

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG HCM
KHOA TOÁN - TIN HỌC



BÁO CÁO ĐỀ TÀI
KHOA HỌC DỮ LIỆU

Đề tài:

CROSS SECTIONAL MOMENTUM

GVHD: TS. Tô Đức Khánh

Sinh viên thực hiện:

<i>Họ và tên</i>	<i>MSSV</i>	<i>Mã lớp</i>
Trần Minh Hiễn	21280018	21KDL1
Nguyễn Thị Yến Như	21280082	21KDL1
Huỳnh Lê Minh Thư	21280110	21KDL1

TPHCM, ngày 02 tháng 6 năm 2024

Mục lục

Lời nói đầu	3
Đặt vấn đề	4
1 Khái quát chung về TSM và CSM	6
1.1 Định nghĩa TSM và CSM	6
1.1.1 Time Series Momentum	6
1.1.2 Cross Sectional Momentum	6
1.2 So sánh CSM và TSM	9
2 Lý thuyết tài chính	10
2.1 Hiệu quả thị trường	10
2.2 Phản ứng thị trường	11
2.2.1 Over-reaction	11
2.2.2 Under-reaction	12
3 Lý thuyết toán học	13
3.1 Lợi suất của chiến lược TSM	13
3.1.1 Độ biến động	13
3.1.2 Lợi suất của chiến lược TSM	14
3.2 Lợi suất theo chiến lược CSM	14
3.2.1 Cross-sectional Momentum cổ điển	15
3.2.2 Regress Then Rank	17
3.2.3 Learning to Rank	18
4 Phương pháp xây dựng	23
4.1 Mục tiêu	23
4.2 Dữ liệu	23
4.3 Mô hình chiến lược	26
4.3.1 Chiến lược dựa theo momentum cổ điển	26
4.3.2 Chiến lược có sự kết hợp mô hình học máy	26
4.4 Mô tả Backtest	27
4.4.1 Cross Sectional Momentum Model Features	27
4.5 Kết quả và thảo luận	27
4.5.1 Thống kê kiểm định	27
4.5.2 So sánh chiến lược từ các phương pháp xếp hạng	31
4.5.3 Thông số đánh giá	33
4.5.4 So sánh chiến lược	35
5 Tổng Kết	37

BÁO CÁO ĐỀ TÀI

Ngày 12 tháng 1 năm 2025

Bảng 1: BẢNG THUẬT NGỮ

Viết tắt	Tiếng Anh	Tiếng Việt
TSM	Time Series Momentum	Động lượng chuỗi thời gian
CSM	Cross Sectional Momentum	Động lượng chéo
EMH	Efficient Market Hypothesis	Giả thuyết thị trường hiệu quả
LTR	Learning to Rank	Học để xếp hạng
	Overreaction	Phản ứng thái quá
	Underreaction	Phản ứng chậm

Lời nói đầu

Năm 1990, Jegadeesh đã giới thiệu khái niệm về động lượng (momentum) trong bài báo **”Evidence of Predictable Behavior of Security Returns.”** Jegadeesh cung cấp bằng chứng thực nghiệm cho thấy có sự tương quan giữa lợi nhuận cổ phiếu trong các khoảng thời gian dài, từ đó chứng minh được khả năng dự đoán lợi nhuận cổ phiếu. Sau khi Jegadeesh công bố vào năm 1990, **Jegadeesh và Titman (1993)** [11] đã phát triển thêm khái niệm động

lượng bằng cách giới thiệu động lượng theo mặt cắt ngang (**Cross-sectional momentum**). Trong các nghiên cứu, Jegadeesh và Titman xem xét mối quan hệ giữa lợi nhuận trong quá khứ và tương lai của các cổ phiếu trên thị trường chứng khoán Mỹ. Theo kết quả của Jegadeesh và Titman (1993) [11], chiến lược này mang lại lợi nhuận dương đáng kể, ngay cả sau khi đã tính đến các yếu tố rủi ro tiêu chuẩn và chi phí giao dịch.

Chiến lược CSM đã trở thành một trong những chủ đề quan trọng và có ảnh hưởng sâu rộng trong lĩnh vực tài chính, đặc biệt là trong quản lý danh mục đầu tư. CSM tập trung vào việc lựa chọn và so sánh hiệu suất giữa các tài sản trong cùng một khoảng thời gian, từ đó tối ưu hóa lợi nhuận bằng cách mua các tài sản có hiệu suất tốt và bán không các tài sản có hiệu suất kém. Chiến lược này không chỉ mang lại lợi nhuận vượt trội mà còn được chứng minh là có thể hoạt động hiệu quả ngay cả khi tính đến các yếu tố rủi ro và chi phí giao dịch. Những nghiên cứu này không chỉ cung cấp cái nhìn sâu sắc về cách thức hoạt động của CSM mà còn giúp các nhà đầu tư hiểu rõ hơn về mối tương quan giữa hiệu suất trong quá khứ và lợi nhuận tương lai. Với nền tảng lý thuyết vững chắc và minh chứng thực nghiệm mạnh mẽ, CSM đã trở thành một yếu tố quan trọng trong việc phân bổ tài sản và xây dựng danh mục đầu tư hiệu quả.

Bài nghiên cứu này sẽ tìm hiểu sâu hơn về Cross-Sectional Momentum, từ lý thuyết cơ bản cho đến các ứng dụng thực tế, cũng như đánh giá hiệu quả của chiến lược này trong bối cảnh đầu tư hiện đại.

Đặt vấn đề

Hiệu ứng động lượng đã được ghi nhận là một trong những hiện tượng tài chính quan trọng và được nghiên cứu rộng rãi trên nhiều thị trường chứng khoán. Tuy nhiên, khả năng áp dụng và hiệu quả của các chiến lược động lượng có thể thay đổi đáng kể khi thị trường chịu ảnh hưởng từ các yếu tố chính trị và kinh tế. Thị trường chứng khoán Mỹ, vốn là một trong những thị trường lớn và có tính thanh khoản cao nhất thế giới, thường trải qua những biến động mạnh do các thay đổi trong chính sách tài khóa, thương mại, và quy định.

Các nghiên cứu trước đây đã chỉ ra rằng sự không chắc chắn có thể tạo ra các biến động lớn trong ngắn hạn, ảnh hưởng đến xu hướng lợi nhuận của các cổ phiếu. Do đó, hiệu ứng động lượng – vốn dựa vào xu hướng hiệu suất trong quá khứ để dự đoán tương lai – có thể bị biến dạng hoặc trở nên khó dự đoán trong giai đoạn này. Cụ thể, chiến lược động lượng theo mặt cắt ngang (cross-sectional momentum), tập trung vào so sánh hiệu suất tương đối của các cổ phiếu trong cùng một khoảng thời gian, có thể gặp thách thức lớn trong việc duy trì hiệu suất ổn định khi sự thay đổi về chính sách và kỳ vọng thị trường ảnh hưởng đến các yếu tố cơ bản của từng cổ phiếu.

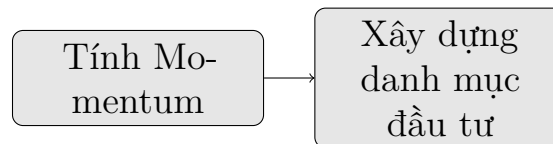
Vấn đề nghiên cứu là **liệu các chiến lược động lượng có thể duy trì hiệu quả trong bối cảnh thị trường bị ảnh hưởng bởi các sự kiện lớn hay không?** Cụ thể hơn, CSM có còn là một chiến lược hợp lý khi các loại tài sản có thể bị ảnh hưởng bởi những thay đổi đột ngột trong điều kiện thị trường? Nghiên cứu này sẽ tập trung phân tích hiệu quả của các chiến lược động lượng, so sánh chúng nhằm cung cấp một cái nhìn toàn diện hơn về cách thông qua nghiên cứu chỉ so sánh bốn phương thức: chỉ long, chỉ short, long short, và trọng số bằng nhau, kết hợp mô hình xếp hạng để tăng hiệu suất.

1 Khái quát chung về TSM và CSM

1.1 Định nghĩa TSM và CSM

1.1.1 Time Series Momentum

- **Time-series momentum (TSM)** là một khái niệm trong tài chính, đề cập đến hiện tượng mà lợi nhuận trong quá khứ của một tài sản có thể dự đoán lợi nhuận trong tương lai của chính nó. Nếu một tài sản có xu hướng tăng giá trong thời gian gần đây, nó có khả năng tiếp tục tăng trong tương lai gần; ngược lại, nếu giá của tài sản giảm trong thời gian gần đây, nó có xu hướng tiếp tục giảm.
- **Quy trình chiến lược:** gồm 2 bước chính

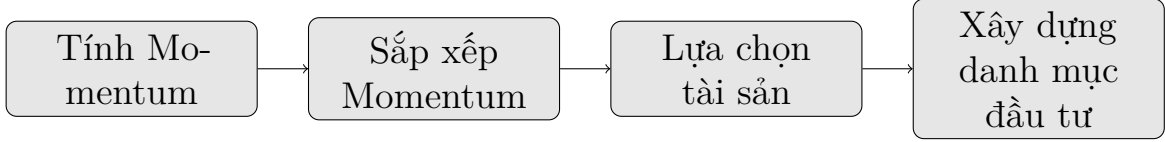


- **Tính Momentum:** ta sử dụng các giá trị lịch sử của tài sản đó để phân tích và đánh giá các đặc điểm của nó. Momentum được tính bằng cách so sánh giá hiện tại với giá trong quá khứ, từ đó đưa ra quyết định giao dịch cho tài sản dựa trên xu hướng giá của nó.
- **Xây dựng danh mục đầu tư:** Danh mục đầu tư được xây dựng bằng cách mua và bán danh mục đầu tư phụ liên quan được hình thành ở bước trước.

1.1.2 Cross Sectional Momentum

- **Cross Sectional Momentum(CSM)** trong tài chính liên quan đến việc phân tích và so sánh các chứng khoán hoặc tiền điện tử từ nhiều ngành và thị trường khác nhau tại một thời điểm cụ thể. Chiến lược này chú trọng vào việc mua các chứng khoán có hiệu suất cao nhất (cổ phiếu thắng) và bán các chứng khoán có hiệu suất kém hơn (cổ phiếu thua lỗ) từ danh sách được sắp xếp theo một số tiêu chí nhất định. Nghiên cứu của Jegadeesh và Titman (1993) [11] đã ghi nhận hiện tượng Cross Sectional Momentum, cho thấy rằng danh mục đầu tư bao gồm các chứng khoán có lợi nhuận cao nhất trong khoảng thời gian từ sáu đến mười hai tháng trước (cổ phiếu thắng) thường hoạt động tốt hơn so với danh mục chứa các chứng khoán có lợi nhuận thấp nhất (cổ phiếu thua lỗ) trong cùng một khoảng thời gian.

- Chiến lược này khác với Time-series momentum, có phạm vi hẹp hơn và thường chỉ tập trung vào một công cụ duy nhất. Moskowitz và các cộng sự (2012) [15] đã đề xuất khái niệm Time-series momentum, liên quan đến việc tổng hợp các chiến lược động lượng theo thang độ biến động đơn biến.
- Quy trình chiến lược:** gồm 4 bước chính



- **Tính Momentum:** Đối với một tập hợp các tài sản có kích thước n , mô hình sẽ phân tích và đánh giá các đặc điểm khác nhau của các tài sản trong danh mục, từ đó xác định Momentum cho mỗi tài sản dựa trên các tiêu chí đã được thiết lập. Momentum này sẽ được sử dụng để phân loại và quyết định xem tài sản nào nên được mua, bán hoặc giữ lại trong danh mục đầu tư.

$$Y_{\tau m}^{(i)} = f(u_{\tau m}^{(i)}), \quad (1.1)$$

Được cung cấp một vectơ đầu vào $u_{\tau m}^{(i)}$ cho tài sản i tại thời điểm τm , mô hình dự đoán của chiến lược f tính toán điểm số tương ứng $Y_{\tau m}^{(i)}$. Đối với một vũ trụ cắt ngang có kích thước $N_{\tau m}$ tại thời điểm τm , danh sách các điểm số của các tài sản được xem xét để giao dịch được biểu diễn bởi vectơ

$$Y_{\tau m} = \{Y_{\tau m}^{(1)}, \dots, Y_{\tau m}^{(N_{\tau m})}\}.$$

- **Sắp xếp Momentum:** sắp xếp lại vectơ Momentum theo một thứ tự nào đó.

$$Z_{\tau m}^{(i)} = R(Y_{\tau m})^{(i)}, \quad (1.2)$$

với $Z_{\tau m}^{(i)} \in \{1, \dots, N_{\tau m}\}$ là chỉ số vị trí của tài sản i sau khi áp dụng toán tử $R(\cdot)$ để sắp xếp các điểm số theo thứ tự tăng dần.

- **Lựa chọn tài sản:** là bước ngưỡng để giữ lại một phần tài sản, nhằm tạo thành danh mục đầu tư dài hạn hoặc ngắn hạn tương ứng. Các kích thước thường được sử dụng bao gồm 10%, 20% cho các tài sản trên cùng và dưới cùng hoặc toàn bộ tài sản trong danh mục đầu tư.

$$X_{\tau m}^{(i)} = \begin{cases} -1 & \text{nếu } Z_{\tau m}^{(i)} \leq \lfloor 0.1 \times N_{\tau m} \rfloor \\ 1 & \text{nếu } Z_{\tau m}^{(i)} > \lfloor 0.9 \times N_{\tau m} \rfloor \\ 0 & \text{khác} \end{cases} \quad (1.3)$$

Việc lựa chọn thường là một bước ngưỡng, trong đó một phần tài sản được giữ lại để tạo thành các danh mục đầu tư long/short tương ứng. Phương trình 1.3 giả định rằng chúng ta đang sử dụng các danh mục đầu tư có kích thước thập phân điển hình cho chiến lược (tức là top và bottom 10)

- **Xây dựng danh mục đầu tư:** Danh mục đầu tư được xây dựng bằng cách mua và bán danh mục đầu tư phụ liên quan được hình thành ở bước trước.

Hầu hết các chiến lược cross-sectional đều tuân theo quy trình này và thường tương tự trong ba bước cuối cùng. Tuy nhiên, chúng rất đa dạng trong việc lựa chọn mô hình dự đoán được sử dụng để tính toán Momentum, từ các phương pháp đơn giản (Jegadeesh & Titman, 1993) [11] đến các kiến trúc neural phức tạp dựa trên nhiều đầu vào vĩ mô (Gu et al., 2018) [9]. Những lựa chọn này đã dẫn đến sự phát triển của các mô hình có thể được chia thành ba loại: Momentum cổ điển (Classical Momentum), Phân tích rồi xếp hạng (Regress-then-Rank), và Học để xếp hạng (Learning to Rank - LTR).

1.2 So sánh CSM và TSM

	Cross Sectional Momentum	Time Series Momentum
Định nghĩa	So sánh hiệu suất của các tài sản khác nhau	Dựa trên hiệu suất quá khứ của chính tài sản đó
Hiệu suất	Tương đối (relative performance)	Tuyệt đối (absolute performance)
Chiến lược	Mua tài sản có hiệu suất tốt nhất và bán tài sản có hiệu suất kém nhất trong một nhóm tài sản	Mua tài sản có hiệu suất tốt và bán tài sản có hiệu suất kém dựa trên hiệu suất quá khứ của chính nó
Ứng dụng	Thường được sử dụng trong các chiến lược đầu tư cổ phiếu và các tài sản khác	Thường được sử dụng trong các chiến lược theo xu hướng (trend-following)
Ví dụ	Mua 10 cổ phiếu có hiệu suất tốt nhất và bán 10 cổ phiếu có hiệu suất kém nhất trong danh mục 100 cổ phiếu	Mua cổ phiếu A nếu nó đã tăng giá trong 12 tháng qua và bán nếu nó đã giảm giá
Rủi ro	Phụ thuộc vào sự khác biệt về hiệu suất giữa các tài sản	Phụ thuộc vào xu hướng thị trường của từng tài sản
Lợi ích	Tận dụng sự khác biệt về hiệu suất giữa các tài sản	Tận dụng xu hướng giá của từng tài sản
Mối quan hệ	Hai loại momentum đều thể hiện sự tự tương quan trong lợi nhuận. TSM có thể giải thích một phần cho CSM thông qua mối liên hệ giữa các yếu tố tự tương quan và cấu trúc tương quan trong lợi nhuận.	

2 Lý thuyết tài chính

2.1 Hiệu quả thị trường

EMH được giới thiệu bởi **Eugene Fama vào năm 1970** [6]. Theo giả thuyết này, các thị trường tài chính có hiệu quả ở chỗ giá tài sản phản ánh tất cả thông tin công khai có sẵn. Fama (1965) [5] chia sự hiệu quả của các thị trường tài chính thành ba dạng:

- **Hiệu quả yếu:** Giá tài sản hiện tại đã phản ánh thông tin từ giá và khối lượng giao dịch trong quá khứ, nên không thể dự đoán giá tương lai dựa vào thông tin lịch sử.
- **Hiệu quả bán mạnh:** Giá tài sản phản ánh tất cả thông tin công khai như báo cáo tài chính, thông báo lợi nhuận, và dữ liệu thị trường.
- **Hiệu quả mạnh:** Giá tài sản đã chứa tất cả thông tin, kể cả thông tin công khai và nội bộ.

Tổng thể, giả thuyết này cho rằng *"không thể vượt trội hơn thị trường một cách liên tục bằng cách sử dụng thông tin công khai vì giá tài sản nhanh chóng tích hợp bất kỳ thông tin nào có thể có lợi cho nhà đầu tư"*. Lý thuyết EMH có những hệ quả quan trọng đối với hiện tượng được gọi là **Overreaction** của thị trường và việc xác định giá cổ phiếu. Overreaction của thị trường xảy ra khi giá tài sản lệch khỏi giá trị cơ bản của chúng phản ứng với thông tin mới. Một số nghiên cứu đã chỉ ra rằng các thị trường có xu hướng phản ứng thái quá với tin tức, dẫn đến các thay đổi giá tạm thời mà cuối cùng sẽ đảo ngược. Hiện tượng này được Marjorie A. Flavin (1983) [8] gọi là "biến động thái quá" và có thể dẫn đến việc nhà đầu tư kiếm được lợi nhuận bất thường trong ngắn hạn. Tuy nhiên, theo EMH, những lợi nhuận bất thường này chỉ bền vững trong ngắn hạn, vì giá sẽ cuối cùng điều chỉnh để phản ánh tất cả thông tin có sẵn. Lý thuyết EMH đã được phát triển thêm bởi Fama và French (1993) [7] bằng cách giới thiệu khái niệm về những dị thường của thị trường. Đề xuất được đưa ra rằng lý thuyết EMH có thể không đủ để giải thích một số dị thường quan sát được trên thị trường, chẳng hạn như hiệu ứng giá trị và kích thước. Hiện tượng được gọi là hiệu ứng giá trị đề cập đến xu hướng các tài sản giá trị hoạt động tốt hơn các tài sản tăng trưởng. Tương tự, hiệu ứng kích thước đề cập đến xu hướng tài sản Small-Cap vượt trội hơn so với tài sản Large-Cap.

- **Cross-sectional momentum** cho thấy rằng việc xếp hạng tài sản dựa trên hiệu suất lịch sử có thể tạo ra lợi nhuận vượt trội trong một khoảng thời gian nhất định. Các nhà đầu tư có thể xây dựng chiến lược đầu tư bằng

cách chọn mua những tài sản có hiệu suất tốt nhất và bán hoặc tránh xa những tài sản có hiệu suất kém, từ đó khai thác các xu hướng này.

- **Hiệu ứng giá trị và kích:** Lý thuyết EMH đã được phát triển thêm bởi Fama và French (1993) [7] thông qua việc giới thiệu khái niệm về các dị thường của thị trường. Họ đã chỉ ra rằng lý thuyết EMH có thể không đủ để giải thích một số dị thường quan sát được trên thị trường, chẳng hạn như hiệu ứng giá trị—xu hướng các tài sản giá trị hoạt động tốt hơn các tài sản tăng trưởng—và hiệu ứng kích thước—xu hướng tài sản nhỏ (Small-Cap) vượt trội hơn so với tài sản lớn (Large-Cap). Những dị thường này cho thấy rằng có khả năng tồn tại các bất cập trên thị trường mà có thể được sử dụng để tạo ra lợi nhuận vượt trội.

2.2 Phản ứng thị trường

2.2.1 Over-reaction

Giao dịch phản hồi và hiệu ứng bầy đàn (Feedback Trading and Herding Effect)

- Khi các nhà đầu tư tham gia vào giao dịch phản hồi (feedback trading), họ mua vào dựa trên kỳ vọng rằng giá sẽ tiếp tục tăng do những nhà đầu tư khác cũng sẽ mua thêm, tạo thành một chu kỳ tự củng cố.
- George Soros đã mô tả khái niệm này thông qua lý thuyết "phản xạ" (reflexivity), nơi giá cả không chỉ phản ánh thực tế mà còn có thể thay đổi hiện thực đó.
- Hiệu ứng bầy đàn (herding) cũng xuất hiện, nơi nhà đầu tư theo đuổi hành động của đám đông thay vì dựa trên đánh giá cá nhân
- Trong chiến lược **CSM**, hiện tượng giao dịch phản hồi có thể tạo ra các tín hiệu sai lệch. Khi nhà đầu tư mua các tài sản đã tăng giá, họ kỳ vọng giá sẽ tiếp tục tăng nhờ sự tham gia của các nhà đầu tư khác. Tuy nhiên, nếu chu kỳ này kéo dài mà không có sự hỗ trợ từ các yếu tố cơ bản, giá có thể bị thổi phồng, dẫn đến các đợt điều chỉnh mạnh sau đó.
 - Ví dụ: Một tài sản có đà tăng giá mạnh mẽ trong 12 tháng qua có thể thu hút nhiều nhà đầu tư theo đà (momentum traders). Điều này đẩy giá lên cao hơn, tạo ra kỳ vọng tăng giá tiếp tục. Tuy nhiên, nếu giá vượt xa giá trị nội tại, điều chỉnh mạnh có thể xảy ra, gây tổn thất cho nhà đầu tư momentum.

Thiên kiến xác nhận và tính đại diện (Confirmation Bias and Representativeness)

- Nhà đầu tư thường tìm kiếm thông tin xác nhận các niềm tin sẵn có của họ (thiên kiến xác nhận), làm tăng mức độ tự tin một cách sai lệch.
- Ví dụ, nếu họ tin rằng giá tài sản sẽ tăng, họ sẽ tìm kiếm thông tin ủng hộ quan điểm đó, dẫn đến việc đầu tư thêm mà không nhận ra giá đã vượt xa giá trị thực tế.
- Tính đại diện cũng là một lỗi tâm lý, nơi nhà đầu tư cho rằng các xu hướng giá gần đây sẽ tiếp tục trong tương lai, thay vì đánh giá một cách khách quan.

2.2.2 Under-reaction

Neo và điều chỉnh (Anchoring and Adjustment)

- **Phản ứng chậm của giá tài sản với thông tin:** Nhà đầu tư thường bị ảnh hưởng bởi các giá trị tham chiếu trước đó (giá lịch sử, kỳ vọng ban đầu) và không điều chỉnh đủ mạnh theo thông tin mới. Khi tài sản nhận được thông tin tích cực (ví dụ, lợi nhuận vượt kỳ vọng), giá trị thực của nó có thể không được phản ánh đầy đủ ngay lập tức. Điều này dẫn đến hiện tượng:
 - **Những tài sản "winner" tiếp tục tăng:** Giá tăng chậm nhưng ổn định do nhà đầu tư dần điều chỉnh nhận thức về giá trị thực của tài sản dựa trên thông tin mới.
- **Cơ hội cho động lượng (Momentum):** Chiến lược *Cross-Sectional Momentum* tận dụng hiện tượng này bằng cách tập trung vào các tài sản có xu hướng tăng giá. Khi các nhà đầu tư tiếp tục điều chỉnh giá kỳ vọng, giá tài sản sẽ duy trì xu hướng tăng trong thời gian dài hơn.
- **Ví dụ minh họa:** Một công ty công bố doanh thu vượt kỳ vọng nhưng nhà đầu tư vẫn đánh giá thấp tác động lâu dài của thông tin này. Trong vài tháng tiếp theo, giá tài sản của công ty này tăng đều đặn khi nhà đầu tư dần điều chỉnh kỳ vọng, tạo cơ hội sinh lời từ chiến lược động lượng.

Hiệu ứng phân bố (Disposition Effect)

- **Áp lực bán tài sản "winner":** Tâm lý sợ mất lợi nhuận khiến nhà đầu tư bán nhanh những tài sản đang có lời. Điều này tạo áp lực bán trong ngắn hạn và làm chậm tốc độ tăng giá của các tài sản này.

- **Giữ tài sản "loser":** Nhà đầu tư giữ lâu các tài sản đang thua lỗ vì không muốn chấp nhận thất bại. Kết quả là giá tài sản *loser* tiếp tục giảm sâu hơn do không có đủ động lực mua vào từ thị trường.
- **Ảnh hưởng đến chiến lược động lượng:**
 - **Winner:** Áp lực bán từ các nhà đầu tư nhỏ lẻ giúp giá tài sản *winner* không tăng đột biến, duy trì xu hướng tăng đều và ổn định trong dài hạn.
 - **Loser:** Việc giữ lại các tài sản thua lỗ kéo dài xu hướng giảm của chúng, tạo ra cơ hội bán khống (*short selling*) trong chiến lược động lượng.
- **Ví dụ minh họa:** tài sản của một công ty vừa vượt qua kỳ vọng doanh thu có xu hướng tăng giá. Tuy nhiên, do tâm lý phân bổ, áp lực bán ngắn hạn của các nhà đầu tư nhỏ lẻ giúp giá tăng từ từ thay vì tăng mạnh. Điều này tạo cơ hội để chiến lược *Cross-Sectional Momentum* tiếp tục hưởng lợi từ xu hướng giá.

3 Lý thuyết toán học

3.1 Lợi suất của chiến lược TSM

3.1.1 Độ biến động

Để đảm bảo có sự so sánh có ý nghĩa giữa các tài sản cho thấy sự khác biệt đáng kể về mức độ biến động, lợi nhuận của chúng có thể được điều chỉnh lại dựa trên mức độ biến động. Sự điều chỉnh này đảm bảo rằng mỗi tài sản đóng góp tương tự vào lợi nhuận chung của danh mục đầu tư. Độ biến động trước của mỗi tài sản tại một thời điểm t cụ thể có thể được ước tính bằng phương trình được trình bày trong Moskowitz và cộng sự (2012) [15], bao gồm mô hình lợi nhuận hàng ngày với bình phương độ trễ có trọng số theo hàm số mũ.

$$\sigma_t^2 = 261 \sum_{i=0}^{\infty} (1 - \delta) \delta^i (r_{t-1-i} - \bar{r}_t)^2 \quad (1)$$

Độ biến động tính theo năm, giá trị 261 biểu thị số ngày giao dịch. Tổng trọng số $(1 - \delta) \delta^i$ bằng 1. \bar{r}_t là lợi nhuận trung bình hàng tháng theo hàm số mũ. Cuối cùng, tham số δ có thể được chọn sao cho trọng tâm của các trọng số là 60. Như đã đề cập trong Moskowitz và cộng sự (2012) [15], có hai lợi thế chính của việc duy trì biến động trước không đổi cho mỗi vị thế trong một chiến lược. Lợi thế đầu tiên là nó tạo điều kiện thuận lợi cho việc kết hợp các công cụ có mức độ biến động khác nhau trong một chiến lược duy nhất. Lợi thế thứ hai là nó giúp

duy trì biến động tương đối ổn định trong chuỗi thời gian, điều này rất cần thiết theo quan điểm kinh tế lượng.

3.1.2 Lợi suất của chiến lược TSM

Chúng ta bắt đầu xem xét từng tài sản riêng biệt, sau đó tập hợp tất cả các tài sản lại với nhau trong danh mục đầu tư. Độ biến động trước hằng năm (ex-ante annualized volatility) là σ_{target} , suy ra quy mô vị thế là $\sigma_{\text{target}}/\sigma_{t-1}$. Trong đó, σ_{t-1} là ước tính độ biến động trước (ex-ante volatility) của từng tài sản. Moskowitz và cộng sự (2012) [15] đề xuất độ biến động trước hằng năm σ_{target} là 40% vì nó xấp xỉ rủi ro trung bình của mỗi một tài sản Hoa Kỳ, vì nghiên cứu của họ dựa trên thị trường chứng khoán Hoa Kỳ. Lợi nhuận TSM cho mỗi tài sản i tại thời điểm t được tính như sau:

$$r_{t,t+1}^{\text{TSM},i} = \text{sign}(r_{t-12,t}^i) \frac{\sigma_{\text{target}}}{\sigma_t^i} r_{t,t+1}^i \quad (2)$$

Biến $r_{t,t+1}^i$ thể hiện lợi nhuận hàng tháng của tài sản i tại thời điểm t và $t+1$. $\text{sign}(r_{t-12,t}^i)$ cho biết lợi nhuận vượt trội trong quá khứ của cổ phiếu là âm hay dương. Do đó, công thức cho lợi nhuận chung của danh mục đầu tư cho các tài sản hiện có sẵn tại thời điểm t là:

$$r_{t,t+1}^{\text{TSM}} = \frac{1}{N} \sum_{i=1}^N \text{sign}(r_{t-12,t}^i) \frac{\sigma_{\text{target}}}{\sigma_t^i} r_{t,t+1}^i \quad (3)$$

3.2 Lợi suất theo chiến lược CSM

Định nghĩa về chiến lược CSM, được đặc trưng bởi một danh mục đầu tư dựa trên cách xếp hạng tương đối của các tài sản với nhau, có thể được xác định như sau:

$$r_{t,t+1}^{\text{CSM}} = \frac{1}{N} \sum_{i=1}^N X_t^i r_{t,t+1}^i \quad (4)$$

Trong đó, $t, t+1 \in \mathcal{T} \subset \{1, \dots, t-1, t_\nu, t+1, \dots, T\}$ và \mathcal{T} là tập hợp các chỉ số trùng với ngày giao dịch cuối cùng của tháng. $r_{t,t+1}^{\text{CSM}}$ là lợi suất danh mục đầu tư đã thực hiện từ tháng t đến $t+1$; N là số tài sản trong danh mục. $X_t^i \in \{-1, 0, 1\}$ đặc trưng cho tín hiệu CSM hoặc quy tắc giao dịch cho tài sản i .

Ngoài ra, Poh et al. (2020) [16] đã phát triển một phương pháp điều chỉnh độ biến động dựa trên các phát hiện của Kim et al. (2016) [13], trong đó lợi suất của mỗi tài sản được điều chỉnh theo độ biến động của danh mục đầu tư. Điều này đảm bảo rằng mỗi tài sản có mức độ tương tự so với độ biến động của danh

mục đầu tư tổng thể. Do đó, CSM đã được điều chỉnh theo độ biến động có thể được biểu diễn như sau:

$$r_{t,t+1}^{\text{CSM}} = \frac{1}{N} \sum_{i=1}^N X_t^i \frac{\sigma_{\text{target}}}{\sigma_t^i} r_{t,t+1}^i \quad (5)$$

Trong đó, mỗi lợi suất $r_{t,t+1}^i$ theo từng tài sản được điều chỉnh theo hệ số điều chỉnh độ biến động $\frac{\sigma_{\text{target}}}{\sigma_t^i}$. Độ biến động hàng năm σ_{target} được cố định ở mức 15% và chuẩn hóa lợi nhuận của tài sản bằng σ_t^i , đây là một ước lượng cho độ biến động hàng tháng trước đó (ex-ante monthly volatility). Trong nghiên cứu này, chúng tôi sử dụng độ lệch chuẩn có trọng số theo cấp số nhân với khoảng thời gian 63 ngày trên lợi nhuận hàng ngày để tính σ_t^i .

Và như đã đề cập trước đó, các chiến lược CSM thường tương tự nhau trong suốt những năm qua, tức là cách mà họ xếp hạng Momentum, lựa chọn và xây dựng danh mục đầu tư. Tuy nhiên, chúng lại rất đa dạng trong sự lựa chọn mô hình dự đoán f được sử dụng để tính toán Momentum, dao động từ các phương pháp đơn giản đến các kiến trúc phức tạp sử dụng một loạt các đầu vào vĩ mô (Gu et al., 2018) [9]. Mặc dù có nhiều kỹ thuật để tính toán điểm số, chúng có thể được chia thành ba nhóm: Classical Momentum, Regresion-then-Rank và phương pháp Learning to Rank.

3.2.1 Cross-sectional Momentum cổ điển

Cross-sectional Momentum cổ điển có xu hướng thiên về việc sử dụng các thủ tục tương đối đơn giản để tính Momentum.

$$Y_t^{(i)} = f(u_t^{(i)}) \quad (6)$$

Với vectơ đầu vào $u_t^{(i)}$ cho tài sản i tại thời điểm t , mô hình dự đoán của chiến lược f tính toán Momentum tương ứng $Y_t^{(i)}$. Đối với một tập hợp các vectơ có kích thước N tại thời điểm t , một danh sách các Momentum của các tài sản được xem xét để giao dịch được biểu diễn bởi vectơ $Y_t^{(i)} = \{Y_t^{(1)}, Y_t^{(2)}, \dots, Y_t^{(N)}\}$.

3.2.1.1 Jegadeesh & Titman (1993) [11]: tính Momentum dựa vào lợi suất tích lũy theo từng tài sản trong vòng 3 đến 12 tháng

$$\text{Tính Momentum: } Y_t^{(i)} = r_{t-252,t}^{(i)} \quad (7)$$

Trong đó, $r_{t-252,t}^{(i)}$ là lợi suất thô trong 252 (12 tháng) ngày trước đó kể từ tại thời điểm t cho tài sản i . Khi đó lợi nhuận chiến lược thực tế từ thời gian t đến thời gian $t + 1$ có thể được định nghĩa như sau:

$$r_{\text{Return}}^{t,t+1} = \frac{1}{n_t} \sum_{i=1}^{n_t} X_t^{(i), \text{Return}} \cdot r_{t,t+1}^{(i)}$$

3.2.1.2 Baz et al. (2015) [1]: sử dụng các chỉ báo MACD chuẩn hóa theo độ biến động như một chỉ báo trung gian, được tính toán trên các khoảng thời gian khác nhau. Chỉ báo được định nghĩa như sau:

$$\widetilde{Y}_t^{(i)} = \frac{\xi_t^{(i)}}{\text{std}(z_{t-252:t}^{(i)})}, \quad (8)$$

$$\xi_t^{(i)} = \frac{\text{MACD}(i, t, S, L)}{\text{std}(p_{t-63:t}^{(i)})}, \quad (9)$$

$$\text{MACD}(i, t, S, L) = m(i, S) - m(i, L), \quad (10)$$

Trong đó, $\text{std}(p_{t-63:t}^{(i)})$ biểu thị độ lệch chuẩn 63 ngày của lợi nhuận của tài sản i vào thời điểm t trong đó:

- $\text{std}(p_{t-63:t}^{(i)})$: Độ lệch chuẩn cuộn trong 63 ngày của chứng khoán i ,
- $m(i, S)$: Trung bình động có trọng số mũ của giá cho tài sản i , với S là hệ số phân rã bán phần $\text{HL} = \frac{\log(0.5)}{\log(1-1/S)}$.

Tín hiệu tổng hợp cuối cùng kết hợp các MACD được chuẩn hóa theo độ biến động trên các thang thời gian khác nhau với một hàm phản hồi $\phi(\cdot)$ và một tập hợp các thang thời gian ngắn và dài $S_k \in \{8, 16, 32\}$ và $L_k \in \{24, 48, 96\}$, như được mô tả trong Baz et al. (2015) [1]:

$$\textbf{Tính Momentum: } \boxed{Y_t^{(i)} = \sum_{k=1}^3 \phi\left(\widetilde{Y}_t^{(i)}(S_k, L_k)\right)} \quad (11)$$

Từ (11), có thể thấy rằng, sau khi tính toán điểm động lực đặc thù MACD, bảng xếp hạng điểm động lực và việc lựa chọn chứng khoán được thực hiện tương tự như các chiến lược CSMOM gốc, trong đó điểm số dựa trên MACD được sắp xếp theo thứ tự giảm dần và sau đó các danh mục đầu tư nhóm top và bottom decile với các vị thế mua tối đa và bán tối đa được hình thành. Cuối cùng, tương tự như khung chiến lược CSMOM chung, lợi nhuận chiến lược thực tế từ thời gian t đến thời gian $t + 1$ có thể được định nghĩa như sau:

$$r_{\text{MACD}}^{t,t+1} = \frac{1}{n_t} \sum_{i=1}^{n_t} X_t^{(i), \text{MACD}} \cdot r_{t,t+1}^{(i)}$$

Trong đó, kích thước vị thế $X_t^{(i), \text{MACD}}$ tương ứng với phương trình trên và tất cả các thành phần chiến lược.

3.2.2 Regress Then Rank

Những công trình nghiên cứu gần đây sử dụng phương pháp dự đoán-sau đó-xếp hạng thường tính toán điểm số thông qua hồi quy tiêu chuẩn:

$$\textbf{Tính Momentum: } \boxed{Y_t^{(i)} = f(u_t^{(i)}; \theta)} \quad (12)$$

trong đó f biểu thị một mô hình dự đoán học máy được tham số hóa bởi θ , với đầu vào là một vector $u_t^{(i)}$. Sử dụng lợi nhuận được chuẩn hóa theo độ biến động làm mục tiêu, mô hình được huấn luyện bằng cách tối thiểu hóa hàm mất mát, thường là MSE:

$$L(\theta) = \frac{1}{M} \sum_{\Omega} \left(Y_t^{(i)} - \frac{r_{t,t+1}^{(i)}}{\sigma_t^{(i)}} \right)^2 \quad (13)$$

$$\Omega = \left\{ \left(Y_1^{(1)}, \frac{r_{1,2}^{(1)}}{\sigma_1^{(1)}} \right), \dots, \left(Y_{t-1}^{(N_{t-1})}, \frac{r_{t-1,t}^{(N_{t-1})}}{\sigma_{t-1}^{(N_{t-1})}} \right) \right\}$$

trong đó Ω đại diện cho tập hợp tất cả các cặp dự báo và mục tiêu khả dĩ M trên tập hợp các công cụ tài chính và các bước thời gian liên quan.

3.2.2.1 Random Forest

Random Forest là một thuật toán ensemble dựa trên việc tạo ra nhiều cây quyết định độc lập (decision trees). và tổng hợp kết quả từ các cây này. Thay vì sử dụng một mô hình hồi quy tuyến tính đơn lẻ f , ta sử dụng RF để dự đoán điểm số:

$$Y_{\tau_m}^{(i)} = \text{RF}(u_{\tau_m}^{(i)}; \Theta)$$

trong đó:

- $\text{RF}(\cdot)$ là mô hình Random Forest,
- Θ là tập hợp các tham số, bao gồm cấu hình cây quyết định, số lượng cây, và các đặc điểm khác của mô hình.

Mục tiêu huấn luyện:

Tương tự như hồi quy tiêu chuẩn, mô hình RF được huấn luyện bằng cách tối thiểu hóa sai số dự báo. Tuy nhiên, RF không yêu cầu tối ưu hóa trực tiếp một hàm mất mát cụ thể, mà thay vào đó giảm thiểu sai số dựa trên trung bình của các dự báo từ nhiều cây quyết định. Hàm mất mát tổng quát có thể viết như sau:

$$L(\Theta) = \frac{1}{M} \sum_{\Omega} \left(Y_{\tau_m}^{(i)} - \frac{r_{\tau_m, \tau_m+1}^{(i)}}{\sigma_{\tau_m}^{(i)}} \right)^2$$

trong đó:

- Ω là tập hợp các cặp dự báo và mục tiêu,
- $r_{\tau_m, \tau_m+1}^{(i)}$ là lợi nhuận dự báo của công cụ tài chính i ,
- $\sigma_{\tau_m}^{(i)}$ là độ biến động chuẩn hóa.

Từ hàm loss trên lợi nhuận chiến lược thực tế từ thời gian t đến thời gian $t + 1$ có thể được định nghĩa như sau:

$$r_{\text{RandomForest}}^{t,t+1} = \frac{1}{n_t} \sum_{i=1}^{n_t} X_t^{(i), \text{RandomForest}} \cdot r_{t,t+1}^{(i)}$$

Đặc điểm nổi bật của RF

- RF sử dụng các mẫu ngẫu nhiên (bootstrap sampling) để huấn luyện từng cây quyết định, giúp giảm overfitting.
- Mỗi cây chỉ xem xét một tập hợp con của các đặc trưng đầu vào, tăng tính đa dạng và giảm sự phụ thuộc giữa các cây.
- Kết quả dự báo cuối cùng được tổng hợp từ trung bình của tất cả các cây:

$$Y_{\tau_m}^{(i)} = \frac{1}{T} \sum_{t=1}^T f_t(u_{\tau_m}^{(i)})$$

trong đó T là số lượng cây trong mô hình RF, và f_t là dự báo từ cây t .

3.2.3 Learning to Rank

Các phương pháp LTR có thể được phân loại thành ba loại: pointwise, pairwise hoặc listwise. Phương pháp pointwise (pairwise) giải quyết vấn đề xếp hạng dưới dạng phân loại, hồi quy hoặc phân loại thứ tự của các mẫu riêng lẻ (cặp mẫu), trong khi phương pháp listwise học mô hình xếp hạng thích hợp bằng cách sử dụng danh sách xếp hạng làm đầu vào. Về hiệu suất xếp hạng, phương pháp pointwise đã được quan sát là kém hơn so với hai phương pháp còn lại (H. Li, 2011) [14]. Thêm vào đó, hàm mất mát không chỉ là sự khác biệt chủ yếu giữa các mô hình này (H. Li, 2011) [14] – việc kết hợp thông tin pairwise và listwise giữa các tài sản khiến các mô hình LTR trở nên khác biệt so với cả các phương pháp cổ điển và phương pháp Regress-then-Rank đã được mô tả ở trên.

3.2.3.1 LamdaMART

LambdaMART (C. J. C. Burges, 2010) [2] là một phương pháp pairwise hiện đại kết hợp giữa thuật toán **LambdaRank** (C. J. C. Burges et al., 2006) [3] và **Multiple Additive Regression Trees (MART)**. Đây là một phương pháp được sử dụng trong các bài toán xếp hạng trong tìm kiếm thông tin và đã trở thành công cụ nổi bật nhờ hiệu quả và hiệu suất của nó.

LambdaRank và λ -gradients LambdaMART xây dựng trên nền tảng của LambdaRank, một thuật toán được thiết kế để tối ưu hóa các hệ thống xếp hạng. Thay vì trực tiếp tối ưu hóa một hàm mất mát, LambdaRank sử dụng các xấp xỉ heuristic của gradient, được gọi là **λ -gradients**. Các gradient này rất hữu ích trong việc huấn luyện các mạng nơ-ron và cho phép thuật toán tránh được những thách thức do các hàm mất mát không khả vi, gián đoạn hoặc phẳng, như **NDCG (Normalized Discounted Cumulative Gain)**, một chỉ số phổ biến trong các bài toán tìm kiếm thông tin (Järvelin & Kekäläinen, 2000) [10].

Công thức λ -gradients Công thức tính λ -gradient cho một cặp tài liệu (d_i, d_j) là:

$$\lambda_{ij} = \frac{1}{Z} \left(\frac{1}{1 + \exp(-(s_i - s_j))} - \frac{1}{1 + \exp(-(s_j - s_i))} \right)$$

Trong đó:

- s_i và s_j là điểm số (score) của tài liệu d_i và d_j lần lượt.
- Z là một hệ số chuẩn hóa.

Hàm này tính toán sự khác biệt giữa điểm số của hai tài liệu, điều chỉnh theo công thức sigmoid để đảm bảo sự chuyển đổi giữa các thứ hạng.

Hàm Mất Mát trong LambdaMART LambdaMART không tối ưu hóa trực tiếp hàm mất mát mà tối ưu hóa các **λ -gradients**. Hàm mất mát trong LambdaMART liên quan đến việc tính toán **loss** cho các cặp tài liệu trong tập huấn luyện. Mất mát này được tính dựa trên việc sử dụng cross-entropy kết hợp với các giá trị **λ -gradients**.

Hàm mất mát có thể được viết dưới dạng:

$$L(\theta) = - \sum_{i,j} \lambda_{ij} \log (P(y_i | d_i) \cdot P(y_j | d_j))$$

Trong đó:

- $P(y_i | d_i)$ là xác suất mà tài liệu d_i có thể xuất hiện ở vị trí y_i trong xếp hạng.
- λ_{ij} là **λ -gradient** tính toán được ở phần trước.

Từ hàm loss trên lợi nhuận chiến lược thực tế từ thời gian t đến thời gian $t + 1$ có thể được định nghĩa như sau:

$$r_{\text{LambdaMART}}^{t,t+1} = \frac{1}{n_t} \sum_{i=1}^{n_t} X_t^{(i), \text{LambdaMART}} \cdot r_{t,t+1}^{(i)}$$

Multiple Additive Regression Trees (MART) MART là một kỹ thuật tree-boosting nổi tiếng với tính linh hoạt và khả năng thích ứng cao. Nó có khả năng xây dựng các mô hình phức tạp trong khi vẫn đảm bảo tính toán hiệu quả. Một tính năng quan trọng của MART là khả năng cân bằng giữa tốc độ và độ chính xác thông qua kỹ thuật **truncation** (cắt ngắn), đây là một yếu tố quan trọng trong các ứng dụng có yêu cầu về thời gian như các công cụ tìm kiếm, nơi mà tốc độ là yếu tố quyết định. MART cũng có tính linh hoạt và độ bền vững, làm cho nó trở thành lựa chọn lý tưởng cho các bài toán xếp hạng (Wu et al., 2010) [18].

Công thức của MART Quá trình huấn luyện MART thực hiện việc tối ưu hóa theo công thức:

$$f_t(x) = f_{t-1}(x) + \eta \cdot h_t(x)$$

Trong đó:

- $f_t(x)$ là mô hình của cây quyết định ở bước huấn luyện thứ t .
- η là hệ số học (learning rate).
- $h_t(x)$ là cây quyết định tại bước huấn luyện thứ t .

Tối Ưu Hóa LambdaMART Kết hợp LambdaRank và MART, LambdaMART tối ưu hóa bằng cách sử dụng một chuỗi các cây quyết định (boosting trees) để học các λ -gradients và cải thiện độ chính xác của xếp hạng. Quá trình huấn luyện của LambdaMART có thể được mô tả dưới dạng tổng quát như sau:

$$\theta^{\text{new}} = \theta^{\text{old}} + \eta \cdot \lambda_{ij}$$

Trong đó:

- θ^{new} là các tham số được cập nhật.
- θ^{old} là các tham số trước khi cập nhật.
- λ_{ij} là λ -gradient tính toán được trong quá trình huấn luyện.

Qua mỗi vòng lặp, mô hình học cách tối ưu hóa các λ -gradients để cải thiện thứ hạng của các tài liệu, và các cây quyết định sẽ được thêm vào để cải thiện hiệu quả mô hình

3.2.3.2 ListNet

ListNet (Z. Cao et al., 2007) [4] định nghĩa một phân phối xác suất trên tập hợp tất cả các thứ tự có thể có của một danh sách. Thay vì chỉ so sánh cặp hai đối tượng hoặc từng đối tượng riêng lẻ, ListNet học toàn bộ danh sách bằng cách tối ưu hóa khoảng cách giữa hai phân phối xác suất:

- Phân phối dự đoán được sinh ra từ mô hình.
- Phân phối tham chiếu được xây dựng từ thứ tự thực tế của danh sách.

Phân phối xác suất Permutation: Để biểu diễn thứ tự của danh sách, ListNet sử dụng hàm Plackett-Luce để định nghĩa xác suất của một thứ tự π :

$$P(\pi|s) = \prod_{i=1}^n \frac{\exp(s_{\pi(i)})}{\sum_{j=i}^n \exp(s_{\pi(j)})}$$

Trong đó:

- $s = (s_1, s_2, \dots, s_n)$ là điểm số mô hình dự đoán cho các đối tượng trong danh sách.
- $\pi(i)$ là đối tượng ở vị trí thứ i trong thứ tự π .

Hàm mất mát: ListNet sử dụng hàm mất mát ở đây sử dụng Jensen-Shannon Divergence (JSD) là một thước đo độ sai khác giữa hai phân phối xác suất.

- Phân phối mục tiêu: Được xây dựng từ nhãn thực tế (ví dụ: điểm số thực hoặc thứ tự xếp hạng).
- Phân phối dự đoán: Sinh ra từ mô hình.

JSD là một phiên bản đối xứng của Kullback-Leibler Divergence (KLD), và được định nghĩa như sau:

$$\text{JSD}(P \parallel Q) = \frac{1}{2} (D_{\text{KL}}(P \parallel M) + D_{\text{KL}}(Q \parallel M))$$

Trong đó:

- P và Q là hai phân phối xác suất, trong bài toán này là các phân phối xác suất thực tế và dự đoán (t và y).
- $M = \frac{1}{2}(P + Q)$ là phân phối trung bình của P và Q .
- D_{KL} là **Kullback-Leibler Divergence**, tính độ sai khác giữa hai phân phối xác suất.

Kullback-Leibler Divergence được định nghĩa như sau:

$$D_{\text{KL}}(P \parallel Q) = \sum_i P(i) \log \left(\frac{P(i)}{Q(i)} \right)$$

Trong đó, $P(i)$ và $Q(i)$ là xác suất của i -th sự kiện trong hai phân phối P và Q . JSD là một thước đo đối xứng, có nghĩa là:

$$\text{JSD}(P \parallel Q) = \text{JSD}(Q \parallel P)$$

Từ hàm loss trên lợi nhuận chiến lược thực tế từ thời gian t đến thời gian $t + 1$ có thể được định nghĩa như sau:

$$r_{\text{ListNet}}^{t,t+1} = \frac{1}{n_t} \sum_{i=1}^{n_t} X_t^{(i), \text{ListNet}} \cdot r_{t,t+1}^{(i)}$$

Jensen-Shannon Divergence có các đặc điểm:

- Khi $P = Q$, thì $\text{JSD}(P \parallel Q) = 0$.
- JSD luôn có giá trị trong khoảng từ 0 đến 1, và do đó có thể coi đây là một thước đo khoảng cách giữa các phân phối xác suất.

JSD được sử dụng phổ biến trong học máy, đặc biệt là trong các bài toán phân loại, để đánh giá độ sai khác giữa phân phối xác suất của các nhãn dự đoán và nhãn thực tế.

4 Phương pháp xây dựng

4.1 Mục tiêu

- **Dự đoán xu hướng** : Sử dụng MACD để phát hiện các dấu hiệu của sự chuyển đổi giữa xu hướng tăng và giảm. MACD giúp nhận diện động lượng bằng cách phân tích sự khác biệt giữa hai đường trung bình động.
- **Tối ưu hóa các chiến lược giao dịch**: So sánh hiệu suất của các chiến lược đầu tư khác nhau (long, short, long-short, trọng số bằng nhau) nhằm xác định chiến lược nào mang lại lợi nhuận cao nhất với mức rủi ro thấp nhất.
- **Quản lý rủi ro**: Xác định thời điểm vào và ra khỏi giao dịch qua các tín hiệu từ MACD nhằm giảm thiểu mức thua lỗ và tối đa hóa lợi nhuận.

4.2 Dữ liệu

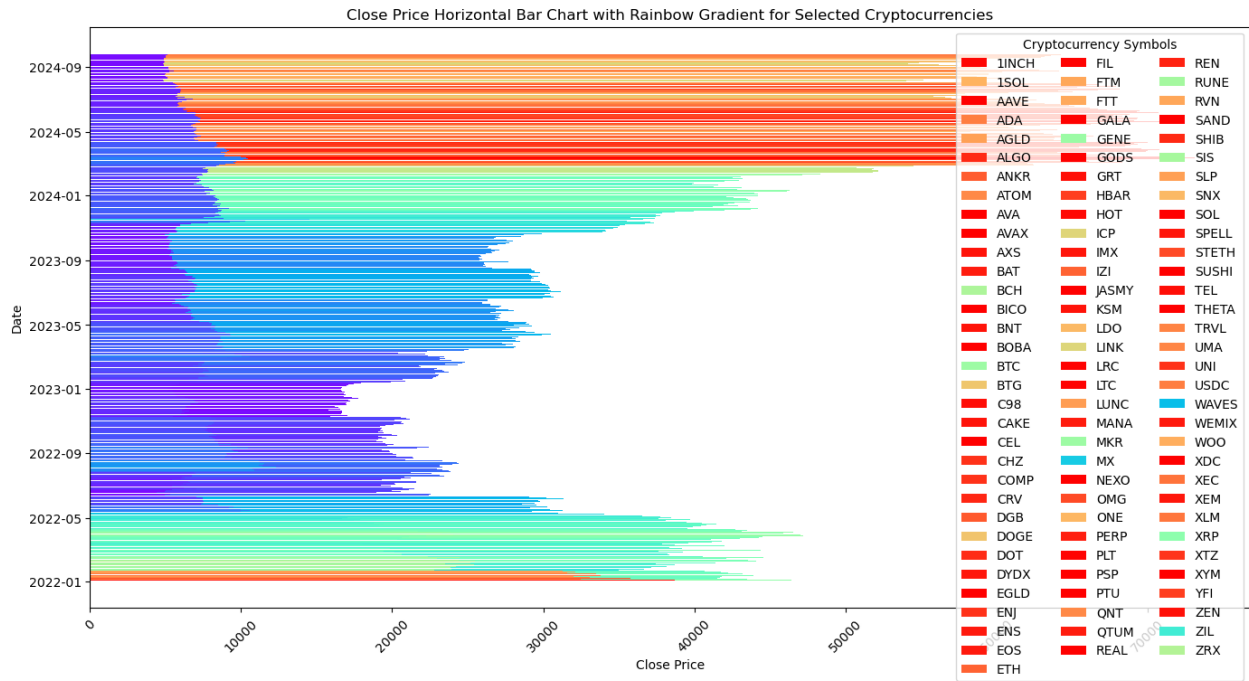
Dữ liệu được thu thập bao gồm các cặp giao dịch tiền mã hóa trên nhiều thị trường, chủ yếu là các cặp giao dịch với USDT, BTC, và USDC. Danh sách này bao gồm tổng cộng 106 cặp giao dịch, được phân loại theo các loại tài sản phổ biến như Bitcoin (BTC), Ethereum (ETH), các loại altcoin (ví dụ: XRP, DOGE, AAVE), và stablecoin (USDT, USDC) trong khoảng thời gian từ 2022 đến nay.

Mục đích phân tích dữ liệu:

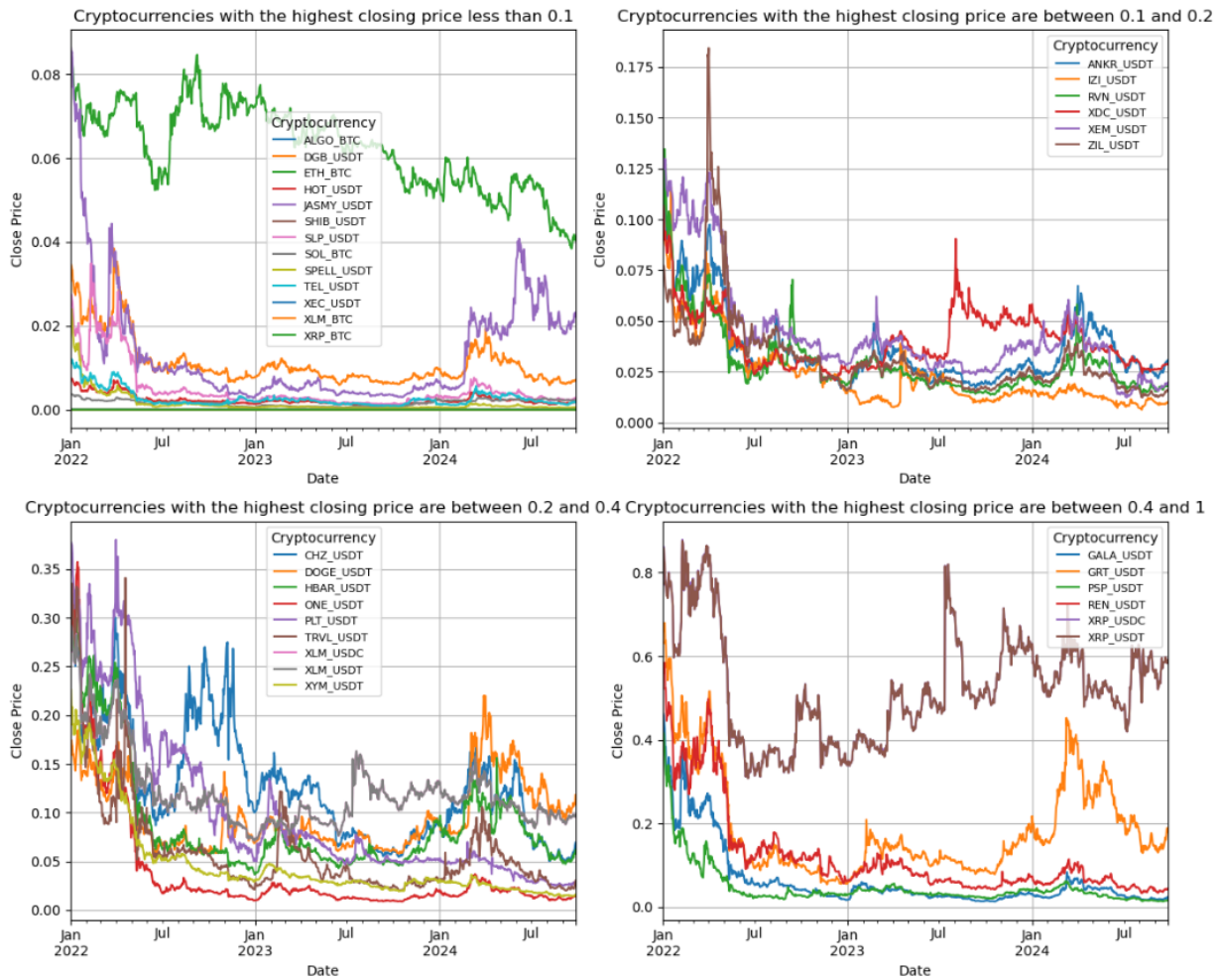
- Đánh giá hiệu suất của các cặp giao dịch theo thời gian.
- Phân tích khối lượng giao dịch và xu hướng giá.
- Xây dựng các chiến lược giao dịch tự động.

STT	Cặp giao dịch	STT	Cặp giao dịch	STT	Cặp giao dịch
1	BTC/USDT	2	ETH/USDT	3	XRP/USDT
4	EOS/USDT	5	ETH/BTC	6	XRP/BTC
7	DOT/USDT	8	XLM/USDT	9	LTC/USDT
10	DOGE/USDT	11	CHZ/USDT	12	AXS/USDT
13	MANA/USDT	14	DYDX/USDT	15	MKR/USDT
16	COMP/USDT	17	AAVE/USDT	18	YFI/USDT
19	LINK/USDT	20	SUSHI/USDT	21	UNI/USDT
22	KSM/USDT	23	ICP/USDT	24	ADA/USDT
25	XTZ/USDT	26	BCH/USDT	27	QNT/USDT
28	USDC/USDT	29	GRT/USDT	30	SOL/USDT

STT	Cặp giao dịch	STT	Cặp giao dịch	STT	Cặp giao dịch
31	FIL/USDT	32	OMG/USDT	33	BAT/USDT
34	ZRX/USDT	35	CRV/USDT	36	AGLD/USDT
37	ANKR/USDT	38	PERP/USDT	39	WAVES/USDT
40	LUNC/USDT	41	SPELL/USDT	42	SHIB/USDT
43	FTM/USDT	44	ATOM/USDT	45	ALGO/USDT
46	ENJ/USDT	47	CBX/USDT	48	SAND/USDT
49	AVAX/USDT	50	WOO/USDT	51	FTT/USDT
52	GODS/USDT	53	IMX/USDT	54	ENS/USDT
55	CAKE/USDT	56	STETH/USDT	57	SLP/USDT
58	C98/USDT	59	GENE/USDT	60	AVA/USDT
61	ONE/USDT	62	PTU/USDT	63	XYM/USDT
64	BOBA/USDT	65	JASMY/USDT	66	GALA/USDT
67	TRVL/USDT	68	WEMIX/USDT	69	XEM/USDT
70	BICO/USDT	71	CEL/USDT	72	UMA/USDT
73	HOT/USDT	74	NEXO/USDT	75	BNT/USDT
76	SNX/USDT	77	REN/USDT	78	1INCH/USDT
79	TEL/USDT	80	SIS/USDT	81	LRC/USDT
82	LDO/USDT	83	REAL/USDT	84	ETH/USDC
85	BTC/USDC	86	1SOL/USDT	87	PLT/USDT
88	IZI/USDT	89	QTUM/USDT	90	ZEN/USDT
91	THETA/USDT	92	MX/USDT	93	DGB/USDT
94	RVN/USDT	95	EGLD/USDT	96	RUNE/USDT
97	XLM/BTC	98	XLM/USDC	99	SOL/USDC
100	XRP/USDC	101	ALGO/BTC	102	SOL/BTC
103	XEC/USDT	104	XDC/USDT	105	BTG/USDT
106	ZIL/USDT	107	HBAR/USDT		



Hình 1: Tổng quan dữ liệu



Hình 2: Xu hướng của dữ liệu

Tỷ Lệ Tiền Điện Tử Theo Giá Đóng Cửa Cao Nhất:

- **Khoảng Giá Thấp** (< 1): Với 34/107 đồng tiền điện tử chiếm khoảng 31.8% tổng số đồng tiền. Điều này cho thấy sự hiện diện tương đối lớn của các đồng tiền có giá thấp.
- **Khoảng Giá Cao** (> 100): Chỉ có 18/107 đồng tiền (khoảng 16.8%). Điều này cho thấy rằng các đồng tiền có giá trị cao hơn là ít phổ biến hơn trong danh sách tổng thể.

Nhận xét:

- Nhiều loại tiền điện tử có xu hướng biến động mạnh trong các khoảng giá khác nhau, đặc biệt là ở các mức giá thấp hơn (dưới 1) so với các mức giá cao hơn.
- Ở các khoảng giá cao hơn (trên 10), sự biến động có thể ít hơn, cho thấy rằng các đồng tiền này có thể đã đạt được sự ổn định hơn hoặc có giá trị thị trường lớn hơn.
- Các đồng tiền có giá thấp có thể mang lại cơ hội lớn, nhưng cũng đi kèm với rủi ro cao hơn do sự biến động lớn. Ngược lại, các đồng tiền có giá cao hơn có sự ổn định hơn.

4.3 Mô hình chiến lược**4.3.1 Chiến lược dựa theo momentum cổ điển**

- Jegadeesh và Titman (1993) theo công thức (7)
- Baz et al. (2015) theo công thức (11)

4.3.2 Chiến lược có sự kết hợp mô hình học máy**4.3.2.1 Regress Then Rank**

- Random Forest ở mục 3.2.2.1

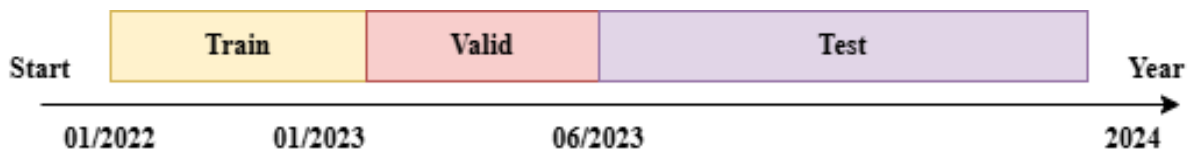
4.3.2.2 Learning to Rank

- LambdaMART (LM) – Mô hình LTR theo cặp của C. J. C. Burges (2010) [2] ở mục 3.2.3.1.
- (LNet) – Mô hình LTR theo danh sách của Z. Cao et al. (2007) [4] ở mục 3.2.3.2.

4.4 Mô tả Backtest

4.4.1 Cross Sectional Momentum Model Features

Tất cả các mô hình còn lại đều được tinh chỉnh lại sau mỗi khoảng thời gian 1 năm. Các trọng số và siêu tham số của các mô hình đã được hiệu chỉnh sau đó được cố định và sử dụng để tái cân bằng danh mục đầu tư ngoài mẫu cho khoảng thời gian 1 năm tiếp theo. Việc tái cân bằng diễn ra vào ngày giao dịch cuối cùng của mỗi tháng. Tập trung vào việc xếp hạng và giao dịch cho hơn 100 cổ phiếu cho mỗi danh mục đầu tư long và short vào mọi thời điểm – chiếm khoảng 10% tổng số cổ phiếu có thể giao dịch trong mỗi khoảng thời gian tái cân bằng.



Hình 3: Train, Test, Valid

Đối với các bộ dữ liệu, sử dụng một sự kết hợp đơn giản của các bộ dữ liệu được sử dụng trong các phương pháp tính điểm cổ điển đã đề cập ở phần trên:

- **Lợi nhuận tích lũy thô** – Lợi nhuận theo phương pháp của Jegadeesh và Titman (1993) [11] trong các khoảng thời gian 3, 6 và 12 tháng.
- **Lợi nhuận chuẩn hóa (Normalized Return)** có thể được tính bằng cách chuẩn hóa lợi nhuận tích lũy theo biến động giá (ví dụ, sử dụng độ lệch chuẩn).
- **Chỉ số MACD** – Giữ lại tín hiệu cuối cùng như được định nghĩa trong Phương trình (11) từ Baz et al. (2015) [1], chúng tôi cũng bổ sung bộ dữ liệu của mình bằng cách bao gồm các tín hiệu trung gian thô $Y_t(S_k, L_k)$ trong Phương trình (8) cho $k = 1, 2, 3$ được tính tại thời điểm t cũng như cho quá khứ.
- **Tín hiệu MACD** được tạo ra từ sự khác biệt giữa EMA12 và EMA26.
- **Chỉ báo Signal Line** được tạo ra từ EMA của MACD.

4.5 Kết quả và thảo luận

4.5.1 Thống kê kiểm định

4.5.1.1 Giả thuyết và Cách hoạt động của ANOVA Giả thuyết:

- **Giả thuyết không (H_0):** Không có sự khác biệt về trung bình lợi nhuận giữa các chiến lược giao dịch.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

Trong đó:

$\mu_1, \mu_2, \mu_3, \mu_4$: Lợi nhuận trung bình của các chiến lược Long-Short, Only Short, Only Long, Equal Weights.

- **Giả thuyết thay thế (H_1):** Ít nhất một trong các trung bình lợi nhuận khác biệt.

Cách hoạt động của ANOVA:

ANOVA phân tích tổng phương sai của dữ liệu bằng cách chia thành hai phần:

1. **Phương sai giữa các nhóm (Between-group variance):** Đo lường sự khác biệt giữa trung bình của từng nhóm so với trung bình tổng thể. Nếu trung bình của một hoặc nhiều nhóm cách xa trung bình tổng thể, điều này ám chỉ có sự khác biệt đáng kể.
2. **Phương sai trong nhóm (Within-group variance):** Đo lường mức độ biến động trong mỗi nhóm riêng lẻ.

Tỷ lệ giữa phương sai giữa các nhóm và phương sai trong nhóm được đo bằng giá trị **F-statistic**:

$$F = \frac{\text{Phương sai giữa các nhóm}}{\text{Phương sai trong nhóm}}$$

Nhắc Step by Step:

- **F lớn:** Chỉ ra có thể có sự khác biệt đáng kể giữa các nhóm.
- **F nhỏ:** Chỉ ra các nhóm có trung bình gần giống nhau.

Quy trình ANOVA:

1. Tính trung bình tổng thể (\bar{X}_{total}): Trung bình của toàn bộ dữ liệu từ tất cả các chiến lược.
2. Tính phương sai giữa các nhóm:

$$SS_{\text{between}} = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X}_{\text{total}})^2$$

Trong đó:

- k : Số nhóm (4 chiến lược trong trường hợp này).
- n_i : Kích thước mẫu của nhóm i .
- \bar{X}_i : Trung bình của nhóm i .

3. Tính phương sai trong nhóm:

$$SS_{\text{within}} = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

4. Tính tổng phương sai:

$$SS_{\text{total}} = SS_{\text{between}} + SS_{\text{within}}$$

5. Tính giá trị F-statistic:

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

Trong đó:

$$MS_{\text{between}} = \frac{SS_{\text{between}}}{k - 1}, \quad MS_{\text{within}} = \frac{SS_{\text{within}}}{N - k}$$

6. Kiểm định giả thuyết: So sánh giá trị F với ngưỡng F-critical từ bảng phân phối F (hoặc sử dụng giá trị p-value). Nếu $p\text{-value} < 0.05$, bác bỏ H_0 : Có sự khác biệt giữa các chiến lược.

4.5.1.2 Kết quả thống kê

Period	Strategy	F-statistic	P-value	Significance
return_3_month	Long-Short	22.435764	3.111694e-14	Significant
return_3_month	Only Short	22.435764	3.111694e-14	Significant
return_3_month	Only Long	22.435764	3.111694e-14	Significant
return_3_month	Equal Weights	22.435764	3.111694e-14	Significant
return_6_month	Long-Short	41.699398	5.880058e-26	Significant
return_6_month	Only Short	41.699398	5.880058e-26	Significant
return_6_month	Only Long	41.699398	5.880058e-26	Significant
return_6_month	Equal Weights	41.699398	5.880058e-26	Significant
return_9_month	Long-Short	52.228031	3.181147e-32	Significant
return_9_month	Only Short	52.228031	3.181147e-32	Significant
return_9_month	Only Long	52.228031	3.181147e-32	Significant
return_9_month	Equal Weights	52.228031	3.181147e-32	Significant
return_12_month	Long-Short	53.645831	4.639975e-33	Significant
return_12_month	Only Short	53.645831	4.639975e-33	Significant
return_12_month	Only Long	53.645831	4.639975e-33	Significant
return_12_month	Equal Weights	53.645831	4.639975e-33	Significant

Bảng 3: Kết quả kiểm định F-statistic và P-value cho các chiến lược giao dịch

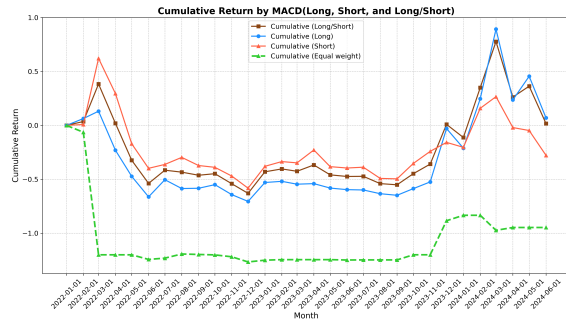
- Tất cả các chiến lược giao dịch qua các khoảng thời gian 3 tháng, 6 tháng, 9 tháng, và 12 tháng đều có giá trị **F-statistic** cao và **P-value** rất nhỏ (gần như bằng 0), cho thấy kết quả kiểm định có ý nghĩa thống kê (**Significant**) với mức độ tin cậy cao.
- **Chiến lược Long-Short** cho từng khung thời gian đều có giá trị **F-statistic** lần lượt là 22.435764 (3 tháng), 41.699398 (6 tháng), 52.228031 (9 tháng), và 53.645831 (12 tháng). Giá trị F-statistic tăng dần khi kéo dài khung thời gian, cho thấy sự tăng cường hiệu quả khi mở rộng khoảng thời gian giao dịch.
- Các chiến lược **Only Long**, **Only Short**, và **Equal Weights** đều có giá trị kiểm định tương tự nhau trong từng khung thời gian, với kết quả nhất quán cho thấy rằng chúng đều có ý nghĩa thống kê.
- **Chiến lược giao dịch 12 tháng** có F-statistic lớn nhất (53.645831), phản ánh mức độ phân biệt mạnh mẽ nhất giữa các nhóm lợi nhuận, và cũng thể hiện hiệu quả tốt hơn khi so sánh với các khung thời gian ngắn hơn.

Kết luận

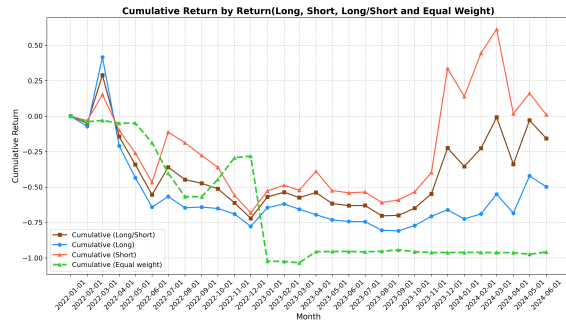
Kết quả kiểm định F-statistic và P-value cho thấy sự khác biệt có ý nghĩa giữa các chiến lược giao dịch trong mỗi khung thời gian. Các chiến lược này đều cho

thấy tính ổn định và hiệu suất mạnh mẽ với mức độ tin cậy cao, đặc biệt khi áp dụng cho chiến lược Long-Short và các khung thời gian dài hơn.

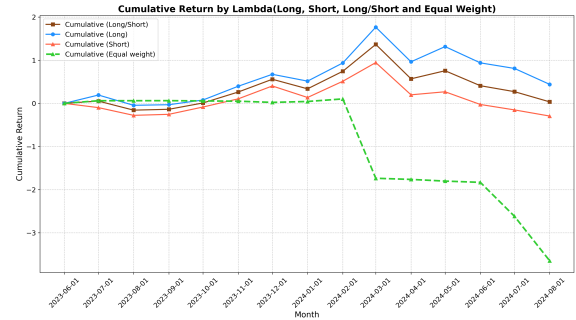
4.5.2 So sánh chiến lược từ các phương pháp xếp hạng



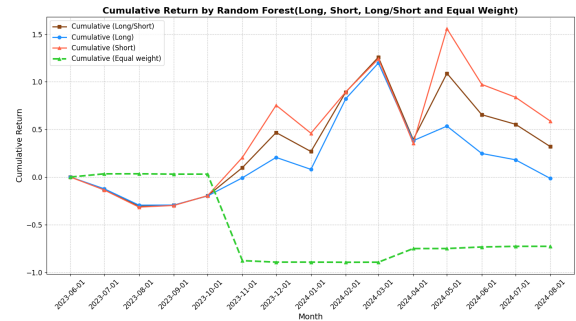
(a) 4 Chiến lược sử dụng chỉ số MACD



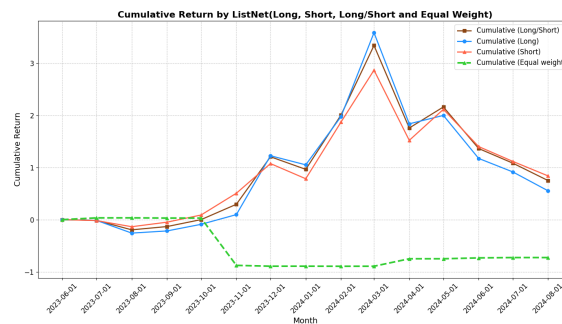
(b) 4 Chiến lược sử dụng chỉ số Return



(c) 4 Chiến lược sử dụng chỉ số mô hình LamdaMART



(d) 4 Chiến lược sử dụng chỉ số Mô hình Random Forest



(e) 4 Chiến lược sử dụng chỉ số mô hình ListNet

Hình 4: So sánh các chiến lược giao dịch từ các phương pháp xếp hạng

Chiến lược chỉ số MACD:

- Lợi suất tích lũy của chiến lược "long-only" có sự biến động lớn, đặc biệt

trong giai đoạn đầu và giữa chu kỳ. Trong khi đó, chiến lược "short-only" cho thấy hiệu suất kém.

- **Chiến lược long-short** có lợi suất tích lũy ổn định, đặc biệt là trong giai đoạn sau của chu kỳ. Việc sử dụng chiến lược long-short có vẻ hiệu quả hơn chiến lược "only long" hay "only short".
- **Trọng số bằng nhau** có lợi suất thấp và không ổn định.

Chiến lược chỉ số Return

- **Chiến lược "long-only"** có sự biến động rất mạnh trong lợi suất tích lũy, tuy nhiên có những thời điểm đạt lợi suất cao.
- **Chiến lược "short-only"** không đem lại lợi suất tốt và có sự giảm sút trong một số giai đoạn.
- **Chiến lược long-short** lại cho thấy sự ổn định hơn trong việc đạt được lợi suất tích lũy tốt hơn các chiến lược còn lại, tương tự như chỉ số MACD.
- **Trọng số bằng nhau** cho kết quả kém hiệu quả với lợi suất thấp và không ổn định.

Chiến lược chỉ số LambdaMART

- **Chiến lược long-only** có sự ổn định nhất trong suốt thời gian, không có sự biến động lớn.
- **Chiến lược short-only** và **trọng số bằng nhau** cho kết quả kém ổn định, với lợi suất thấp và giảm sút.
- **Chiến lược long-short** cho kết quả không vượt trội như chiến lược long-only, nhưng cũng có tính ổn định và giảm thiểu được rủi ro.

Chiến lược chỉ số Random Forest

- **Chiến lược long-only** cho kết quả khá ổn định, mặc dù lợi suất có sự biến động ở giai đoạn sau.
- **Chiến lược short-only** cho thấy hiệu suất khá kém và giảm sút rõ rệt trong lợi suất tích lũy.
- **Chiến lược long-short** cho thấy một sự phân bố lợi suất hợp lý, với hiệu quả không quá tốt nhưng ổn định.
- **Trọng số bằng nhau** cho thấy hiệu quả thấp và không có sự tăng trưởng tích cực.

Tổng kết

- **Chiến lược long-short** thường có hiệu suất tốt hơn, giúp tối ưu hóa lợi suất và giảm thiểu rủi ro trong nhiều chiến lược.
- **Chiến lược long-only** có thể mang lại lợi suất tốt trong một số tình huống, nhưng cần có sự quản lý rủi ro chặt chẽ hơn.
- **Chiến lược short-only** và **trọng số bằng nhau** không cho thấy hiệu quả rõ ràng, đặc biệt là khi so với các chiến lược long-short.

4.5.3 Thông số đánh giá

4.5.3.1 Sharpe Ratio Một thước đo hiệu suất đầu tư đã điều chỉnh rủi ro, được tính bằng cách so sánh lợi nhuận của khoản đầu tư đó với lợi nhuận của tài sản không rủi ro

$$SR = \frac{R_p - R_f}{\sigma_p}$$

trong đó:

- R_p là lợi nhuận của danh mục đầu tư
- R_f là tỷ lệ phi rủi ro
- σ_p độ lệch chuẩn vượt quá của danh mục đầu tư

Điều kiện:

Tỷ lệ Sharpe = 1 : Lợi nhuận của khoản đầu tư

Tỷ lệ Sharpe > 1 : Lợi nhuận tốt hơn so với rủi ro đã chấp nhận

Tỷ lệ Sharpe < 1 : Lợi nhuận không xứng đáng với mức độ rủi ro

4.5.3.2 IR Ratio Là thước đo lường lợi nhuận danh mục đầu tư vượt quá lợi nhuận của điểm chuẩn, thường là chỉ số, so với sự biến động của những lợi nhuận đó. Điểm chuẩn thường là một chỉ số đại diện cho thị trường hoặc một lĩnh vực hoặc ngành cụ thể.

$$IR = \frac{R_p - R_b}{\sigma_{(R_p - R_b)}}$$

trong đó:

- R_p là lợi nhuận của danh mục đầu tư,
- R_b là lợi nhuận của chỉ số tham chiếu,

- $\sigma_{(R_p - R_b)}$ là sai số theo dõi, tức là độ lệch chuẩn của sự khác biệt giữa lợi nhuận danh mục và lợi nhuận của chỉ số tham chiếu.

Tỷ lệ Sharpe so sánh lợi nhuận của tài sản với tỷ lệ phi rủi ro (như trái phiếu kho bạc), không tính đến mối tương quan với các tài sản khác. **Tỷ lệ IR (Information Ratio)** đo lường lợi nhuận so với một điểm chuẩn (như S&P 500) và đánh giá tính nhất quán của hiệu suất đầu tư.

4.5.3.3 Expected Return Lợi nhuận kỳ vọng được sử dụng để đánh giá mức sinh lời trung bình của các tài sản trong danh mục. Trong CSM, lợi nhuận kỳ vọng giúp xác định tài sản nào có xu hướng mang lại lợi nhuận tốt hơn trong quá khứ:

$$E(R) = \frac{1}{T} \sum_{t=1}^T R_t \quad (14)$$

Tại đây, R_t là lợi nhuận của tài sản tại thời điểm t , và T là tổng số thời điểm quan sát.

4.5.3.4 Volatility :

Volatility đo lường rủi ro của các tài sản thông qua độ lệch chuẩn của lợi nhuận. Trong CSM, biến động có thể được dùng để chuẩn hóa lợi nhuận, giúp loại bỏ các yếu tố nhiễu:

$$\sigma = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (R_t - \mu)^2} \quad (15)$$

Tại đây, μ là lợi nhuận trung bình.

4.5.3.5 Max Drawdown :

Mức giảm sâu lớn nhất phản ánh mức lỗ lớn nhất từ đỉnh đến đáy của giá tài sản trong một giai đoạn. Chỉ số này giúp đánh giá rủi ro sụt giảm vốn trong chiến lược CSM:

$$\text{Max Drawdown} = \frac{\text{Peak Value} - \text{Trough Value}}{\text{Peak Value}} \quad (16)$$

4.5.3.6 Hit Rate Tỷ lệ thắng đo lường xác suất một giao dịch trong danh mục CSM có lợi nhuận. Chỉ số này giúp đánh giá khả năng thành công của chiến lược:

$$\text{Hit Rate} = \frac{\text{Số giao dịch có lợi nhuận}}{\text{Tổng số giao dịch}} \quad (17)$$

4.5.3.7 Average Profit Lợi nhuận trung bình của các giao dịch có lời giúp xác định mức sinh lời trung bình trong chiến lược CSM:

$$\text{Average Profit} = \frac{\sum_{i=1}^N P_i}{N} \quad (18)$$

Tại đây, P_i là lợi nhuận của giao dịch thứ i , và N là tổng số giao dịch có lợi nhuận.

4.5.3.8 Average Loss Lỗ trung bình của các giao dịch thua lỗ giúp đánh giá mức độ tổn thất khi chiến lược không thành công:

$$\text{Average Loss} = \frac{\sum_{i=1}^N L_i}{N} \quad (19)$$

Tại đây, L_i là lỗ của giao dịch thứ i , và N là tổng số giao dịch thua lỗ.

4.5.4 So sánh chiến lược

	Long & Short Return	Equal Weights Return	Long & Short MACD	Equal Weights MACD	Long & Short RF	Long & Short LN	Long & Short LDM
Expected Return	0.002019	0.001423	0.000235	0.001808	-0.003711	0.014360	0.016680
Volatility	0.065884	0.001623	0.000169	0.057271	0.049202	0.069686	0.045773
Max Drawdown	-1.016486	-0.000957	-0.001710	-1.008714	-0.997679	-0.970837	-0.864153
Sharpe Ratio	0.030649	0.876742	1.387331	0.031599	-0.075423	0.206073	0.364417
Sortino Ratio	0.000000	1.000000	1.000000	0.000000	0.000000	0.000000	1.000000
Hit Rate	0.472670	0.685115	0.530899	0.451691	0.459034	0.433544	0.454651
Average Profit	0.132063	0.004222	0.000656	0.136490	0.113985	0.182731	0.141753
Average Loss	-0.111308	-0.000161	-0.000040	-0.103653	-0.088296	-0.103604	-0.059178

Bảng 4: Comparison of Strategies

Expected Return (Lợi nhuận kỳ vọng)

- Chiến lược **Long & Short LDM** có lợi nhuận kỳ vọng cao nhất (0.016680), cho thấy khả năng sinh lợi ổn định nhất trong số các chiến lược.
- Ngược lại, chiến lược **Long & Short RF** có lợi nhuận kỳ vọng âm (-0.003711), cho thấy hiệu suất kém với khả năng thua lỗ.

Volatility (Độ biến động)

- Chiến lược **Long & Short LN** có độ biến động cao nhất (0.069686), đồng nghĩa với mức độ rủi ro lớn hơn.

- Trong khi đó, chiến lược **Equal Weights Return** có độ biến động thấp nhất (0.001623), thể hiện tính ổn định cao hơn.

Max Drawdown (Sụt giảm tối đa)

- Chiến lược **Long & Short RF** và **Long & Short LN** đều có mức sụt giảm lớn, lần lượt khoảng -0.997679 và
- 0.970837, cho thấy khả năng chịu rủi ro lớn với các khoản lỗ tiềm năng nghiêm trọng.

Sharpe Ratio

- Chiến lược **Long & Short MACD** có Sharpe Ratio cao nhất (1.387331), biểu thị hiệu suất điều chỉnh rủi ro tốt nhất.
- Chiến lược **Long & Short RF** có Sharpe Ratio âm (-0.075423), cho thấy lợi nhuận không bù đắp được mức rủi ro.

Sortino Ratio

- Tương tự Sharpe Ratio, Sortino Ratio cao nhất thuộc về **Long & Short MACD** (2.623033), thể hiện khả năng kiểm soát rủi ro sụt giảm tốt nhất.

Hit Rate (Tỷ lệ chiến thắng)

- Chiến lược **Equal Weights Return** có tỷ lệ chiến thắng cao nhất (0.685115), trong khi các chiến lược khác dao động từ 0.43 đến 0.47.

Average Profit (Lợi nhuận trung bình) và Average Loss (Thua lỗ trung bình)

- Chiến lược **Long & Short LN** có lợi nhuận trung bình cao nhất (0.182731) nhưng cũng đi kèm với thua lỗ trung bình lớn nhất (-0.103604).
- Chiến lược **Equal Weights MACD** và **Equal Weights Return** có mức lợi nhuận và thua lỗ trung bình nhỏ, cho thấy khả năng kiểm soát biến động giá tốt hơn nhưng với lợi nhuận thấp.

Kết luận:

Tổng thể, các chiến lược có độ hiệu quả khác nhau với lợi nhuận và rủi ro tương ứng. **Long & Short LDM** và **Long & Short MACD** nổi bật về hiệu suất và quản lý rủi ro, trong khi **Long & Short RF** thể hiện hiệu suất kém nhất.

5 Tổng Kết

Thời gian Train-Valid-Test

- Train: Từ tháng 1/2022 đến tháng 12/2022
- Valid: Từ tháng 1/2023 đến tháng 6/2023
- Test: Từ tháng 7/2023 đến tháng 9/2024

Các Mô Hình Được Sử Dụng

- Momentum Cổ Điển:
 - Return: Dựa trên tỷ suất sinh lợi lịch sử.
 - MACD: Sử dụng chỉ báo MACD để xác định xu hướng.
- Regress-Then-Rank: dùng mô hình hồi quy machine learning để dự đoán và xếp hạng điểm số. Mô hình được chọn là Random Forest (dựa vào decision tree)
- Learning to Rank (LTR):
 - LamdaMART: Kết hợp giữa hồi quy và xếp hạng để tối ưu hóa thứ tự cổ phiếu.
 - ListNet: Mô hình xếp hạng dựa trên xác suất.

4 Chiến Lược Thực Hiện

- Only Long: Mua vào 10% các cổ phiếu có xu hướng tăng.
- Only Short: Bán khống 10% các cổ phiếu có xu hướng giảm.
- Long-Short: Kết hợp mua vào 10% và bán khống 10% để tối ưu hóa lợi nhuận.
- Equal-Weight: Đầu tư đều vào tất cả các cổ phiếu trong danh mục.

Kết Quả

- Chiến lược long-short có hiệu suất tốt nhất, giúp tối ưu hóa lợi suất và giảm thiểu rủi ro trong nhiều chiến lược.
- 2 mô hình tốt nhất là Momentum Cổ Điển sử dụng chỉ số MACD và LamdaMART.

Tóm lại, kết quả cho thấy chiến lược Long-Short, kết hợp với mô hình MACD và LamdaMART, mang lại hiệu suất vượt trội so với các mô hình và chiến lược còn lại. Điều này cho thấy tầm quan trọng của việc kết hợp các phương pháp phân tích truyền thống với các kỹ thuật machine learning hiện đại trong việc tối ưu hóa lợi nhuận đầu tư.

6 Tài liệu tham khảo

Tài liệu

- [1] Baz, J., Granger, N. M., Harvey, C. R., Le Roux, N., & Rattray, S. (2015). Dissecting investment strategies in the cross section and time series. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2695101>
- [2] Burges, C. J. C. (2010). From RankNet to LambdaRank to LambdaMART: An overview (tech. rep. MSR-TR-2010-82) *Microsoft Research*. https://www.researchgate.net/publication/228936665_From_ranknet_to_lambdarank_to_lambdamart_An_overview
- [3] Burges, C. J. C., Ragno, R., & Le, Q. V. (2006). Learning to rank with nonsmooth cost functions. *Proceedings of the 19th International Conference on Neural Information Processing Systems*. <https://doi.org/10.7551/mitpress/7503.003.0029>
- [4] Cao, Z., Qin, T., Liu, T.-Y., Tsai, M.-F., & Li, H. (2007). Learning to rank: From pairwise approach to listwise approach. *Proceedings of the 24th International Conference on Machine Learning, ICML '07*(129), 129–136. <https://doi.org/10.1145/1273496.1273513>
- [5] Fama, E. (1965). *The Behavior of Stock-Market Prices*. *The Journal of Business*, 38 (1), 34-105. <https://doi.org/10.1086/294743>
- [6] Fama, E. (1970). *Efficient Capital Markets: A Review of Theory and Empirical Work*. *Journal of Finance*, 25(2), 383-417. <https://doi.org/10.2307/2325486>
- [7] Fama, E. F., & French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33(1), 3–56. [https://doi.org/10.1016/0304-405X\(93\)90023-5](https://doi.org/10.1016/0304-405X(93)90023-5)
- [8] Flavin, M. A. (1983). *Excess Volatility in the Financial Markets: A Re-assessment of the Empirical Evidence*. *Journal of Political Economy*, 91(6) <https://doi.org/10.1086/261194>
- [9] Gu, S., Kelly, B., & Xiu, D. (2018). Empirical Asset Pricing via Machine Learning (tech. rep. w25398). *National Bureau of Economic Research*. Cambridge, MA. <https://doi.org/10.3386/w25398>

- [10] Järvelin, K., & Kekäläinen, J. (2000). IR evaluation methods for retrieving highly relevant documents. *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR'00*, 41–48. <https://doi.org/10.1145/345508.345545>
- [11] Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), 65–91. <https://doi.org/10.1111/j.1540-6261.1993.tb04702.x>
- [12] Kampe, N. L. (2023). Deep neural network construction for enhanced multi-asset time series and cross-sectional momentum strategies. *Master's thesis, University of St. Gallen*.
- [13] Kim, A. Y., Tse, Y., & Wald, J. K. (2016). Time series momentum and volatility scaling. *Journal of Financial Markets*, 30(100), 103–124. <https://doi.org/10.1016/j.finmar.2016.05.003>
- [14] Li, H. (2011). Learning to Rank for Information Retrieval and Natural Language Processing. *Synthesis Lectures on Human Language Technologies*, 4(1), 1–113. <https://doi.org/10.2200/S00348ED1V01Y201104HLT012>
- [15] Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, 104(2), 228–250. <https://doi.org/10.1016/j.jfineco.2011.11.003>
- [16] Poh, D., Lim, B., Zohren, S., & Roberts, S. (2020). Building cross-sectional systematic strategies by learning to rank. *The Journal of Financial Data Science*, 3(2). <https://doi.org/10.48550/arXiv.2012.07149>
- [17] Poh, D. L. C. (2023). Constructing cross-sectional trading strategies: A machine learning approach. *Doctoral dissertation, University of Oxford*.
- [18] Wu, Q., Burges, C. J. C., Svore, K. M., & Gao, J. (2010). Adapting boosting for information retrieval measures. *Information Retrieval*, 13(3), 254–270. <https://doi.org/10.1007/s10791-009-9112-1>