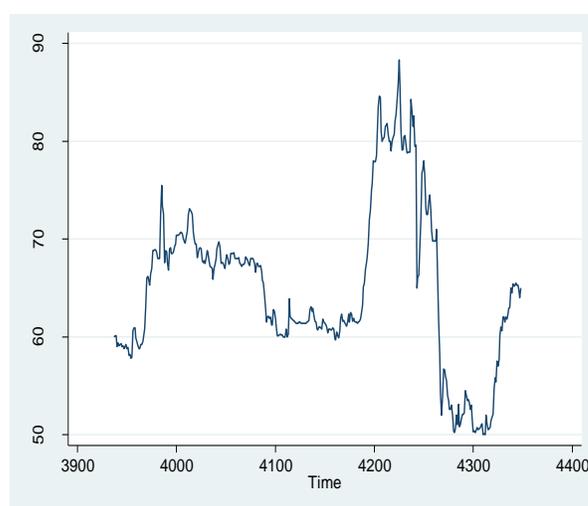
Mã chứng khoán	D = 4			D = 5			D = 6		
	Entropy hoán vị	Entropy hoán vị chuẩn hóa	Xếp hạng	Entropy hoán vị	Entropy hoán vị chuẩn hóa	Xếp hạng	Entropy hoán vị	Entropy hoán vị chuẩn hóa	Xếp hạng
VRE	3,0984	0,9749	3	4,3536	0,9094	1	5,0189	0,7628	1

Nguồn: Tác giả tính toán từ dữ liệu thu thập được

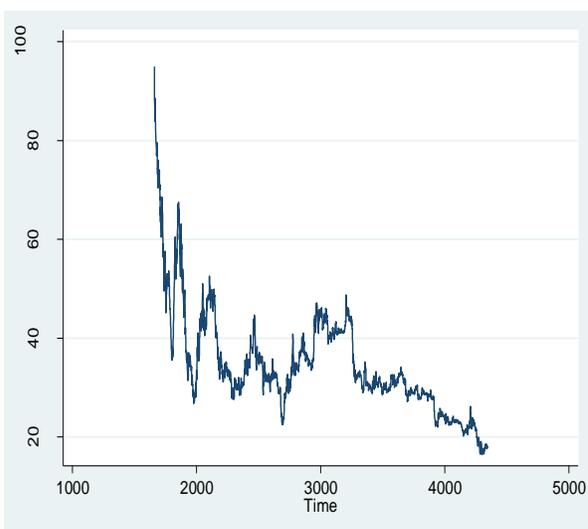
Hình 1 cho đến Hình 4 của bài viết thể hiện chuỗi giá đóng cửa những cổ phiếu nằm trong nhóm có độ phức tạp cao nhất và thấp nhất tính bằng entropy hoán vị như cổ phiếu ROS, cổ phiếu NVL, cổ phiếu VCB, và cổ phiếu DPM. Tuy nhiên, hình ảnh trên đồ thị không thể hiện được độ phức tạp của chuỗi thời gian vì xu hướng biến động của các cổ phiếu thể hiện trên đồ thị đều không có xu hướng rõ rệt. Vì vậy, việc sử dụng các đại lượng định lượng như entropy hoán vị là cần thiết để định lượng về độ phức tạp của chuỗi.



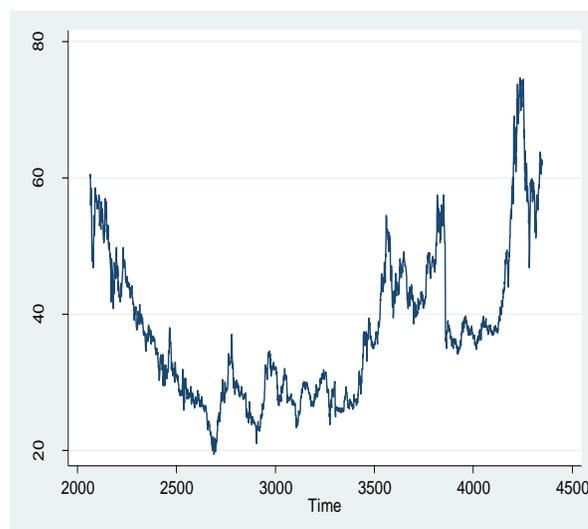
Hình 1. Giá đóng cửa của của cổ phiếu ROS



Hình 2. Giá đóng cửa của của cổ phiếu NVL



Hình 3. Giá đóng cửa của của cổ phiếu DPM



Hình 4. Giá đóng cửa của của cổ phiếu VCB

Đề đo lường mối liên hệ giữa tỷ suất sinh lợi và các entropy hoán vị, bài viết tiến hành hồi quy biến phụ thuộc là tỷ suất sinh lợi trung bình của cổ phiếu theo biến độc lập lần lượt là độ lệch chuẩn và các entropy hoán vị của chuỗi giá. Kết quả hồi quy cùng với hệ số xác định của mô hình được thể hiện trên Bảng 5.

Bảng 5

Kết quả hồi quy tỷ suất sinh lợi theo độ lệch chuẩn và theo các entropy hoán vị

Biến độc lập		Hệ số góc	Thông kê t	R ²
Độ lệch chuẩn		0.1241	0.5	0.0087
Giá đóng cửa	PerEn với D = 4	-0,3804***	-3,94	0,3572***
	PerEn với D = 5	-0,1909***	-3,49	0,3035***
	PerEn với D = 6	-0,09024***	-2,9	0,2305***
Tỷ suất sinh lợi	PerEn với D = 4	-1,2287*	-2,14	0,1405*
	PerEn với D = 5	-0,2027	-1,48	0,0723
	PerEn với D = 6	-0,0447	-1,35	0,0608

***, **, *: có ý nghĩa thống kê tương ứng ở mức ý nghĩa 1%, 5%, và 10%

Nguồn: tác giả tính toán từ dữ liệu thu thập được

Kết quả hồi quy cho thấy rằng mối liên hệ giữa tỷ suất sinh lợi với độ lệch chuẩn rất yếu ớt và không có ý nghĩa thống kê. Độ lệch chuẩn vẫn là đại lượng truyền thống để đo lường độ rủi ro trong cổ phiếu nhưng hệ số xác định của phương trình hồi quy tỷ suất sinh lợi trung bình theo độ lệch chuẩn rất thấp và không có ý nghĩa thống kê cho thấy việc đo lường rủi ro bằng độ lệch chuẩn không giải thích được cho biến động của tỷ suất sinh lợi trung bình của các cổ phiếu thuộc danh mục VN30. Trong khi đó, hệ số xác định khi hồi quy tỷ suất sinh lợi theo các entropy hoán vị cao hơn rất nhiều, đặc biệt là entropy hoán vị theo giá đóng cửa. Hệ số xác định cao nhất được ghi nhận với hàm hồi quy theo entropy hoán vị của giá đóng cửa với độ trễ 4, lên đến 35,72%. Tiếp theo là hệ số xác định của hàm hồi quy theo entropy hoán vị chuỗi giá với độ trễ D = 5 với R² là 30,35%.

4.2. Thảo luận kết quả nghiên cứu

Kết quả phân tích tỷ suất sinh lợi cho thấy hơn một nửa các cổ phiếu trong danh mục VN30 của thị trường chứng khoán Việt Nam có tỷ suất sinh lợi trung bình mang dấu âm trong giai đoạn dữ liệu được thu thập. Các entropy hoán vị tính toán theo chuỗi giá đóng cửa của các cổ phiếu khá xa giá trị 1 cho thấy tính không hiệu quả của thị trường. Ngoài ra, kết quả thực hiện hồi quy để định lượng mối liên hệ giữa rủi ro của chứng khoán và tỷ suất sinh lợi cho thấy rằng việc đo lường rủi ro bằng độ lệch chuẩn không giải thích được cho biến động của tỷ suất sinh lợi trung bình của các cổ phiếu thuộc danh mục VN30, trong khi đó, việc sử dụng các entropy hoán vị để đo lường rủi ro lại giải thích tốt hơn cho chuỗi tỷ suất sinh lợi.

5. Kết luận và gợi ý chính sách

Bài viết áp dụng phương pháp Bandt và Pompe trên dữ liệu giá đóng cửa và tỷ suất sinh lợi của các cổ phiếu thuộc danh mục VN30 trên thị trường chứng khoán Việt Nam. Kết quả tính toán entropy hoán vị và entropy hoán vị chuẩn hóa cho thấy mức độ phức tạp cũng như tính

ngẫu nhiên trong biến động của các cổ phiếu khác nhau rất khác nhau. Entropy hoán vị càng cao thì độ phức tạp của chuỗi càng tăng, tính ngẫu nhiên trong chuỗi càng rõ rệt và càng khó dự đoán diễn biến của chuỗi.

Nếu xét theo entropy hoán vị chuẩn hóa của chuỗi giá đóng cửa chứng khoán vì có thể thấy độ phức tạp của cổ phiếu ROS và cổ phiếu NVL là nhỏ nhất. Trong khi đó các cổ phiếu DPM, cổ phiếu VCB, cổ phiếu PNJ và cổ phiếu CII là những cổ phiếu có độ phức tạp của chuỗi giá đóng cửa cao nhất tính bằng entropy hoán vị. Tuy vậy, entropy hoán vị chuẩn hóa của các cổ phiếu vẫn không gần 1, cho thấy biến động của các chuỗi chưa thực sự ngẫu nhiên, còn có tính hình mẫu và có thể dự đoán một phần bằng các hình mẫu hoán vị.

Nếu xét theo entropy hoán vị chuẩn hóa của chuỗi tỷ suất sinh lợi thì có thể thấy biến động của chuỗi tỷ suất sinh lợi mang tính ngẫu nhiên nhiều hơn, và vì thế ít có thể dự đoán được hơn so với chuỗi giá. Điều này thể hiện ở việc entropy hoán vị chuẩn hóa của chuỗi tỷ suất sinh lợi ở hầu hết các cổ phiếu đều nhỏ hơn 1. Theo kết quả entropy hoán vị tính được, các cổ phiếu VCB, HGS và CII có độ phức tạp trong chuỗi giá chứng khoán cao nhất; trong khi các cổ phiếu có độ phức tạp chuỗi tỷ suất sinh lợi thấp nhất bao gồm VJC và VRE.

Ngoài ra, bài viết còn cho thấy kết quả rằng việc dự báo trên chuỗi giá cổ phiếu dựa trên hình mẫu hoán vị sẽ dễ dàng hơn rất nhiều so với dự báo chuỗi tỷ suất sinh lợi. Bên cạnh đó, entropy hoán vị của chuỗi giá giải thích tốt hơn cho biến động của tỷ suất sinh lợi trung bình so với dùng entropy của chuỗi tỷ suất sinh lợi hoặc dùng độ đo rủi ro thông thường là độ lệch chuẩn của tỷ suất sinh lợi. Thực nghiệm với số liệu VN30 còn cho thấy rằng rủi ro đo lường bằng độ lệch chuẩn không giải thích được cho biến động của tỷ suất sinh lợi trung bình.

Từ kết quả nghiên cứu đã cho, có thể thấy rằng nhà đầu tư có thể dùng hình mẫu hoán vị để dự đoán xu hướng biến động trên giá cổ phiếu và nên đưa ra các dự đoán bằng tính hình mẫu hoán vị với chuỗi giá để khả năng dự đoán thành hiện thực sẽ cao hơn so với dự đoán trên chuỗi tỷ suất sinh lợi. Dựa vào tính hình mẫu, nhà đầu tư cũng có thể xây dựng chiến lược đầu tư cho riêng mình. Ngoài ra, nhà đầu tư còn có thể dựa trên entropy hoán vị của giá đóng cửa các cổ phiếu để lựa chọn danh mục đầu tư phù hợp.

Tài liệu tham khảo

- Bandt, C., & Pompe, B. (2002). Permutation entropy: A natural complexity measure for time series. *Physical Review Letters*, 88(17), 174102.
- Kantz, H., & Schreiber, T. (2003). *Nonlinear time series analysis*. Cambridge, UK: Cambridge University Press.
- Lempel, A., & Ziv, J. (1976). On the complexity of finite sequences. *IEEE Transactions on Information Theory*, 22(1), 75-81.
- Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 88(6), 2297-2301.
- Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate and sample entropy. *American Journal of Physiology: Heart And Circulatory Physiology*, 278(6), 2039-2049.

- Riedl, M., Muller, A., & Wessel, N. (2013). Practical considerations of permutation entropy. *The European Physical Journal Special Topics*, 222(2), 249-262.
- Risso, W. A. (2008). The informational efficiency and the financial crashes. *Research in International Business and Finance*, 22(3), 396-408.
- Shannon, C. E. A. (1948). Mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423.