**Tóm tắt (Abstract)**

Trong bài nghiên cứu này, chúng tôi khai thác sức mạnh của học máy để xem xét tính hữu dụng của các thước đo vi mô thị trường kinh điển trong kỷ nguyên giao dịch bằng máy móc. Cụ thể, chúng tôi phân tích mức độ các đặc trưng vi mô giải thích và dự báo những đặc tính then chốt của chuỗi lợi nhuận như biến động thực hiện, phân phối lợi nhuận, và chênh lệch mua–bán.

Trên bộ dữ liệu gồm 87 hợp đồng tương lai thanh khoản trong giai đoạn 5 năm, chúng tôi huấn luyện các mô hình rừng ngẫu nhiên để đánh giá tầm quan trọng của từng thước đo vi mô. Kết quả cho thấy nhiều đặc trưng truyền thống vẫn có sức mạnh giải thích trong mẫu, nhưng sức mạnh dự báo ngoài mẫu lại khác biệt đáng kể. Đặc biệt, **VPIN** (Volume-synchronized Probability of Informed Trading) nổi bật như một thước đo có sức mạnh dự báo ngoài mẫu vượt trội đối với nhiều biến vi mô, trong khi các thước đo khác (như Amihud, Roll, Kyle) thể hiện vai trò trong một số bối cảnh riêng biệt.

Kết quả này khẳng định giá trị bền vững của các thước đo vi mô trong môi trường thị trường hiện đại, đồng thời nhấn mạnh lợi ích khi kết hợp chúng với các kỹ thuật học máy để phân tích và dự báo động lực giá.

**Giới thiệu (Introduction)**

Trong bối cảnh ngày nay, phần lớn giao dịch tài chính được thực hiện bởi máy tính thay vì con người. Thuật toán giao dịch, hệ thống khớp lệnh điện tử, và dữ liệu tốc độ cao đã định hình lại cách thị trường vận hành ở cấp độ vi mô. Điều này đặt ra một câu hỏi quan trọng: **liệu các thước đo vi mô truyền thống – vốn được phát triển từ thời thị trường chủ yếu dựa vào con người – có còn giữ giá trị trong “thời đại máy móc” không?**

Ở mức trực giác, có thể kỳ vọng rằng tầm quan trọng của vi mô thị trường sẽ giảm đi. Khi dữ liệu trở nên dồi dào và việc thực thi diễn ra trong khoảng thời gian mili–micro giây, dường như những yếu tố nhỏ lẻ (như ma sát trong đặt lệnh, sự bất cân xứng thông tin) sẽ bị che mờ trong “biển” dữ liệu khổng lồ. Tuy nhiên, thực tế lại cho thấy điều ngược lại: càng ở thang thời gian ngắn, các cơ chế vi mô của thị trường càng trở nên quan trọng để hiểu động lực giá.

Sự phát triển nhanh chóng của giao dịch tần suất cao (HFT) và thuật toán đã làm thay đổi bản chất cạnh tranh trên thị trường. Các chiến lược không còn đơn giản dựa trên xu hướng giá mà phải tính đến cấu trúc sổ lệnh, mức độ thanh khoản, và dòng lệnh (order flow). Những yếu tố này – vốn là đối tượng nghiên cứu của kinh tế học vi mô thị trường – do đó trở nên thiết yếu trong việc xây dựng, đánh giá và giám sát chiến lược giao dịch.

Trong nghiên cứu này, chúng tôi sử dụng học máy – cụ thể là **rừng ngẫu nhiên** – để đánh giá mức độ hữu ích của các thước đo vi mô trong việc giải thích và dự báo các hiện tượng vi mô quan trọng. Chúng tôi phân loại các thước đo này thành ba nhóm:

1. **Dựa trên giá** (ví dụ: Roll spread).
2. **Dựa trên giá–khối lượng** (ví dụ: Kyle’s lambda, Amihud illiquidity).
3. **Dựa trên dòng lệnh/giao dịch** (ví dụ: PIN, VPIN).

Mục tiêu là kiểm định liệu những thước đo kinh điển này có còn giữ giá trị trong môi trường giao dịch hiện đại, và xác định đặc trưng nào có sức mạnh giải thích trong mẫu, đặc trưng nào có sức mạnh dự báo ngoài mẫu. Kết quả sẽ cung cấp bằng chứng về vai trò bền vững (hoặc giới hạn) của vi mô thị trường trong một thế giới nơi máy móc chiếm ưu thế.

**Phương pháp nghiên cứu (Methodology)**

**1. Dữ liệu**

Chúng tôi sử dụng bộ dữ liệu gồm **87 hợp đồng tương lai thanh khoản cao**, trải dài trên nhiều lớp tài sản (cổ phiếu, trái phiếu, hàng hóa, tiền tệ). Dữ liệu bao gồm thông tin chi tiết ở cấp **tick** trong khoảng thời gian **5 năm**.

Để giảm thiểu sự không đồng nhất về thời gian và khối lượng, dữ liệu được tổng hợp thành **thanh khối lượng đô-la (dollar volume bars)**, tức là mỗi thanh đại diện cho một giá trị giao dịch bằng đô-la cố định, thay vì một khoảng thời gian cố định. Cách tiếp cận này giúp đồng bộ hóa đơn vị quan sát với cường độ thông tin mà thị trường hấp thụ.

**2. Các thước đo vi mô**

Chúng tôi xem xét nhiều thước đo vi mô tiêu biểu thuộc ba nhóm:

1. **Dựa trên giá**
   * **Roll’s spread**: đo chênh lệch mua–bán ẩn dựa trên hiệp phương sai của thay đổi giá.
2. **Dựa trên giá–khối lượng**
   * **Lambda của Kyle**: hệ số đo độ nhạy của giá đối với khối lượng ròng.
   * **Amihud illiquidity**: tỷ lệ tuyệt đối lợi nhuận trên khối lượng đô-la, phản ánh chi phí giao dịch.
3. **Dựa trên giao dịch/dòng lệnh**
   * **PIN (Probability of Informed Trading)**: xác suất giao dịch đến từ nhà đầu tư thông tin.
   * **VPIN (Volume-synchronized PIN)**: biến thể của PIN, đồng bộ theo khối lượng thay vì theo thời gian.

Mỗi thước đo được tính trên một cửa sổ nhìn lại cố định (\(W\)) và được chuẩn hóa để so sánh.

**3. Khung học máy**

Chúng tôi sử dụng mô hình **rừng ngẫu nhiên (Random Forests)** với 100 cây, bootstrap mẫu quan sát, và số đặc trưng xem xét tại mỗi nút là \(\lfloor \sqrt{d} \rfloor\) (với \(d\) là số đặc trưng).

Chúng tôi đo lường **tầm quan trọng của đặc trưng (feature importance)** theo hai cách:

* **MDI (Mean Decrease Impurity)**: mức giảm trung bình độ hỗn tạp (impurity) trong tập huấn luyện, phản ánh sức mạnh giải thích trong mẫu.
* **MDA (Mean Decrease Accuracy)**: mức giảm chính xác khi tráo đổi (permutation) đặc trưng trên tập kiểm tra, phản ánh sức mạnh dự báo ngoài mẫu.

**4. Biến đích**

Các biến được dự báo bao gồm nhiều khía cạnh vi mô then chốt:

* **Chênh lệch mua–bán** (ước lượng theo Corwin–Schultz).
* **Biến động thực hiện** (realized volatility).
* **Đặc tính phân phối lợi nhuận**: độ lệch (skewness), độ nhọn (kurtosis), kiểm định Jarque–Bera (chuẩn tính), và tương quan tuần tự (autocorrelation).

Nhãn dự báo được xác định bằng **dấu của thay đổi** trong các biến đích tại một chân trời dự báo cố định (ví dụ: 250 thanh).

**Kết quả nghiên cứu (Results)**

**1. Tầm quan trọng đặc trưng trong mẫu (MDI)**

Khi đánh giá theo thước đo **MDI**, nhiều chỉ báo vi mô cổ điển thể hiện sức mạnh giải thích cao trong mẫu. Cụ thể:

* **Amihud illiquidity** và **VPIN** thường xuyên nằm trong nhóm đặc trưng quan trọng nhất khi giải thích biến động, chênh lệch mua–bán và đặc tính phân phối lợi nhuận.
* Một số thước đo dựa trên giá đơn thuần như **Roll’s spread** có khả năng giải thích đáng kể trong mẫu, song sức mạnh này không nhất thiết được duy trì trong dự báo ngoài mẫu.

**2. Tầm quan trọng đặc trưng ngoài mẫu (MDA)**

Khi xem xét theo chuẩn **MDA** (dự báo ngoài mẫu), bức tranh thay đổi rõ rệt:

* **VPIN** nổi bật là đặc trưng dự báo hiệu quả nhất đối với phần lớn biến đích, đặc biệt là biến động thực hiện, chênh lệch mua–bán và đuôi phân phối lợi nhuận (kurtosis, Jarque–Bera).
* Các thước đo như **Amihud** và **Kyle’s lambda** vẫn hữu ích trong một số tình huống, song sức mạnh dự báo nhìn chung thấp hơn VPIN.
* **Roll’s spread**, mặc dù kém hiệu quả trong dự báo biến động, lại thể hiện ưu thế nhất định khi dự báo **tương quan tuần tự (autocorrelation)**.

**3. Tính ổn định qua các điều kiện khác nhau**

Kết quả cũng cho thấy thứ hạng tầm quan trọng đặc trưng khá bền vững:

* **Thay đổi siêu tham số** (số cây, độ sâu cây) không làm xáo trộn đáng kể thứ hạng của VPIN.
* **Thay đổi phương pháp xây dựng thanh dữ liệu** (thanh khối lượng đô-la so với thanh thời gian) cho thấy VPIN vẫn giữ vị thế nổi bật.
* **Thay đổi theo giai đoạn thời gian** (sub-sample analysis) khẳng định tính ổn định của VPIN như một chỉ báo mạnh ngoài mẫu.

**4. Ý nghĩa tổng quan**

Những phát hiện này cho thấy:

* Không phải thước đo nào có sức mạnh giải thích trong mẫu đều có giá trị dự báo ngoài mẫu.
* **VPIN** có khả năng phản ánh sự mất cân đối dòng lệnh và mức độ “độc hại” của giao dịch, từ đó dự báo tốt các hiện tượng vi mô quan trọng.
* Các chỉ báo khác (Amihud, Roll, Kyle) không bị loại bỏ, mà đóng vai trò bổ trợ, hữu ích trong các bối cảnh chuyên biệt.

**Kết luận và Hàm ý (Conclusion)**

Kết quả nghiên cứu cho thấy rằng, trái với quan niệm phổ biến, các thước đo vi mô thị trường truyền thống **không hề mất đi giá trị** trong bối cảnh giao dịch hiện đại do máy móc thống trị. Thay vào đó, khi được đánh giá bằng khung học máy phi tham số, nhiều chỉ báo vi mô vẫn chứng minh được sức mạnh giải thích và dự báo đáng kể.

Đặc biệt, **VPIN (Volume-synchronized Probability of Informed Trading)** nổi bật như thước đo có sức mạnh dự báo ngoài mẫu vượt trội đối với hầu hết các biến vi mô then chốt. Điều này cho thấy VPIN nắm bắt hiệu quả sự mất cân đối dòng lệnh và mức độ bất cân xứng thông tin, những yếu tố trung tâm trong việc lý giải và dự báo động lực giá.

Tuy nhiên, nghiên cứu cũng chỉ ra rằng các thước đo khác không nên bị bỏ qua. Ví dụ, Amihud illiquidity và Roll’s spread thể hiện vai trò hữu ích trong một số khía cạnh cụ thể, như dự báo tương quan tuần tự hay đặc tính thanh khoản. Như vậy, **một bộ công cụ đa dạng các thước đo vi mô** sẽ hữu ích hơn so với việc chỉ dựa vào một chỉ báo duy nhất.

**Hàm ý thực tiễn**

* Đối với **nhà quản trị rủi ro**, VPIN và các thước đo vi mô khác cung cấp tín hiệu quan trọng về mức độ độc hại của dòng lệnh, hỗ trợ quản lý thanh khoản và rủi ro giao dịch.
* Với **nhà giao dịch và tạo lập thị trường**, những chỉ báo này giúp cải thiện chiến lược thực thi, điều chỉnh hành vi đặt lệnh trong môi trường tần suất cao.
* Đối với **nghiên cứu học thuật**, kết quả nhấn mạnh sự cần thiết phải kết hợp kinh tế lượng vi mô với các kỹ thuật học máy để nâng cao khả năng dự báo trong bối cảnh dữ liệu lớn và thị trường phức tạp.

Tóm lại, trong “thời đại máy móc”, vi mô thị trường không những không mất đi tính quan trọng mà còn trở thành yếu tố then chốt để hiểu và dự báo động lực giá. Các thước đo cổ điển, khi được đặt trong khung phân tích hiện đại, tiếp tục phát huy giá trị và cung cấp nền tảng hữu ích cho cả nghiên cứu lẫn thực tiễn.