

# Flow Toxicity and Liquidity in a High-Frequency World

## 1 Một số khái niệm

### 1.1 Order Flow Toxicity

Order flow toxicity được định nghĩa là khi các nhà tạo lập thị trường (market makers) bị bất lợi do dòng lệnh từ các nhà giao dịch có thông tin. Họ cung cấp thanh khoản nhưng không ý thức được rằng mình đang giao dịch ở mức bất lợi, dẫn đến thua lỗ.

**Tác động:** Giảm động lực của market makers, ảnh hưởng tiêu cực đến thanh khoản thị trường và gia tăng biến động giá.

### 1.2 VPIN: Volume-Synchronized Probability of Informed Trading

VPIN dựa trên sự mất cân bằng của khối lượng giao dịch (imbalance volume) và cường độ giao dịch (trade intensity) được tính trong Volume Time (thời gian dựa trên khối lượng thay vì thời gian thực).

Không yêu cầu ước lượng tham số quan sát, giúp nó phù hợp với các thị trường tần suất cao.

VPIN hoạt động như một chỉ báo hiệu quả cho biến động ngắn hạn do độc tính của dòng lệnh (toxicity-induced volatility) và cải thiện khả năng dự đoán biến động hơn so với các phương pháp truyền thống.

### 1.3 Bulk Volume Classification (BVC)

**Khái niệm:** Quy trình phân loại giao dịch thành *buy* hoặc *sell* dựa trên khối lượng lệnh trong một khoảng thời gian cụ thể.

**Ưu điểm:** Được tối ưu hóa cho các thị trường tần suất cao, nơi các phương pháp phân loại truyền thống như *tick test* hoặc *quote test* có thể không phù hợp.

**Bid price:** Giá mua - giá cao nhất mà người mua sẵn sàng trả.

**Ask price:** Giá bán - giá thấp nhất mà người bán sẵn sàng chấp nhận.

**Spread:**  $\text{Spread} = \text{Ask Price} - \text{Bid Price}$

### 1.4 High-Frequency Trading Firms

Dù chỉ chiếm khoảng 2% trong tổng số gần 20.000 công ty giao dịch, các HF trading firms lại chiếm:

- Hơn 70% khối lượng giao dịch trên thị trường chứng khoán Mỹ (từ 2009).

- Gần 50% khối lượng giao dịch trên thị trường hợp đồng tương lai.

Hoạt động chính của HF trading firms thường là tạo lập thị trường (market making), cung cấp thanh khoản thông qua lệnh thụ động (passive orders). Họ không đặt cược theo hướng giá cả (no directional bets), mà kiếm lợi nhuận từ biên độ nhỏ thông qua khối lượng giao dịch lớn.

## 1.5 Volume Bucketing

Dữ liệu được chia thành các khoảng sao cho mỗi *bucket* có khối lượng giao dịch bằng nhau, giúp giảm tác động của thời gian hoặc biến động cụm.

Ví dụ: Nếu tổng khối lượng giao dịch là 1 triệu cổ phiếu và bạn muốn 10 *buckets*, mỗi bucket sẽ chứa 100.000 cổ phiếu, bất kể thời gian khớp lệnh kéo dài bao lâu.

**Time Bar:** Biểu đồ mà mỗi thanh (bar) đại diện cho một khoảng thời gian cố định (1 phút, 5 phút, 1 giờ, v.v.).

**Bulk Bar:** Biểu đồ mà mỗi thanh chỉ được tạo ra khi đạt đến một lượng giao dịch nhất định, thay vì dựa trên thời gian cố định.

## Bố Cục Bài Báo

- **Phần 1:** Khung lý thuyết và cách PIN ảnh hưởng đến chênh lệch giá mua-bán.
- **Phần 2:** Quy trình ước tính chỉ số VPIN.
- **Phần 3:** Đánh giá độ tin cậy của VPIN.
- **Phần 4:** Ước tính VPIN cho chỉ số cổ phiếu và hợp đồng tương lai dầu.
- **Phần 5:** Khả năng dự báo biến động của VPIN.
- **Phần 6:** Tóm tắt các phát hiện.

## 2 Nội dung chính của bài báo

### 2.1 The Model - Mô hình

Trong phần này mô tả mô hình cơ bản để suy ra mức độ độc hại của dòng lệnh. Bắt đầu với một mô hình vi cấu trúc thị trường tiêu chuẩn, từ đó xây dựng thước đo mức độ độc hại của dòng lệnh, PIN, và sau đó điều chỉnh PIN để áp dụng cho thị trường giao dịch tần suất cao.

Một loạt nghiên cứu (Easley và O'Hara 1987, 1992; Easley, Kiefer, O'Hara và Paperman 1996; Easley, Engle, O'Hara và Wu 2008) đã chứng minh cách một mô hình vi cấu trúc thị trường có thể được ước tính cho từng cổ phiếu riêng lẻ bằng cách sử dụng dữ liệu giao dịch để xác định xác suất giao dịch dựa trên thông tin (**PIN**).

Mô hình vi cấu trúc này xem giao dịch như một trò chơi giữa các nhà cung cấp thanh khoản và các nhà giao dịch (position takers), được lặp lại qua các khoảng thời gian giao dịch  $i = 1, \dots, I$ . Ở đầu mỗi khoảng giao dịch, một sự kiện thông tin có thể xảy ra hoặc không, với xác suất  $\alpha$ . Nếu thông tin là tin tốt, các nhà giao dịch có thông tin biết rằng vào cuối khoảng thời gian giao dịch, tài sản sẽ có giá  $\bar{S}_i$ , và nếu thông tin là tin xấu, tài

sản sẽ có giá  $\underline{S}_i$ , với  $\bar{S}_i > \underline{S}_i$ . Tin tốt xảy ra với xác suất  $(1 - \delta)$ , trong khi tin xấu xảy ra với xác suất còn lại  $\delta$ .

$\alpha$  là xác suất mà một sự kiện thông tin (information event) xảy ra ở đầu mỗi khoảng thời gian giao dịch. Nếu một sự kiện thông tin xảy ra, nó có thể là tin tốt (good news) hoặc tin xấu (bad news). **Xác suất tin tốt:**  $(1 - \delta)$  **Xác suất tin xấu:**  $\delta$ . Nếu không có sự kiện thông tin xảy ra (xác suất  $1 - \alpha$ ), thị trường chỉ có giao dịch từ các nhà giao dịch không có thông tin (uninformed traders).

Sau khi sự kiện thông tin xảy ra hoặc không xảy ra, hoạt động giao dịch trong khoảng thời gian đó bắt đầu, với các nhà giao dịch xuất hiện theo **quá trình Poisson** trong suốt khoảng thời gian giao dịch. Trong những khoảng có sự kiện thông tin, các lệnh từ nhà giao dịch có thông tin đến với tỷ lệ  $\mu$ . Những nhà giao dịch này sẽ **mua nếu có tin tốt** và **bán nếu có tin xấu**. Trong mỗi khoảng thời gian, các lệnh mua và bán từ những nhà giao dịch không có thông tin đến với tỷ lệ  $\varepsilon$  cho cả hai phía.

Mô hình cấu trúc liên kết các kết quả quan sát được trên thị trường (tức là các lệnh mua và bán) với các thông tin không quan sát được và các quy trình đặt lệnh cơ bản. Các nghiên cứu trước đây tập trung vào việc ước lượng các tham số xác định những quy trình này bằng phương pháp **hợp lý cực đại** (maximum likelihood method).

Một cách trực quan, mô hình diễn giải mức độ mua và bán bình thường của một cổ phiếu là giao dịch không có thông tin và sử dụng dữ liệu để xác định tỷ lệ dòng lệnh không có thông tin,  $\varepsilon$ . Các giao dịch mua hoặc bán bất thường được hiểu là giao dịch dựa trên thông tin và được sử dụng để xác định  $\mu$ . Số khoảng thời gian xuất hiện giao dịch mua hoặc bán bất thường được dùng để xác định  $\alpha$  và  $\delta$ .

Một nhà cung cấp thanh khoản sử dụng kiến thức về các tham số này để xác định mức giá mà anh ta sẵn sàng mua vào (bid) và mức giá mà anh ta sẵn sàng bán ra (ask). Hai mức giá này khác nhau, tạo ra chênh lệch giá mua-bán (bid-ask spread), vì nhà cung cấp thanh khoản không biết liệu đối tác giao dịch của mình có phải là nhà giao dịch có thông tin hay không.

Khoảng chênh lệch này chính là sự khác biệt giữa giá trị kỳ vọng của tài sản khi có người mua từ nhà cung cấp thanh khoản và giá trị kỳ vọng của tài sản khi có người bán cho nhà cung cấp thanh khoản. Hai kỳ vọng có điều kiện này khác nhau do vấn đề lựa chọn bất lợi (adverse selection), được gây ra bởi khả năng xuất hiện các nhà giao dịch có thông tin tốt hơn.

Khi giao dịch diễn ra, các nhà cung cấp thanh khoản quan sát các lệnh và được mô hình hóa theo cách sử dụng quy tắc Bayes để cập nhật niềm tin của họ về mức độ độc hại của dòng lệnh, được mô tả trong mô hình bằng các ước lượng tham số. Gọi  $P(t) = (P_n(t), P_b(t), P_g(t))$  là niềm tin của nhà cung cấp thanh khoản về các sự kiện "không có tin tức" (n), "tin xấu" (b), "tin tốt" (g) tại thời điểm  $t$ . Niềm tin ban đầu của họ tại thời điểm  $t = 0$  là  $P(0) = (1 - \alpha, \alpha\delta, \alpha(1 - \delta))$ .

Để xác định giá mua hoặc giá bán tại thời điểm  $t$ , nhà cung cấp thanh khoản cập nhật niềm tin của mình dựa trên sự xuất hiện của một lệnh giao dịch có liên quan. Giá trị kỳ vọng của tài sản tại thời điểm  $t$ , với điều kiện dựa trên lịch sử giao dịch trước đó, được tính như sau:

$$E[S_i|t] = P_n(t)S_i^* + P_b(t)S_i + P_g(t)\bar{S}_i,$$

trong đó

$$S_i^* = \delta\underline{S}_i + (1 - \delta)\bar{S}_i,$$

là giá trị kỳ vọng ban đầu của tài sản.

Giá mua  $B(t)$  là giá trị kỳ vọng của tài sản có điều kiện khi có người muốn bán tài sản cho nhà cung cấp thanh khoản. Do đó, ta có:

$$B(t) = E[S_i|t] - \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)} (E[S_i|t] - \underline{S}_i).$$

Tương tự, giá bán  $A(t)$  là giá trị kỳ vọng của tài sản có điều kiện khi có người muốn mua tài sản từ nhà cung cấp thanh khoản. Do đó, ta có:

$$A(t) = E[S_i|t] + \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)} (\bar{S}_i - E[S_i|t]).$$

Những phương trình này minh họa vai trò rõ ràng của sự xuất hiện của các nhà giao dịch có thông tin và không có thông tin trong việc ảnh hưởng đến báo giá. Nếu không có nhà giao dịch có thông tin ( $\mu = 0$ ), thì giao dịch không mang thông tin, và do đó giá mua và giá bán đều bằng với giá trị kỳ vọng ban đầu của tài sản.

Ngược lại, nếu không có nhà giao dịch không có thông tin ( $\varepsilon = 0$ ), thì giá mua và giá bán lần lượt đạt mức giá tối thiểu và tối đa. Ở các mức giá này, không có nhà giao dịch có thông tin nào thực hiện giao dịch, và thị trường về cơ bản sẽ ngừng hoạt động.

Thông thường, cả nhà giao dịch có thông tin và không có thông tin đều có mặt trên thị trường, do đó giá mua sẽ thấp hơn  $E[S_i|t]$  và giá bán sẽ cao hơn  $E[S_i|t]$ .

Chênh lệch giá mua-bán tại thời điểm  $t$  được ký hiệu là  $\Sigma(t) = A(t) - B(t)$ . Chênh lệch này được tính bằng:

$$\Sigma(t) = \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)} (\bar{S}_i - E[S_i|t]) + \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)} (E[S_i|t] - \underline{S}_i).$$

Thuật ngữ đầu tiên trong phương trình chênh lệch biểu thị xác suất rằng một lệnh mua là một giao dịch dựa trên thông tin, nhân với mức lỗ kỳ vọng khi giao dịch với một người mua có thông tin. Thuật ngữ thứ hai là một biểu thức đối xứng cho các lệnh bán.

Chênh lệch cho các báo giá ban đầu trong khoảng giao dịch,  $\Sigma$ , có một dạng đặc biệt đơn giản trong trường hợp tự nhiên khi các sự kiện tin tốt và tin xấu có xác suất bằng nhau. Nghĩa là, nếu  $\delta = 1 - \delta$ , thì:

$$\Sigma = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} (\bar{S}_i - \underline{S}_i).$$

Thành phần quan trọng của mô hình này là xác suất một lệnh được đặt bởi một nhà giao dịch có thông tin, được gọi là PIN. Có thể dễ dàng chứng minh rằng xác suất giao dịch đầu tiên trong một khoảng thời gian là giao dịch dựa trên thông tin được tính bằng:

$$PIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon},$$

trong đó  $\alpha\mu + 2\varepsilon$  là tỷ lệ đến của tất cả các lệnh và  $\alpha\mu$  là tỷ lệ đến của các lệnh dựa trên thông tin. Do đó, PIN đo lường tỷ lệ lệnh đến từ các nhà giao dịch có thông tin so với tổng dòng lệnh, và phương trình chênh lệch giá chỉ ra rằng đây là yếu tố chính quyết định mức chênh lệch giá mua-bán.

Những phương trình này minh họa rằng các nhà cung cấp thanh khoản cần ước tính chính xác PIN để xác định mức tối ưu để tham gia thị trường. Việc PIN tăng lên bất ngờ sẽ dẫn đến tổn thất cho các nhà cung cấp thanh khoản không điều chỉnh giá của họ.

## 2.2 Chỉ số VPIN và Ước lượng Tham số

Cách tiếp cận tiêu chuẩn để tính toán mô hình PIN sử dụng phương pháp ước lượng hợp lý cực đại (maximum likelihood) để ước tính các tham số không quan sát được  $(\alpha, \mu, \delta, \varepsilon)$ , vốn quyết định quá trình ngẫu nhiên của các giao dịch, và sau đó suy ra giá trị PIN từ các ước lượng tham số này.

Trong phần này đề xuất một phương pháp ước lượng trực tiếp về mức độ độc hại (toxicity), không yêu cầu quá trình ước lượng số trung gian đối với các tham số không quan sát được. Chúng tôi cập nhật thước đo này trong thời gian dựa trên khối lượng giao dịch (volume time) nhằm cố gắng điều chỉnh theo tốc độ xuất hiện của thông tin mới trên thị trường.

Cách tiếp cận dựa trên khối lượng này, mà chúng tôi gọi là **VPIN**, cung cấp một chỉ số đơn giản để đo lường mức độ độc hại của dòng lệnh trong môi trường giao dịch tần suất cao. Trước tiên, chúng tôi sẽ thảo luận về vai trò của thông tin và thời gian trong giao dịch tần suất cao.

### 2.2.1 Bản chất của thông tin và thời gian

Trong mô hình giao dịch tuần tự tiêu chuẩn, thông tin thường được xem như dữ liệu cung cấp dấu hiệu về giá trị tương lai của tài sản. Trong bối cảnh thị trường cổ phiếu, thông tin có thể liên quan đến các sự kiện tương lai, chẳng hạn như triển vọng của công ty hoặc thị trường đối với sản phẩm của công ty.

Trong một thị trường hiệu quả, giá trị của tài sản sẽ hội tụ về giá trị thông tin đầy đủ của nó, khi các nhà giao dịch có thông tin tìm cách thu lợi nhuận từ thông tin của họ bằng cách giao dịch. Vì các nhà tạo lập thị trường có thể giữ vị thế mua hoặc bán cổ phiếu, nên biến động giá trị tương lai của tài sản ảnh hưởng đến lợi nhuận của họ. Do đó, họ cố gắng suy luận bất kỳ thông tin mới tiềm ẩn nào từ các mô hình giao dịch. Chính niềm tin được cập nhật của họ sẽ được phản ánh vào giá mua và giá bán của họ.

Trong một thế giới giao dịch tần suất cao, các nhà tạo lập thị trường phải đối mặt với cùng một vấn đề cơ bản, mặc dù khung thời gian mà họ hoạt động làm thay đổi một số khía cạnh theo những cách thú vị.

Một nhà tạo lập thị trường tần suất cao, người dự đoán sẽ nắm giữ cổ phiếu trong vài phút, sẽ bị ảnh hưởng bởi thông tin tác động đến giá trị của cổ phiếu trong khoảng thời gian đó. Thông tin này có thể liên quan đến các yếu tố cơ bản của tài sản, nhưng cũng có thể phản ánh các yếu tố liên quan đến bản chất của hoạt động giao dịch trên toàn thị trường hoặc nhu cầu thanh khoản cụ thể trong một khoảng thời gian nhất định. Ví dụ, trong một hợp đồng tương lai, thông tin làm gia tăng nhu cầu phòng ngừa rủi ro (hedging demand) đối với hợp đồng thường sẽ ảnh hưởng đến giá hợp đồng tương lai, và do đó, nó có ý nghĩa đối với một nhà tạo lập thị trường. Định nghĩa rộng hơn về thông tin này có nghĩa là các sự kiện thông tin có thể xảy ra thường xuyên trong ngày và chúng có thể có mức độ quan trọng khác nhau đối với biên độ biến động giá trong tương lai. Tuy nhiên, sự tồn tại của chúng vẫn được phản ánh thông qua bản chất và thời điểm của các giao dịch.

Khía cạnh quan trọng nhất của mô hình giao dịch tần suất cao là các giao dịch không diễn ra cách đều nhau theo thời gian. Các lệnh giao dịch xuất hiện với tần suất không đều, và một số giao dịch quan trọng hơn những giao dịch khác vì chúng tiết lộ lượng thông tin khác nhau.

Ví dụ, như Hình 1 minh họa, hoạt động giao dịch trong hợp đồng tương lai E-mini S&P 500 (đường màu xanh và trục tỷ lệ bên trái của biểu đồ) và hợp đồng tương lai EUR/USD

(đường màu đỏ và trục tỷ lệ bên phải của biểu đồ) thể hiện tính chu kỳ trong ngày khác nhau. Sự xuất hiện của thông tin mới trên thị trường sẽ kích hoạt làn sóng các quyết định được đưa ra, dẫn đến sự bùng nổ về khối lượng giao dịch. Thông tin liên quan đến các sản phẩm khác nhau xuất hiện vào những thời điểm khác nhau, do đó tạo ra sự biến động khối lượng giao dịch theo chu kỳ trong ngày là khác biệt.

Trong nghiên cứu này, thay vì mô hình hóa thời gian theo đồng hồ, chúng tôi làm việc với thời gian theo khối lượng (\*volume time\*). Easley và O'Hara (1992) đã phát triển ý tưởng rằng khoảng thời gian giữa các giao dịch có mối tương quan với sự xuất hiện của thông tin mới, tạo cơ sở để chúng tôi xem xét thời gian giao dịch thay vì thời gian đồng hồ. Có vẻ hợp lý khi cho rằng một thông tin càng quan trọng thì càng thu hút nhiều khối lượng giao dịch hơn. Bằng cách lấy mẫu mỗi khi thị trường trao đổi một lượng khối lượng cố định, chúng tôi cố gắng mô phỏng sự xuất hiện của thông tin có mức độ quan trọng tương đương trên thị trường. Nếu một thông tin cụ thể tạo ra khối lượng giao dịch gấp đôi so với một thông tin khác, chúng tôi sẽ lấy số quan sát gấp đôi, qua đó nhân đôi trọng số của nó trong mẫu.

### 2.2.2 Phân chia theo khối lượng (Volume Bucketing)

Trong ví dụ trên, nếu chúng tôi lấy một mẫu hợp đồng tương lai E-mini S&P 500 sau mỗi 200.000 hợp đồng được giao dịch, trung bình chúng tôi sẽ thu được khoảng chín mẫu mỗi ngày. Vào những ngày có hoạt động giao dịch sôi động, số lượng mẫu có thể là bội số lớn của chín, trong khi vào những ngày ít giao dịch hơn, số điểm dữ liệu thu thập được sẽ ít hơn. Vì hợp đồng tương lai EUR/USD có khối lượng giao dịch trung bình hàng ngày chỉ bằng khoảng một phần mười so với hợp đồng tương lai E-mini S&P 500, nên để đạt được mục tiêu lấy chín mẫu mỗi ngày, khoảng cách khối lượng giữa hai lần lấy mẫu sẽ cần được giảm xuống khoảng 20.000 hợp đồng. Do mô hình giao dịch trong ngày khác nhau giữa hai hợp đồng này, khi chúng tôi thu thập mẫu quan sát đầu tiên của ngày đối với hợp đồng tương lai E-mini S&P 500, thì gần như đồng thời chúng tôi đã thu được mẫu quan sát thứ tư của ngày đối với hợp đồng tương lai EUR/USD.

Để thực hiện phương pháp lấy mẫu dựa trên khối lượng này, chúng tôi nhóm các giao dịch liên tiếp vào các khối lượng bằng nhau, gọi là volume buckets, với kích thước  $V$  được xác định ngoại sinh. Một volume bucket là tập hợp các giao dịch có tổng khối lượng bằng  $V$ . Nếu giao dịch cuối cùng để hoàn thành một bucket có khối lượng lớn hơn mức yêu cầu, phần dư thừa sẽ được chuyển sang bucket tiếp theo. Chúng tôi ký hiệu  $r$  là chỉ số của các volume buckets có khối lượng bằng nhau. Một thuật toán chi tiết cho quá trình đóng gói khối lượng này được trình bày trong Phụ lục trực tuyến. Phương pháp lấy mẫu theo volume buckets cho phép chúng tôi chia phiên giao dịch thành các khoảng thời gian có nội dung thông tin tương đương, trong đó sự mất cân bằng giao dịch có tác động kinh tế đáng kể đối với các nhà cung cấp thanh khoản.

### 2.2.3 Phân loại khối lượng mua và khối lượng bán

Một vấn đề mà chúng tôi chưa đề cập là cách phân biệt giữa khối lượng mua và khối lượng bán. Nhắc lại rằng khối lượng có dấu (\*signed volume\*) là cần thiết vì khả năng tương quan của nó với mức độ độc hại của dòng lệnh (\*order toxicity\*). Mặc dù tổng khối lượng giao dịch có thể báo hiệu sự xuất hiện của thông tin mới, nhưng hướng của khối lượng giao dịch lại thể hiện ý nghĩa của nó đối với sự thay đổi giá. Do đó, sự áp đảo của khối lượng mua (bán) có thể cho thấy mức độ độc hại của dòng lệnh phát sinh từ sự xuất hiện của tin tốt (xấu).

Nghiên cứu vi cấu trúc thị trường (\*microstructure research\*) đã dựa vào các thuật toán dựa trên dấu tick (\*tick-based algorithms\*) để xác định hướng giao dịch. Tuy nhiên, việc phân loại giao dịch luôn là một vấn đề nan giải.

Một trong những vấn đề là cách báo cáo giao dịch trên các thị trường có thể xử lý lệnh mua và lệnh bán khác nhau. Ví dụ, Sở Giao dịch Chứng khoán New York (\*NYSE\*) chỉ báo cáo một giao dịch nếu một lệnh bán lớn được khớp với nhiều lệnh mua trên sổ lệnh, nhưng sẽ báo cáo nhiều giao dịch nếu một lệnh mua lớn được khớp với nhiều lệnh bán. Tương tự, việc chia nhỏ các lệnh lớn thành nhiều lệnh nhỏ hơn có nghĩa là các giao dịch diễn ra trong khoảng thời gian ngắn thực tế không phải là những quan sát độc lập. Để giải quyết những vấn đề này, các nhà nghiên cứu thực nghiệm đã sử dụng phương pháp tổng hợp các giao dịch cùng phía thị trường trong khoảng thời gian ngắn thành một quan sát duy nhất.

Một khó khăn thứ hai là việc xác định hướng giao dịch cũng đòi hỏi phải so sánh giá giao dịch với báo giá hiện hành. Các nhà giao dịch thực hiện lệnh tại giá mua (\*bid\*) của nhà tạo lập thị trường được coi là người bán, trong khi những người thực hiện lệnh tại giá bán (\*ask\*) được coi là người mua. Các giao dịch nằm giữa khoảng giá này thường được phân loại bằng thuật toán dựa trên dấu tick (\*tick-based algorithm\*).

Thuật toán Lee-Ready (1991) cũng đề xuất sử dụng độ trễ năm giây giữa báo giá được ghi nhận và giá giao dịch để phản ánh thực tế rằng cơ chế báo giá lên hệ thống không giống với cơ chế báo cáo giao dịch. Ngay cả trong một môi trường giao dịch đơn giản hơn với các nhà giao dịch chuyên biệt (\*specialist trading\*), lỗi trong phân loại giao dịch vẫn khá đáng kể.

Trong môi trường giao dịch tần suất cao, việc phân loại giao dịch trở nên phức tạp hơn nhiều. Trong các thị trường hợp đồng tương lai mà chúng tôi nghiên cứu, không có nhà giao dịch chuyên biệt (\*specialist\*), và thanh khoản được cung cấp thông qua một sổ lệnh điện tử chứa các lệnh giới hạn được đặt bởi nhiều loại nhà giao dịch khác nhau. Trong thị trường điện tử này, một nhà giao dịch có thể khớp lệnh tại mức giá giống với giao dịch trước đó hoặc có thể gửi một lệnh giới hạn cải thiện giá giao dịch trước đó, khiến quy tắc tick có thể gán sai hướng của giao dịch. Ngoài ra, việc chia nhỏ lệnh là tiêu chuẩn, việc hủy báo giá và lệnh diễn ra tràn lan, và khối lượng giao dịch lớn đến mức áp đảo. Sử dụng dữ liệu hợp đồng tương lai E-mini S&P 500 từ tháng 5 năm 2010, chúng tôi nhận thấy rằng trung bình một ngày có 2.650.391 thay đổi báo giá tốt nhất (\*best-bid-or-offer - BBO\*) do việc thêm hoặc hủy lệnh, và 789.676 thay đổi báo giá do các giao dịch. Vì BBO thay đổi nhiều lần giữa các giao dịch, nhiều hợp đồng được giao dịch tại cùng một mức giá thực tế đã diễn ra cả ở giá mua và giá bán. Trong thế giới giao dịch tần suất cao này, việc áp dụng các thuật toán tiêu chuẩn lên từng giao dịch riêng lẻ trở nên có vấn đề.

$$V_{\tau}^B = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i \cdot Z \left( \frac{P_i - P_{i-1}}{\sigma_{\Delta P}} \right)$$

$$V_{\tau}^S = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i \cdot \left[ 1 - Z \left( \frac{P_i - P_{i-1}}{\sigma_{\Delta P}} \right) \right] = V - V_{\tau}^B, \quad (7)$$

trong đó  $t(\tau)$  là chỉ số của thanh thời gian cuối cùng được bao gồm trong \*volume bucket\* thứ  $\tau$ ,  $Z$  là hàm phân phối tích lũy (\*CDF - Cumulative Distribution Function\*) của phân phối chuẩn chuẩn tắc, và  $\sigma_{\Delta P}$  là ước lượng độ lệch chuẩn của sự thay đổi giá giữa các thanh thời gian. Quy trình của chúng tôi chia đều khối lượng giao dịch trong một thanh thời gian thành khối lượng mua và khối lượng bán nếu không có sự thay đổi giá từ

đầu đến cuối thanh thời gian. Ngược lại, nếu giá tăng, khối lượng sẽ được phân bổ nhiều hơn về phía mua so với bán, và mức phân bổ này phụ thuộc vào độ lớn của sự thay đổi giá so với phân phối của các biến động giá.

Một điểm khác biệt quan trọng giữa phương pháp phân loại khối lượng (\*bulk classification\*) và thuật toán Lee-Ready là thuật toán Lee-Ready phân loại toàn bộ khối lượng giao dịch thành hoặc mua hoặc bán, trong khi phương pháp phân loại khối lượng chia một phần khối lượng là mua và phần còn lại là bán. Nói cách khác, thuật toán Lee-Ready cung cấp một phân loại rời rạc, trong khi thuật toán phân loại khối lượng là liên tục. Điều này có nghĩa là ngay cả trong trường hợp cực đoan khi một thanh thời gian duy nhất lấp đầy một \*volume bucket\*, khối lượng vẫn có thể được cân bằng hoàn toàn theo phương pháp phân loại khối lượng, phụ thuộc vào  $\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}$ .

Mục đích chính của chúng tôi khi sử dụng khối lượng giao dịch là để tính toán sự mất cân bằng lệnh (\*order imbalance\*). Gọi  $OI_\tau = |V_\tau^B - V_\tau^S|$  là sự mất cân bằng lệnh trong \*volume bucket\*  $\tau$ . Thước đo này, tất nhiên, chỉ là một xấp xỉ của mất cân bằng lệnh thực tế vì nó dựa trên phương pháp phân loại khối lượng giao dịch theo xác suất.

Chúng tôi trước tiên xem xét cách  $E[OI_\tau]$  liên quan đến tốc độ giao dịch bằng cách chứng minh rằng nó không bị ảnh hưởng bởi một phép co giãn đơn giản của khối lượng giao dịch. Giả sử rằng khối lượng của mỗi thanh thời gian được co giãn theo một hệ số  $\beta > 0$ , tức là  $V_i^* = \beta V_i$ , và sự mất cân bằng khối lượng được phân phối đồng đều trong \*bucket\*. Khi đó, số lượng thanh thời gian cần thiết để lấp đầy một \*bucket\* sẽ tỷ lệ nghịch với  $\beta$ , tức là  $\frac{t(\tau) - t(\tau-1)}{\beta}$ . Từ phương trình (7), giá trị kỳ vọng của mất cân bằng lệnh vẫn không đổi:

$$E[OI_\tau^*] = E[|V_\tau^{*B} - V_\tau^{*S}|] = \frac{1}{\beta} E[|\beta V_\tau^B - \beta V_\tau^S|] = E[OI_\tau]. \quad (8)$$

Thứ hai, chúng tôi đặt câu hỏi liệu trong một giới hạn hợp lý, khoảng thời gian của một thanh thời gian có ảnh hưởng đến thước đo mất cân bằng lệnh hay không. Để xác định điều này, chúng tôi tính toán mất cân bằng lệnh cho hợp đồng tương lai E-mini S&P 500 trong giai đoạn từ ngày 1 tháng 1 năm 2008 đến ngày 1 tháng 8 năm 2011, sử dụng các thanh thời gian có độ dài từ 1 đến 240 phút. Với mỗi quy định về thanh thời gian, chúng tôi sử dụng 50 \*volume buckets\* mỗi ngày và tính toán tỷ lệ giữa mất cân bằng lệnh so với kích thước \*bucket\*, được đo lường bằng khối lượng trong mỗi \*bucket\*. Chúng tôi nhận thấy rằng tỷ lệ mất cân bằng lệnh so với kích thước \*bucket\* (như một hàm của số lượng trung bình các thanh thời gian trên mỗi \*bucket\*) tăng dần khi số thanh thời gian trong mỗi \*bucket\* giảm, nhưng không bao giờ tiến gần đến giá trị 1, và cuối cùng ổn định khi sử dụng các thanh thời gian quá dài một cách phi thực tế. Do đó, đối với các thanh thời gian có độ dài hợp lý, thời lượng của một thanh thời gian có rất ít tác động đến việc đo lường sự mất cân bằng lệnh.

Phương pháp này sẽ dẫn đến một số trường hợp phân loại sai khối lượng giao dịch. Tuy nhiên, mục tiêu của chúng tôi không phải là phân loại chính xác từng giao dịch riêng lẻ (một nhiệm vụ gần như bất khả thi), mà là phát triển một chỉ báo về sự mất cân bằng tổng thể của giao dịch, giúp xây dựng một thước đo mức độ độc hại của dòng lệnh (\*toxicity\*). Chúng tôi sử dụng các thanh thời gian (\*time bars\*) để tạo khoảng thời gian cho giá thị trường điều chỉnh theo hướng giao dịch, thông tin mà chúng tôi cố gắng khôi phục thông qua phương pháp phân loại khối lượng (\*bulk classification\*). Sau đó, trong bài báo này, chúng tôi sẽ trình bày bằng chứng cho thấy phương pháp phân loại khối lượng của chúng tôi mang lại kết quả hữu ích hơn trong việc ước tính mức độ độc hại của dòng lệnh so với phương pháp phân loại chi tiết dựa trên dữ liệu giao dịch thô.



#### 2.2.4 Xác suất giao dịch có thông tin đồng bộ theo khối lượng (Chỉ số độc hại dòng lệnh VPIN)

Mô hình PIN tiêu chuẩn chỉ xem xét số lượng lệnh mua và bán để suy ra thông tin về cấu trúc thông tin cơ bản; trong đó không có vai trò rõ ràng của khối lượng giao dịch. Trong các thị trường tần suất cao mà chúng tôi phân tích, số lượng giao dịch là một vấn đề phức tạp. Quay lại nền tảng lý thuyết của PIN, điều chúng tôi thực sự muốn là thông tin về ý định giao dịch phát sinh từ các nhà giao dịch có hoặc không có thông tin. Mỗi liên kết giữa các ý định giao dịch này và dữ liệu giao dịch thực tế có rất nhiều nhiễu, vì ý định giao dịch có thể bị chia nhỏ thành nhiều phần để giảm tác động thị trường, một lệnh có thể tạo ra nhiều giao dịch khớp lệnh, và các giao dịch dựa trên thông tin có thể được thực hiện dưới nhiều hình thức lệnh khác nhau. Vì những lý do này, chúng tôi xem mỗi giao dịch được báo cáo như một sự tổng hợp của các giao dịch có kích thước đơn vị (tức là một giao dịch gồm năm hợp đồng ở một mức giá  $p$  sẽ được xem như năm giao dịch, mỗi giao dịch một hợp đồng tại mức giá  $p$ ). Quy ước này đưa cường độ giao dịch vào phân tích một cách rõ ràng.

Chúng tôi biết từ nghiên cứu của Easley, Engle, O'Hara và Wu (2008) rằng đối với mỗi khoảng thời gian, giá trị kỳ vọng của mất cân bằng lệnh là

$$E[|V_\tau^S - V_\tau^B|] \approx \alpha\mu$$

và tổng số lệnh kỳ vọng là

$$E[V_\tau^B + V_\tau^S] = \alpha\mu + 2\varepsilon.$$

Phương pháp phân chia theo khối lượng (\*volume bucketing\*) giúp chúng tôi ước lượng mô hình này một cách đơn giản. Cụ thể, chúng tôi chia ngày giao dịch thành các \*volume buckets\* có kích thước bằng nhau và xem mỗi \*bucket\* là tương đương với một khoảng thời gian tiếp nhận thông tin. Điều này có nghĩa là  $V_\tau^B + V_\tau^S$  là một hằng số và bằng  $V$  cho mọi  $\tau$ . Chúng tôi sau đó xấp xỉ mất cân bằng lệnh kỳ vọng bằng cách tính trung bình mất cân bằng lệnh qua  $n$  buckets.

Từ các giá trị được tính toán ở trên, chúng tôi có thể viết xác suất giao dịch có thông tin đồng bộ theo khối lượng (\*volume-synchronized probability of informed trading\*), hay chỉ số độc hại dòng lệnh VPIN, như sau:

$$VPIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} \approx \frac{\sum_{\tau=1}^n |V_\tau^S - V_\tau^B|}{nV}. \quad (9)$$

Ước lượng chỉ số VPIN yêu cầu lựa chọn  $V$ , tức là khối lượng trong mỗi \*bucket\*, và  $n$ , số lượng \*buckets\* được sử dụng để xấp xỉ mất cân bằng lệnh kỳ vọng. Trong thiết lập ban đầu, chúng tôi tập trung vào giá trị  $V$  bằng một phần năm mươi của khối lượng giao dịch trung bình hàng ngày. Nếu sau đó chúng tôi chọn  $n = 50$ , thì chỉ số VPIN sẽ được tính toán trên 50 \*buckets\*, và trong một ngày có khối lượng giao dịch trung bình, điều này tương ứng với việc tìm chỉ số VPIN hàng ngày. Kết quả của chúng tôi vẫn nhất quán với nhiều lựa chọn khác nhau của  $V$  và  $n$ , như sẽ được thảo luận trong Mục 5. Chỉ số VPIN được cập nhật sau mỗi \*volume bucket\*. Do đó, khi \*bucket\* thứ 51 được lấp đầy, chúng tôi loại bỏ \*bucket\* thứ 1 và tính toán VPIN mới dựa trên các \*buckets\* từ 2 đến 51. Chúng tôi cập nhật chỉ số VPIN theo thời gian khối lượng (\*volume time\*) vì hai lý do. Thứ nhất, chúng tôi muốn tốc độ cập nhật VPIN phản ánh tốc độ thông tin đến thị trường. Chúng tôi sử dụng khối lượng giao dịch như một đại diện cho tốc độ

tiếp nhận thông tin để đạt được mục tiêu này. Thứ hai, chúng tôi mong muốn mỗi lần cập nhật dựa trên một lượng thông tin tương đương. Trong các giai đoạn giao dịch có ít sự tham gia, khối lượng giao dịch có thể rất mất cân bằng, và trong những giai đoạn có khối lượng thấp như vậy, dường như ít có khả năng xuất hiện thông tin mới. Vì vậy, nếu cập nhật chỉ số VPIN theo thời gian đồng hồ (\*clock time\*), các lần cập nhật có thể dựa trên các lượng thông tin không đồng nhất.

Ví dụ, hãy xem xét giao dịch hợp đồng tương lai E-mini S&P 500 vào ngày 6 tháng 5 năm 2010. Khối lượng giao dịch trong ngày này (được nhớ đến vì sự kiện \*flash crash\*) cực kỳ cao, do đó quy trình của chúng tôi tạo ra 137 ước lượng của chỉ số VPIN, so với mức trung bình 50 ước lượng hàng ngày. Vì độ dài mẫu ( $n$ ) cũng là 50, nên khoảng thời gian được sử dụng để tính toán một số ước lượng của chỉ số VPIN vào ngày 6 tháng 5 năm 2010 chỉ kéo dài vài giờ, so với mức trung bình là 24 giờ.

Hình 2 minh họa cách phạm vi thời gian trở nên "co giãn" tùy thuộc vào cường độ giao dịch, vốn được xem như một đại diện cho tốc độ tiếp nhận thông tin. Vào lúc 9:30 sáng (EST), dữ liệu được sử dụng để tính toán VPIN bao phủ gần như toàn bộ một ngày.

Tuy nhiên, khi Sở Giao dịch Chứng khoán New York (\*NYSE\*) mở cửa vào ngày 6 tháng 5 năm 2010, thuật toán của chúng tôi cập nhật chỉ số VPIN thường xuyên hơn và dựa trên một khoảng thời gian ngắn hơn theo thời gian đồng hồ. Đến 12:17 chiều, VPIN được tính toán dựa trên khoảng thời gian chỉ bằng một nửa ngày. Lưu ý rằng việc giảm phạm vi thời gian của mẫu dữ liệu không dẫn đến các ước lượng nhiễu hơn. Ngược lại, chỉ số VPIN tiếp tục thay đổi theo một xu hướng liên tục. Lý do là vì các khoảng thời gian không chứa cùng một lượng thông tin có thể so sánh được. Thay vào đó, chính các khoảng khối lượng (\*volume ranges\*) mới tạo ra lượng thông tin tương đương cho mỗi lần cập nhật.

Các mô hình GARCH cung cấp một cách tiếp cận thay thế để xử lý hiện tượng cụm biến động (\*volatility clustering\*), đặc trưng của dữ liệu tần suất cao khi lấy mẫu theo thời gian đồng hồ (\*clock time\*). Làm việc theo thời gian khối lượng (\*volume time\*) giúp giảm tác động của hiện tượng cụm biến động, vì chúng tôi tạo ra các ước lượng dựa trên các mẫu có khối lượng bằng nhau. Vì những biến động giá lớn thường đi kèm với khối lượng giao dịch lớn, nên việc lấy mẫu theo khối lượng có thể được xem như một đại diện cho việc lấy mẫu theo biến động giá. Kết quả thu được là một tập hợp các quan sát có phân phối gần với phân phối chuẩn hơn và ít có tính phương sai không đồng nhất (\*heteroscedasticity\*) hơn so với khi lấy mẫu đồng đều theo thời gian đồng hồ. Do đó, làm việc theo thời gian khối lượng có thể được xem như một phương pháp thay thế đơn giản cho việc sử dụng mô hình GARCH.

Để thấy cách biến đổi này cho phép khôi phục một phần tính bình thường, chúng tôi xem xét mẫu dữ liệu một phút của hợp đồng tương lai E-mini S&P 500 từ ngày 1 tháng 1 năm 2008 đến ngày 1 tháng 8 năm 2011. Chúng tôi lấy trung bình năm mươi quan sát giá mỗi ngày, chia đều theo thời gian trong trường hợp đầu tiên (thời gian tuần tự) và chia đều theo khối lượng trong trường hợp thứ hai (thời gian theo khối lượng). Tiếp theo, chúng tôi tính toán sai phân bậc nhất và chuẩn hóa từng mẫu. Cả hai mẫu đều có độ lệch âm và đuôi dày, nhưng mẫu theo thời gian khối lượng gần với phân phối chuẩn hơn, ít tương quan chuỗi hơn và ít dị phương sai hơn. Điều này trở nên rõ ràng hơn khi tần suất lấy mẫu tăng lên (so sánh các bảng với năm mươi và một trăm lần lấy mẫu mỗi ngày). Bảng 1 cung cấp các thống kê kết quả, và Hình 3 minh họa sự thay đổi giá đã được chuẩn hóa.

## 2.3 Tính ổn định của ước lượng VPIN

Ước lượng chỉ số VPIN liên quan đến nhiều vấn đề đặc tả. Trong phần này, chúng tôi sẽ chỉ ra tính ổn định của chỉ số VPIN đối với hai trong số những vấn đề quan trọng nhất - đó là các phương pháp phân loại khối lượng thay thế và sự thay đổi trong hồ sơ giao dịch. Các phép tính của chúng tôi dựa trên dữ liệu hợp đồng tương lai E-mini S&P 500 với các thanh dữ liệu mỗi phút, từ ngày 1 tháng 1 năm 2008 đến ngày 15 tháng 8 năm 2011, cho kích thước nhóm tương thích với trung bình năm mươi nhóm khối lượng mỗi ngày, và độ dài mẫu của năm mươi nhóm.

### 2.3.1 Tính ổn định dưới các phương pháp phân loại khối lượng khác nhau

Lựa chọn cách phân loại khối lượng có ảnh hưởng quan trọng đến ước lượng của VPIN. Cụ thể, vì VPIN liên quan đến việc xem xét sự mất cân bằng giao dịch và cường độ, việc tổng hợp theo các thanh thời gian sẽ làm giảm độ nhiễu trong biến này cũng như thay đổi tỷ lệ của nó. Trong Easley, López de Prado, và O'Hara (2011a), chúng tôi đã lập luận rằng điều này đòi hỏi phải xem xét các mức độ tương đối của VPIN như được mô tả bởi hàm phân phối tích lũy của chúng thay vì các mức tuyệt đối của VPIN.

Điều này có thể được minh họa bằng cách xem xét hành vi của VPIN trong một ngày cụ thể với các thuật toán phân loại khối lượng khác nhau. "Sự sụp đổ đột ngột" vào ngày 6 tháng 5 đặc biệt quan trọng trong các thị trường hợp đồng tương lai (và cổ phiếu), vì vậy chúng tôi minh họa hành vi của VPIN trong ngày này bằng ba phương pháp phân loại khối lượng thay thế. Hình 4(a) hiển thị VPIN được tính toán bằng cách phân loại khối lượng theo thanh thời gian một phút; 4(b) sử dụng phân loại khối lượng theo thanh thời gian mười giây; và 4(c) sử dụng phân loại giao dịch theo từng lệnh của Lee-Ready. Cả ba phương pháp đều đồng ý trong việc chỉ ra một mức độ cực đoan cho chỉ số VPIN về tính độc hại của dòng chảy ít nhất hai giờ trước khi xảy ra sự cố (xem đường chấm CDF(VPIN) cắt ngưỡng 0.9). Hai phương pháp phân loại theo thanh thời gian một phút và mười giây đưa ra các kết quả tương tự về mức độ tương đối của VPIN. Trong cả hai trường hợp, VPIN tăng trước khi xảy ra sự cố và duy trì mức cao trong suốt phần còn lại của ngày. Tuy nhiên, kết quả phân loại giao dịch theo từng lệnh lại khác. Ở đây, VPIN ước tính tăng trước khi xảy ra sự cố, mặc dù không mạnh mẽ như với các thanh thời gian, nhưng lại giảm xuống mức thấp bất thường ngay sau khi xảy ra sự cố và duy trì mức thấp trong phần còn lại của ngày. Tuy nhiên, sự ì động này diễn ra trong một giai đoạn mà thị trường đã tăng hơn 4%. Có vẻ như rất khó tin rằng khối lượng thực sự được cân bằng trong suốt giai đoạn này. Chúng tôi nghi ngờ rằng đây chủ yếu là kết quả của việc phân loại giao dịch sai hơn là bất kỳ yếu tố nào khác, vì vậy chúng tôi không sử dụng phân loại giao dịch theo từng lệnh. Thay vào đó, chúng tôi báo cáo kết quả cho các thanh thời gian một phút, đồng thời lưu ý rằng các kết quả tương đối với các thanh thời gian mười giây là rất giống nhau.

### 2.3.2 Tính ổn định đối với sự thay đổi trong hồ sơ giao dịch

Một thử nghiệm ổn định thứ hai liên quan đến tác động của sự thay đổi trong hồ sơ giao dịch đối với ước lượng VPIN. Những sự khác biệt nhỏ trong hồ sơ giao dịch, chẳng hạn như các giao dịch bị thiếu, tạo ra hai tác động. Thứ nhất, các giao dịch bị thiếu có thể làm thay đổi sự mất cân bằng khối lượng. Tuy nhiên, vì chúng tôi sử dụng một lượng khối lượng tương đương với một phiên giao dịch đầy đủ, tác động này được kỳ vọng là không đáng kể (một giao dịch điển hình chỉ liên quan đến một vài hợp đồng, so với trung

bình hàng ngày hơn hai triệu hợp đồng giao dịch trên hợp đồng tương lai E-mini S&P 500 trong năm 2010). Tác động thứ hai của các giao dịch bị thiếu là sự thay đổi trong quỹ đạo VPIN. Tác động này có thể được đánh giá bằng cách thay đổi điểm bắt đầu của các quỹ đạo VPIN và tính toán độ lệch chuẩn theo chiều ngang theo thời gian.

Hình 5 cho thấy độ lệch chuẩn theo chiều ngang của sự khác biệt trên 1.000 quỹ đạo VPIN là không đáng kể. Dãy thời gian đó có giá trị trung bình là 0.015, trong khi giá trị trung bình của VPIN lớn hơn nhiều (khoảng 0.23). Nhưng như chúng tôi đã lập luận trước đó, việc các nhóm của mỗi quỹ đạo VPIN trong số 1.000 nhóm không được hoàn thành đồng thời có nghĩa là độ lệch chuẩn theo chiều ngang thực tế còn nhỏ hơn giá trị trung bình 0.015 này. Các đỉnh trong độ lệch chuẩn theo chiều ngang mà chúng ta thấy trong Hình 5 trùng với các đỉnh trong giá trị VPIN. Ví dụ, vào ngày 6 tháng 5 năm 2010, độ lệch chuẩn theo chiều ngang chỉ vượt qua 0.02, vào một ngày mà VPIN đạt mức gần 0.5.

Kết luận, sự khác biệt trong hồ sơ giao dịch do các giao dịch bị thiếu hoặc các điểm bắt đầu thay thế dường như không ảnh hưởng đáng kể đến ước lượng VPIN. Để thấy được điểm này, hãy xem xét hai ước lượng của VPIN, một là 0.45, cái còn lại là 0.5. Sự khác biệt này vượt xa những gì chúng ta thấy trong Hình 5. Mặc dù sự khác biệt giữa hai ước lượng VPIN này có vẻ lớn, nhưng chúng không khác biệt đáng kể vì hai ước lượng VPIN này ở gần như cùng một điểm trên CDF của VPIN (0.991 so với 0.995).

## 2.4 Ước lượng Chỉ số VPIN trên Hợp đồng Tương lai

Sau khi đã thiết lập tính ổn định của các phương pháp ước lượng của chúng tôi cho chỉ số độc hại VPIN, chúng tôi tiếp tục minh họa ứng dụng của nó đối với hai hợp đồng tương lai giao dịch sôi động nhất: hợp đồng E-mini S&P 500 (giao dịch trên CME) và hợp đồng tương lai dầu thô WTI (giao dịch trên NYMEX). Khoảng thời gian mẫu của chúng tôi là từ ngày 1 tháng 1 năm 2008 đến ngày 6 tháng 6 năm 2011, sử dụng tại mỗi thời điểm hết hạn có khối lượng giao dịch cao nhất trong ngày, kéo dài tới tương lai. Chúng tôi sử dụng kích thước nhóm bằng  $1/50$  khối lượng giao dịch trung bình trong khoảng thời gian mẫu của mình ( $V$ ). Các tham số được ước lượng trên của số di động có kích thước mẫu  $n = 50$  (tương đương với một ngày giao dịch khối lượng trung bình). Chúng tôi sử dụng toàn bộ khoảng thời gian mẫu để xác định hàm phân phối tích lũy của các tham số ước tính. Bảng 2 cung cấp các thống kê cơ bản của các ước lượng chỉ số VPIN cho các hợp đồng này.

Một mối quan ngại tự nhiên phát sinh từ các ước lượng VPIN của chúng tôi là các hệ số AR(1) của nó rất gần với 1. Điều này có thể gợi ý rằng VPIN có một gốc đơn vị, với các tác động tương ứng là VPIN không ổn định và do đó, hàm phân phối tích lũy của nó trong bất kỳ khoảng thời gian mẫu nào sẽ không hữu ích trong việc đánh giá khả năng xảy ra các giá trị trong tương lai. May mắn thay, đây không phải là trường hợp. Chúng tôi đã chỉ ra rằng VPIN không bị ảnh hưởng bởi sự thay đổi trong ngày bắt đầu của mẫu mà chúng tôi xem xét trong Hình 5. Bảng 3 cũng chỉ ra rằng các CDF của VPIN thu được bằng cách chia mẫu của chúng tôi thành trước và sau sự cố "flash crash" (ngày 6 tháng 5 năm 2011) gần như là giống hệt nhau. Do đó, VPIN thực tế rất ổn định. Hệ số AR(1) cao và tính ổn định của VPIN xảy ra vì cùng một lý do. Chỉ số VPIN của chúng tôi được tính toán sử dụng năm mươi nhóm. Khi nó được cập nhật, nhóm đầu tiên trong số năm mươi nhóm hiện có sẽ bị loại bỏ và nhóm mới nhất sẽ được thêm vào. Quá trình trung bình này khiến VPIN có sự tự tương quan cao, nhưng cũng đảm bảo rằng quá trình này không có trí nhớ dài, vì tại mỗi thời điểm, giá trị hiện tại của VPIN không thể phụ

thuộc vào giá trị mà VPIN đã có trong năm mươi một nhóm trước đó.

#### 2.4.1 S&P 500 (CME)

Hình 6 cho thấy sự tiến triển của hợp đồng tương lai E-mini S&P 500 (đường màu đỏ, được biểu diễn theo giá trị thị trường) và giá trị chỉ số VPIN của nó (đường màu xanh). Chỉ số VPIN nhìn chung ổn định, mặc dù nó rõ ràng thể hiện sự biến động đáng kể. Chúng tôi nhận thấy rằng chỉ số VPIN đạt mức cao nhất trong mẫu này vào ngày 6 tháng 5 năm 2010, ngày xảy ra sự cố "flash crash". Những mức độ VPIN cao như vậy phù hợp với việc phần lớn dòng lệnh có một chiều trong suốt thời gian tương đương với một ngày giao dịch. Như đã thảo luận trong Easley, López de Prado, và O'Hara (2011a), tính độc hại quá mức này đã khiến một số nhà tạo lập thị trường trở thành người tiêu thụ thanh khoản thay vì cung cấp thanh khoản khi họ đóng cửa hoạt động của mình. Một sự kiện gần đây của tính độc hại cực đoan đã xảy ra sau trận động đất ở Nhật Bản. Mặc dù trận động đất lớn Tohoku và sóng thần xảy ra vào sáng sớm ngày 11 tháng 3 năm 2011, chỉ số S&P 500 không trải qua một biến động lớn cho đến khi cuộc khủng hoảng hạt nhân Fukushima bùng phát vào ngày 14 tháng 3 năm 2011. Vào ngày này, S&P 500 đã ghi nhận một mức độ độc hại dòng lệnh cực đoan khác. Không giống như vào ngày 6 tháng 5 năm 2010, sự cố vào ngày 14 tháng 3 năm 2011 xảy ra với khối lượng giao dịch thấp, trong phiên giao dịch đêm (từ 6 giờ chiều đến 11 giờ tối theo giờ EST). Sau khi chỉ có 287.360 hợp đồng được giao dịch, chỉ số này đã mất khoảng 2,5% giá trị của nó. Hình 7 cho thấy CDF(VPIN) đã đạt ngưỡng 0.97 từ 3 giờ chiều, minh họa rằng tính độc hại dòng chảy cũng xảy ra trong những trường hợp có cường độ giao dịch giảm.

#### 2.4.2 WTI Dầu Thô (NYMEX)

Dầu thô là mặt hàng được giao dịch nhiều nhất và vai trò chiến lược của nó trong nền kinh tế thế giới khiến nó trở thành lựa chọn lý tưởng để thực hiện các động thái địa chính trị và kinh tế vĩ mô. Các hợp đồng cá cược. Hợp đồng tương lai năng lượng cũng là nơi mà các nhà tạo lập thị trường phải đối mặt với sự biến động cực độ trong dòng lệnh. Như được thể hiện trong Hình 8, mức độ độc tính dòng cao nhất cho hợp đồng này xảy ra vào ngày 6 tháng 5 năm 2010. Hành vi này phù hợp với thực tế rằng mặc dù các vấn đề vào ngày 6 tháng 5 không liên quan đến năng lượng, nhưng hợp đồng tương lai dầu vẫn bị ảnh hưởng bởi sự lan truyền của thanh khoản và độc tính trên các thị trường. Ngoài ngày xảy ra sự cố "flash crash", mức độ độc tính cao tiếp theo cho hợp đồng này xảy ra vào ngày 5 tháng 5.

Vào đầu tháng 5 năm 2011, CFTC đã báo cáo vị thế đầu cơ mua lớn nhất trong lịch sử của các nhà giao dịch dầu thô. Từ New York Times cho rằng các vị thế lớn này xuất phát từ niềm tin của các nhà giao dịch rằng giá năng lượng sẽ tăng mạnh, được thúc đẩy bởi bạo lực lan rộng khắp Bắc Phi và Trung Đông. Một số nhà giao dịch này đã quyết định chốt lời vào ngày 5 tháng 5 năm 2011. Việc thoái lui các vị thế khổng lồ của họ đã khiến họ tìm kiếm thanh khoản, nhưng khi các nhà tạo lập thị trường nhận ra áp lực bán kéo dài, họ bắt đầu rút lui, điều này lại càng làm tăng sự tập trung của dòng độc tính trong tổng khối lượng giao dịch. Đến 9:53 sáng, chỉ số CDF(VPIN) đã vượt ngưỡng 0,9 và duy trì ở đó trong suốt phần còn lại của ngày. Trong vài giờ đó, WTI đã mất hơn 8%.

## 2.5 Độc tính và Biến động Giá trong Tương lai

Trong một thị trường giao dịch tần suất cao, các nhà tạo lập thị trường có thể sử dụng chỉ số VPIN được tính toán ở trên để đo lường mức độ độc tính của dòng lệnh. Vì độc tính ảnh hưởng đến lợi nhuận của các nhà tạo lập thị trường, nên nó cũng sẽ ảnh hưởng đến hành vi của họ và, theo đó, ảnh hưởng đến tính thanh khoản trên thị trường. Trong phần này, chúng tôi sẽ trình bày chi tiết hơn về mối liên hệ giữa độc tính và biến động giá trong tương lai.

Như đã đề cập trước đó trong bài viết, thời gian không phải là một khái niệm có ý nghĩa đặc biệt đối với các nhà tạo lập thị trường tần suất cao. Vì các nhà tạo lập thị trường là những người giao dịch thụ động, họ phải chờ dòng lệnh đến với họ, nên không phải thời gian mà khối lượng mới là thước đo chính. Cùng một thước đo khối lượng này cũng có liên quan khi xem xét mối liên hệ trong tương lai giữa độc tính và biến động giá. Một nhà tạo lập thị trường cần biết cách độc tính sẽ ảnh hưởng đến hành vi giá như thế nào khi họ nắm giữ một vị thế.

Các nhà tạo lập thị trường tìm cách xoay vòng các vị thế của mình nhiều lần trong ngày, nhưng tần suất họ có thể làm điều đó phụ thuộc vào khối lượng giao dịch. Do đó, đối với nhà tạo lập thị trường, có hai câu hỏi quan trọng. Thứ nhất, độc tính cao ảnh hưởng như thế nào đến hành vi giá trong suốt thời gian nắm giữ này? Và thứ hai, sự kéo dài của mức độc tính cao ảnh hưởng như thế nào đến hành vi giá?

Mặc dù các câu hỏi này có vẻ đơn giản, nhưng việc trả lời chúng thì không. Một khó khăn là các mô hình vi cấu trúc chuẩn không phù hợp để mô tả hành vi trong thế giới giao dịch tần suất cao mới. Chúng ta đơn giản là chưa có các mô hình về nhiều nhà tạo lập thị trường cạnh tranh, phải đối mặt với các hạn chế về thông tin và tồn kho, và quản lý rủi ro bằng cách di chuyển giữa các thị trường hoặc thậm chí rút hoàn toàn khỏi chúng trong vài micro giây. Do đó, lý thuyết chưa cung cấp được mối liên hệ chính xác giữa độc tính, thanh khoản và sự biến động mà chúng ta tìm kiếm. Một khó khăn khác là các phương pháp kinh tế lượng để phân tích thanh khoản và sự biến động trong thế giới này vẫn còn sơ khai, phản ánh nhiều đặc điểm độc đáo của dữ liệu tần suất cao đã được đề cập trong suốt bài viết này.

Để giải quyết các câu hỏi này, chúng tôi dựa trên các mối quan hệ cơ bản để xem xét mối liên hệ giữa độc tính, thanh khoản và sự biến động. Trước tiên, chúng tôi xem xét mối quan hệ giữa độc tính và biến động giá được xác định trong các khoảng khối lượng giao dịch tiếp theo, mà chúng tôi cho rằng đây là khoảng thời gian phù hợp từ góc nhìn của nhà tạo lập thị trường. Sau đó, chúng tôi xem xét cách mà sự kéo dài của độc tính ảnh hưởng đến hành vi lợi nhuận trong các khoảng thời gian dài hơn. Nhìn chung, chúng ta biết rằng khi độc tính tăng lên, các nhà tạo lập thị trường phải đối mặt với nguy cơ thua lỗ tiềm tàng và có thể chọn cách giảm bớt hoặc thậm chí từ bỏ các hoạt động tạo lập thị trường. Sự giảm thanh khoản này, đến lượt nó, gợi ý rằng mức VPIN cao có thể báo trước sự biến động giá lớn hơn.

### 2.5.1 5.1 Bề mặt tương quan

Chúng ta bắt đầu bằng cách đặt ra một câu hỏi đơn giản: Liệu chỉ số biến động VPIN có tương quan với các biến động giá trong tương lai không? Để đo lường mối quan hệ này, chúng ta sử dụng hệ số tương quan Pearson giữa logarit tự nhiên của VPIN và mức sinh lời tuyệt đối của giá trong khoảng khối lượng tiếp theo,  $\rho\left((VPIN)_{\tau-1}, \left|\frac{P_{\tau}}{P_{\tau-1}} - 1\right|\right)$ , trong đó  $\tau$  là các khoảng khối lượng.

Vì VPIN có thể được ước tính bằng các kết hợp khác nhau của số lượng khoảng khối lượng mỗi ngày và độ dài mẫu, chúng ta sẽ xem xét cách các tham số ước tính này ảnh hưởng đến mối quan hệ giữa chỉ số VPIN và các biến động giá trong tương lai.

Đối với hợp đồng tương lai E-mini S&P 500, VPIN có tương quan dương với độ biến động giá trong tương lai. Mối quan hệ này được minh họa trong Hình 10, trong đó mỗi điểm trên đồ thị được tính toán bằng các mẫu lớn, có khi lên đến hơn 44.000 quan sát. Hình này cho thấy rằng tương quan giữa VPIN và mức sinh lời tuyệt đối trong tương lai thay đổi một cách mượt mà khi thay đổi các tham số. Nói chung, tăng độ dài mẫu sẽ làm tăng hệ số tương quan, tương tự như việc tăng số lượng khoảng khối lượng mỗi ngày, dù ở mức độ thấp hơn.

Kết hợp (50,250) có vẻ là một cặp hợp lý cho E-mini S&P 500, và nó có thể được hiểu đơn giản là “một tuần” dữ liệu (năm mươi khoảng khối lượng mỗi ngày và năm ngày giao dịch mỗi tuần). Với sự kết hợp này, chúng ta thu được hệ số tương quan  $\rho\left(VPIN_{\tau-1}, \left|\frac{P_{\tau}}{P_{\tau-1}} - 1\right|\right) = 0.400$  trên 44.537 quan sát.

Hình 10 gợi ý rằng không có nhiều lợi thế khi “quá khớp” các tham số này, vì có một vùng rộng các kết hợp tham số mang lại sức mạnh dự đoán tương tự.

Hình 11 cung cấp biểu đồ về mức sinh lời của hợp đồng tương lai E-mini S&P 500 trong khoảng khối lượng tiếp theo (1/50 của khối lượng trung bình mỗi ngày sử dụng kết hợp (50,250)) được sắp xếp theo mức VPIN trước đó. Biểu đồ trải dài theo chiều dọc khi VPIN tăng, minh họa rằng mức độ độc tính cao hơn, được đo bằng VPIN cao hơn, dẫn đến mức sinh lời tuyệt đối lớn hơn.

Những phát hiện này về mặt tương quan mang tính gợi ý, nhưng tương quan đơn giản là một tiêu chí hẹp về sự phụ thuộc. Thực tế, VPIN thể hiện sự tương quan nối tiếp đáng kể, điều này khiến cho việc rút ra kết luận từ một mối tương quan đơn giản trở nên có vấn đề. Một phương án thay thế là ước tính mô hình chuỗi thời gian của quá trình kết hợp giữa VPIN và mức sinh lời. Chúng tôi không theo đuổi cách tiếp cận này ở đây vì giả thuyết của chúng tôi không phải là mức VPIN cao dẫn đến mức sinh lời tuyệt đối cao trong một khoảng thời gian cụ thể. Thay vào đó, giả thuyết của chúng tôi là các mức VPIN cao kéo dài có tác động đến hành vi của các nhà tạo lập thị trường, điều này lại có tác động đến mức sinh lời tuyệt đối được đo bằng đồng hồ khối lượng. Để làm sáng tỏ giả thuyết này, chúng tôi sử dụng một khung phân tích không dựa trên mô hình, dựa trên xác suất có điều kiện. Sau đó, chúng tôi đặt ra hai câu hỏi cơ bản: (i) Khi VPIN cao, mức sinh lời tuyệt đối sẽ có hành vi như thế nào? (ii) Khi mức sinh lời tuyệt đối cao, mức VPIN trước đó là bao nhiêu?

## 2.5.2 Xác suất có điều kiện

Để hiển thị các xác suất có điều kiện này, chúng ta cần tính toán phân phối kết hợp của VPIN và mức sinh lời tuyệt đối. Chúng ta thực hiện điều này bằng cách trước tiên nhóm các giá trị VPIN thành 5 phân vị và mức sinh lời tuyệt đối thành các khoảng có kích thước 0.25% để có thể hiển thị các phân phối rời rạc. Sau đó, chúng ta tính toán phân phối kết hợp của  $\left(VPIN_{\tau-1}, \left|\frac{P_{\tau}}{P_{\tau-1}} - 1\right|\right)$ . Từ phân phối kết hợp này, chúng ta suy ra hai phân phối xác suất có điều kiện.

Trước tiên, chúng ta xem xét phân phối của mức sinh lời tuyệt đối trong khoảng khối lượng tiếp theo với điều kiện VPIN nằm trong mỗi nhóm phần trăm vị của chúng ta. Điều này dẫn đến hai mươi phân phối có điều kiện, mỗi phân phối tương ứng với một nhóm VPIN, được hiển thị trong Bảng 4a. Mỗi hàng của bảng này đại diện cho một phân

phối có điều kiện của mức sinh lời tuyệt đối, được điều kiện hóa theo mức VPIN trước đó.

Có ba kết quả quan trọng cần lưu ý. Thứ nhất, khi VPIN thấp, mức sinh lời tuyệt đối tiếp theo cũng thấp. Cụ thể, khi VPIN nằm trong nhóm phần tư thấp nhất của nó, mức sinh lời tuyệt đối tiếp theo nằm trong khoảng 0% đến 0.25% hơn 90% thời gian. Thứ hai, khi VPIN cao, phân phối có điều kiện của mức sinh lời tiếp theo phân tán hơn nhiều. Cụ thể, mức sinh lời tuyệt đối cao (trên 1.5%) đôi khi xảy ra trong khoảng khối lượng tiếp theo, trong khi chúng không bao giờ xảy ra sau khi VPIN rất thấp. Thứ ba, ngay cả khi VPIN ở mức cao, mức sinh lời tuyệt đối trong khoảng khối lượng tiếp theo phần lớn không cao. Chúng tôi sẽ làm rõ hơn về điểm này trong phần tiếp theo, nơi chúng tôi lập luận rằng cần có mức VPIN cao kéo dài để có thể tạo ra mức sinh lời tuyệt đối lớn một cách đáng tin cậy.

Tiếp theo, trong Bảng 4b, chúng ta xem xét phân phối của VPIN trong khoảng  $\tau - 1$  với điều kiện là mức sinh lời tuyệt đối giữa các khoảng  $\tau - 1$  và  $\tau$ . Mỗi cột của bảng này cung cấp phân phối của các giá trị VPIN trước đó với điều kiện là mức sinh lời tuyệt đối trong mỗi khoảng có kích thước 0.25%. Kết quả quan trọng ở đây là khi mức sinh lời tuyệt đối lớn, VPIN ngay trước đó hiếm khi thấp. Cụ thể, phần tư trên của phân phối này chứa hơn 84% của tất cả các mức sinh lời tuyệt đối lớn hơn 0.75%. Thực tế này gợi ý rằng VPIN có một số giá trị bảo hiểm chống lại sự biến động giá cực đoan.

Để tính toán các xác suất có điều kiện này, chúng ta đã sử dụng kết hợp tham số chuẩn (50,250). Sự lựa chọn tham số này tối đa hóa hệ số tương quan giữa VPIN và mức sinh lời tuyệt đối, nhưng không nhất thiết phải tối đa hóa một "kịch bản rủi ro" cụ thể nào đó — chẳng hạn như:

$$\text{Prob} \left( \text{CDF}(VPIN_{\tau-1}) > \frac{3}{4} \left| \frac{P_{\tau}}{P_{\tau-1}} - 1 \right| > 0.75\% \right).$$

Hình tiếp theo cho thấy ảnh hưởng của sự lựa chọn tham số đối với giá trị bảo hiểm của VPIN chống lại các mức sinh lời tuyệt đối lớn hơn 0.75%.

Hình 12 cho thấy rằng miễn là số lượng khoảng mỗi ngày không quá lớn và độ dài mẫu không quá nhỏ, xác suất rằng VPIN nằm trong phần tư trên một khoảng khối lượng trước khi có mức sinh lời tuyệt đối từ 0.75% trở lên sẽ nằm trong khoảng từ 80% đến 90%. Kết quả này không chỉ có nghĩa là VPIN dự báo một tỷ lệ lớn các sự kiện biến động cực đoan, mà còn cho thấy rằng sự biến động do độc tính gây ra dường như là một nguồn đáng kể của sự biến động chung.

### 2.5.3 Liệu sự biến động cực đoan luôn đi trước bởi VPIN cao?

Bảng 4a cho thấy rằng đối với hợp đồng tương lai E-mini S&P 500, các mức sinh lời tuyệt đối lớn, chẳng hạn như trên 2%, rất khó xảy ra nếu VPIN trước đó thấp. Tuy nhiên, điều này không phải là không thể, và việc tìm kiếm trên các hợp đồng khác chắc chắn sẽ cung cấp các ví dụ trong đó chúng xảy ra. Một ví dụ gần đây về sự biến động giá cực đoan trong hợp đồng tương lai khí tự nhiên minh họa rõ khả năng này. Theo Financial Times, vào ngày 8 tháng 6 năm 2011: “Thị trường giao dịch của Sở Giao dịch Hàng hóa New York đã đóng cửa hơn năm giờ khi vào cuối ngày thứ Tư, giá khí tự nhiên kỳ hạn tháng 7 trên Nymex giảm 39 cent, tương đương 8.1%, xuống còn 4.510 đô la mỗi triệu đơn vị nhiệt Anh (British thermal unit). Chỉ sau vài giây, giá đã bật trở lại.”

Một lời giải thích đã được đưa ra trên Financial Times vào ngày 9 tháng 6 năm 2011: “Một số nhà quan sát thị trường cho rằng sự sụt giảm này là do lỗi ‘ngón tay béo’, khi



một nhà giao dịch vô tình gõ thêm một số 0 vào cuối lệnh, tăng kích thước của nó lên gấp 10 lần. Những người khác đổ lỗi cho sự cố trong các thuật toán giao dịch hợp đồng tương lai. Khối lượng giao dịch thấp, có nghĩa là bất kỳ lệnh lớn nào cũng sẽ có tác động đáng kể và có thể kích hoạt bán tự động.”

Nếu lời giải thích của Financial Times là đúng, thì VPIN lẽ ra không nên cao trước khi giá giảm. Hình 13 cho thấy điều này đã xảy ra. Đã có một sự sụt giảm đột ngột về giá theo sau là sự phục hồi ngay lập tức, tất cả đều diễn ra ở mức độ tính tương đối thấp. Ví dụ này minh họa hai điểm hữu ích: Không phải tất cả sự biến động đều do độc tính gây ra, và việc biết được liệu một đợt giảm giá là do độc tính hay do các nguyên nhân tiềm năng khác có thể sẽ hữu ích cho các cơ quan quản lý.

#### 2.5.4 Liệu sự biến động cực đoan luôn xảy ra khi VPIN cao?

Chúng ta biết từ Bảng 4b rằng các mức sinh lời tuyệt đối lớn không nhất thiết xảy ra trong khoảng khối lượng tiếp theo khi VPIN cao. Thực tế, phần lớn các mức sinh lời tuyệt đối ngay sau khi quan sát thấy VPIN cao không phải là lớn. Tuy nhiên, quan sát này phù hợp với giả thuyết của chúng ta rằng các mức VPIN cao kéo dài sẽ dẫn đến sự biến động.

Để kiểm chứng giả thuyết của chúng ta, chúng ta cần định lượng mức độ biến động tối đa mà một nhà tạo lập thị trường phải đối mặt khi VPIN đạt đến một mức tới hạn nào đó và duy trì ở mức đó hoặc cao hơn. Sự cố "flash crash" là một minh họa rõ ràng cho lý do tại sao sự khác biệt giữa biến động ngay lập tức và biến động tiếp theo lại quan trọng. Trong sự cố "flash crash", VPIN vẫn ở mức cao từ trước khi sự cố bắt đầu cho đến khi nó kết thúc và giá bắt đầu phục hồi một phần. Nếu chúng ta tập trung vào khoảng thời gian bắt đầu khi VPIN đạt đến một mức cao và kết thúc khi nó giảm xuống, thì chỉ có một sự sụt giảm giá nhỏ. Thực tế này sẽ không mang lại nhiều sự an ủi cho nhiều nhà tạo lập thị trường đã rời khỏi thị trường trong thời gian xảy ra sự cố. Họ đã bị ảnh hưởng bởi sự biến động trung gian lớn xảy ra khi VPIN ở mức cao.

Để kiểm tra mức độ biến động trung gian tối đa mà một nhà tạo lập thị trường phải đối mặt khi VPIN cao, chúng ta tính toán các sự kiện biến động khi VPIN vẫn nằm trong bất kỳ nhóm phần trăm vị nào. Cụ thể, mỗi khi VPIN chuyển từ một nhóm phần trăm vị này sang nhóm khác, chúng ta tính toán mức sinh lời tuyệt đối lớn nhất xảy ra giữa bất kỳ hai khoảng trung gian nào (không nhất thiết phải liên tiếp), cho đến khi VPIN chuyển sang nhóm phần trăm vị khác.

Chẳng hạn, giả sử chỉ số VPIN chuyển vào nhóm phần trăm vị thứ tám mươi lăm, và bốn khoảng khối lượng sau đó, nó chuyển sang nhóm phần trăm vị thứ chín mươi. Chúng ta sẽ tính mức sinh lời tuyệt đối giữa các khoảng 1 và 2, 1 và 3, và 1 và 4, cũng như giữa 2 và 3, 2 và 4, và 3 và 4. Chúng ta quan tâm đến mức sinh lời tuyệt đối lớn nhất trong số này, điều mà chúng ta xem như mức độ biến động tối đa mà nhà tạo lập thị trường phải đối mặt. Chúng ta thực hiện điều này cho mọi lần VPIN vượt qua như vậy.

Để chính xác hơn, chúng ta ký hiệu  $i$  là một chỉ số được tăng lên một mỗi khi VPIN chuyển từ một nhóm phần trăm vị này sang nhóm khác, và  $\tau(i)$  là khoảng tương ứng với lần chuyển thứ  $i$ . Điều này có nghĩa là VPIN vẫn ở trong cùng một nhóm phần trăm vị trong  $\tau(i+1) - \tau(i)$  khoảng; trong ví dụ trên của chúng ta, đây là khoảng bốn khối lượng trong đó VPIN vẫn ở nhóm phần trăm vị thứ tám mươi lăm.

Mức sinh lời tuyệt đối lớn nhất giữa bất kỳ hai khoảng trung gian nào khi VPIN vẫn nằm trong một nhóm phần trăm vị cụ thể (tức là, từ khoảng mà nó thực hiện lần chuyển thứ  $i$  cho đến khi nó thực hiện lần chuyển thứ  $i+1$ ) được xác định bởi:

$$\max_{\substack{\tau(i) \leq m < l \\ \tau(i) < l \leq \tau(i+1)}} \left| \frac{P_l}{P_m} - 1 \right|.$$

Trong ví dụ trên của chúng ta, đây là mức sinh lời tuyệt đối lớn nhất trong số sáu mức sinh lời trung gian trong khoảng thời gian mà VPIN vẫn ở nhóm phần trăm vị thứ tám mươi lăm.

Phân tích này phong phú hơn so với các xác suất có điều kiện được minh họa trong Bảng 4a ở hai khía cạnh. Thứ nhất, chúng ta không chỉ xem xét mức sinh lời tuyệt đối xảy ra ngay sau khi VPIN vượt qua một nhóm, mà còn xem xét mức sinh lời tuyệt đối lớn nhất xảy ra khi VPIN vẫn ở trong một nhóm. Điều này phù hợp với lý thuyết vì cấu trúc của chúng ta rằng sự biến động xuất hiện khi độc tính đã đạt đến một điểm bão hòa vượt quá mức chịu đựng của các nhà tạo lập thị trường.

Thứ hai, nó nắm bắt sự biến động giá qua tất cả các chuỗi khoảng trung gian, do đó bao gồm cả tác động của sự trôi giá cũng như sự phục hồi giá. Điều này quan trọng vì sự trôi giá chậm và sự phục hồi giá đều có thể che giấu sự biến động cực đoan.

Tuy nhiên, phân tích này cũng có hạn chế ở chỗ nó không nắm bắt được sự gia tăng độc tính kéo dài qua nhiều nhóm phần trăm vị của VPIN. Cụ thể, chúng ta không nắm bắt được điều gì xảy ra khi độc tính tăng từ nhóm phần trăm vị thứ bảy mươi lăm lên tám mươi, sau đó lên tám mươi lăm, chín mươi, v.v. Do đó, ngưỡng mà chúng ta đặt ra ở đây chắc chắn sẽ đánh giá thấp tác động của sự biến động do độc tính gây ra.

Hình 14 biểu diễn xác suất rằng mức sinh lời tuyệt đối lớn nhất vượt quá 0.75% khi VPIN vẫn nằm trong bất kỳ nhóm phần trăm vị nào thuộc phần tư trên của phân phối của nó (tức là, VPIN nằm trong các nhóm 0.75-0.80, 0.80-0.85, 0.85-0.90, 0.90-0.95 hoặc 0.95-1.00) với các kết hợp khác nhau của số lượng khoảng mỗi ngày và độ dài mẫu.

Đối với kết hợp chuẩn của chúng ta (50, 250), chúng ta thấy rằng 51.84% số lần VPIN đi vào một nhóm phần trăm vị trong phần tư trên, sẽ có ít nhất một mức sinh lời trung gian vượt quá 0.75% trước khi VPIN rời khỏi nhóm đó.

Kết hợp tham số tối đa hóa sức mạnh dự báo của VPIN là {10, 350} — nghĩa là, mười khoảng mỗi ngày với độ dài mẫu là 350 (khoảng 1.6 tháng). Đối với kết hợp tham số này, Bảng 5 cho thấy rằng 78.57% số lần VPIN đi vào một nhóm phần trăm vị trong phần tư trên của phân phối của nó, sự biến động do độc tính gây ra thường sẽ đáng kể (mức sinh lời tuyệt đối vượt quá 0.75%) trước khi VPIN rời khỏi nhóm đó.

## 2.6 Kết luận

Bài viết này trình bày một quy trình mới để ước tính xác suất giao dịch có thông tin được đồng bộ hóa theo khối lượng, hay chỉ số độc tính dòng VPIN. Một điểm quan trọng một lợi thế của chỉ số độc tính VPIN và quy trình ước tính liên quan là nó được cập nhật trong ngày với tần suất phù hợp với khối lượng để bắt kịp tốc độ của sự xuất hiện thông tin. Ngoài ra, đây còn là một quy trình ước tính phân tích trực tiếp, không dựa vào việc ước tính trung gian các tham số không thể quan sát hoặc các phương pháp số. Chúng tôi đã chỉ ra rằng chỉ số VPIN có sức mạnh dự báo đáng kể đối với sự biến động do độc tính gây ra, và rằng nó mang lại giá trị bảo hiểm chống lại các mức sinh lời tuyệt đối cao trong tương lai. Chính thuộc tính sau cùng này khiến chúng tôi tin rằng VPIN là một công cụ quản lý rủi ro cho thế giới giao dịch tần suất cao mới. Việc cung cấp thanh khoản hiện nay là một quá trình phức tạp, và các mức độc tính ảnh hưởng đến cả quy mô và phạm vi hoạt động của các nhà tạo lập thị trường. Các mức VPIN cao biểu thị

một rủi ro cao về các biến động giá lớn tiếp theo, xuất phát từ tác động của độc tính lên việc cung cấp thanh khoản.

Rủi ro dựa trên thanh khoản này rất quan trọng đối với các nhà tạo lập thị trường, những người trực tiếp chịu ảnh hưởng của độc tính, nhưng nó cũng có ý nghĩa đối với các nhà giao dịch phải đối mặt với khả năng các biến động giá lớn do độc tính gây ra. Việc phát triển các thuật toán có thể thay đổi mô hình thực thi lệnh dựa trên mức độ độc tính sẽ cho phép các nhà giao dịch giảm thiểu rủi ro này.

Các sàn giao dịch cũng có thể áp dụng VPIN để phân bổ tài nguyên máy theo cách tăng tốc giao dịch ở phía có thanh khoản lớn hơn trong khi làm chậm giao dịch ở phía bị tấn công (một dạng ngắt mạch động), từ đó cho phép các nhà tạo lập thị trường tiếp tục hoạt động. Chúng tôi tin rằng đây là một lĩnh vực quan trọng cần được nghiên cứu thêm trong tương lai.