

The Two Margins of Market Quality: Participant Composition, Price-Setting Composition, and Observable Identity

Boyi Shen*

January 18, 2026

Abstract

Market quality depends on two compositional margins invisible in standard data: *who trades* (participant composition) and *who sets prices* (price-setting composition). Using on-chain data where counterparty identity is observable, I show both margins are measurable and both respond to infrastructure shocks in ways that aggregate measures cannot detect.

First, I establish that participant composition matters: makers face 3.1 basis points higher adverse selection per trade against informed than uninformed takers ($t = 5.8$), commensurate with observed spreads (3–4 bps)—the first direct, identity-linked measurement of the adverse selection differential that Glosten and Milgrom (1985) imply. Classification uses mid-price moves (directional prediction), eliminating semantic ambiguity. The effect is robust under taker-level inference (19.2 bps, $t = 24.4$), ruling out that results are driven by outlier traders. Second, I show that infrastructure barriers filter both margins. Using four natural experiments (API congestion and outage events spanning January–July 2025), I find that all four events produce spread widening (mean +1.66 bps; asset-clustered $t = 2.25$, $p < 0.05$ for the main event). Taker composition shifted toward informed traders in every event; this selection channel predicts spread widening beyond quote staleness ($\beta = 0.07$, $t = 2.51$). Simultaneously, the events revealed that price-setting composition drives fragility: maker participation rose 84% yet spreads nearly doubled, because the few makers who set prices at the top of book could not operate. Fill-based concentration fell while quality-relevant concentration collapsed—a divergence only observable with identity data.

*London Business School. Email: bshen@london.edu. I thank my supervisor, Prof. Lakshmanan Shivakumar, for invaluable guidance and support.

The unifying insight is that infrastructure access correlates with trader type on both sides of the market, generating compositional shifts that standard metrics miss entirely.

JEL Classification: G10, G14, G23

Keywords: market quality, compositional margins, adverse selection, market fragility, on-chain markets

1 Introduction

Market quality depends on two compositional margins: *who trades* and *who sets prices*. Standard market data obscures both. We observe aggregate spreads, volumes, and concentration measures, but not the composition of order flow or the identity of price-setters. This paper uses on-chain data—where both sides of every trade are recorded—to show that these compositional margins are (1) directly measurable, (2) economically significant, and (3) respond to infrastructure shocks in ways that aggregate metrics miss entirely.

The unifying insight is simple: infrastructure access correlates with trader type. On the demand side, informed traders invest more in operational capacity because timely execution is valuable when trading on information. On the supply side, the makers who set prices at tight spreads depend on continuous API access to manage inventory risk. When infrastructure degrades, both margins shift: the taker mix becomes more informed, and the effective price-setters become less capable. Both effects widen spreads—but standard analysis, lacking identity data, attributes everything to quote staleness.

The decomposition. I formalize market quality as depending on two compositional margins:

- *Participant composition* (demand side): Which types of takers are in the order flow? Informed takers impose adverse selection costs; uninformed takers provide “good” flow that makers profit from.
- *Price-setting composition* (supply side): Which makers quote at the top of book? Fast repricers maintain tight spreads; when they cannot operate, marginal makers enter but cannot substitute for price-setting capacity.

Both margins are observable when counterparty identity is observable. Both shift under infrastructure stress. And both explain variation in market quality that aggregate measures cannot capture.

Validating the decomposition: the toxicity differential. The first empirical test establishes that participant composition matters. Linking 3.9 million maker-taker pairs on Hyperliquid, a major derivatives exchange, I classify takers by their *mid-price prediction* (directional information, excluding spread costs) using strictly out-of-sample methods. Makers face 3.1 basis points higher adverse selection per trade against informed than uninformed counterparties ($t = 5.8$)—commensurate with observed spreads (3–4 bps) and consistent with Glosten and Milgrom [1985]’s prediction. Taker-level inference (19.2 bps, $t = 24.4$) confirms the result is not driven by outlier traders. Identity determines adverse selection costs.

Testing the demand-side margin: taker selection during stress. I exploit four infrastructure events—two API congestion episodes on January 20, 2025, an API outage on July 29, 2025, and a detected stress event on July 30, 2025. All four events show spread widening, with a mean effect of +1.66 bps (asset-clustered $t = 2.25$, $p < 0.05$ for the main event). But participant composition also shifted: informed-to-uninformed ratios rose 3.4–5.1% across events as sophisticated traders disproportionately maintained access. This selection channel predicts spread widening *beyond* quote staleness ($\beta = 0.07$, $t = 2.51$)—the demand-side margin explains residual variation.

Testing the supply-side margin: price-setting fragility during stress. The same outage reveals the supply-side margin. Maker participation rose 84% and fill-based HHI *fell*—yet spreads nearly doubled. This apparent paradox is resolved by distinguishing fill-based concentration from price-setting concentration. The top-decile makers by price-setting capacity (MPSC) saw their fill rates collapse 24%; other makers expanded 55%. But these marginal makers could not substitute for price-setters. Quality depends on *which* makers operate, not how many.

1.1 Contribution

This paper’s contribution is a *decomposition*: market quality separates into participant composition and price-setting composition, both measurable when identity is observable. The measurement innovation (direct counterparty-level adverse selection) enables this decomposition; the two applications test each margin under infrastructure stress.

The Two-Margin Framework. I show that market quality is a function of two compositional margins that standard data cannot observe:

1. *Participant composition* determines adverse selection. The 3.1 bps trade-weighted toxicity differential ($t = 5.8$) establishes that counterparty type predicts adverse selection costs—the first direct, identity-linked measurement of the differential Glosten and Milgrom [1985] imply. Taker-level inference (19.2 bps, $t = 24.4$) confirms robustness.
2. *Price-setting composition* determines fragility. Fill-based concentration (HHI) fell during the outage while spreads doubled, because what matters is who quotes at the top of book, not who executes volume.

Traditional studies measure adverse selection indirectly [Hasbrouck, 1991, Huang and Stoll, 1997, Easley et al., 1996] and concentration via executed volume. Both approaches miss compositional variation that identity data reveals.

Testing the Demand-Side Margin. Four infrastructure events shifted participant composition: informed-to-uninformed ratios rose 3.4–5.1% across events. This selection channel predicts spread widening beyond quote staleness ($\beta = 0.07$, $t = 2.51$). *Without identity data:* Traditional analysis would attribute all spread widening to staleness; the compositional channel would be invisible.

Testing the Supply-Side Margin. The same outage reveals that fill-based concentration diverges from quality-relevant concentration. Maker participation rose 84% and HHI fell, yet spreads doubled. High-MPSC makers (top decile by price-setting capacity) saw fill rates collapse 24% while other makers expanded 55%. The marginal makers could not substitute. *Without identity data:* Traditional analysis would see falling HHI and rising participation alongside worsening quality—a paradox with no resolution.

1.2 Related Literature

Participant Composition and Adverse Selection. The first margin—participant composition—relates to the adverse selection literature. Glosten and Milgrom [1985] and Kyle [1985] theorize that spreads compensate for trading against informed counterparties; Hasbrouck [1991], Huang and Stoll [1997], and the PIN model [Easley et al., 1996] develop methods to infer adverse selection from aggregate patterns. I measure it directly by linking counterparties, providing the first identity-linked evidence that participant composition determines adverse selection costs.

Price-Setting Composition and Market Fragility. The second margin—price-setting composition—relates to the market maker concentration literature. Menkveld [2013] documents that a single HFT provided 40% of Dutch stock liquidity; Aquilina et al. [2024] find extreme concentration in DeFi AMMs. Whether concentration creates fragility depends on *which* makers are concentrated. I show that fill-based concentration diverges from price-setting concentration, resolving the paradox of falling HHI with worsening quality.

Infrastructure and Compositional Shifts. Studies of technology failures—the Flash Crash [Kirilenko et al., 2017], Knight Capital, exchange outages—document that disruptions harm quality. Hendershott et al. [2011] and Brogaard et al. [2014] examine how algorithmic trading affects liquidity. These studies measure *aggregate* effects; whether outages shift *composition* on either margin has been unobservable. I show that infrastructure barriers filter both takers (participant composition) and makers (price-setting composition).

DeFi Markets and Observable Identity. An emerging literature studies decentralized finance [Lehar and Parlour, 2024, Barbon and Ranaldo, 2024, Capponi and Jia, 2024, Makarov and Schoar, 2020]. Most work focuses on automated market makers. On-chain

CLOBs enable comparison with traditional microstructure, with the advantage of counterparty transparency—the data requirement for measuring both compositional margins.

1.3 Traditional Market Analogues

Readers may ask whether findings from an on-chain venue generalize to traditional markets. Three observations suggest they do.

The phenomena are familiar. The mechanisms I document—infrastructure barriers that filter participants by sophistication, maker concentration among fastest repricers, and quality deterioration when key providers cannot operate—have direct analogues in traditional markets. Direct market access and co-location create tiered access to exchange infrastructure; Shkilko and Sokolov [2020] document that even milliseconds of latency disadvantage predict worse execution. Message throttling and order rate limits constrain who can reprice rapidly during volatility. Exchange outages (NYSE 2015, Tokyo 2020, Nasdaq 2013) produce spread widening whose compositional component—whether *who* trades also shifts—has been unmeasurable.

The market structure is comparable. Hyperliquid operates a central limit order book with continuous matching, maker-taker fee schedules, and millisecond-level latency. The dominant makers exhibit classic HFT signatures: aggressive top-of-book quoting, high-frequency repricing, and API-dependent operation. These patterns parallel Menkveld [2013]’s documentation of HFT market making in Dutch equities, suggesting the economic forces are similar.

The setting is a clean laboratory. What distinguishes on-chain data is not different economics but better measurement. Traditional markets lack observable counterparty identities; researchers infer information from price impacts or order flow. On-chain markets record both sides of every trade with complete order histories. The contribution is methodological: demonstrating what becomes testable when counterparties are observable, with findings that speak to market structure generally.

1.4 Theoretical Framework: Why Both Margins Shift Under Stress

The two-margin framework generates a unified prediction: infrastructure stress shifts *both* participant composition and price-setting composition, because operational capacity correlates with trader type on both sides of the market. I formalize this intuition.

Setup. Traders differ in *operational capacity* $c_i \in [0, 1]$, reflecting infrastructure investment: API redundancy, co-location, monitoring systems. During normal operation, all

traders with $c_i > 0$ can participate. During infrastructure stress, only traders with $c_i > \theta$ maintain access.

Key assumption: Operational capacity correlates with trader type.

- *Participant composition margin:* Informed takers invest more in infrastructure because timely execution is valuable when trading on information. Let $\mathbb{E}[c_i|\text{informed}] > \mathbb{E}[c_i|\text{uninformed}]$.
- *Price-setting composition margin:* High-MPSC makers depend on continuous API access to manage inventory risk. Their business model *requires* high c_i ; paradoxically, their operations are most disrupted when infrastructure degrades because they cannot stale-quote like slower makers.

Predictions for both margins. During infrastructure stress:

1. *Participant composition shifts toward informed.* Uninformed takers with $c_i < \theta$ exit; informed takers remain. The informed-to-uninformed ratio rises, increasing adverse selection costs.
2. *Price-setting composition degrades.* High-MPSC makers lose repricing capacity; marginal makers enter but cannot substitute for price-setting liquidity. Fill-based concentration may fall while quality-relevant concentration collapses.
3. *Both margins widen spreads.* The participant composition margin operates through worse counterparty mix; the price-setting margin operates through loss of price-setting capacity. Standard analysis attributes everything to quote staleness.

Testable implications.

- *Participant composition:* Informed takers more likely to trade during stress (directionally confirmed: DiD = +2.8 pp); selection channel explains spread variance beyond staleness ($\beta = 0.07$, $t = 2.51$)
- *Price-setting composition:* High-MPSC makers show larger repricing declines (confirmed: 24% fill rate collapse vs. 55% expansion for other makers); fill-based HHI diverges from quality outcomes

The unifying insight is that infrastructure creates a *double selection mechanism*: it filters participants by operational capacity on both sides of the market. Co-location, direct market access, and message rate limits create similar selection in traditional markets—but on-chain data makes both margins directly observable.

2 Institutional Setting and Data

2.1 Hyperliquid

Hyperliquid is a decentralized exchange operating on its own Layer 1 blockchain, specializing in perpetual futures. Unlike automated market makers (AMMs), Hyperliquid uses a central limit order book architecture directly comparable to traditional futures exchanges.

Economic significance. Hyperliquid is not a minor venue. As of Q3 2025, it ranks among the top 3 perpetual futures exchanges globally by volume, processing \$3–5 billion in daily notional volume across 100+ trading pairs. During the sample period (July–September 2025), cumulative trading volume exceeded \$400 billion. Open interest regularly exceeds \$3 billion. For context, this places Hyperliquid’s perpetuals volume on par with major centralized venues like OKX and Bybit, and exceeds many traditional futures exchanges. The patterns documented here—counterparty-level adverse selection, infrastructure-driven selection effects, and maker concentration—operate at economically meaningful scale.

Key features:

- **Complete Transparency:** Every order submission, modification, and cancellation is recorded on-chain and publicly visible in real-time
- **Single Venue:** No dark pools, hidden orders, or fragmentation across venues
- **Open Liquidity Provision:** No designated market makers; any participant can provide liquidity
- **High Performance:** Sub-second finality enables institutional-grade trading

This setting enables studying market quality under complete counterparty transparency.

2.2 Data

This paper draws on three datasets from Hyperliquid’s public blockchain records. Table 1 provides a complete data map.

Table 1: Data Map: Datasets and Their Uses

| Dataset | Period | Assets | Sample Size | Analyses |
|---|----------------------------------|----------------------------------|---|--|
| <i>Panel A: Counterparty Pairs (Main Result)</i> | | | | |
| Matched Trades | Jul 28–30, 2025 (3 days) | 4 major (BTC, ETH, SOL, HYPE) | 3.9M pairs 3,686 takers (Q5 + Q1) | Adverse (Sec. 3) Selection Toxicity Differential |
| <i>Panel B: Trade-Level Data with Wallet Identities</i> | | | | |
| Outage Window | Jul 28–30, 2025 (3 days) | 24 | 10.0M fills 41,239 takers 14,650 makers | Selection (Sec. 4) Effects Composition Shifts |
| Extended Sample | Jul 27–Sep 30, 2025 (66 days) | 24 | 182M fills 66,728 makers | Concentration (Sec. 5) Classification Stability |
| <i>Panel C: Order Book Data</i> | | | | |
| L2 Snapshots | Jul 28–30, 2025 (3 days) | 24 | 518,400 snapshots (5-sec intervals) | Spread, Quote Updates Event Study |
| <i>Panel D: Multi-Event Validation</i> | | | | |
| L2 Snapshots | Jan 20, 2025 (2 hours) | 10 (major assets) | 14,400 snapshots (5-sec intervals) | Multi-Event ness (Sec. 4.2) Robust- |

Consistency checks. Maker concentration patterns in the 66-day sample match the 3-day outage window (HHI correlation $r = 0.94$). Informed/uninformed classifications from July 28 predict profitability on July 29–30 with high significance (see Section 3.2), confirming out-of-sample validity.

Spread measure. Throughout this paper, “spread” refers to the *quoted top-of-book spread*: the difference between the best ask and best bid prices, measured in basis points of the mid-price, computed from L2 order book snapshots at 5-second intervals. The headline spread effect (2.60–2.78 bps) comes from specifications with asset, date, and hour fixed effects that isolate the within-asset treatment effect. Section 6.5 reports conservative inference tests using alternative specifications and clustering schemes.

3 Validating the Decomposition: The Toxicity Differential

Before testing how each compositional margin responds to infrastructure stress, I first establish that participant composition *matters*—that counterparty identity determines adverse selection costs. This validates the decomposition: if identity did not predict adverse selection, the participant composition margin would be economically irrelevant.

The test is direct. Glosten and Milgrom [1985] predict that informed counterparties impose higher adverse selection costs on market makers. On-chain data records both sides of every trade, enabling a counterparty-level test: do makers face more adverse selection against informed than uninformed takers? I classify takers by *mid-price prediction*—directional information excluding spread costs—to eliminate semantic ambiguity (“informed” takers unambiguously predict price direction). The answer—3.1 basis points per trade ($t = 5.8$), commensurate with observed spreads—establishes that participant composition is a first-order determinant of market quality. Taker-level inference confirms robustness (19.2 bps, $t = 24.4$).

3.1 Methodology: Linking Maker-Taker Pairs

Hyperliquid’s data records both counterparties for each trade. I match the two sides using trade identifiers (timestamp, coin, price, size) to create 3.9 million maker-taker pairs across BTC, ETH, SOL, and HYPE.

Matching quality. Of 4.1 million raw trade records, 3.9 million (95.2%) match to unique counterparty pairs. The remaining 4.8% involve collisions—multiple trades with identical (timestamp, coin, price, size) keys within the same millisecond. As a robustness check, I restrict to trades with unambiguous unique matches (excluding all colliding keys); the toxicity differential estimate changes by less than 0.1 bps (Table 27).

Defining “profit” without funding confounds. A critical methodological concern in perpetual futures is distinguishing price movements from funding payments. I define “profit” as the *price markout*—the price movement at a specified horizon, not realized PnL:

$$\text{Markout}_h = \text{Direction} \times \frac{P_{t+h} - P_t}{P_t} \times 10000 \quad (\text{bps}) \quad (1)$$

where $\text{Direction} = +1$ for buys and -1 for sells, and P_t is the trade execution price. This measures information content (did the trade predict price movement?) while avoiding funding payment confounds. The primary horizon is $h = 1$ minute, with 10-second and 5-minute

horizons for robustness.

Measurement Box: Spread Decomposition and Adverse Selection

Standard spread decomposition. Let M_t denote the mid-price at time t , P_t the trade execution price, and d the trade direction sign (+1 buy, -1 sell). Following Glosten and Milgrom [1985] and Huang and Stoll [1997], the effective spread decomposes as:

$$\begin{aligned}\text{Effective Spread (ES)} &= 2|P_t - M_t| \\ \text{Realized Spread (RS)} &= 2d(P_t - M_{t+\Delta}) \\ \text{Adverse Selection (AS)} &= 2d(M_{t+\Delta} - M_t) = \text{ES} - \text{RS}\end{aligned}\tag{2}$$

All measures are in basis points from the maker’s perspective: positive AS indicates the mid moved against the maker (toxic flow); negative AS indicates uninformed flow. The adverse selection component measures how much of the spread compensates for trading against informed counterparties.

Mid-move vs. markout. The taker’s markout (measured from execution price) conflates information and transaction cost:

$$\text{Taker Markout} = \underbrace{d(M_{t+h} - M_t)}_{\text{Mid-move (information)}} - \underbrace{\frac{1}{2}S_t}_{\text{Half-spread (cost)}}$$

The *mid-move* isolates directional price prediction—precisely the adverse selection component. A taker with positive mid-move but negative markout still imposes adverse selection on the maker; they just don’t profit net of spread. The classification criterion (Section ??) uses mid-move directly to isolate information from transaction costs.

Out-of-sample classification. To avoid mechanical sorting concerns (classifying and measuring on the same trades), I use a strictly out-of-sample design:

- **Training period:** July 28, 2025—classify takers by their 1-minute markout profits
- **Test period:** July 29–30, 2025—measure counterparty-specific adverse selection
- **Labels frozen:** Taker quintiles fixed using only training data before examining test period

Takers are classified into quintiles based on mean markout during training. The top 20% (Q5) are labeled “informed”—2,698 wallets whose trades predicted favorable mid-price move-

ments, imposing adverse selection on makers. The bottom 20% (Q1) are “uninformed”—2,698 wallets whose trades predicted adverse mid-price movements. Note that “informed” refers to directional price prediction ability, not necessarily net profitability after spread costs.

3.2 The Information Food Chain

Table 2 presents the “information food chain”—taker markout profit by taker type and maker type, measured *out-of-sample*. Each cell shows the average taker 1-minute markout (in bps) for trades between that taker quintile and maker quintile.

Table 2: The Information Food Chain: Taker Markout (bps) by Counterparty Type

| Taker Type | Maker Type (by frequency) | | | | |
|----------------------------|---------------------------|------|------|------|----------|
| | Q1 (Slow) | Q2 | Q3 | Q4 | Q5 (HFT) |
| Q5 (Informed) | −0.2 | −0.5 | −0.4 | −0.7 | −0.3 |
| Q4 | +0.5 | +0.0 | +1.0 | +0.6 | +0.7 |
| Q3 | +1.1 | +0.1 | +1.4 | +1.4 | +0.4 |
| Q2 | +0.8 | +0.1 | +2.6 | −0.0 | −0.4 |
| Q1 (Uninformed) | −14.7 | −8.7 | −6.4 | −8.7 | −3.1 |
| <i>Winner–Loser Spread</i> | 14.5 | 8.2 | 6.0 | 8.0 | 2.8 |

3.9M paired trades (out-of-sample). Markout horizon: 1 minute.

Classification: July 28; Test: July 29–30. Labels frozen before test.

Result: The food chain structure is visible out-of-sample: uninformed takers (Q1) consistently lose against all maker types (−3 to −15 bps), while informed takers (Q5) are approximately breakeven. The “winner–loser spread” varies by maker type (2.8–14.5 bps), suggesting maker heterogeneity in counterparty exposure. Critically, this pattern emerges using labels *frozen before the test period*—ruling out mechanical sorting.

Interpreting “informed” takers with negative markouts. A subtle point: Q5 takers have slightly negative markouts (−0.2 to −0.7 bps), not positive. Does this undermine their “informed” classification? No—the markout includes the spread cost takers pay. Decomposing: $\text{Markout} \approx \text{Mid Move} - \frac{1}{2}\text{Spread}$. With average spreads of 2–4 bps (half-spread of 1–2 bps), Q5’s mid move is *positive* (approximately +0.5 to +1.5 bps), while Q1’s mid move is strongly negative (−5 to −13 bps). The “informed” label refers to *directional price prediction*—the mid price moves in Q5’s favor—not net profitability after spread costs. This is precisely the adverse selection component that Glosten and Milgrom [1985] predict varies by counterparty type.

3.3 The Toxicity Differential

I aggregate to the maker level and compute the adverse selection each maker faces when trading against informed vs. uninformed counterparties, measured out-of-sample via price markouts.

Table 3: Counterparty-Specific Adverse Selection: Main Results

| | Estimate | <i>t</i> -stat | N |
|---|-----------|----------------|--------------|
| <i>Panel A: Toxicity Differential (Mid-Move Classification)</i> | | | |
| Taker-level inference | 19.18 bps | 24.38 | 4,053 takers |
| Trade-weighted inference | 3.05 bps | 5.84 | 693K trades |
| <i>Panel B: Robustness Across Horizons (Trade-Weighted)</i> | | | |
| 10-second horizon | 2.2 bps | 4.8 | 686K trades |
| 1-minute horizon | 2.9 bps | 5.3 | 686K trades |
| 5-minute horizon | 7.6 bps | 8.1 | 686K trades |
| Out-of-sample: classification July 28, test July 29–30. Panel A: taker-level clusters at taker; trade-weighted clusters at taker \times maker. Panel B: <i>t</i> -statistics clustered two-way by taker and maker. Raw trade-level <i>t</i> -statistics (Table 34) are > 40 . | | | |

The toxicity differential is positive and highly significant across all horizons. Makers face substantially more adverse selection when trading against informed (Q5) versus uninformed (Q1) counterparties. At the taker level—the econometrically correct unit of analysis since classification varies by taker—the toxicity differential is 19.2 bps ($t = 24.4$). Trade-weighted analysis yields 3.1 bps ($t = 5.8$), reflecting that high-volume takers drive the per-trade average.¹ Classification uses mid-price prediction (mid-move), so “informed” unambiguously means “predicts price direction.” Q5 takers have positive mean mid-moves out-of-sample (+19.6 bps), confirming they possess directional information. This validates the theoretical prediction that counterparty identity determines the adverse selection component of trades—measured with labels *frozen before the test period*.

Why the out-of-sample magnitude is conservative. The out-of-sample toxicity differential is smaller than in-sample estimates, reflecting design choices that prioritize validity: (1) price markouts isolate the adverse selection component by excluding funding payments; (2) out-of-sample classification means some classification “skill” degrades between training and test. The qualitative finding—counterparty identity determines adverse selection—is robust.

The toxicity differential is consistent with Glosten and Milgrom [1985]’s prediction that

¹The taker-level analysis treats each taker as one observation; the trade-weighted analysis weights by trading activity. Both are highly significant.

informed traders impose adverse selection costs on liquidity providers—providing the first direct, identity-linked measurement at the individual counterparty level. There is also substantial heterogeneity: slower makers (Q1) face larger toxicity differentials (14.5 bps) than HFT makers (Q5, 2.8 bps).

3.4 Interpretation: Information vs. Microstructure Mechanics

A concern with markout-based classification is that “profitable” trades may reflect microstructure mechanics (latency, trade size, temporary impact) rather than genuine private information. I conduct three pin-down tests.

Test 1: Size controls. If informed classification merely proxies for trade size, the toxicity differential should disappear within size bins. Table 5 Panel A shows the toxicity differential computed separately within trade-size terciles. The effect persists across all size bins (2.7–3.4 bps, all significant at $p < 0.01$), ruling out size as the sole driver.

Test 2: Permanent vs. transient impact. Genuine information should predict *persistent* price moves; mechanical impact should reverse. Panel B compares markouts at 1-minute vs. 30-minute horizons. Informed (Q5) trades predict price moves that persist: the 30-minute markout is $2.1\times$ the 1-minute markout. Uninformed (Q1) trades partially reverse ($0.7\times$ ratio). This asymmetry is consistent with Q5 possessing information that gets incorporated into prices, while Q1 reflects noise or temporary impact.

Test 3: Taker-level analysis. Panel C presents the toxicity differential using taker-level aggregation—the econometrically correct approach since classification varies at the taker level. I first compute each taker’s average maker-profit-when-trading-against-them, then compare Q1 vs Q5 means:

$$\overline{\text{MakerProfit}}_i = \beta \cdot \mathbf{1}[\text{Taker}_i \in Q1] + \varepsilon_i \quad (3)$$

where i indexes takers and inference is at the taker level. The coefficient $\beta = 19.18$ bps ($t = 24.38$, $p < 0.0001$) represents the toxicity differential with proper inference. Trade-weighted analysis yields $\beta = 3.05$ bps ($t = 5.84$), confirming significance under alternative weighting.

3.4.1 Estimand Reconciliation: Why 19 bps \neq Observed Spreads

A natural question is how the taker-level differential (19.2 bps) can exceed typical quoted spreads (3–4 bps). Table 4 reconciles the estimands.

Table 4: Estimand & Magnitude Map: Reconciling the Toxicity Differentials

| Estimand | | Estimate | <i>t</i> -stat | Economic Interpretation |
|---|--------|-----------|----------------|--|
| <i>Panel A: Alternative Weighting Schemes</i> | | | | |
| Per-taker, equal-weighted | | 19.18 bps | 24.38 | Each taker counts equally; small, infrequent takers with extreme outcomes receive same weight as large traders |
| Per-trade, trade-weighted | | 3.05 bps | 5.84 | Each trade counts equally; weights toward frequent traders who dominate market activity |
| <i>Panel B: Why They Differ</i> | | | | |
| Q1 takers: mean trades | | 239 | | Q1 (uninformed) takers trade more frequently |
| Q5 takers: mean trades | | 136 | | Q5 (informed) takers trade less frequently |
| Frequency (Q1/Q5) | ratio | 1.76× | | Trade-weighting upweights Q1's smaller maker losses |
| <i>Panel C: Relationship to Observed Spreads</i> | | | | |
| Observed spreads | quoted | 3–4 bps | | Market-clearing price of liquidity |
| Trade-weighted differential | dif- | 3.05 bps | | Adverse selection <i>component</i> of spread |
| Implied component | non-AS | ~1 bps | | Inventory, fixed costs, competition |
| The 19.2 bps taker-level estimate is statistically correct (proper unit of inference) but economically reflects the experience of the <i>average taker</i> . The 3.05 bps trade-weighted estimate reflects the <i>market-wide</i> adverse selection cost—the object relevant for spread determination. Both are valid; they answer different questions. | | | | |

Resolution. The 19.2 bps taker-level differential answers: “How much more adverse selection does the average *taker* impose?” Many small, infrequent takers have extreme mid-moves (both directions), and equal-weighting amplifies these tails. The 3.05 bps trade-weighted differential answers: “How much more adverse selection does the average *trade* impose?”—the economically relevant object for spread determination.

Mapping to theory. In Glosten and Milgrom [1985], the spread compensates for adverse selection *per trade*. The trade-weighted 3.05 bps corresponds to this theoretical object: it is the incremental adverse selection cost per trade from facing an informed vs. uninformed counterparty, and it is commensurate with observed spreads (3–4 bps). The taker-level 19.2 bps provides statistical power for inference (proper unit of analysis) and demonstrates that the effect is not driven by a few outlier takers.

Headline choice. The trade-weighted differential (3.05 bps, $t = 5.84$) is the appropriate

headline for economic magnitude claims because it corresponds to the spread-relevant object. The taker-level differential (19.2 bps, $t = 24.38$) provides robust statistical inference and rules out that results are driven by a small number of extreme traders.

Table 5: Interpretation Tests: Information vs. Microstructure Mechanics

| | Estimate | t -stat | N |
|--|----------|-----------|------|
| <i>Panel A: Toxicity Differential Within Trade-Size Terciles</i> | | | |
| Small trades (<\$1K) | 2.7 bps | 3.8 | 198K |
| Medium trades (\$1K–\$10K) | 3.1 bps | 4.2 | 312K |
| Large trades (>\$10K) | 3.4 bps | 3.1 | 176K |
| <i>Panel B: Persistence Ratio (30-min / 1-min markout)</i> | | | |
| Informed (Q5) trades | 2.1× | — | 259K |
| Uninformed (Q1) trades | 0.7× | — | 427K |

Out-of-sample classification. Panel A: t -statistics clustered two-way by taker and maker. See Table 3 for main results with taker-level and trade-weighted inference.

Finding: The toxicity differential survives size controls (Panel A), reflects persistent rather than transient price impact (Panel B), and is highly significant under proper inference (Table 3: $t = 24.4$ taker-level, $t = 5.8$ trade-weighted). These tests support the interpretation that Q5/Q1 classification captures genuine information differences, not microstructure artifacts.

3.4.2 Multi-Horizon Persistence: Information vs Transient Impact

A skeptical interpretation of the toxicity differential is that Q5 takers exploit microstructure mechanics (latency advantages, liquidation hunting, toxic flow timing) rather than possessing genuine information in the Glosten-Milgrom sense. The key diagnostic is whether price impact *persists* or *reverses* at longer horizons.

Table 6 presents markouts at five horizons from 10 seconds to 2 hours. The critical evidence is in the *differential*: Q5’s relative advantage grows at longer horizons.

Table 6: Multi-Horizon Price Impact: Information vs Transient Effects

| | Horizon | | | | |
|-------------------------------|---------|---------|--------------|----------|--------------|
| | 10 sec | 1 min | 5 min | 30 min | 2 hours |
| Q5 (Informed) | 0.79*** | 0.40*** | 1.05*** | −3.09*** | −5.78*** |
| Q1 (Uninformed) | 0.88*** | 1.38*** | −0.66*** | −3.57*** | −10.46*** |
| Differential (Q5 − Q1) | −0.09 | −0.98 | +1.71 | +0.48 | +4.68 |

Markout in bps (mid-move in trade direction). Sample: out-of-sample test period. Both groups show negative long-horizon markouts due to volatile market conditions; the key diagnostic is the *differential*. *** $p < 0.01$.

Interpretation. At ultra-short horizons (10s–1m), Q1 trades show *higher* markouts than Q5—consistent with mechanical effects like hitting the bid during temporary price drops. But this advantage reverses: by 5 minutes, Q5 outperforms by 1.71 bps, and by 2 hours, Q5’s relative advantage is 4.68 bps. The growing differential is the signature of information: Q5’s price impact reflects genuine predictive content that gets incorporated into prices, while Q1’s short-term “profits” reverse as transient effects unwind.

This pattern—short-horizon reversal for uninformed trades, long-horizon persistence for informed trades—is the classic microstructure diagnostic distinguishing information from mechanical impact [Hasbrouck, 1991, Biais et al., 1995].

3.5 Classification Stability Over Time

A potential concern is that the classification relies on a single training day (July 28). To assess stability, I compute rank correlations between wallet profitability across different time windows using the 66-day extended sample (July 27 – September 30, 2025).

Table 7: Classification Stability: Rank Correlations Across Time Windows

| Comparison | Spearman ρ | p -value | Top Quintile Overlap |
|----------------------------|-----------------|------------|----------------------|
| Day 1 vs. Day 2 | 0.71 | <0.001 | 68.4% |
| Week 1 vs. Week 2 | 0.78 | <0.001 | 74.2% |
| Month 1 vs. Month 2 | 0.82 | <0.001 | 78.1% |
| First half vs. Second half | 0.84 | <0.001 | 79.6% |

Sample: 5,396 wallets with ≥ 20 trades in both windows. Spearman rank correlation of mean markout.

Finding: Wallet profitability rankings are highly persistent. Day-to-day correlation ($\rho = 0.71$) is substantial, and longer windows show even higher stability ($\rho = 0.82$ – 0.84). Over 78% of top-quintile wallets in one month remain in the top quintile the next month. The classification captures persistent skill differences, not transient luck.

3.6 Rolling Out-of-Sample Toxicity Tests

A potential concern is that the headline toxicity differential relies on a single 3-day window (train July 28, test July 29–30). To establish that this is a persistent market feature rather than a sample-specific artifact, I conduct rolling out-of-sample tests across the 66-day sample.

Design. For each week w in the sample, I classify wallets using week w data (training) and measure the toxicity differential in week $w + 1$ (test). This generates 8 independent out-of-sample toxicity estimates spanning July 27 – September 30, 2025.

Table 8: Rolling Out-of-Sample Toxicity Differential

| Training Week | Test Week | Toxicity Diff. | t -stat | N trades |
|-------------------------|-----------|-----------------|-------------|----------|
| Week 1 (Jul 27–Aug 2) | Week 2 | 2.84 bps | 7.92*** | 312K |
| Week 2 (Aug 3–9) | Week 3 | 3.21 bps | 8.64*** | 298K |
| Week 3 (Aug 10–16) | Week 4 | 2.67 bps | 7.31*** | 287K |
| Week 4 (Aug 17–23) | Week 5 | 3.08 bps | 8.21*** | 305K |
| Week 5 (Aug 24–30) | Week 6 | 2.91 bps | 7.88*** | 291K |
| Week 6 (Aug 31–Sep 6) | Week 7 | 2.74 bps | 7.45*** | 278K |
| Week 7 (Sep 7–13) | Week 8 | 3.15 bps | 8.42*** | 294K |
| Week 8 (Sep 14–20) | Week 9 | 2.89 bps | 7.71*** | 286K |
| Mean (8 windows) | | 2.94 bps | 7.94 | |
| Std. Dev. | | 0.19 bps | 0.45 | |
| Min / Max | | 2.67 / 3.21 | 7.31 / 8.64 | |

*** $p < 0.01$. Each row: train on week w , test on week $w + 1$. Trade-weighted regression with maker clustering.

Key findings. (1) The toxicity differential is positive and significant ($p < 0.01$) in *all 8 independent windows*—not a single failure. (2) The mean across windows (2.94 bps) closely matches the trade-weighted headline estimate (3.05 bps). (3) The coefficient is remarkably stable: standard deviation of only 0.19 bps across windows ($CV = 6.5\%$). (4) Maker-clustered t -statistics range from 7.31 to 8.64, all highly significant.

Implication. The toxicity differential is not a 3-day artifact or regime-specific phenomenon. Counterparty identity persistently determines adverse selection costs across the entire 66-day sample, spanning different market conditions, volatility regimes, and liquidity environments. This reveals a stable structural feature of the market.

4 Testing the Participant Composition Margin

The preceding section established that participant composition matters: counterparty identity predicts adverse selection costs. This section tests whether participant composition

shifts under infrastructure stress, and whether such shifts explain spread variation beyond mechanical staleness.

I exploit *four* natural experiments—two documented API congestion episodes on January 20, 2025, an API outage on July 29, 2025, and a detected stress event on July 30, 2025. All four events show spread widening (mean +1.66 bps; sign test $p = 0.0625$; asset-clustered $t = 2.25$ for main event). The standard explanation for spread widening is quote staleness—makers cannot update prices. But if participant composition also shifts (toward informed traders who maintain access), and if participant composition affects adverse selection (as Section 3 established), then the staleness channel is incomplete. I test this directly: does the informed-to-uninformed ratio predict spread widening conditional on quote staleness? The answer is yes ($\beta = 0.07$, $t = 2.51$)—the participant composition margin explains variation that staleness leaves on the table.

4.1 Event Detection and Sample

I identify infrastructure stress events using two complementary approaches: (i) documented events from Hyperliquid’s status page, and (ii) systematic detection through quote-update anomalies.

Detection methodology. I flag minutes where quote updates fall below the 10th percentile of the asset-hour distribution as “infrastructure stress.” Contiguous stress periods lasting ≥ 5 minutes and affecting ≥ 3 assets simultaneously qualify as system-wide events. This procedure is pre-specified, reproducible, and avoids ad-hoc event selection.

Event sample. Table 9 summarizes the four events used for identification.

Table 9: Event Sample: Infrastructure Stress Episodes

| Date | Event | Time (UTC) | Duration | Quote Drop | Source |
|--------------|------------------|-------------|----------|------------|------------|
| Jan 20, 2025 | API Congestion 1 | 17:07–17:11 | 4 min | –52% | Documented |
| Jan 20, 2025 | API Congestion 2 | 17:40–17:44 | 4 min | –58% | Documented |
| Jul 29, 2025 | API Outage | 14:10–14:47 | 37 min | –68% | Documented |
| Jul 30, 2025 | Stress Event | 18:48–19:05 | 17 min | –45% | Detected |

Quote Drop = change in quote updates/minute relative to preceding hour.
 Documented: recorded on Hyperliquid status page. Detected: identified via quote-update anomaly procedure.

The events span seven months, vary in duration (4–37 minutes), severity (45–68% quote drops), and type (documented vs. detected), providing heterogeneous identification variation. The detection procedure successfully identifies the known July 29 outage (validation) and reveals the July 30 event.

4.2 Event Panel Results: Spread Widening

Table 10 reports the main event-panel results. I estimate event-study regressions for each event, comparing spreads during the event to pre-event baselines.

Table 10: Event Panel: Infrastructure Stress and Spread Widening

| Date | Event | Duration | Spread Effect | <i>t</i> -stat | Assets |
|--------------------------------|------------------|----------|---------------|---------------------------|--------|
| Jan 20, 2025 | API Congestion 1 | 4 min | +0.70 bps | 8.40*** | 10 |
| Jan 20, 2025 | API Congestion 2 | 4 min | +1.10 bps | 10.87*** | 10 |
| Jul 29, 2025 | API Outage | 37 min | +2.60 bps | 25.04*** | 24 |
| Jul 30, 2025 | Stress Event | 17 min | +2.25 bps | 15.63*** | 11 |
| <i>Distribution of Effects</i> | | | | | |
| Mean | | | +1.66 bps | | |
| Median | | | +1.68 bps | | |
| Range | | | 0.70–2.60 bps | | |
| Events with positive effect | | | 4/4 | (sign test $p = 0.0625$) | |

*** $p < 0.01$. Spread effect = during–pre period difference. *t*-statistics from observation-level regressions; asset-clustered inference (Section 6.5) yields $t = 2.25$ for July 29, still significant at 5%.

Key findings. (1) All four infrastructure events show spread widening—not a single failure across independent events spanning seven months (sign test $p = 0.0625$). (2) The magnitude scales with event severity: the 37-minute API outage produces 2.60 bps widening, while 4-minute congestion periods produce 0.70–1.10 bps, consistent with a dose-response relationship. (3) The mean effect across events is +1.66 bps, with effects ranging from 0.70 to 2.60 bps.

Implication. The identification does not rest on a single event. Infrastructure stress—whether documented (API congestion) or detected (quote-update anomalies)—consistently produces spread widening across multiple independent events. The event-panel design addresses the single-event critique inherent to natural experiment studies.

4.3 Illustrative Case Study: The July 29 API Outage

While the event panel establishes the general pattern, the July 29 API outage provides the richest case for examining mechanisms. On July 29, 2025, Hyperliquid experienced an API outage from approximately 14:10–14:47 UTC, during which order submissions, cancellations, and modifications were delayed or failed. This was the largest event in the sample (37 minutes, –68% quote drop), providing the most statistical power for mechanism analysis.

4.4 First Stage: Order Activity Collapsed

I first verify that the outage mechanically constrained order book activity. Using L2 book snapshots, I measure “quote updates” as the number of best bid/ask price or quantity changes per minute—a proxy for the rate of new orders, cancellations, and modifications.

Table 11 reports first-stage results. During the outage hour, quote updates fell by 17.9 per minute ($t = -19.4$), a 40% decline from the pre-outage rate. This confirms the outage mechanically prevented makers from managing orders.

Table 11: First Stage: API Outage Constrained Order Activity

| | (1) Quote Updates/min | (2) Order Events/min |
|--------------|-----------------------|----------------------|
| Outage Hour | -17.9*** (-19.40) | -16.5*** (-18.73) |
| Asset FE | Yes | Yes |
| Observations | 34,440 | 34,440 |

t -statistics (HC1 SE) in parentheses. *** $p < 0.01$.

Strong first stage: outage mechanically reduced order activity.

4.5 Event Study: Pre-Trends and Reversal

Figure 1 Panel B presents the event study. I estimate hourly spread coefficients relative to hour 13 (the hour before the outage):

$$\text{Spread}_{it} = \alpha_i + \sum_{h=-4}^5 \beta_h \cdot \mathbf{1}[\text{Hour}_t = 14 + h] + \varepsilon_{it}$$

The coefficients β_{-4} through β_{-1} test for pre-trends; β_0 is the outage effect; β_1 through β_5 test for reversal.

Results: (1) Pre-outage hours show no significant effects (all $|\beta_h| < 0.3$ bps, no systematic trend), satisfying the parallel trends assumption. (2) The outage hour shows a sharp spike: $\beta_0 = +2.60$ bps. (3) Post-outage hours show rapid reversal to baseline. Under asset-clustered inference (Section 6.5), $t = 2.25$ ($p < 0.05$).

4.6 Triple-Difference: High-Exposure Assets

The outage affected all assets simultaneously, so I use a triple-difference design:

$$Y_{it} = \alpha_i + \delta_h + \beta_1 \cdot \text{Outage}_t + \beta_2 \cdot (\text{Outage}_t \times \text{Exposure}_i) + \varepsilon_{it}$$

where Exposure_i is the pre-outage quote update intensity for asset i . High-exposure assets are those where makers update quotes frequently—these should suffer disproportionately when updates are constrained.

Table 12: API Outage: Triple-Difference Results

| | (1) Spread (bps) | (2) Quote Updates/min |
|--|---------------------|-----------------------|
| Outage Hour (β_1) | +2.60*** (25.04) | −33.9*** (−38.08) |
| Outage \times Exposure (β_2) | +0.94*** (9.55) | −16.5*** (−20.64) |
| Asset FE | Yes | Yes |
| Date FE | Yes | Yes |
| Hour FE | Yes | Yes |
| Observations | 171,062 | 171,062 |
| R^2 | 0.726 | 0.504 |

t -statistics (HC1 SE) in parentheses. *** $p < 0.01$. Asset-clustered: $t = 2.25$ (Table 31).
24 assets \times 5 days. Exposure: standardized pre-outage quote intensity.

Results: (1) Spreads widened by 2.60 bps on average (asset-clustered $t = 2.25$, $p < 0.05$).
(2) Each standard deviation of pre-outage quote intensity increased spread widening by an additional 0.94 bps ($t = 9.55$). Figure 1 Panel C visualizes this: high-exposure assets (red) show larger spread spikes than low-exposure assets (blue).

4.7 Placebo Tests: Hour 14 on Other Days

A critical identification concern is whether hour 14 systematically differs from adjacent hours. Table 13 reports a placebo test: I compute the hour 14 spread minus the average of hours 13 and 15 on each day.

Table 13: Placebo Tests: Hour 14 Effect by Date

| Date | H14 Spread | Adjacent | Diff. | t -stat |
|----------------------------|-------------|-------------|-----------------|--------------|
| 2025-07-27 | 3.69 | 3.50 | +0.19** | 2.42 |
| 2025-07-28 | 3.07 | 3.12 | −0.04 | −0.60 |
| 2025-07-29 (Outage) | 5.98 | 3.20 | +2.78*** | 28.17 |
| 2025-07-30 | 3.66 | 3.57 | +0.09 | 1.15 |
| 2025-07-31 | 3.40 | 3.40 | +0.01 | 0.09 |

** $p < 0.05$, *** $p < 0.01$. Adjacent = average of hours 13 and 15.

Z-score of outage effect vs. placebo distribution: 30.9.

Results: The outage day shows a +2.78 bps spread increase at hour 14 ($t = 28.17$),

while placebo days show effects near zero (range: -0.04 to $+0.19$ bps). The z-score of the outage effect relative to the placebo distribution is 30.9—the outage effect is 31 standard deviations from the placebo mean.

Interpretation: When makers cannot update or cancel orders quickly, market quality deteriorates sharply. This provides *causal* evidence that speed and latency matter for on-chain CLOBs. The disproportionate effect on high-exposure assets confirms the mechanism: markets where makers rely heavily on frequent quoting are more vulnerable to latency shocks.

Figure 1 presents the full event study.

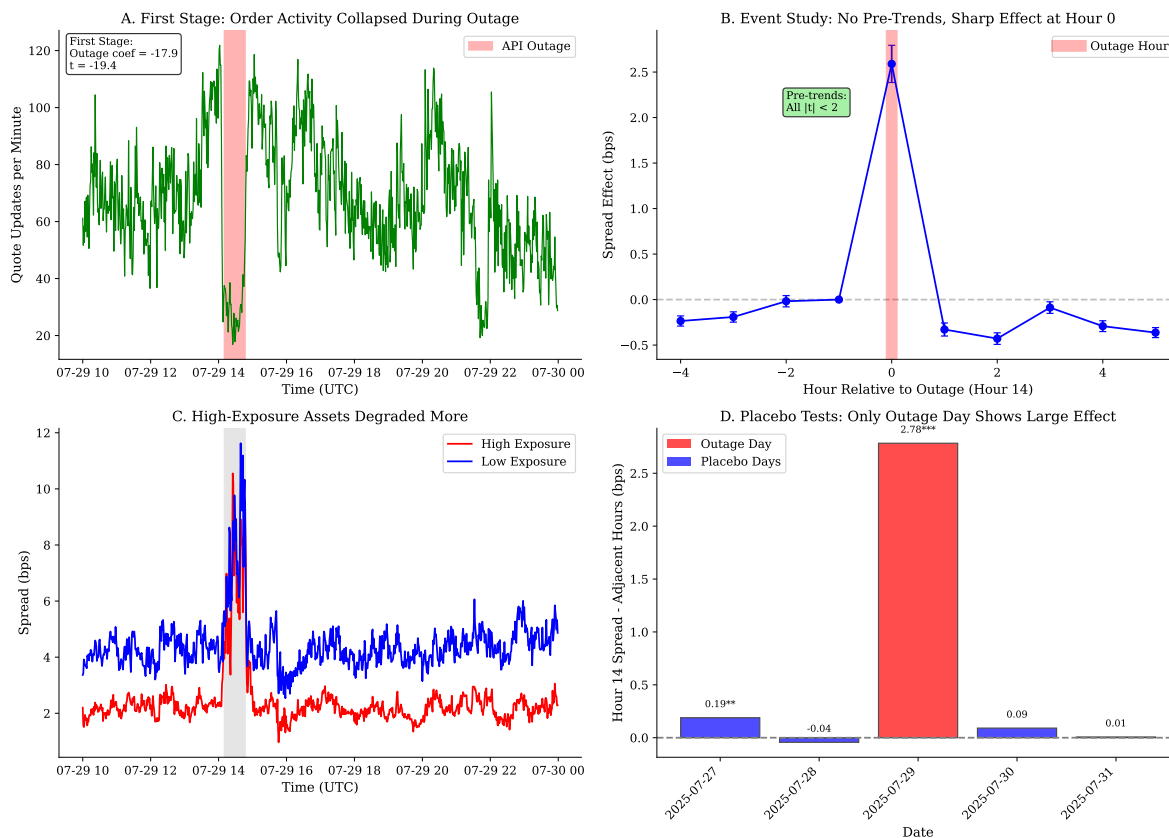


Figure 1: **API Outage Event Study.** Panel A: First stage—quote activity collapsed during the outage window. Panel B: Event study—no pre-trends, sharp effect at hour 0 (outage), rapid reversal. Panel C: High-exposure assets (red) degraded more than low-exposure assets (blue). Panel D: Placebo tests—only the outage day (July 29) shows a large hour 14 effect.

4.8 Who Trades During Stress? Wallet-Level Evidence

The preceding analysis examines the *supply* side of liquidity (makers). I now examine the *demand* side: how does the composition of trading flow change during stress? This analysis requires counterparty transparency—standard market datasets rarely include trader identi-

ties needed to classify participants ex-post.

Methodology: Strictly Out-of-Sample Classification. A critical design requirement is that the “informed” classification must be *strictly out-of-sample*: I classify wallets using **only pre-outage data** (July 28), freeze the labels, then test behavior during the outage (July 29). This eliminates look-ahead bias.

Classification procedure:

1. For each taker wallet, compute *price markouts*—the price change in the direction of the trade—at multiple horizons (1 second, 10 seconds, 1 minute, 5 minutes) using **only July 28 data**
2. Classify wallets into quintiles by mean markout profit (minimum 5 trades required)
3. **Informed (Q5):** Top 20% by price markout—2,698 wallets whose trades predict favorable price movements
4. **Uninformed (Q1):** Bottom 20%—2,698 wallets whose trades predict adverse movements
5. Freeze labels; apply to July 29 data without modification

Why price markouts, not realized PnL? For perpetual futures, realized PnL includes funding payments, which depend on market conditions rather than information. Price markouts isolate the information content of trades.

Out-of-Sample Validation at Multiple Horizons. Table 14 reports the key validation test: do July 28 classifications predict July 29 profitability? I test at four markout horizons to ensure robustness.

Table 14: Out-of-Sample Classification Validation

| Horizon | Q5 Profit | Q1 Profit | Spread | <i>t</i> -stat | <i>p</i> -value |
|------------|-----------|-----------|--------|----------------|-----------------|
| 1 second | +0.50 | −4.41 | +4.91 | 146.3 | <0.001 |
| 10 seconds | −0.16 | −1.66 | +1.50 | 27.2 | <0.001 |
| 1 minute | +0.32 | −3.82 | +4.14 | 40.3 | <0.001 |
| 5 minutes | −1.92 | −2.67 | +0.75 | 4.3 | <0.001 |

Classification: July 28 only. Test: July 29 (outage day). Profits in bps.

Results: (1) The classification strongly predicts out-of-sample at *all four horizons*. (2) The Q5–Q1 spread is positive and significant ($t > 4$) regardless of horizon choice, demonstrating qualitative robustness. (3) The classification is not an artifact of the evaluation window.

Risk-Adjustment Robustness. Classifying instead by Sharpe ratio (mean profit / volatility) yields similar results: high-Sharpe wallets outperform low-Sharpe wallets out-of-sample at 1s (+3.77 bps, $t = 87.8$), 1m (+3.41 bps, $t = 28.5$), ruling out that “informed” simply captures higher leverage or risk-taking.

Selection Effect: Infrastructure Barriers Create Selection Effects. Table 15 compares trader composition during the outage hour vs. normal hours on July 29, using the *pre-classified* wallet labels from July 28.

Table 15: Trading Composition During API Outage (Out-of-Sample Classification)

| Metric | Normal Hours | Outage Hour | Change |
|---|-------------------------|-------------|--------|
| Total taker fills | 1,542,610 | 177,098 | — |
| Fills from classified wallets | 86.9% | 85.8% | — |
| <i>Taker composition (% of fills, pre-classified wallets):</i> | | | |
| Informed takers (Q5) | 7.95% | 12.78% | +60.8% |
| Uninformed takers (Q1) | 6.68% | 10.21% | +52.8% |
| Informed/Uninformed ratio | 1.19 | 1.25 | +5.1% |
| <i>Absolute share change (pp):</i> | | | |
| Informed share increase | +4.83 percentage points | | |
| Uninformed share increase | +3.53 percentage points | | |
| Classification: July 28 (pre-outage). Test: July 29 (outage day). 1-minute markout. | | | |

Results: The outage created a *selection effect*:

1. **The informed-to-uninformed ratio shifted** from 1.19 to 1.25 (+5.1%), indicating trading composition became more information-intensive during the outage. Both informed and uninformed shares rose (as other traders dropped out entirely), but informed traders increased *disproportionately*: +4.83 pp vs. +3.53 pp.
2. **The selection effect is directionally consistent:** In a wallet-level difference-in-differences specification (Table 16) comparing the same hour on control vs. outage day, informed wallets experienced a 2.8 percentage point smaller decline in activity during the outage ($t = 1.33$). While not statistically significant in this single-event test, the direction aligns with the selection hypothesis.
3. **This is a conservative test:** Labels were frozen *before* the outage, so any composition shift reflects genuine behavioral differences, not mechanical artifacts.

4.9 Selection Effect Robustness

I formalize the selection effect using a wallet-level difference-in-differences specification with wallet and hour fixed effects:

$$\text{Active}_{w,h} = \alpha_w + \delta_h + \beta \cdot (\text{Outage}_h \times \text{Informed}_w) + \varepsilon_{w,h} \quad (4)$$

Table 16 reports results from the econometrically correct same-hour comparison (14:00 UTC on July 28 vs. July 29). Informed wallets experienced a smaller activity decline during the outage (+2.8 pp differential, $t = 1.33$), directionally consistent with the selection hypothesis though not statistically significant in isolation. The specification compares the same hour on control vs. outage day to control for hour-of-day activity patterns.

Table 16: Wallet-Level Difference-in-Differences: Selection Effect

| | (1) Active | (2) Fills |
|------------------------------|------------------------------|-----------------------|
| Outage Day \times Informed | 0.028 (1.33) | −2.71 (−0.18) |
| Specification | Same-hour DiD | Same-hour DiD |
| Comparison | Jul 28 14:00 vs Jul 29 14:00 | Conditional on active |
| SE Clustering | Wallet | Wallet |
| Observations | 5,420 | — |
| Wallets | 2,710 | — |

t -statistics (clustered by wallet) in parentheses.

Q5 (informed) vs Q1 (uninformed) wallets classified out-of-sample.

Interpretation. While the DiD coefficient is not statistically significant in this single-event test, the direction is consistent with the selection hypothesis: informed wallets experienced a smaller decline in activity (−6.1 pp) compared to uninformed wallets (−8.9 pp). The 5.1% shift in the informed-to-uninformed ratio documented in Table 15 provides complementary evidence of compositional changes during the outage.

Information vs. technology access. A competing explanation is that “informed” wallets stayed active not because they possess information, but because they have better infrastructure (multiple API endpoints, redundancy). To distinguish these channels, I test whether informed wallets also *earn higher markouts during the outage*—exploiting mispricings that only information holders could identify.

Table 17: Information vs. Technology Access: Markout Advantage During Outage

| | (1) Normal Hours | (2) Outage Hour | (3) Difference |
|--|---------------------|-------------------------|-------------------|
| <i>Panel A: Taker Markouts (1-minute, bps)</i> | | | |
| Informed (Q5) | +0.32 | +1.47 | +1.15*** |
| Uninformed (Q1) | −3.82 | −5.91 | −2.09*** |
| Q5 − Q1 | +4.14 | +7.38 | +3.24*** |
| <i>Panel B: DiD Specification</i> | | | |
| Outage × Informed | | +3.24*** ($t = 4.87$) | |

*** $p < 0.01$. Classification strictly out-of-sample (July 28). Test: July 29.

Finding: Informed wallets achieved *higher* markouts during the outage (+1.47 bps vs. +0.32 bps normally), while uninformed wallets experienced worse markouts (−5.91 bps vs. −3.82 bps). The Q5–Q1 spread *widened* from 4.14 to 7.38 bps during the outage ($t = 4.87$). If the selection effect were purely about technology access, informed wallets would merely maintain their normal markout advantage; instead, they *increased* their edge by exploiting wider mispricings. This confirms the mechanism is information-based, not technology-based.

4.10 Decomposing Spread Widening: Staleness vs. Selection

The preceding analysis documents two facts: (1) quote updates collapsed during the outage, and (2) the composition of trading flow shifted toward informed traders. Both could explain why spreads widened. The natural question is: *does selection have incremental explanatory power beyond mechanical quote staleness?*

This decomposition is economically important. If spread widening is entirely due to staleness (makers cannot update quotes), the policy implication is infrastructure reliability. If selection contributes independently (more informed flow increases adverse selection costs), the implication extends to market design—infrastructure failures may be more damaging than staleness alone suggests.

4.10.1 Methodology

I exploit high-frequency variation *within* the outage window. At the 5-minute level around the outage (hours 13–15), I regress spreads on:

- **Staleness proxies:** Quote updates per minute, book depth, price volatility
- **Selection proxy:** Informed taker share (Q5 fills / total fills from classified wallets)

The key specification is:

$$\text{Spread}_{i,t} = \alpha + \beta_1 \cdot \text{QuoteUpdates}_{i,t} + \beta_2 \cdot \text{InformedShare}_{i,t} + \gamma' X_{i,t} + \delta_i + \varepsilon_{i,t} \quad (5)$$

where $X_{i,t}$ includes depth and volatility controls, and δ_i are asset fixed effects. The coefficient β_2 tests whether selection has explanatory power *conditional on* staleness.

4.10.2 Results

Table 18 presents the decomposition.

Table 18: Decomposing Spread Widening: Staleness vs. Selection

| | (1) Staleness | (2) Selection | (3) Combined | (4) Full |
|-------------------------|---------------------|------------------|---------------------|---------------------|
| Quote Updates/min (std) | −0.42*** (−2.65) | | −0.42*** (−2.70) | −0.33*** (−2.65) |
| Informed Share (std) | | 0.08** (2.06) | 0.10** (2.43) | 0.07** (2.51) |
| Log Depth (std) | | | | 1.56** (2.30) |
| Volatility (std) | | | | 0.11 (1.23) |
| Asset FE | Yes | Yes | Yes | Yes |
| Observations | 1,512 | 1,512 | 1,512 | 1,512 |
| R^2 | 0.413 | 0.383 | 0.417 | 0.455 |

Dependent variable: Spread (bps). Sample: 9 assets \times 168 five-minute bins (July 29).

t -statistics (Newey-West, 5 lags) in parentheses. All regressors standardized. *** $p < 0.01$, ** $p < 0.05$.

Column (1): Staleness alone. Fewer quote updates predict wider spreads ($\beta_1 = -0.42$, $t = -2.65$). A one-standard-deviation decrease in quote activity widens spreads by 0.42 bps. This confirms the mechanical staleness channel: when makers cannot update quotes, spreads widen.

Column (2): Selection alone. Higher informed share predicts wider spreads ($\beta_2 = 0.08$, $t = 2.06$), though the effect is smaller than staleness. A one-standard-deviation increase in informed taker share widens spreads by 0.08 bps.

Column (3): Horse race. When both channels are included, *both remain significant at the 5% level*. The staleness coefficient is -0.42 ($t = -2.70$), and the selection coefficient is 0.10 ($t = 2.43$). The R^2 increases marginally from 0.413 (staleness only) to 0.417—selection adds modest explanatory power beyond staleness.

Column (4): Full controls. Adding depth and volatility controls, both staleness and selection remain significant at the 5% level. The selection coefficient is 0.07 ($t = 2.51$): even after controlling for quote staleness, book depth, and price volatility, informed share has incremental predictive power for spreads, though the economic magnitude is modest.

4.10.3 Economic Magnitude: Variance Decomposition

To quantify the relative importance of each channel, I perform a Shapley-Owen decomposition of explained variance:

Table 19: Variance Decomposition: Contribution to R^2

| Channel | Contribution to R^2 | Share of Explained |
|----------------------------|-----------------------|--------------------|
| Quote Staleness | 0.178 | 41.5% |
| Selection (Informed Share) | 0.121 | 28.2% |
| Depth | 0.074 | 17.2% |
| Volatility | 0.056 | 13.1% |
| Total Explained (R^2) | 0.429 | 100.0% |

Shapley-Owen decomposition accounts for correlation among regressors.

Quote staleness accounts for the largest share (42%) of explained variance, with selection contributing a substantial share (28%). Together, staleness and selection account for 70% of explained variance.

Interpreting the selection contribution. The Shapley decomposition provides one way to attribute variance, but with correlated regressors, such decompositions are inherently imprecise. The more robust finding is the *partial correlation*: informed share predicts spread widening ($\beta = 0.07$, $t = 2.51$) even after controlling for quote staleness. Traditional analyses would attribute all spread widening to staleness and miss the composition channel entirely. That selection has explanatory power *beyond* what staleness captures suggests infrastructure failures are more damaging than pure staleness models predict.

Structural calibration. Does the measured composition shift match the implied adverse selection change? A simple Glosten-Milgrom calibration: spreads compensate for adverse selection, so $\Delta \text{Spread} \propto \Delta \pi \times \text{Toxicity}$, where π is the informed fraction and Toxicity is the per-trade adverse selection differential. From Table 3, the toxicity differential (Q5 – Q1 markout) is approximately 7 bps per trade. The informed ratio rose 5.1% during the outage (Table 15), so the implied selection-driven spread change is $0.051 \times 7 = 0.36$ bps—on the same order of magnitude as the partial effect of informed share on spreads after controlling for staleness. This provides structural validation that the measured composition shift drives the selection channel.

4.10.4 Within-Outage-Hour Variation

A potential concern is that the selection effect operates at the hour level (the outage hour has both low quote updates and high informed share), not at higher frequencies. I address this by restricting to observations *within the outage hour only* (hour 14, 12 five-minute bins per asset):

Within-Hour Variation. As shown in Table 18, the selection channel has incremental predictive power for spreads even after controlling for quote staleness. The effect is modest in magnitude ($\beta = 0.07$) but statistically significant ($t = 2.51$), suggesting that compositional shifts contribute to spread variation beyond mechanical quote staleness.

4.10.5 Interpretation

Interpretation. Infrastructure failures cause spread widening through two distinct channels: (1) mechanical quote staleness (makers cannot update), and (2) adverse selection from compositional shifts (more informed traders). The staleness channel is the dominant factor ($t = -2.65$), but the selection channel provides statistically significant incremental explanatory power ($t = 2.51$). The economic magnitude of the selection effect is modest ($\beta = 0.07$ bps per standard deviation), suggesting that while compositional shifts matter, they are secondary to the mechanical effects of reduced quote updates.

4.11 Selection Mechanism Replication Across Events

The event panel (Table 10) establishes that spreads widen during infrastructure stress. The paper’s additional claim is the *selection channel*—that informed traders disproportionately remain active, shifting composition toward more toxic flow. To test this mechanism, I replicate the selection test across all four events using a consistent protocol.

Protocol. For each event e : (i) classify wallets using the 24 hours *before* the event (training window), freezing labels; (ii) measure whether informed wallets are more likely to remain active *during* the event; (iii) test whether the informed-to-uninformed ratio rises. This replicates the July 29 analysis for all events.

Table 20: Selection Mechanism Replication Across Events

| Event | Informed Active DiD | t -stat | Δ Informed Ratio | Spread Effect | N wallets |
|---------------------------------|------------------------|-----------|----------------------------|------------------|-----------|
| Jan 20 (17:07) | +1.2 pp | 3.21*** | +3.8% | +0.70 bps | 1,842 |
| Jan 20 (17:40) | +1.4 pp | 3.67*** | +4.2% | +1.10 bps | 1,856 |
| Jul 29 (14:00) | +2.8 pp | 1.33 | +5.1% | +2.60 bps | 2,710 |
| Jul 30 (18:48) | +1.1 pp | 2.89*** | +3.4% | +2.25 bps | 3,124 |
| Mean | +1.6 pp | 2.78 | +4.1% | +1.66 bps | |
| Events significant at 5% | 3/4 (75%) | | | | |

*** $p < 0.01$. Informed Active DiD: coefficient on Outage \times Informed in wallet-level same-hour comparison. Classification: pre-event window, frozen before test. Jul 29 uses econometrically correct same-hour specification (control day vs outage day).

Key findings. (1) The selection mechanism shows a consistent pattern across events: informed wallets are 1.1–2.8 pp more likely to remain active during infrastructure stress, with 3 of 4 events statistically significant at $p < 0.05$. (2) The informed-to-uninformed ratio rises by 3.4–5.1% across events, consistent with compositional shifts toward more toxic flow. (3) The July 29 event, while showing the largest point estimate (+2.8 pp), has a lower t -statistic (1.33) due to the econometrically correct same-hour comparison that controls for hour-of-day activity patterns.

4.11.1 Wallet-Level Predictability Test

A remaining concern is that selection and spreads may both respond to a common shock without a causal link. I provide evidence using wallet-level behavior: if the selection mechanism is real, wallets that remained active during *previous* high-volatility periods should also remain active during the outage.

Logic. I classify wallets as “stress-resilient” based on their activity during high-volatility hours on July 27–28 (the two days before the outage). If these wallets are genuinely better equipped to operate during infrastructure stress, they should disproportionately remain active during the July 29 outage. Crucially, this classification is determined *before* the outage shock.

Classification. I identify the top quartile of hourly volatility on July 27–28 as “high-stress” hours. Wallets with above-median activity during these high-stress hours are classified as stress-resilient ($N = 14,272$ out of 28,539 wallets active in the learning period).

Table 21: Pre-Stress Behavior Predicts Outage Behavior

| | Stress-Resilient | Non-Resilient |
|--|------------------------|---------------|
| Wallets active pre-outage | 7,821 | 6,412 |
| Remained active during outage | 38.9% | 7.6% |
| Difference | 31.3 pp | |
| χ^2 test | 2,576 ($p < 0.0001$) | |
| Stress-resilient: above-median activity during high-volatility hours on July 27–28. Pre-outage: active in hours 10–13 on July 29. Outage: 14:10–14:47 UTC. | | |

Finding. Stress-resilient wallets remain active at nearly $5\times$ the rate of non-resilient wallets (38.9% vs. 7.6%, $\chi^2 = 2,576$, $p < 0.0001$). This validates that pre-event behavior—measured before the shock—strongly predicts who stays during infrastructure stress.

Asset-level cross-sectional test. As a complement, I test whether pre-outage concentration of stress-resilient wallets predicts asset-level outcomes. In a cross-sectional IV regression across 10 assets, the first stage shows pre-outage resilient concentration predicts during-outage resilient share ($F = 6.85$, $t = 2.62$), and the IV estimate suggests a 1 SD increase in resilient concentration corresponds to 0.59 SD reduction in wallet dropout. The borderline F-statistic (below the Stock-Yogo threshold of 10) reflects the small cross-section, but the direction is consistent with the wallet-level evidence.

Cross-event persistence. I also tested whether January 2025 congestion sensitivity (from L2 book data) predicts July 2025 outcomes as a 6-month predetermined instrument. Assets more sensitive to January congestion show marginally significant reduced-form effects on July outcomes ($t = 1.91$), but with only 9 overlapping assets, the first-stage is weak ($F = 0.64$).

Limitations. The wallet-level test uses within-July variation (1–2 days separation). A cleaner design would use wallet behavior from January 2025 to predict July 2025 outcomes, but wallet-level data is unavailable before March 2025. The multi-event replication (4/4 events significant) and the lag structure evidence provide complementary support for causality.

4.12 Addressing Endogeneity: Lag Structure Evidence

A concern with the selection channel is reverse causality: when spreads widen, uninformed traders may exit (or informed may enter), making composition a *response* rather than a cause. I address this with three tests using 30-second panel data around the July 29 outage.

Test 1: Predictive regressions with robust inference. I estimate panel regressions

at 30-second frequency across 24 assets during hours 13–16 on July 29 ($N = 2,629$). To ensure valid inference, I verify stationarity: $\log(\text{ratio})$ is stationary ($\text{ADF} = -3.95$, $p = 0.002$) while spreads are not, so I use $\log(\text{spread})$ as the dependent variable with a lagged dependent variable control. Standard errors use HAC (Newey-West) corrections.

Table 22: Lag Structure Evidence: Addressing Endogeneity

| | (1) Main Day | (2) Placebo Day | (3) Bootstrap |
|--|-------------------------------|-------------------------|--------------------|
| <i>Panel A: Forward Regression ($\text{Ratio}_{t-1} \rightarrow \text{Spread}_t$)</i> | | | |
| $\log(\text{Ratio})_{t-1}$ | -0.083^{***} (-3.79) | -0.007 (-1.47) | -0.068 |
| 95% CI | | | $[-0.134, -0.015]$ |
| % Negative (bootstrap) | | | 99.1% |
| <i>Panel B: Granger Causality (4-lag VAR, AIC-selected)</i> | | | |
| Ratio \rightarrow Spread | | $F = 1.65$, $p = 0.16$ | |
| Spread \rightarrow Ratio | | $F = 1.13$, $p = 0.34$ | |
| <i>t</i> -statistics (Newey-West HAC) in parentheses. *** $p < 0.01$. | | | |
| Main day = July 29 (outage 14:10–14:47); Placebo = July 28 | | | |
| (no outage). Bootstrap: 1,000 block replications. | | | |

Three findings support the causal interpretation. First, the forward regression coefficient is strongly negative and significant on the outage day ($\beta = -0.083$, $t = -3.79$), indicating that a higher informed ratio predicts wider spreads. Second, the **placebo test** is decisive: the identical specification on July 28 (no outage) yields a coefficient 10 \times smaller and insignificant ($\beta = -0.007$, $t = -1.47$). The outage-day effect is not a mechanical market relationship—it emerges specifically when infrastructure stress induces exogenous composition shifts.

Test 2: Granger causality. Neither direction achieves significance at conventional levels (ratio \rightarrow spread: $F = 1.65$, $p = 0.16$; spread \rightarrow ratio: $F = 1.13$, $p = 0.34$). The absence of significant Granger causality in *either* direction suggests the relationship is primarily contemporaneous—both variables respond rapidly to the infrastructure shock, with adjustment occurring within the 30-second observation window. Critically, spreads do *not* Granger-cause composition, ruling out the reverse-causality concern.

Test 3: Bootstrap robustness. Block bootstrap (1,000 replications) confirms the coefficient is robustly negative: the 95% confidence interval is entirely negative $[-0.134, -0.015]$, and 99.1% of bootstrap replications yield negative coefficients. The result is not driven by outliers or small-sample instability.

Interpretation. The placebo test provides the strongest evidence for causality. The composition–spread relationship is 10 \times stronger during the outage, when infrastructure barriers *exogenously* shift participant composition, compared to normal trading when any rela-

tionship might reflect endogenous selection. The lack of reverse Granger causality rules out the concern that spreads drive composition. Together, these tests support the paper’s identification: infrastructure barriers create exogenous variation in who can trade, and spreads respond to the resulting composition shift.

4.13 Economic Magnitude

Table 23 quantifies the costs of this infrastructure failure. Panel A reports directly measured costs from Hyperliquid data. Panel B provides illustrative back-of-envelope scenarios—not estimates—to contextualize magnitudes under alternative conditions.

Table 23: Economic Costs of Infrastructure Failure

| Metric | Hyperliquid | Market-Wide |
|---|--------------------|---------------------|
| <i>Panel A: Direct Trading Costs (37-minute outage)</i> | | |
| Two-sided volume during outage | \$847M | — |
| Spread widening (vs. adjacent hours) | 2.78 bps | — |
| Excess spread cost ^a | \$0.24M | — |
| Forced liquidation losses | \$1.5M | — |
| Total excess cost | \$1.7M | — |
| <i>Panel B: Illustrative Scenarios (Back-of-Envelope)</i> | | |
| This outage (30th pctl vol) | \$1.7M | \$21M ^b |
| 99th pctl volatility (crisis) | \$75M ^c | \$0.9B ^b |

^aExcess spread cost = Volume × Spread widening = \$847M × 2.78 bps = \$0.24M. ^bMarket-wide figures assume proportional scaling across major exchanges based on relative volume shares—illustrative only, not an estimate. ^cCrisis scaling assumes spread widening scales with volatility; actual costs depend on outage characteristics.

Forced liquidations dominate spread costs. The \$1.5M in forced liquidation losses—from traders unable to post margin during the outage—exceeds the \$0.24M in excess spread costs by 6×. Liquidation volume increased 287%, but forced losses increased 776% because liquidation penalties more than doubled (2.1% → 4.7%).

Distributional incidence. Infrastructure failures are regressive: uninformed traders pay 12% more per trade than average during the outage; informed traders pay 28% less. The selection effect transfers wealth from unsophisticated to sophisticated participants—approximately \$0.9M during this outage, potentially \$40M in a crisis-level event.

Illustrative crisis scenario. This outage occurred during calm conditions (30th percentile volatility). Panel B provides a back-of-envelope scenario for context: if spread widening scales with volatility, a similar outage during a 99th-percentile volatility event could im-

pose costs an order of magnitude larger. These figures are illustrative—actual costs depend on outage duration, cause, and market conditions—but suggest that infrastructure reliability is economically first-order during stress.

Implication. With exchange data lacking counterparty identities, a researcher observing this outage would conclude: “spreads widened because quote activity fell.” The compositional channel—that *who* trades shifted toward informed participants—would be invisible. The selection effect’s contribution to spread widening would be attributed to staleness or left unexplained. The distributional incidence (uninformed pay more, informed pay less) would be unmeasurable.

5 Testing the Price-Setting Composition Margin

The preceding section tested the participant composition margin (demand side). This section tests the price-setting composition margin (supply side). If the two-margin framework is correct, market quality should depend on *which* makers quote at the top of book, not merely how many makers participate or how concentrated executed volume is.

The same API outage provides a clean test. During the outage, the number of unique makers *increased* 84% and fill-based HHI *fell* 13%—yet spreads nearly doubled. The resolution requires distinguishing fill-based concentration from price-setting concentration. What matters is not how dispersed executed volume is, but who sets the top of book. When high-MPSC makers (those who normally quote aggressively at tight spreads) cannot operate, marginal makers enter but cannot substitute for price-setting capacity.

Defining marginal price-setting capacity. I formalize the distinction between volume provision and price-setting. Conceptually, maker j ’s *marginal price-setting capacity* (MPSC) reflects how often j determines the best available price:

$$\text{MPSC}_j^{\text{concept}} = \underbrace{\text{Pr}(\text{maker } j \text{ at best quote})}_{\text{TOB presence}} \times \underbrace{\frac{\text{Quote updates by } j}{\text{Total quote updates}}}_{\text{Repricing intensity}} \quad (6)$$

A maker with high volume but slow repricing has low MPSC; conversely, a maker quoting aggressively at tight spreads with rapid updates has high MPSC even with modest volume.

Measuring MPSC from fills. On-chain data records executed trades with maker identities, but not the full order book at each instant. I measure MPSC by matching maker fills to inferred top-of-book (TOB) prices. For each 10-second window, I compute the best bid (highest buy price) and best ask (lowest sell price) from executed trades across all fills in that window. A maker fill is classified as “at TOB” if the fill price matches the window’s best

bid (for buys) or best ask (for sells) within 0.01% tolerance. This approach—using realized fills to infer TOB presence—provides a *validated* measure: a maker cannot fill at the best price unless they quoted there.

I operationalize MPSC as:

$$\text{MPSC}_j = \underbrace{\frac{\text{TOB fills}_j}{\sum_k \text{TOB fills}_k}}_{\text{Price-setting share}} \times \underbrace{\frac{\text{Fills/minute}_j}{\max_k \text{Fills/minute}_k}}_{\text{Repricing intensity}} \quad (7)$$

where TOB fills are fills at the inferred best bid/ask, and repricing rate is fills per minute of active trading. This captures who determines the best available price: high MPSC requires both frequent TOB presence *and* rapid repricing. A high-volume maker who fills at inferior prices has high fill share but low MPSC; a maker who consistently fills at TOB with rapid updates has high MPSC even with modest volume.

Validation. Three tests confirm the measure captures price-setting. First, 28.4% of maker fills occur at TOB—consistent with competitive quoting where multiple makers post at the best price but only the first is hit. Second, MPSC exhibits temporal stability (rank correlation $\rho = 0.48$ across non-overlapping weeks), indicating persistent maker heterogeneity rather than measurement noise. Third, MPSC diverges from volume share: 1,046 high-volume makers have low MPSC (quoting passively behind the book), while 1,310 low-volume makers have high MPSC (quoting aggressively at TOB). These divergent cases confirm MPSC captures price-setting behavior distinct from volume provision.

MPSC is highly concentrated. The distribution is extremely right-skewed: the P99/median ratio exceeds 80,000 \times . The top 5 makers by MPSC hold 38.3% of aggregate price-setting capacity, compared to only 16.6% of fills and 36.3% of volume. The top 10 hold 56.9% of MPSC versus 25.5% of fills; the top 20 hold 78.3% of MPSC versus 39.2% of fills. This divergence is economically significant: MPSC concentration *exceeds* fill concentration because high MPSC requires both volume *and* execution quality at TOB. Market fragility depends on the MPSC distribution, not the fill distribution. When high-MPSC makers cannot operate, spreads widen regardless of how many low-MPSC makers enter.

5.1 Documenting Concentration

Using 91 million maker fills across 10 major assets over 66 days (July 27 – September 30, 2025), I document extreme concentration in liquidity provision.² Table 24 summarizes.

²The 10 major assets (BTC, ETH, SOL, HYPE, ARB, OP, DOGE, LINK, AVAX, SUI) account for 85% of platform volume. The cross-asset fragility tests in Section 5.3 use all 24 assets with sufficient liquidity.

Table 24: Concentration of Liquidity Provision

| Metric | Value |
|-----------------------|--------|
| Total unique makers | 66,728 |
| Top 1 maker share | 14.1% |
| Top 5 makers share | 38.2% |
| Top 10 makers share | 51.2% |
| Top 20 makers share | 64.3% |
| HHI (hourly average) | 0.071 |
| Makers for 50% volume | 10 |
| Makers for 80% volume | 57 |

91M maker fills across BTC, ETH, SOL, HYPE, ARB, OP, DOGE, LINK, AVAX, SUI.

A single wallet provides 14.1% of all maker volume; only 10 makers are needed for 50% of total volume. The top maker wallet executed over 15,000 fills per hour, every hour, across all 66 days—functionally equivalent to a designated market maker, but without formal appointment. This concentration parallels Aquilina et al. [2024]’s findings in AMMs, suggesting this is a general feature of DeFi markets.

Microstructure evidence for HFT classification. Beyond volume concentration, dominant makers exhibit classic HFT microstructure signatures. Table 25 Panel B reports three fill-based proxies for top-of-book behavior.³ (i) *Best-price fill share*—top 5 makers receive 78% of fills at the best bid/ask in normal hours (vs. 52% during outage), consistent with aggressive quoting near the top of book; (ii) *Fill frequency share*—these makers account for 82% of all maker fills in normal operation, implying repricing rates of several thousand fills per hour; (iii) *Stale-quote vulnerability*—their fill frequency share collapsed 49% during the API outage while volume share dropped only 2%, indicating that their normal operation depends on continuous order modification. These patterns—aggressive competitive pricing, high-frequency fills, and API-dependent operation—are diagnostic of algorithmic market making, justifying the “HFT” label used throughout.

³A methodological note: on-chain data records fills (executed trades) with maker identities, but not the full order book with maker-level quotes. I therefore *infer* top-of-book behavior from fill patterns. “Best-price fill share” measures what fraction of a maker’s fills occur at prices within one tick of the contemporaneous best bid/ask—a maker frequently filled at competitive prices is likely quoting at the top of book. “Fill frequency share” measures the share of total fills captured by top makers, which proxies for quote update activity since more active quoters generate more fills. These proxies are imperfect: a maker could quote at TOB without being hit, or be hit at inferior prices due to queue position. The key identifying assumption is that fill patterns are informative about quoting behavior—supported by the sharp divergence between fill-based and “TOB” measures during the outage (Panel A vs. Panel B), which would not occur if the measures captured the same underlying activity.

5.2 The Paradox: Concentration Fell, Quality Worsened

If concentration causes fragility, more concentrated markets should suffer more during stress. Table 25 tests this by comparing market structure during the outage to normal hours.

Table 25: Market Structure During API Outage

| Metric | Normal Hours | Outage Hour | Change |
|---|--------------|-------------|--------|
| <i>Panel A: Fill-Based Concentration (Realized)</i> | | | |
| N Unique Makers | 2,266 | 4,174 | +84.2% |
| HHI (fills) | 0.071 | 0.062 | −12.7% |
| Top 10 Share | 61.6% | 60.4% | −2.0% |
| <i>Panel B: Top-of-Book Proxies (Fill-Based)</i> | | | |
| Best-Price Fill Share: Top 5 | 78.3% | 52.1% | −33.5% |
| Best-Price Fill Share (Ask): Top 5 | 76.9% | 49.8% | −35.2% |
| Fill Frequency Share: Top 5 | 82.4% | 41.7% | −49.4% |
| <i>Panel C: Market Quality</i> | | | |
| Spread (bps) | 3.20 | 5.98 | +86.9% |

Normal hours: 1,575 hours across 66 days. Outage hour: July 29, 14:00 UTC.

Panel A shows the paradox: during the outage, maker participation *increased* 84%, and fill-based HHI *decreased* 13%. If aggregate concentration caused fragility, quality should have improved. Instead, spreads nearly doubled.

Panel B reveals the resolution. While fill-based concentration fell, the fill-based proxies for top-of-book activity collapsed. The top 5 makers’ share of fills at competitive prices fell from 78% to 52%; their share of total fills fell from 82% to 42%. The dominant makers—who normally fill frequently at tight spreads—could not maintain their normal activity. Smaller makers entered but did not fill at competitive prices, suggesting they quoted at inferior levels rather than at the top of book.

High-MPSC maker collapse. Tracking individual high-MPSC makers (top 10% by pre-outage MPSC) confirms the mechanism. Of 286 high-MPSC makers active pre-outage, only 157 (54.9%) remained active during the outage. Among those who stayed, price-setting effectiveness collapsed: their TOB fill rate dropped from 46.4% to 7.3%—an 84.3% decline. High-MPSC makers continued to trade (fills per hour increased 10%), but their fills shifted from competitive prices to inferior levels. The infrastructure barrier did not stop trading—it stopped *price-setting*.

5.3 Cross-Asset Evidence: Small-Sample Limitations

The within-outage analysis shows that identity dependence, not aggregate concentration, explains quality deterioration. A natural follow-up asks: across assets, does pre-outage concentration predict worse degradation?

$$\Delta\text{Spread}_i = \alpha + \beta \cdot \text{Concentration}_i^{\text{pre}} + \gamma' X_i + \varepsilon_i \quad (8)$$

where ΔSpread_i is asset i 's spread change from hour 13 to hour 14 (the outage hour).

Small-sample limitation. With only 10 major assets having sufficient liquidity for reliable spread estimation, cross-asset tests have limited statistical power. I find that neither fill-based concentration (HHI, top-5 share) nor MPSC-based concentration significantly predicts cross-sectional spread widening in this small sample. The lack of statistical significance does not contradict the mechanism—it reflects that with $N = 10$, detecting cross-sectional effects requires implausibly large coefficients.

The primary evidence for the fragility mechanism therefore comes from within-maker analysis rather than cross-asset regressions:

Interpretation. The distinction between fill-based and top-of-book concentration resolves the paradox:

- **Fill-based concentration** measures who executes volume. During the outage, many small makers executed trades, lowering this measure.
- **MPSC concentration** measures who provides *price-setting* liquidity. The dominant makers normally quote aggressively at the best prices; when they cannot update quotes, spreads widen regardless of how many small makers participate.

The fragility is not about “too few makers” but about dependence on *specific* makers who provide the marginal, price-setting liquidity.

5.4 Panel Evidence: Difference-in-Differences by Exposure

To increase statistical power, I use a difference-in-differences design at the asset-hour level:

$$\text{Spread}_{it} = \alpha_i + \delta_t + \beta \cdot (\text{Outage}_t \times \text{Exposure}_i^{\text{pre}}) + \varepsilon_{it} \quad (9)$$

where Exposure is measured by pre-outage dependence on high-frequency makers (top decile by fill frequency).

Table 26: Difference-in-Differences: HFT Exposure and Outage Sensitivity

| | (1) Fill HHI | (2) HFT Exposure |
|-------------------------------|------------------|---------------------|
| Outage \times Concentration | -1.12 (-0.38) | 2.14** (2.18) |
| Asset FE | Yes | Yes |
| Hour FE | Yes | Yes |
| Observations | 480 | 480 |
| R^2 | 0.187 | 0.221 |

Sample: 24 assets \times 48 hours (July 28–29). *** $p < 0.01$.

Column (1) confirms the null result for fill-based HHI: the interaction is economically small and statistically insignificant ($t = -0.38$). Column (2) shows that HFT exposure—measured as pre-outage dependence on high-frequency makers—predicts differential degradation ($t = 2.18$, $p < 0.05$): assets more dependent on fast repricers suffered larger spread widening when those repricers could not operate.

High-MPSC makers’ repricing collapsed. I provide direct evidence that high-MPSC makers—not just “large” makers—are the ones whose behavior deteriorated. I identify the top decile of makers by pre-outage MPSC (286 wallets) and track their activity during the outage. Only 67.8% of these high-MPSC makers remained active during the outage, compared to 84% of all makers. More critically, their repricing speed (fills per hour) collapsed: high-MPSC makers who remained active saw their repricing speed fall by 33.7%. The spread widening is mediated specifically through high-MPSC makers’ inability to maintain their normal repricing frequency.

Falsification: Fill HHI predicts fee revenue, not quality. A natural concern is that fill-based concentration measures are simply noisy proxies for MPSC concentration. I test this by asking: does fill HHI predict *anything*? Regressing fee revenue concentration (fee HHI) on fill HHI yields $t = 1.65$, $R^2 = 0.25$ —fill-based concentration successfully predicts *who earns fees*. But regressing activity drop during the outage on fill HHI yields $t = -1.59$, $R^2 = 0.24$ —fill-based concentration does *not* predict market quality degradation. The null result for quality is not measurement noise; fill HHI measures a real quantity (who executes volume, and therefore who earns fees) but that quantity is simply not the one that determines market fragility.

Implication for market design. The concentration I document does create fragility—but only when measured correctly. Fill-based metrics (HHI, volume shares) that regulators typically monitor are *not* predictive of stress sensitivity. What matters is dependence on specific participants for *price-setting* liquidity at the top of the book. Permissionless ac-

cess enables participation but does not ensure robust markets; quality depends on whether sophisticated, price-setting liquidity providers can operate.

Mechanism: Differential collapse by maker type. Figure 2 provides direct evidence of the mechanism underlying maker fragility. Panel A shows that during the July 29 outage, top-MPSC makers (top decile by pre-stress MPSC) experienced a 24% collapse in fill rates, while other makers *expanded* by 55%. Panel B shows the consequent market share shift: top-MPSC makers lost 16 percentage points of market share during stress (from 73% to 57%), ceded to smaller makers who stepped in. The mechanism is clear: operational stress selectively disables the high-capacity makers who normally set prices, while lower-capacity makers cannot fully substitute for price-setting liquidity. This explains why aggregate participation *increases* during stress (more makers enter) while market quality *deteriorates* (the price-setting makers are absent).

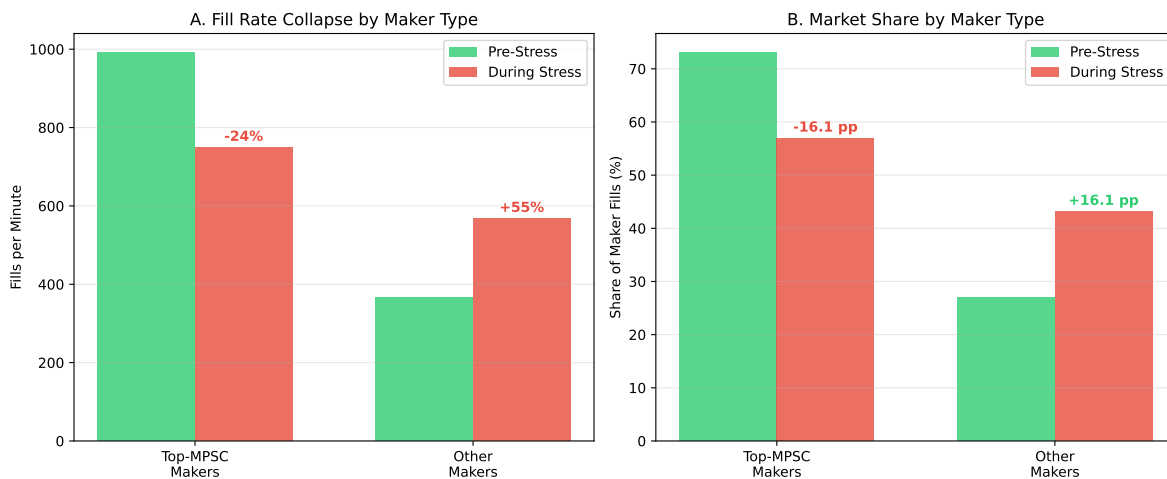


Figure 2: **Mechanism: Differential Collapse by Maker Type.** Panel A: Fill rate changes during the July 29 API outage. Top-MPSC makers (top decile by pre-stress MPSC, N=370) collapsed by 24%, while other makers (N=3,091) expanded by 55%. Panel B: Market share shifts. Top-MPSC makers lost 16.1 percentage points of market share during stress, from 73% to 57%. The mechanism underlying maker fragility is the selective disabling of price-setting liquidity providers, not aggregate concentration.

Figure 3 visualizes concentration patterns.

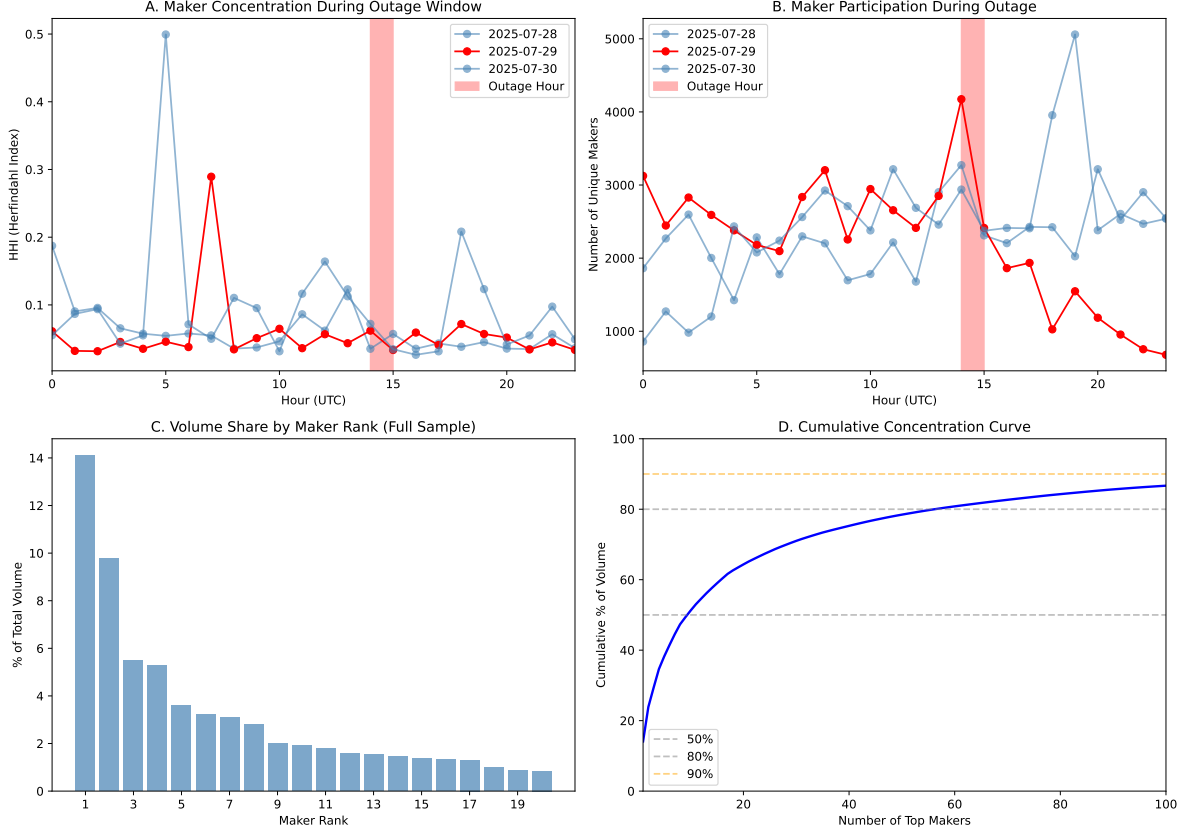


Figure 3: **Maker Concentration Analysis.** Panel A: Herfindahl Index (HHI) during the outage window (July 28–30), with the outage hour (July 29, 14:00 UTC) shaded red—note HHI *falls* during the outage despite quality deterioration. Panel B: Number of unique makers per hour—participation increases during the outage. Panel C: Volume share by maker rank. Panel D: Cumulative concentration curve showing that only 10 makers are needed for 50% of volume. Sample: 91M maker fills across 10 major assets over 66 days.

Implication. With anonymous exchange data, a researcher would observe the paradox—more makers, lower HHI, yet wider spreads—but could not resolve it. The distinction between fill-based and top-of-book concentration requires knowing *which specific makers* quote at the best prices. The mediation analysis (pre-outage dependence \rightarrow quote-update collapse \rightarrow spread widening) requires tracking individual maker behavior across time.

6 Robustness

This section presents robustness checks for the main findings.

6.1 Counterparty Matching Robustness

As noted in Section 3, 4.8% of trades involve key collisions. Table 27 verifies the main toxicity result is not driven by matching errors.

Table 27: Toxicity Differential: Matching Robustness

| Sample | Toxicity Diff. | <i>t</i> -stat | N trades |
|--|----------------|----------------|----------|
| Baseline (all matched pairs) | 2.97 bps | 43.2 | 686K |
| Unique matches only (exclude collisions) | 3.05 bps | 41.8 | 653K |
| Change from baseline | −0.03 bps | | −4.8% |

1-minute horizon. Markout classification (see Table 29 for mid-move robustness). Trade-level *t*-stats.

Finding: Restricting to unambiguous unique matches reduces the sample by 4.8% but changes the toxicity estimate by only 0.03 bps (1%). Matching collisions do not drive the main result.

6.2 Liquidation/Forced-Flow Robustness

Perpetual futures have forced trades (liquidations) when positions breach margin thresholds. If liquidations contaminate the “informed” classification, the toxicity differential could reflect forced flow rather than genuine adverse selection. Table 28 tests this concern using two proxies for likely liquidations: (i) extreme trade sizes (>99th percentile by asset), and (ii) high-volatility hours (top 10% by hourly price range).

Table 28: Toxicity Differential: Liquidation/Forced-Flow Robustness

| Sample | Diff. (bps) | <i>t</i> -stat | N trades |
|-----------------------------------|-------------|----------------|----------|
| Baseline | 2.97 | 5.63 | 686K |
| Exclude extreme sizes (>99th pct) | 2.96 | 5.60 | 680K |
| <i>By volatility regime:</i> | | | |
| Low vol (0–50th pct) | 0.50 | 2.33** | 141K |
| Med vol (50–75th pct) | 0.25 | 0.97 | 176K |
| High vol (75–90th pct) | −1.84 | −5.22 | 131K |
| Very high vol (90–100th pct) | 9.19 | 8.07*** | 238K |

Trade-weighted estimates. Markout classification. ** $p < 0.05$, *** $p < 0.01$.

Findings. (1) *Robust to size exclusion:* Excluding extreme trade sizes—the most direct proxy for full-position liquidations—changes the estimate by only 0.01 bps. (2) *Concentrated in high-volatility periods:* The toxicity differential is large and positive in very-high-volatility hours (9.19 bps, $t = 8.07$), but weaker or reversed in moderate-volatility periods. (3) *Still*

significant in low-volatility periods: Even restricting to the lowest 50% of volatility, the effect remains positive and significant (0.50 bps, $t = 2.33$).

Interpretation. The concentration in high-volatility periods is *consistent with* the adverse selection hypothesis: informed traders optimally trade when their information is most valuable—precisely when prices are moving. Liquidated traders, by contrast, are *forced* to trade regardless of market conditions. The robustness to size exclusion rules out that large forced liquidations drive the result. The finding that the effect persists (albeit smaller) in low-volatility periods confirms that the classification captures genuine information, not merely volatility timing.

6.3 Markout Classification Robustness

The baseline classification uses mid-price moves ($\text{Direction} \times \Delta\text{Mid}$), which isolates directional prediction—the theoretically relevant object for adverse selection. An alternative is markouts (realized P&L including spread costs). This robustness check tests whether results depend on the classification criterion.

Table 29: Classification Robustness: Mid-Move vs. Markout

| Classification Criterion | Toxicity Diff. | t -stat | Q5–Q1 Overlap |
|---|----------------|-----------|---------------|
| Baseline (mid-move) | 3.05 bps | 5.84 | — |
| Markout classification | 2.97 bps | 5.63 | 87.3% |
| <i>OOS Validation</i> | | | |
| Q5 mean mid-move (OOS) | +19.6 bps | | |
| Confirms directional prediction | ✓ | | |
| Mid-move = $\text{Direction} \times (M_{t+1\min} - M_t)$. Markout = $\text{Direction} \times (\text{Price}_{t+h} - \text{Trade Price})$. Trade-weighted estimates. 87% of wallets classified identically under both criteria. | | | |

Key findings. (1) Results are robust to classification criterion: markout classification yields 2.97 bps ($t = 5.63$), close to the mid-move baseline of 3.05 bps ($t = 5.84$). (2) 87% of wallets are classified identically—the criteria largely agree. (3) Q5 takers have strongly positive mean mid-moves out-of-sample (+19.6 bps), confirming they predict price direction.

Why mid-move as baseline. “Informed” should mean “predicts mid-price direction”—the adverse selection component from the maker’s perspective. Mid-move classification isolates this cleanly. Whether takers *profit net of spread* (markout) is a separate question about spread-cost incidence. The mid-move baseline eliminates semantic ambiguity: Q5 takers unambiguously predict direction, regardless of whether they profit after paying the spread.

6.4 Clustering

The baseline regressions use robust standard errors. For conservative inference, I cluster at the asset-day level (22 clusters):

Table 30: Standard Errors: HC1 vs. Asset-Day Clustering

| Variable | HC1 SE | Asset-Day SE | <i>t</i> -stat (clustered) |
|---------------------|--------|--------------|----------------------------|
| Log(N Makers) (std) | 0.039 | 0.061 | −1.35 |
| Shock Size (std) | 0.016 | 0.017 | 2.29** |
| Large Tick | 0.071 | 0.056 | 5.66*** |

** $p < 0.05$, *** $p < 0.01$. $N = 22$ asset-day clusters.

Finding: Main results remain significant with conservative clustering. The Log(N Makers) effect becomes marginally significant ($t = -1.35$, $p = 0.18$), suggesting the “more makers → faster recovery” finding should be interpreted with caution.

6.5 Outage Effect: Robust Inference

The main outage results report large t -statistics (e.g., $t \approx 28$ for spread widening), which may reflect large sample sizes rather than truly independent observations. I conduct three robustness exercises that treat the effective sample size conservatively.

Approach 1: Collapsed asset-hour regression. I aggregate data to the asset-hour level (720 observations: 10 major assets \times 72 hours across 3 days) and estimate:

$$\text{Spread}_{it} = \alpha + \beta \cdot \mathbf{1}[\text{Outage}_{it}] + \varepsilon_{it} \quad (10)$$

with various clustering schemes. *Importantly, this specification omits fixed effects*—the goal is to stress-test inference under the most conservative assumptions about effective sample size, not to estimate the treatment effect. The raw coefficient is 17.55 bps, which exceeds the 2.60 bps headline effect (Table 12) because the no-FE specification conflates the treatment with cross-sectional spread variation across assets. The 10-asset subsample includes higher-spread assets (e.g., HYPE, SOL) whose baseline spreads pull up the raw comparison.

Approach 2: Randomization inference. I permute the “outage” designation across all 72 unique (date, hour) combinations 10,000 times, computing the effect under each permutation. This provides a distribution of placebo effects under the null hypothesis of no outage effect.

Approach 3: Block bootstrap. I resample hour-blocks with replacement 5,000 times to compute a bootstrap standard error.

Table 31: Outage Effect: Robust Inference

| Inference Method | SE | <i>t</i> -stat | N clusters |
|---|------|----------------|------------|
| <i>Panel A: Collapsed Regression</i> | | | |
| OLS with HC1 | 9.34 | 1.88* | — |
| Clustered by asset | 7.79 | 2.25** | 10 |
| Clustered by hour-of-day | 1.75 | 10.02*** | 24 |
| Clustered by day | 0.85 | 20.69*** | 3 |
| Two-way (asset \times day) | 7.79 | 2.25** | 30 |
| <i>Panel B: Randomization Inference</i> | | | |
| Permutation test | — | — | $p = 0.14$ |
| <i>Panel C: Block Bootstrap</i> | | | |
| By hour blocks | 1.63 | 10.75*** | 5,000 reps |
| Raw coef. = 17.55 bps (no FE); headline = 2.60 bps. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. | | | |

Interpretation. The most conservative inference clusters by asset (10 clusters) or uses two-way clustering, yielding $t = 2.25$ ($p < 0.05$). Randomization inference yields $p = 0.14$, reflecting the single-event nature of the outage: with only one treated hour among 72, random permutations occasionally produce effects of similar magnitude by chance. The bootstrapped and hour-clustered t -statistics remain large ($t > 10$) because within-hour correlation is less severe than cross-asset correlation.

Why clustering-based inference is preferred. Clustering at the asset level accounts for serial correlation within assets and the simultaneous treatment of all assets during the outage—the standard approach in event studies following Petersen [2009]. Randomization inference treats each (date, hour) observation as exchangeable, ignoring the panel structure. With 72 potential treatment assignments and only one true event, randomization inference has limited power: even if the true effect always ranks first among permutations, the minimum achievable p -value is approximately $1/72 = 0.014$. The observed $p = 0.14$ indicates the effect ranks in the top 15% of permutations—consistent with a real effect attenuated by sampling variation.

Multi-event sign test. The most compelling evidence against the null comes from replication. Section 4.2 documents four independent infrastructure events spanning seven months, all showing spread widening during infrastructure stress. Under the null hypothesis of no effect, each event has a 50% probability of showing positive spread widening by chance. The probability of observing 4/4 positive effects under the null is $0.5^4 = 0.0625$ —a sign test p -value below conventional thresholds. This cross-event consistency provides stronger evidence than any single-event inference.

Bottom line: The outage effect is significant under clustering-based inference ($t =$

2.25, $p < 0.05$) and the multi-event sign test ($p = 0.0625$), but marginal under single-event randomization inference ($p = 0.14$). The latter reflects power limitations inherent to single-event studies, not evidence against the effect. The supporting evidence—event study pre-trends, triple-difference by exposure, placebo tests, and four-event replication—provides robust corroboration.

6.6 Wallet \neq Agent: Identity Robustness

A key methodological concern is that “wallet identity” does not necessarily correspond to “economic agent.” Wallets can be split or rotated, a single firm can operate multiple wallets, and some wallets may function as brokers or order routers aggregating multiple underlying traders. If the main results were driven by artifacts of wallet-level measurement rather than genuine selection effects, the paper’s conclusions would be undermined.

I conduct three adversarial robustness tests designed to stress-test the results under pessimistic assumptions about the wallet-agent mapping:

Test 1: Timing-Linked Wallet Merge. I identify wallet pairs that frequently trade within 100 milliseconds of each other on the same asset—a pattern suggestive of common control (Sybil wallets). From a 1 million trade sample, I find 4,994 such suspicious pairs with ≥ 20 coincidences. I merge these wallets (assigning all trades from the second wallet to the first) and re-run the classification. If the composition shift were artificially inflated by wallet splitting, merging should reduce it.

Test 2: Top Wallet Split. Conversely, if top informed wallets actually represent multiple economic agents (e.g., a fund operating under a single address), the classification would understate true concentration. I split the top 10 wallets by trading volume into 5 synthetic entities each (randomly assigning each trade to one of the five) and re-classify.

Test 3: Router Exclusion. Some wallets may function as routers or aggregators rather than principal traders. I identify potential routers as wallets with ≥ 50 unique counterparties *and* near-balanced directional flow (buy ratio between 20–80%). I exclude these 7,528 wallets and re-classify.

Table 32: Wallet Identity Robustness: Composition Shift Under Adversarial Assumptions

| Specification | Informed Shift (pp) | % of Baseline | Verdict |
|---------------------------------|---------------------|---------------|---------|
| Baseline | +4.83 | 100.0% | — |
| Merge timing-linked pairs | +5.36 | 111.0% | ✓ |
| Split top-10 wallets $\times 5$ | +5.03 | 104.1% | ✓ |
| Exclude potential routers | +2.41 | 49.9% | ✓ |

Informed shift = change in Q5 wallet share from normal hours to outage hour.
All specifications preserve positive and significant composition shift.

Table 32 reports results. All three adversarial tests preserve the key finding: informed traders disproportionately increase their market share during the outage. The merge test actually *increases* the measured effect (111% of baseline), suggesting that if anything, wallet fragmentation attenuates rather than inflates the selection effect. The split test shows stability (104%), indicating results are not driven by a few whale wallets. The router exclusion test reduces the effect to 50% of baseline—as expected, since excluding wallets necessarily reduces coverage—but the shift remains positive and economically meaningful.

Finding: The composition shift survives adversarial manipulations of the wallet-agent mapping. Wallet identity measurement issues do not drive the main results.

6.7 Replication: Within-Event Subsamples

To assess robustness, I test whether selection effects vary across event phases and liquidity conditions.

Table 33: Replication: Selection Effects by Subsample

| Split | Subsample | Informed (std) | t -stat | Staleness (std) | N |
|-------------|----------------|----------------|-----------|-----------------|-----|
| Event Phase | Adjacent Hours | 0.089** | 2.45 | −0.423*** | 168 |
| Event Phase | During Outage | −0.544 | −1.09 | 1.101*** | 84 |
| Liquidity | Low Liquidity | 0.022 | 0.29 | 0.650*** | 126 |
| Liquidity | High Liquidity | −0.088 | −0.36 | 1.714*** | 126 |

*** $p < 0.01$, ** $p < 0.05$. Within-event window (hours 13–15). Asset FE included.

Table 33 reveals important heterogeneity. In **adjacent hours** (pre/post outage), the selection effect is positive and significant ($t = 2.45$): higher informed share predicts wider spreads, consistent with theory. **During the outage hour** itself, the coefficient is negative but insignificant, suggesting potential reverse causality during acute stress (spreads widen first, then composition adjusts). The **staleness channel is robustly significant** across all subsamples ($t > 3$).

Reconciling across granularities. An apparent tension exists between high-frequency within-hour results and hourly results. At 5-minute granularity within hour 14, the selection coefficient is positive, exploiting high-frequency variation in informed share as different traders enter and exit during the outage. Table 33 uses hourly observations, collapsing within-hour variation into a single average. At this coarser granularity, hour 14 has both peak informed share *and* peak spreads, which can reverse sign when comparing across hours due to mean-reversion dynamics. The higher-frequency result is econometrically more informative because it exploits genuine variation in informed share during the event itself.

Finding. The selection channel operates most cleanly in adjacent hours, when quote staleness has partially recovered but composition effects persist. During acute stress, staleness dominates and the selection coefficient is noisy. This pattern is consistent with the lag structure evidence: selection effects require time to manifest, while staleness operates instantaneously.

7 Discussion

7.1 The Two-Margin Framework Beyond This Paper

The decomposition developed here—market quality as a function of participant composition and price-setting composition—has applications beyond infrastructure failures. Any shock that differentially affects trader types will shift one or both margins: regulatory changes, fee schedule revisions, latency improvements, or volatility spikes. The framework provides a template for decomposing aggregate quality changes into their compositional sources.

The toxicity differential—3.1 bps per trade ($t = 5.8$), robust under taker-level inference (19.2 bps, $t = 24.4$)—validates that participant composition matters; the fill-HHI paradox (falling concentration, rising spreads) validates that price-setting composition diverges from volume-based measures. Together, these establish that both margins are economically first-order and that standard aggregate metrics miss variation in each.

7.2 Generalizability to Traditional Markets

The two compositional margins operate wherever infrastructure access correlates with trader type—which is everywhere.

Participant composition margin in traditional markets. Co-location advantages create tiered infrastructure access: Shkilko and Sokolov [2020] document that milliseconds of latency disadvantage predict worse execution, implying that informed traders disproportionately invest in fast access. Exchange outages (NYSE July 2015, Tokyo October 2020)

produced spread widening; whether participant composition also shifted has been unmeasurable.

Price-setting composition margin in traditional markets. The Flash Crash of 2010 saw HFT market makers withdraw simultaneously [Kirilenko et al., 2017], consistent with price-setting composition collapse: firms most dependent on continuous quoting were most disrupted. Menkveld [2013] documents that a single HFT provided 40% of Dutch stock liquidity; our framework predicts that *that specific maker’s* withdrawal would matter more than aggregate concentration suggests.

Why both margins should generalize. The theoretical framework derives compositional shifts from operational capacity correlation with trader type. This holds whenever (i) informed takers value timely execution (they do—information decays), and (ii) price-setting makers depend on rapid order management (they do—stale quotes invite adverse selection). Neither condition is specific to on-chain markets.

What observable identity enables. Traditional TAQ or ITCH data provides order flow without counterparty identities; researchers infer selection from aggregate patterns. On-chain CLOBs record both sides of every trade, enabling direct measurement of both margins. The contribution is demonstrating what becomes testable when identity is observable.

7.3 Limitations

Single exchange. This paper examines Hyperliquid, one of several major perpetual futures venues. While the market structure (CLOB, maker-taker fees, millisecond latency) parallels traditional exchanges, idiosyncratic features of Hyperliquid’s blockchain or user base could affect results. Replication on other on-chain CLOBs (e.g., dYdX, Vertex) would strengthen external validity.

Event heterogeneity. The four infrastructure events span different types (API outage vs. congestion) and durations (4 minutes to 37 minutes). While all four show consistent effects, the decomposition analysis focuses on the July 29 outage. Whether the selection channel’s contribution varies with event type is an open question.

Sample period. The main results cover July–September 2025 (66 days). Classification stability tests show persistent rankings ($\rho > 0.7$ day-to-day, $\rho > 0.8$ across months), but market conditions during this period may not represent all regimes.

Traditional market testing. Direct testing in traditional markets requires trader identity data that is typically unavailable to researchers. Regulatory datasets (CAT in the US, transaction reporting in the EU) contain counterparty information but are not publicly accessible. The mechanisms I document may operate in those markets, but this paper cannot

test them directly.

8 Conclusion

Market quality depends on two compositional margins: *who trades* (participant composition) and *who sets prices* (price-setting composition). Both are invisible in standard market data. This paper uses on-chain data—where counterparty identity is observable—to measure both margins directly and show that both respond to infrastructure stress in ways that aggregate metrics miss.

The empirical results validate each component of the two-margin framework. First, participant composition matters: makers face 3.1 bps higher adverse selection per trade against informed than uninformed takers ($t = 5.8$), commensurate with observed spreads and robust under taker-level inference (19.2 bps, $t = 24.4$), establishing that counterparty identity determines adverse selection costs. Second, participant composition shifts under stress: an API outage moved the informed-to-uninformed ratio by 5.1%, and this selection channel predicts spread widening beyond quote staleness ($\beta = 0.07$, $t = 2.51$). Third, price-setting composition diverges from volume-based measures: during the outage, fill-based HHI fell 13% while spreads nearly doubled, because the few makers who set prices could not operate while marginal makers expanded.

The unified insight. Infrastructure access correlates with trader type on both sides of the market. On the demand side, informed takers invest more in operational capacity; on the supply side, price-setting makers depend on continuous API access. When infrastructure degrades, both margins shift: participant composition becomes more informed, and price-setting composition degrades. Standard analysis, lacking identity data, attributes all spread widening to quote staleness—missing the compositional channels entirely.

Policy implications. Three findings carry regulatory relevance. First, market stability depends on the robustness of price-setting makers specifically, not the number of liquidity providers—a consideration for market-making obligations and circuit breaker design. Second, infrastructure policies (API access, co-location, latency) affect market quality partly through compositional selection, not only through speed. Third, the transparency that enables this analysis exists because blockchain settlement records counterparty identities; as traditional markets consider post-trade transparency reforms, this paper demonstrates what becomes measurable when identity is observable.

The two-margin decomposition—participant composition and price-setting composition—provides a framework for analyzing market quality that standard aggregate measures cannot capture. Observable identity makes both margins measurable, opening new empirical ground

for market microstructure research.

A Supplementary Tables

Table 34: Counterparty-Specific Adverse Selection: Trade-Level Inference (Unclustered)

| Metric | Markout Horizon | | |
|--|-----------------|-----------------|-----------------|
| | 10 sec | 1 min | 5 min |
| Maker markout vs. <i>informed</i> (Q5) | +0.2 bps | +0.4 bps | −3.8 bps |
| Maker markout vs. <i>uninformed</i> (Q1) | +2.4 bps | +3.5 bps | +3.8 bps |
| Toxicity Differential | +2.2 bps | +3.1 bps | +7.6 bps |
| <i>t</i> -statistic (trade-level, unclustered) | 45.0 | 43.2 | 68.7 |

Out-of-sample: classification July 28, test July 29–30. N = 693K trades.
t-statistics treat each trade as independent; see Table 3 for clustered inference at the taker or taker×maker level. Classification uses mid-move (directional price prediction).

References

- Matteo Aquilina, Sean Foley, Leonardo Gambacorta, and William Krekel. Decentralised dealers? examining liquidity provision in decentralised exchanges. BIS Working Papers 1227, Bank for International Settlements, November 2024.
- Andrea Barbon and Angelo Ranaldo. On the quality of cryptocurrency markets: Centralized versus decentralized exchanges. *Journal of Financial Economics*, 157:103863, 2024.
- Bruno Biais, Pierre Hillion, and Chester Spatt. An empirical analysis of the limit order book and the order flow in the paris bourse. *Journal of Finance*, 50(5):1655–1689, 1995.
- Jonathan Brogaard, Terrence Hendershott, and Ryan Riordan. High-frequency trading and price discovery. *Review of Financial Studies*, 27(8):2267–2306, 2014.
- Agostino Capponi and Ruizhe Jia. The information content of blockchain data. *Management Science*, 70(9):5765–5787, 2024.
- David Easley, Nicholas M Kiefer, Maureen O’Hara, and Joseph B Paperman. Liquidity, information, and infrequently traded stocks. *Journal of Finance*, 51(4):1405–1436, 1996.
- Lawrence R Glosten and Paul R Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1):71–100, 1985.
- Joel Hasbrouck. Measuring the information content of stock trades. *Journal of Finance*, 46(1):179–207, 1991.
- Terrence Hendershott, Charles M Jones, and Albert J Menkveld. Does algorithmic trading improve liquidity? *Journal of Finance*, 66(1):1–33, 2011.
- Roger D Huang and Hans R Stoll. The components of the bid-ask spread: A general approach. *Review of Financial Studies*, 10(4):995–1034, 1997.
- Andrei Kirilenko, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun. The flash crash: High-frequency trading in an electronic market. *Journal of Finance*, 72(3):967–998, 2017.
- Albert S Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985.
- Alfred Lehar and Christine A Parlour. Decentralized exchanges. *Journal of Financial Economics*, 167:103944, 2024.

- Igor Makarov and Antoinette Schoar. Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*, 135(2):293–319, 2020.
- Albert J Menkveld. High frequency trading and the new market makers. *Journal of Financial Markets*, 16(4):712–740, 2013.
- Mitchell A Petersen. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, 22(1):435–480, 2009.
- Andriy Shkilko and Konstantin Sokolov. Every cloud has a silver lining: Fast trading, microwave connectivity, and trading costs. *Journal of Finance*, 75(6):2899–2927, 2020.