Retail Trading Network and Cross-Predictability in the US Equity Market

QIHONG RUAN*

ABSTRACT

I investigate the network structure of retail investor trading and its consequences for the U.S. equity market. Using Nasdaq's Retail Trading Activity Tracker (RTAT) from 2016 to 2024, I construct a directed network of stocks based on lead-lag correlations in retail trading volume. The methodology is first validated using the iShares Bitcoin Trust (IBIT) as a prototype. This approach successfully identifies economically meaningful peers, including crypto-related firms, growth tech stocks, large-cap ETFs, and inflation-hedging ETFs, whose metrics predict IBIT's market dynamics.

Applying this network framework to U.S. common stocks, I find that the average return of a stock's retail trading peers robustly predicts its future returns. I demonstrate that the effect operates through two primary channels: (1) **price discovery**, where retail networks aggregate dispersed information, evidenced by stronger effects when peer average ret is positive, the peer-own return gap is large, during "buy the dip" episodes, and in weaker institutional information environments; and (2) **liquidity provision**, where coordinated retail buying systematically provides depth to the market, particularly on the bid side, leading market makers to widen spreads to compensate for adverse selection risk. The return predictability is persistent and does not reverse, suggesting the permanent incorporation of new information rather than temporary price pressure.

*Ph.D., Cornell University, 2025. Email: qr33@cornell.edu. I am grateful to Maureen O'Hara for her valuable comments. All errors are my own.

Finally, I extend the analysis across the full CRSP universe, revealing that the peer effect is a pervasive phenomenon. It is statistically significant across various asset types—including common stocks, foreign stocks, REITs, ETPs, and American Depositary Receipts (ADRs)—and exchanges. The effect's magnitude varies logically across market segments, proving stronger for assets with higher information asymmetry, such as ADRs, and more pronounced on NASDAQ than on the NYSE. In contrast, volatility-linked products on the Cboe exchange do not exhibit significant cross-predictability.

I. Introduction

The past decade has seen a significant expansion in U.S. retail trading, characterized by rapid growth in brokerage accounts and assets under administration. Data from major retail brokers—Charles Schwab, Fidelity, Interactive Brokers, and Robinhood—show total accounts increasing from 30.1 million in 2016 to 104.2 million in 2024 (Figure 1), with combined assets under administration rising from 8.565 trillion to 25.9 trillion (Figure 2). This growth accelerated sharply between 2019 and 2021, exemplified by Robinhood's funded accounts, which surged 345% from 5.1 million to 22.7 million, before stabilizing at 25.2 million by 2024. The period also featured notable industry consolidation, including Charles Schwab's 2020 acquisition of TD Ameritrade, briefly positioning it as the largest broker by account count until Fidelity's organic growth reclaimed the lead with 39.2 million customers by 2024.[1]

[Figure 1 about here.]

[Figure 2 about here.]

Given the inherent interconnectedness of financial markets, as highlighted by Easley and Kleinberg (2010), retail trading is no exception to network phenomena. Evidence of such connections is prevalent, from brokerage features like Robinhood's "People also own" section (Figure 3) to discussions on platforms like Stocktwits and Reddit's WallStreetBets,

[1]See Appendix A for further discussion.

and extends to market episodes such as the GameStop short squeeze or the rise of AI-themed stocks. Such abundant examples of network connections among stocks naturally motivate an inquiry into the network structure of retail trading. This logically leads to several research questions: For a given security, which stocks are in its retail trading network? Which stocks are being followed by others in terms of retail trading activity? Further, does the network of retail trading provide economically valuable information for the focal stock?

[Figure 3 about here.]

Constructing a network of retail trading requires a broad measure of investor activity. Common proxies often capture specific segments, such as off-exchange trades identified by sub-penny price improvements (Boehmer, Jones, Zhang, and Zhang, 2021), activity on a single platform like Robintrack, or fractional share volume. To ensure a generalizable analysis, I use the Nasdaq Retail Trading Activity Tracker (RTAT). Nasdaq provides this daily data, where RTAT for a security is its retail dollar volume as a share of the total retail dollar volume across all securities:

$$RTAT_{i,t} = \frac{\text{Retail Volume}_{i,t}}{\sum_j \text{Retail Volume}_{j,t}}.$$

Developed by Nasdaq using a proprietary algorithm, the dataset represents approximately 40% of retail trading activity and $30 billion in daily volume, covering the period from January 2016 to December 2024.[2] To validate its use, I confirm that RTAT activity is highly correlated with other common retail trading metrics, including those from Boehmer, Jones, Zhang, and Zhang (2021), Robintrack data (Welch, 2022; Barber, Huang, Odean, and Schwarz, 2022; Eaton, Green, Roseman, and Wu, 2022), and data on fractional shares (Da, Fang, and Lin, 2024; Bartlett, McCrary, and O'Hara, 2023, 2024) and odd-lot trades. RTAT is also employed as a benchmark for representative retail trading in recent research (Even-Tov, Munevar, Kogan, and So, 2022; Kogan, Makarov, Niessner, and Schoar, 2024).

I construct the retail trading network by modeling the comovement in RTAT time

[2] I provide a more detailed description of the RTAT data in Section III.

series between stocks. A prerequisite for using correlation is stationarity. I test the daily RTAT series for over 10,000 tickers and find that 94% reject the null hypothesis of a unit root, which supports this approach.

I define a peer relationship based on two conditions. First, the lagged correlation in trading activity must be significant. I designate stock j as a peer of stock i if:

$$\text{Corr}(RTAT_{i,t}, RTAT_{j,t-1}) \geq 0.4.$$

This threshold is selected based on simulations to ensure a low false positive rate, meaning it is unlikely that a spurious relationship is incorrectly identified as a peer link. Second, to focus on economically important relationships, stock j must also rank among the top 20 such peers by average market capitalization.

The resulting network is well-suited for analysis. On average, each stock has about 13 peers, indicating the top-20 cutoff is not overly restrictive. Furthermore, the network structure provides the necessary variation for econometric analysis. The average overlap between the peer sets of any two securities is only 6%. This low overlap is crucial for identification; if all securities shared the same set of peers, any peer-based regressor would be constant across the sample, rendering its coefficient unidentifiable. The observed heterogeneity in peer groups ensures that this is not a concern.

This structural validation provides the foundation for my primary empirical analysis. I examine if the retail trading network method captures valuable economic linkage, using the recently launched Bitcoin ETF as a prototype for my peer network analyses. The reasons are as follows. First, using a specific prototype like the Bitcoin ETF allows me to see precisely which securities emerge as peers. This direct examination of the peer group is crucial for evaluating whether my correlation-based approach identifies valid economic linkages. I then test if peer metrics derived from this network can predict bitcoin market dynamics. Second, Bitcoin ETF lacks the corporate fundamentals that recent studies (e.g. Feng, Huo, Liu, Mao, and Xiang (2025)) in the cross-predictability literature build on. The retail trading footprint on the Bitcoin ETF, however, allows me to identify peers through correlations in trading activity, demonstrating the versatility of a market-based

approach. This focus on the linkage between an ETF and common stocks also differs from recent literature with a focus on common stocks. Third, since its launch on January 11, 2024, the Bitcoin ETF (IBIT) has grown its market cap to over $60 billion, making the sources of its demand a timely and relevant question.

I find that IBIT's retail trading activity follows that of securities in the cryptocurrency ecosystem—such as MicroStrategy, Coinbase, Robinhood, and the Fidelity Wise Origin Bitcoin Fund (FBTC)—confirming that my correlation-based method captures economic linkages. The peer group also contains growth technology firms like Palantir, a large-cap ETF (SCHX) which is a passive investment instrument, and Schwab U.S. TIPS ETF (SCHP). The strong correlation with SCHP provides evidence that U.S. retail investors use Bitcoin as an inflation hedge. From these results, Bitcoin ETF has been a composite of a cryptocurrency, growth tech stock, passive investment, and inflation hedge for retail investors in the US equity market. Furthermore, my regression analyses demonstrate that the market metrics of these peers have significant predictive power for IBIT's pricing and liquidity dynamics.

I then extend my analyses to the common stocks sample in the US equity market. The results show that a peer group's average return is positively and significantly associated with the focal stock's return on the next trading day. This holds after controlling for market-based variables, company fundamentals, market sentiments, and risks, and is robust in both fixed effects models and Fama-MacBeth regressions. The peer effect is also economically significant. For instance, in the fixed effects model with all controls, the coefficient of peer average return is 0.126. This translates to a 15.88% annualized return for two portfolios that have a 5% difference in peer average return, given the portfolios are otherwise equivalent.

Further, I investigate the economic mechanisms underlying the cross-predictability of returns within retail investor trading networks. Having established a robust, positive correlation between the average daily return of a stock's peers ($PeerAvgReturn_{i,t}$) and the stock's own future return ($Return_{i,t+1}$), I test two primary hypotheses for this phenomenon: **price discovery** and **liquidity provision**. I motivate both hypotheses

with the institutional feature that retail investors are persistent net buyers in U.S. equity markets (Mackintosh, 2021, 2025). I propose that peer networks aggregate and amplify this tendency, allowing retail investors to collectively perform fundamental market functions.

I find systematic evidence supporting a price discovery channel, where retail networks facilitate a "wisdom of crowds" effect (Welch, 2022). First, I show the predictive effect is asymmetric and concentrated on the buy-side; I find the total effect when peer returns are positive is 0.0244, nearly eight times the effect of 0.0031 when peer returns are negative. Second, I demonstrate the effect strengthens with the information gap between peers and the focal stock. Third, I find the peer effect is amplified when the focal stock experiences a large negative return, a "buy the dip" phenomenon, where I calculate the relevant interaction coefficient to be large and significant (0.6553, t=11.81). Fourth, I establish that the effect is stronger in weaker information environments, showing a negative interaction with institutional ownership and analyst coverage, which leads me to conclude that retail information is a substitute for traditional information sources. In further tests, I show that the peer effect is partially incorporated during the trading day, improves measures of price efficiency, and generates spillovers to options market volatility. Crucially, I find the return predictability does not reverse. The coefficient on $PeerAvgReturn_{i,t}$ remains positive and significant in my regressions for predicting cumulative returns up to 126 trading days, which I interpret as evidence for the permanent incorporation of new information. This is different from temporary price pressure from attention-induced trading in Barber, Huang, Odean, and Schwarz (2022) and pricing bubbles associated with fractional share trading in Da, Fang, and Lin (2024).

Complementing the price discovery mechanism, I also find strong evidence for a liquidity provision channel, consistent with findings that sophisticated retail investors can improve market quality (Eaton, Green, Roseman, and Wu, 2022). I show that an increase in peer returns predicts higher future trading activity ($\beta = 0.3446$ for log dollar volume) and a significant increase in market depth. I find this liquidity provision is asymmetric, with a substantially larger effect on bid depth than offer depth in my estimations (coefficients of 0.1170 vs. 0.0477, respectively), consistent with retail's net-buying behavior. I observe

that this informed liquidity flow leads market makers to widen spreads; in my analysis, both quoted and effective spreads increase significantly following high peer returns. As I predicted, these liquidity effects are stronger for less liquid stocks, which I measure using wider spreads, lower market capitalization, and lower depth. I also demonstrate that this liquidity provision is persistent, with a positive and significant effect on bid depth observed in my results at horizons up to 126 days. Collectively, I argue these findings provide a unified, microstructure-based explanation for the observed cross-predictability. Retail trading networks facilitate a process where investors are compensated for the dual roles of discovering information and providing systematic liquidity to the market.

Finally, to test the generality of my findings, I extend the analysis beyond common stocks. Motivated by the wide-reaching footprint of retail trading activity across asset types and exchanges, I examine the full cross-section of over 12 million observations across all securities in the CRSP database. The results show that the peer effect is not an isolated phenomenon but is pervasive across the market. The peer return coefficient is positive and significant across all asset types: common stocks (0.0121), foreign stocks (0.0178), REITs (0.0038), ETPs (0.0033), and is largest for American Depositary Receipts (ADRs) at 0.0212, suggesting the effect is stronger in assets with higher information asymmetry. Analysis by exchange reveals heterogeneity in market structure effects: the coefficient on NASDAQ (0.0153) is nearly double that on NYSE (0.0085), while NYSE Arca, which specializes in ETPs, shows a coefficient (0.0039) similar to that of ETPs as an asset class. Most importantly, this analysis provides a powerful placebo test. For volatility-linked products on the Cboe exchange (e.g., VXX, UVXY), whose returns are mechanically driven, the peer effect is economically and statistically zero (0.0001, t-stat=0.10). This null result provides strong evidence that the peer effects I document are a genuine network phenomenon, not a mechanical artifact.

II. Related Literature

My work makes several contributions to the literature across four areas. First, I introduce and validate a network-based approach to analyzing retail investor behavior. Second, I

establish this retail trading network as a novel and robust channel for the cross-predictability of asset returns and provide a market microstructure mechanism that explains this cross-predictability. Third, I show that peer average returns from the retail trading network serve as new predictors for market microstructure variables and establish a new stock-option cross-market impact channel. Fourth, I apply this methodology to the new asset class of Bitcoin ETFs, offering unique insights into their economic role.

A. Retail Trading and Network Structure

My first contribution is to the literature on retail trading. I depart from prior work that has primarily focused on how the volume of retail activity in a single security affects its own market dynamics—such as pricing (Barber, Huang, Odean, and Schwarz, 2022), bubble formation (Da, Fang, and Lin, 2024), volatility (Bartlett, McCrary, and O'Hara, 2024), or associated options markets (Bryzgalova, Pavlova, and Sikorskaya, 2023; Eaton, Green, Roseman, and Wu, 2024). Instead, I analyze the *network structure* that arises from correlated retail trading activity across different securities.

To do so, I conduct a comprehensive analysis of Nasdaq's Retail Trading Activity Tracker (RTAT) from 2016 to 2024. This official dataset is particularly relevant for studying the modern, post-Robinhood era and is increasingly used as a benchmark for retail activity (Even-Tov, Munevar, Kogan, and So, 2022; Kogan, Makarov, Niessner, and Schoar, 2024; Liang and Sun, 2024; Liang, Najand, Selover, and Sun, 2024). Its use is timely, as recent research highlights identification challenges with traditional retail trading proxies (Barber, Huang, Jorion, Odean, and Schwarz, 2024; Battalio, Jennings, Saglam, and Wu, 2022). On a technical level, I also contribute by refining the fractional share identification method of Bartlett, McCrary, and O'Hara (2023) and Bartlett, McCrary, and O'Hara (2024), improving its precision by lowering the latency bound from 20ms to 10ms. My central contribution, however, is to employ these data to uncover and empirically validate a network of retail trading co-movement and test its impact on cross-market predictability.

B. Cross-Predictability Through Retail Trading Links

My second contribution is to the asset pricing literature on cross-predictability. I identify and test a new channel for this phenomenon: economic linkages based on retail trading co-movement. This channel is distinct from previously documented linkages, such as those based on common industry (Moskowitz and Grinblatt, 1999; Cohen and Lou, 2012; Hoberg and Phillips, 2018), supply chains (Cohen and Frazzini, 2008; Menzly and Ozbas, 2010), international industry (Huang, 2015), institutional ownership (Anton and Polk, 2014; Ying, 2024), multinational connections (Finke and Weigert, 2017), analyst coverage (Ali and Hirshleifer, 2020), geography (Parsons, Sabbatucci, and Titman, 2020), firm characteristics (Müller, 2019), business descriptions (Breitung and Müller, 2025), bond rating comovements (Feng, Huo, Liu, Mao, and Xiang, 2025), or similar technologies (Lee, Sun, Wang, and Zhang, 2019). I find that the predictive power of peer groups identified through retail trading correlations is robust, holding after controlling for these known factors, firm characteristics, and fixed effects. This establishes retail co-movement as a novel information channel.

I find this phenomenon is pervasive across the U.S. market and not limited to common stocks. The peer effect is positive and significant across various asset classes, including foreign stocks, ETPs, and REITs. The effect is particularly strong for assets with higher information asymmetry, such as American Depositary Receipts (ADRs), and is more pronounced on the NASDAQ exchange compared to the NYSE. As a powerful placebo test, I show the peer effect is absent for mechanically-driven volatility products on the Cboe exchange, supporting the interpretation that my finding is a genuine network phenomenon, not a mechanical artifact.

Finally, I provide a market microstructure-based explanation for *why* this cross-predictability exists, a further contribution to this literature. I find the channel operates through retail investors' price discovery and liquidity provision, particularly on the buy side. Specifically, positive returns among peer stocks lead to increased net buying of the focal stock by retail investors. This targeted buying pressure enhances the focal stock's liquidity and contributes to price discovery, thereby transmitting the economic information

contained in the peer returns.

C. *Market Microstructure and Cross-Market Effects*

My work also contributes to the market microstructure literature by identifying a new set of predictors for future market quality. While Easley, López de Prado, O'Hara, and Zhang (2021) use machine learning to predict microstructure changes and Easley, O'Hara, Yang, and Zhang (2024) document cross-asset effects in specific markets like cryptocurrency, my analysis is distinct. I formally establish a new empirical regularity within the broad U.S. equity market: the average return of peer stocks, as defined by the retail trading network, can forecast a focal stock's future market microstructure variables. Specifically, I show that peer returns have predictive power for a stock's subsequent trading activity, volatility, liquidity, and information efficiency. This finding provides a new toolset for understanding and anticipating changes in market quality based on network dynamics, and creates a connection between stock and options markets through peer stock returns in the retail trading network.

D. *Cryptocurrency and Bitcoin ETFs*

This study contributes to the cryptocurrency literature, for which John, O'Hara, and Saleh (2022); John, Kogan, and Saleh (2023) provide comprehensive reviews. Specifically, my work adds to the emerging research on Bitcoin exchange-traded funds (ETFs) (Mazur and Polyzos, 2025; Hornback and Whaley, 2025; Krause, 2024). I introduce a framework grounded in retail investor preferences to understand the demand for Bitcoin ETFs within the broader U.S. equity market. By analyzing retail trading activity, I propose a market-based methodology for identifying the economic profile of a new asset. My analysis reveals that retail investors perceive the peers of the iShares Bitcoin Trust (IBIT) to be a portfolio of high-growth technology stocks, firms with direct crypto exposure (such as Coinbase and Robinhood), and broad market ETFs. This evidence-based approach complements the theoretical analyses of Budish (2025) and Cong, Li, and Wang (2021).

Notably, I document a significant correlation in retail trading between IBIT and the

Schwab U.S. TIPS ETF (SCHP). This finding provides direct, market-based evidence that retail investors may use Bitcoin as an inflation hedge. Furthermore, I demonstrate that the market metrics of these identified peers can predict the subsequent returns and market dynamics of the Bitcoin ETF. My approach is distinct from platform-specific analyses, such as Kogan, Makarov, Niessner, and Schoar (2024)'s study of eToro, where underlying stock and cryptocurrency markets remain segmented. By examining a Bitcoin ETF, which trades within traditional equity infrastructure, my analysis facilitates a direct comparison of Bitcoin with other securities under a unified market system.

III. Nasdaq's Retail Trading Activity Tracker (RTAT)

The starting point of my analyses is the **Nasdaq Retail Trading Activity Tracker (RTAT)** dataset, derived from the premium U.S. Retail Equity Flows (UREF) dataset.

While sub-decimal pricing can identify retail trades, Mackintosh (2021) notes this approach misses retail orders resting on exchanges and off-exchange orders executed at decimal prices. For their analysis, Mackintosh (2021) employs RTAT, which uses a proprietary algorithm applied to Securities Information Processor (SIP) end-of-day data to capture approximately 45% of U.S. retail order flow, tracking over $30 billion in daily volume across more than 9,500 securities.[3]

Retail trading volume increased from $18 billion daily pre-2020 to over $30 billion during the pandemic (Mackintosh, 2021). Retail investors exhibit persistent net buying behavior, contributing $100 billion annually to equity markets. Weekly net purchases rose from $1 billion in 2020 to $3 billion in 2021, continuing through 2025 despite market volatility (Mackintosh, 2021, 2025). Retail investors purchase ETFs consistently each month while alternating between buying and selling individual stocks, representing 4% of total market value and 6.4% of ETF trading volume by 2025 (Mackintosh, 2025).

[3]See https://data.nasdaq.com/databases/RTAT and https://data.nasdaq.com/databases/UREF for specifications. I reached out to a Nasdaq representative and learned that the algorithm effectively identifies retail flows using consolidated market feeds. In this study, I sign up for RTAT rather than UREF, as the latter requires premium institutional access.

I employ the RTAT 'activity' metric, defined as:

$$\text{Activity}_{i,t} = \frac{\text{Retail Volume}_{i,t}}{\sum_{j=1}^{N_t} \text{Retail Volume}_{j,t}},$$

where $\text{Retail Volume}_{i,t}$ is the dollar volume traded by retail investors in security i on day t, and N_t is the total number of securities on date t. This normalization controls for aggregate fluctuations in retail participation and captures the significance of the retail trading activity of a security in the market. Given a specific trading day, RTAT 'activity' contains full information of the relative intensity of retail trading activity between securities. Being detrended by the total market volume helps alleviate the concerns about non-stationarity when computing correlations between two RTAT activity time series. Therefore, the daily RTAT activity measure provides sufficient statistics for constructing retail trading networks based on retail trading activity correlations between securities.

The sample spans January 4, 2016, to December 31, 2024, encompassing 15,624 unique tickers. Retail trading activity is present in 89.84% of CRSP security-day observations, illustrating the pervasive nature of retail participation across the equity market.[4] This broad coverage, combined with the temporal and cross-sectional patterns illustrated in Figures 4 and 5, demonstrates that retail participation is a structural feature of modern financial markets, rather than a phenomenon confined to specific market conditions or asset classes.

[Figure 4 about here.]

[Figure 5 about here.]

Figure 4 reveals the stage-agnostic characteristic of retail trading, with investors maintaining consistent presence across common stocks,[5] ETPs, ADRs, foreign stocks, and

[4]Nasdaq documents data collection issues at https://static.quandl.com/RTAT/RTAT_missing_dates_and_tickers.csv, concentrated during January 3-10, 2020 and September 10, 2020. These missing observations constitute 0.067% of the full CRSP sample and are therefore inconsequential for my analyses.

[5]Following Bartlett, McCrary, and O'Hara (2024), I use CRSP share code 11 for common stocks throughout this analysis, as the traditional distinction between share codes 10 and 11 has been consolidated into 11 in current CRSP dataset.

REITs throughout various market cycles. This persistent multi-asset participation suggests that the mechanisms through which retail activity influences asset prices are continuously in effect, creating permanent channels for cross-asset spillovers and information transmission.

Figure 5 illustrates the interconnectedness of retail trading across different trading venues, with retail order flow simultaneously present on NYSE, NASDAQ, NYSE Arca, and Cboe. The gradual migration of activity from NYSE to NASDAQ over the sample period occurs while maintaining substantial volumes on all major exchanges, indicating that retail investors act as a conduit linking these market centers. Such parallel trading activities highlight how retail investors transmit information and liquidity shocks across seemingly disparate parts of the market infrastructure. The footprints of retail investors are, therefore, not isolated but form a web of economic links across assets and trading venues.

[Table 1 about here.]

A cross-sectional analysis reinforces this view of widespread retail engagement. Table I presents the average proportion of retail trading dollar volume distributed across major U.S. exchanges and security types. The data reveals that retail activity is highly concentrated in common stocks, which account for approximately 64.4% of the total retail dollar volume traded. This activity is primarily split between NASDAQ (36.9%) and NYSE (27.5%), establishing these traditional equity markets as central nodes in the retail trading network. Beyond common equities, Exchange Traded Products (ETPs) represent the second-largest category, comprising 23.9% of retail flow. Notably, retail ETP trading is predominantly concentrated on NYSE Arca (17.4%), indicating exchange specialization shapes retail trading patterns and creates natural bridges between market segments.

The broad distribution of retail trading across various exchanges and diverse security categories—including foreign stocks, REITs, and ADRs—underscores the far-reaching presence of individual investors in the U.S. equity market. Across all asset types, NASDAQ captures the largest share of retail activity (45.3%), followed by NYSE (36.1%), NYSE Arca (17.7%), and Cboe (0.9%). This multi-venue, multi-asset participation pattern has profound implications for market dynamics. When retail investors adjust their portfolios in

response to market events or sentiment shifts, their simultaneous presence across 18 distinct asset-exchange combinations enables rapid propagation of these adjustments throughout the market system. The $30 billion daily volume tracked by RTAT, representing 45% of total U.S. retail flow, constitutes an economically significant force capable of generating and transmitting comovement patterns across market segments.

These empirical patterns suggest that retail trading activity serves as a fundamental mechanism linking disparate parts of the financial market. The stage-agnostic and wide-reaching nature of retail participation creates a persistent infrastructure for cross-asset and cross-exchange information flows, potentially explaining observed correlations in returns and volatility across seemingly unrelated securities. Understanding these retail-driven linkages is essential for comprehending modern market microstructure and the channels through which shocks propagate across the financial system.

A. Comparing RTAT with Other Activity Measures

To understand how RTAT 'activity' relates to other measures of trading volume, I compare it against two benchmarks: the raw dollar volume share (daily ticker dollar volume divided by total market dollar volume on that date) and the RHDW dollar value share (derived from estimates of fractional shares trading from Robinhood and Drivewealth, available in my dataset from November 1, 2021, to December 2024). The total market dollar volume averages $436.4 billion daily throughout the sample period, while RHDW fractional share activity averages $29.8 million daily.

Table II presents the top 10 tickers based on their average market share across these three metrics.

[Table 2 about here.]

As shown in Table II, there is significant **market concentration**. The top 10 stocks in terms of RTAT account for 22.30% of all retail activity measured by RTAT. Similarly, the top 10 by dollar volume share make up 21.39% of the total market volume. Notably, the RHDW measure shows even higher concentration, with its top 10 tickers comprising 54.20% of its total value.

Several tickers, namely **TSLA, AAPL, NVDA, SPY, QQQ, AMZN, and META (formerly FB)**, feature prominently in the top 10 across all three measures, indicating their broad significance in both overall market activity and retail trading. However, some tickers are unique leaders in specific categories: **AMD and TQQQ** appear in the RTAT top 10 but not the others shown, **IWM** is unique to the Dollar Volume top 10, and **NFLX and VOO** are unique to the RHDW top 10.

The correlation between these measures provides further insight. The `rtat_activity` has a strong positive correlation with `dv_share` (0.773) and a moderately strong positive correlation with `rhdw_share` (0.619). The `dv_share` and `rhdw_share` also share a positive correlation of 0.631. These correlations suggest that while RTAT captures a distinct aspect of retail trading, it aligns considerably with broader market activity and other retail-focused measures.

B. Time Series Dynamics of RTAT Activity

To visualize the temporal behavior of RTAT, I plot the time series of RTAT 'activity' for a selection of prominent tickers. Figure 6 illustrates these dynamics.

[Figure 6 about here.]

Figure 6 reveals distinct patterns. In **Panel A (Technology Leaders)**, both TSLA and NVDA exhibit substantial RTAT activity, particularly surging from 2020 onwards, with NVDA showing a dramatic increase in retail attention more recently. **Panel B (Index ETFs)** shows that SPY and QQQ maintain consistent, albeit fluctuating, retail interest. Spikes often correspond to periods of heightened market volatility. **Panel C (Leveraged ETFs)** demonstrates that TQQQ (leveraged Nasdaq 100) and SOXL (leveraged semiconductor ETF) also attract significant retail trading, with their activity levels appearing to amplify broader market trends and sentiment shifts. Perhaps most strikingly, **Panel D (Meme Stocks)** captures the extraordinary episodes of retail focus on GME (GameStop) and AMC. AMC, for instance, reached an RTAT 'activity' of 0.2615 on June 2, 2021, meaning over 26% of all tracked retail dollar volume on that day was

concentrated in this single stock. This was the highest single-day RTAT 'activity' value observed for any stock in my dataset. For context, TSLA's peak RTAT was 0.2226 on January 27, 2023.

C. Comparison with Dollar Volume Share

While RTAT provides a direct measure of retail attention, comparing it with a security's share of the total market dollar volume offers a more nuanced perspective on trading behavior. A security's dollar volume share is defined as the total dollar value of trades in that security divided by the total dollar value of trades across all securities in the market. A divergence between these two metrics can reveal whether retail interest in a stock is proportional to its overall trading activity. This section explores these divergences across several asset classes.

The analysis reveals a set of distinct and informative patterns. For leading technology stocks, meme stocks, and leveraged Exchange Traded Funds (ETFs), RTAT consistently exceeds the dollar volume share, indicating a focused retail interest that is disproportionately high compared to their share of overall market activity. Conversely, for broad-based index ETFs and certain foreign listings, the dollar volume share can surpass RTAT.

[Figure 7 about here.]

For prominent technology leaders such as TSLA and NVDA, retail trade activity as measured by RTAT is consistently higher than their respective shares of total market dollar volume (Figure 7). This shows that while these are heavily traded stocks, they command an even greater share of attention from retail investors relative to their share of the overall market's dollar volume. This heightened retail focus may be attributable to their high media visibility and strong narrative appeal.

[Figure 8 about here.]

In contrast, for large, broad-market index ETFs like SPY and QQQ, the dollar volume share is substantially higher than RTAT (Figure 8). The divergence is particularly

pronounced for SPY. This pattern indicates that while these ETFs are a component of retail activity, their share of the total market's dollar volume is significantly larger. The trading of these instruments, when measured as a fraction of total market value, is greater than their share of tracked retail dollar volume.

[Figure 9 about here.]

Leveraged ETFs, such as TQQQ and SOXL, exhibit a similar pattern to technology leaders, where RTAT is consistently higher than dollar volume share (Figure 9). This finding highlights a revealed preference of retail investors for products offering magnified market exposure. The pronounced retail interest in these instruments aligns with observations by Nasdaq that retail investors show significant interest in Exchange Traded Products (ETPs) that provide sophisticated exposures alongside individual stocks.This is consistent with the observations in Mackintosh (2021). The higher RTAT suggests these leveraged funds are key instruments for retail traders implementing tactical, short-term market views.

[Figure 10 about here.]

The most striking divergence between the two metrics is observed in "meme stocks" like GME and AMC (Figure 10). For these securities, RTAT dramatically exceeds the dollar volume share, particularly during periods of intense social media activity. This demonstrates that the meme stock phenomenon was fundamentally a retail-driven event. The outsized RTAT values confirm that these stocks captured a vast proportion of the retail community's attention and capital, far exceeding their importance in the broader market as measured by their share of total dollar volume.

[Figure 11 about here.]

Finally, examining foreign companies listed in U.S. markets reveals that trading patterns are not monolithic. Figure 11 contrasts ASML, a Dutch firm traded via an American Depositary Receipt (ADR), with SHOP, a Canadian company with a direct U.S. listing. For ASML, the dollar volume share is higher than its RTAT, a pattern similar to that of broad index ETFs. In contrast, SHOP exhibits a higher RTAT relative to its dollar volume

share, a profile more aligned with U.S. domestic technology leaders. This suggests that a foreign-domiciled firm with a high-visibility, consumer-facing business model (SHOP) can attract a disproportionately large share of retail attention, while an industrial technology firm (ASML) may have a larger footprint in the overall market's trading volume than in the specific retail segment.

D. Statistical Properties of RTAT Time Series

To delve deeper into the characteristics of retail trading behavior, I analyze the statistical properties of approximately 10,000 RTAT 'activity' time series for tickers with sufficient observations, following standard practice in time series analyses as in (Tsay, 2013; Hamilton, 2020).

D.1. Autocorrelation

First, I examine the persistence of retail trading attention by calculating the first-order autocorrelation for each ticker's RTAT 'activity' series (for 10,046 tickers with at least 120 observations and non-zero variance). The **mean first-order autocorrelation is 0.338, and the median is 0.335**. A substantial 81.1% of these stocks display positive autocorrelation in their RTAT 'activity'. Furthermore, 54.2% show moderate to strong persistence (autocorrelation ≥ 0.3). This indicates that retail trading activity tends to have momentum; higher (or lower) than usual retail interest in a stock today is often followed by similarly higher (or lower) interest tomorrow.

D.2. Stationarity

Next, I assess the stationarity of these RTAT time series using Augmented Dickey-Fuller (ADF) tests. Stationarity is a crucial property for reliable correlation analysis. The results are encouraging: at the 10% significance level, **94.2% of the 10,046 tested series reject the null hypothesis of a unit root**, suggesting they are stationary. At the 1% significance level, 88.5% reject the unit root null.

For the small minority (less than 6%) of series where the unit root null hypothesis is not rejected, it is important to consider the inherent nature of the RTAT 'activity' metric. RTAT is a ratio representing a stock's share of total retail dollar volume on a given day. As such, it is naturally **bounded between 0 and 1**. This boundedness, combined with the fact that it is detrended by the daily total market retail volume, means its variance cannot grow infinitely, a key concern with non-stationary series. Therefore, even if some series exhibit trend-like behavior (perhaps reflecting evolving market interest or structural shifts), their correlations can still yield economically meaningful insights. For example, if a rapidly growing company consistently captures an increasing share of retail attention, its RTAT series might appear non-stationary by an ADF test but still accurately reflect a real economic trend. The co-movement of such a series with another that exhibits similar evolution in retail interest remains interpretable.

E. Constructing a Lead-Lag Retail Trading Network

Armed with largely stationary and economically interpretable RTAT time series, I proceed to investigate potential lead-lag relationships in retail trading activity across different stocks. I aim to identify "peer" stocks where high retail activity in one stock today is associated with high retail activity in another stock tomorrow.

Specifically, stock j is defined as a peer for stock i if the Pearson correlation coefficient between stock i's RTAT 'activity' on day t ($RTAT_{i,t}$) and stock j's RTAT 'activity' on day $t-1$ ($RTAT_{j,t-1}$) is **greater than 0.4**. Among such qualifying stocks, I select up to the **top 20 peers for stock i based on stock j's market capitalization**. This approach prioritizes influential (larger cap) lagged peers.

This analysis, conducted on 10,046 pre-filtered stocks, yielded over 75 million valid lead-lag correlation coefficients. The full distribution of these lead-lag correlations has a **mean of approximately 0.0068** and a median of -0.0059. While the average correlation is close to zero, the distribution has notable tails, and my focus is on the positive tail.

Using the 0.4 correlation threshold and the market cap filter, I successfully identified peer groups for **9,913 firms (98.7% of those analyzed)**. On average, firms in my

network have **13.4 peers**, with a median of 14 peers. The maximum number of peers is, by construction, capped at 20. These results validate my threshold choices. First, the near-complete coverage demonstrates that retail trading correlations are pervasive across the market, not limited to specific sectors or firm types. Second, the average firm having 13-14 peers—rather than hitting the 20-firm cap—indicates my thresholds strike an appropriate balance. They are strict enough to exclude weakly correlated firms but not so restrictive as to leave firms isolated. This produces economically meaningful peer groups of manageable size for the subsequent cross-predictability analysis.

E.1. Reciprocity and Network Asymmetry

The retail trading activity network exhibits strong asymmetric properties. From a universe of 99,650,306 potential directed relationships, I identify 133,011 edges where lagged retail trading activity in one stock predicts current activity in another, yielding a network density of 0.13%. The structure of these connections is predominantly unidirectional. I find that only 6,215 edges are reciprocated, corresponding to a reciprocity rate of just 4.7%. This indicates that activity flows, rather than conversations, characterize the network.

Initial evidence suggests this asymmetry is systematically related to firm size. A majority of influence flows from larger to smaller firms, with 61.7% of all directed edges originating from a larger market capitalization stock and pointing to a smaller one. The median market cap ratio is 2.28, indicating that the typical source stock is more than twice the size of its target.

Figure 12 provides visual evidence for the network construction process and the resulting peer group structure. The left panel shows that while most lead-lag correlations cluster near zero, a substantial right tail exists above the 0.4 threshold, justifying this cutoff for identifying meaningful relationships. The right panel reveals that the resulting peer groups are well-populated, with most firms having between 10 and 20 peers after applying the market capitalization filter.

[Figure 12 about here.]

To formally test whether this asymmetry reflects systematic differences in firm size, I proceed with a regression analysis.

Model Specification I model the probability of a directed edge from a source firm j to a target firm i using a Logit specification. The model is specified as:

$$P(\text{Edge}_{j\to i} = 1|\mathbf{X}_{ij}) = \Lambda\left(\beta_0 + \beta_1 \log\left(\frac{\text{MCap}_j}{\text{MCap}_i}\right) + \boldsymbol{\beta}'\mathbf{Z}_{ij} + \epsilon_{ij}\right),$$

where $\Lambda(\cdot)$ is the logistic function. The key explanatory variable, $\log(\text{MCap}_j/\text{MCap}_i)$, captures the relative market capitalization between the source and target. The vector $\mathbf{Z}_{ij}$ contains a set of control variables for both firms, including stock prices, trading volumes, volatilities, average retail trading activity (RTAT), and a binary indicator for same-industry pairs. A positive and significant β_1 would confirm that trading activity systematically flows from larger to smaller firms.

Regression Results Table III presents the Logit regression estimates with 10% random sample of all possible directed pairs. The coefficient for the log market capitalization ratio is positive and highly significant across all specifications, confirming the central hypothesis.

[Table 3 about here.]

In the full specification (Model 3), the coefficient of 0.0948 on the log market cap ratio is statistically significant at the 1% level. This result provides robust evidence that the probability of an activity link forming increases with the relative size of the source firm to the target. Additionally, same-industry pairs are substantially more likely to form edges (coefficient of 0.514), and high-volume stocks are more likely to both send and receive connections. Notably, average retail trading activity in the source firm strongly predicts edge formation (coefficient of 26.596), while the target's RTAT has no significant effect.

Economic Significance and Robustness To assess the economic magnitude, I calculate the marginal effect at the mean of the explanatory variables. The marginal effect of the log market cap ratio on the edge probability is 0.000125. This implies that a one-unit

increase in the log-ratio—equivalent to the source firm being approximately 2.7 times larger than the target—increases the probability of an edge forming by 0.0125 percentage points, representing a 9.6% increase relative to the baseline edge probability of 0.13%.

Table IV demonstrates the stability of this finding. I re-estimate the full model on random subsamples of the data, and the coefficient on the log market cap ratio remains positive, statistically significant, and stable in magnitude. This confirms that the asymmetric flow of activity from larger to smaller firms is a fundamental and robust feature of the network.

[Table 4 about here.]

These findings establish that retail trading attention flows systematically from larger to smaller stocks in a hierarchical pattern. Combined with the low reciprocity rate, this evidence supports the use of directed, lead-lag correlations in constructing peer networks for cross-predictability analysis. The asymmetric structure suggests that monitoring trading activity in large-capitalization stocks provides valuable signals about future activity in smaller, related stocks.

F. Justifying the Peer Selection Criteria

The chosen criteria for peer identification—a lead-lag correlation greater than 0.4 and selection of up to the top 20 peers by market capitalization—are deliberate choices aimed at ensuring both statistical robustness and economic relevance.

F.1. Statistical Significance of the 0.4 Correlation Threshold

A primary concern when identifying correlations is the risk of finding spurious relationships, especially with financial time series that can exhibit autocorrelation. To mitigate Type I errors—also known as false positives, where a relationship is concluded to exist when it does not—I use a threshold where the correlation coefficient, r, must be 0.4 or greater.

This 0.4 threshold is higher than the value required for statistical significance in large samples. The datasets I use contain 2,264 daily observations, where low correlation values,

such as 0.1, can be statistically significant. For comparison, with a sample size of 1,000, the critical value for significance at an α level of 0.001 is 0.104. The 0.4 threshold is therefore set to ensure that identified relationships are not only statistically significant but also economically meaningful in magnitude.

More importantly, I considered the impact of autocorrelation. The RTAT series in my sample have a mean first-order autocorrelation of approximately 0.338. Simulation studies demonstrate the robustness of the 0.4 threshold even in the presence of such persistence. For example, when simulating pairs of unrelated AR(1) time series, each with an autocorrelation coefficient of $\rho = 0.3$ (similar to the average observed RTAT autocorrelation) and a length of $N = 1000$ to $N = 2500$ observations (mirroring my data), the probability of observing an absolute sample correlation $|r| \geq 0.4$ purely by chance is effectively **zero (0.0000)**. Even if I assume a much higher underlying autocorrelation, say $\rho = 0.8$, the probability of a spurious correlation reaching $|r| \geq 0.4$ for $N = 120$ is only 0.0268, and this probability also drops to 0.0000 for $N \geq 500$. Given that the RTAT series in my study has 2264 days, much longer than 500 observations, the 0.4 threshold ensures a very low false positive rate and helps to filter out noise from genuine lead-lag signals.

F.2. Rationale for Top 20 Market Capitalization Peers

The selection of peers based on market capitalization (up to the top 20) serves two main purposes: ensuring **importance** and facilitating **identification** in subsequent econometric models.

1. **Importance**: For a peer's trading activity to plausibly influence another stock's retail interest, the peer itself should ideally be significant or visible. Market capitalization is a common proxy for a firm's size, visibility, and potential market impact. Filtering by market cap helps to ensure that the identified lagged peers are substantial enough to be considered influential.

2. **Uniqueness and Identification**: If every stock had an identical, very large set of peers, it would be difficult to disentangle specific peer effects from common market-wide movements. Limiting the number of peers (to a maximum of 20) helps

in creating more distinct peer groups for each focal stock, which is crucial for the econometric identification of peer effects. If peer groups overlap too much, the peer influence variable becomes too similar across observations, akin to a common factor, making it hard to estimate its coefficient precisely.

My choice of capping at 20 peers is empirically supported. The average number of peers identified per firm using the 0.4 threshold is 13.4. Allowing up to 20 peers ensures I capture most of these significant relationships, as it sits at the right tail of the peer group size distribution (as seen in the right panel of Figure 12). This approach avoids being overly restrictive while still promoting peer group diversity. Indeed, the analysis of peer group overlap (using Jaccard similarity) shows a **mean overlap of only 0.068 (6.8%)**. This low overlap indicates that the resulting peer portfolios are largely unique to each focal stock. This results in approximately 9,243 “effective" unique peer groups from the 9,913 focal stocks, which provides sufficient cross-subsectional variation for robustly identifying peer effects in econometric models.

IV. IBIT: Market Dynamics and Network Analysis

This section provides a case-study validation of the retail trading network method that is central to my main empirical analysis. I use the iShares Bitcoin Trust (IBIT), a spot Bitcoin ETF launched on January 11, 2024, as a prototype to test if my method captures valuable economic linkages.

The rationale for selecting IBIT is threefold.

First, the analysis of a specific prototype is essential for validation. It allows for a transparent assessment of the peer group’s composition, which is a critical test of whether the method identifies sensible economic connections. After constructing the network, I evaluate its practical utility by testing if peer-derived metrics can predict IBIT’s market dynamics.

Second, my market-based approach is distinct from and, in certain contexts, more versatile than methods that rely on corporate fundamentals (e.g., Feng, Huo, Liu, Mao,

and Xiang (2025)). By using correlations in retail trading, my method can construct a network for an asset like IBIT whose valuation is not based on traditional fundamentals. This focus on an ETF-to-common-stock linkage further differentiates my approach from the common stock-to-stock networks studied in prior literature.

Third, the analysis is economically relevant. The immense growth and significant market capitalization achieved by IBIT since its recent inception make understanding its investor base and price dynamics a timely and pertinent research question.

My analysis unfolds by first examining IBIT's market capitalization and trading activity, drawing comparisons with MicroStrategy (MSTR). I then detail the peer selection methodology and describe the characteristics of the resulting network.

A. IBIT's Market Capitalization and Price Dynamics

Since its inception, IBIT has exhibited extraordinary growth in market capitalization. Figure 13 plots the market capitalization of IBIT and MSTR, along with IBIT's price, from IBIT's launch through December 31, 2024. IBIT's market capitalization surged from approximately $11 million at inception to $52,171 million by the end of the year, representing a growth of over 489,000%. For comparison, MSTR's market capitalization grew from $7,991 million to $60,935 million over the same period relative to IBIT's inception, a 663% increase. The right axis of Figure 13 displays IBIT's price, which, as expected for a spot Bitcoin ETF, closely tracks the price movements of Bitcoin. The comovement observed between IBIT's price and MSTR's market capitalization underscores the strong influence of Bitcoin's valuation on both assets.

[Figure 13 about here.]

B. IBIT's Retail Trading Activity

Figure 14 displays the daily RTAT for IBIT and MSTR. Both series exhibit considerable fluctuations, with notable spikes in activity that often coincide. The contemporaneous correlation between IBIT's RTAT and MSTR's RTAT over the sample period is 0.750,

indicating a strong similarity in retail trading patterns for these two Bitcoin-related assets. This high correlation confirms the effectiveness of RTAT comovement to capture economic links between securities and justify the approach of using RTAT activity correlations as a primary criterion for identifying IBIT's peers.

[Figure 14 about here.]

C. Peer Selection Methodology

I employ a two-stage process to select a relevant peer group for IBIT from a universe of 10,630 unique tickers with available data. The selection aims to identify firms whose retail trading dynamics might lead or influence IBIT's.

Stage 1: RTAT Correlation Filtering. The first stage involves calculating the correlation between IBIT's current-day RTAT ($RTAT_{IBIT,t}$) and each potential peer j's previous-day RTAT ($RTAT_{j,t-1}$). This lagged correlation captures potential leading behavior from the peers. A minimum of 20 paired daily observations is required for a correlation to be considered. From the 10,440 tickers with RTAT data, I identify peers with a correlation coefficient greater than 0.4. Figure 15 illustrates this filtering process. The left panel shows the distribution of these lagged correlations for all potential peers; the red dashed line indicates the 0.4 threshold, which clearly isolates a group of tickers with stronger leading RTAT relationships. The right panel shows these correlations against the number of paired observations used in their calculation, with the green shaded area representing the qualified region (correlation > 0.4 and observations ≥ 20). This stage yielded 81 qualified peers.

[Figure 15 about here.]

Stage 2: Market Capitalization Ranking. From the 81 qualified peers identified in Stage 1, I select the top 20 based on their average market capitalization over the sample period.[6] Figure 16 visualizes these 81 qualified peers (light blue circles) by plotting their

[6]Top 20 market cap is an ex-ante design based on importance and uniqueness of the peer group. Recent study Feng, Huo, Liu, Mao, and Xiang (2025) also use market cap as a criteria to decide peer

average market capitalization (log scale) against their RTAT correlation with IBIT. The final 20 selected peers are highlighted as dark green circles. This approach ensures that the selected peers are not only correlated in terms of retail activity but also represent firms with substantial market presence.

[Figure 16 about here.]

Table V lists the final 20 selected peers, along with their lagged RTAT correlation with IBIT and their average market capitalization. The selected peers exhibit an average correlation of 0.472, ranging from 0.401 to 0.676. Their average market capitalization is approximately $24.6 billion.

[Table 5 about here.]

D. Selected Peer Network Composition

The composition of the 20 selected peers, visualized in the network graph in Figure 17, is diverse. It includes firms directly involved in the cryptocurrency ecosystem, such as MicroStrategy (MSTR, correlation 0.676), Coinbase (COIN, correlation 0.525), and Robinhood (HOOD, correlation 0.450). Another spot Bitcoin ETF, FBTC (correlation 0.436), is also present. The peer group also contains large-cap ETFs like SCHX (Schwab U.S. Large-Cap ETF, correlation 0.416) and SCHP (Schwab U.S. TIPS ETF, correlation 0.532). The inclusion of SCHP is particularly noteworthy, as TIPS are often considered an inflation hedge; its strong RTAT correlation with IBIT suggests that retail investors view Bitcoin, and by extension IBIT, through a similar inflation-hedging lens. Technology firms like Palantir (PLTR, correlation 0.401) and Monolithic Power Systems (MPWR, correlation 0.434), along with companies from other sectors such as Dow Inc. (DOW, correlation 0.491), are also part of this network. The overall network density of this peer group is 0.490, with an average degree of 19.62, indicating a relatively interconnected group. IBIT itself has 13 outgoing RTAT-based connections (implying its RTAT is led

group. Average market cap is also ex-ante design to aggregate information during the period. The results hold when using initial market cap, or entirely removing the market cap filter.

by these peers) and is connected to all 20 selected peers when considering bidirectional influence or general comovement.

[Figure 17 about here.]

This diverse yet connected peer group, identified through retail trading activity patterns and market significance, forms the basis for my subsequent analysis to investigate the flow of predictive information within this network concerning IBIT's market dynamics.

E. Peer Network Effects on IBIT Market Dynamics

In this section, I examine the extent to which the market dynamics of IBIT can be predicted by the contemporaneous dynamics of its peer network. I estimate a series of OLS regressions where various market quality and return metrics for IBIT at time $t+1$ are regressed on peer group average metrics at time t, controlling for IBIT's own market conditions at time t.

The general specification for the regressions is:

$$IBIT_Metric_{t+1} = \alpha + \beta \cdot Peer_Metric_t + \sum_{k=1}^{K} \delta_k \cdot Control_{k,t} + \epsilon_{t+1},$$

where $IBIT_Metric_{t+1}$ is the dependent variable for IBIT on the next trading day. $Peer_Metric_t$ is one of the four aggregate metrics derived from the peer group on the current trading day: $Peer_Avg_Return$, $Peer_Avg_Return^2$ (the average of squared returns across peers), $Peer_Avg_Dollar_Volume$ (log-transformed), or $Peer_Avg_RTAT_Activity$. The vector $Controls_t$ includes IBIT's own lagged characteristics: its return on day t ($Return_t$), the log of its price on day t ($Log(Price)_t$), the log of its trading volume on day t ($Log(Volume)_t$), and its cumulative return over the prior five trading days ($Return_{[t-5,t-1]}$). $Return_t$ controls for the reversal/momentum effect of the last day; $Return_{[t-5,t-1]}$ stands for the reversal/momentum effect of past-week return on future returns; $Log(Price)_t$ is a known factor affecting retail trades and hence return dynamics; $Log(Volume)_t$ is related to liquidity, trading activity, and asset returns. These control variables are commonly seen in literature (Ruan, 2019; Eaton, Green, Roseman, and Wu, 2022; Barber, Huang,

Odean, and Schwarz, 2022; Da, Fang, and Lin, 2024) in return-related studies. Market capitalization is not included as a control due to IBIT's outstanding shares changing significantly with fund flows, making its market capitalization dynamics distinct from typical equities. All regressions use data spanning 243 trading days.

E.1. Return Predictability

Table VI presents the results for predicting IBIT's next-day returns, both close-to-close ($Return_{t+1}$) and open-to-close ($Open_to_Close_Return_{t+1}$).

I find that peer metrics exhibit significant predictive power. For instance, a one-unit increase in the logarithm of $Peer_Avg_Dollar_Volume$ is associated with a 0.0230 increase in IBIT's next-day return (Column 3), an effect that is statistically significant at the 1% level ($p < 0.01$). This suggests that heightened trading activity within the peer group positively predicts IBIT's subsequent returns. Similarly, $Peer_Avg_Return^2$ (Column 2) also significantly predicts IBIT's next-day return, with a coefficient of 4.1298 ($p < 0.05$).

Focusing on intraday dynamics, Column (4) shows that $Peer_Avg_RTAT_Activity$ is a strong predictor of IBIT's next-day $Open_to_Close_Return$. The coefficient of 7.7399 is highly significant ($p < 0.01$), implying that increased retail trading attention within the peer group on day t foreshadows a higher open-to-close return for IBIT on day $t + 1$. These results collectively support the hypothesis that information or sentiment conveyed through peer network activity can anticipate IBIT's short-term price movements. The R^2 values for these return prediction models range from 0.0542 to 0.0763.

[Table 6 about here.]

E.2. Volatility, Trading Activity, and Market Quality Predictability

Table VII extends my analysis to other dimensions of IBIT's market dynamics: volatility, trading activity, and market depth.

I find that peer metrics demonstrate notable predictive capabilities in these areas as well. Column (5) shows that $Peer_Avg_Return^2$ significantly predicts IBIT's next-day

$Return^2$ (a proxy for volatility), with a coefficient of 0.4449 ($p < 0.01$). This suggests a spillover of volatility from the peer group to IBIT.

A more direct measure of intraday price dispersion, $Log(Intraday_Volatility_Q)_{t+1}$, is examined in Column (6). $Peer_Avg_RTAT_Activity$ has a substantial and highly significant positive coefficient of 143.1771 ($p < 0.01$), indicating that higher retail trading activity in the peer network is associated with a sharp increase in IBIT's intraday volatility on the following day.

Column (7) focuses on $Log(Dollar_Volume)_{t+1}$ as a measure of trading activity. $Peer_Avg_Dollar_Volume$ (log-transformed) significantly predicts IBIT's next-day $Log(Dollar_Volume)$ with a coefficient of 0.1935 ($p < 0.05$). This implies that trading activity levels tend to comove between IBIT and its peers. The R^2 for this model is notably high at 0.7107, suggesting a strong overall fit primarily driven by the control variables related to IBIT's own lagged price and volume, but the peer metric still adds significant explanatory power.

Finally, Column (8) investigates the predictability of IBIT's market depth, using $Log(Depth)_{t+1}$ as the dependent variable. $Peer_Avg_RTAT_Activity$ is found to have a significant negative impact on next-day depth, with a coefficient of -88.4194 ($p < 0.05$). This suggests that heightened retail activity in the peer group may lead to a reduction in quoted depth for IBIT, potentially reflecting increased uncertainty or market maker reticence.

In summary, my results presented in Tables VI and VII consistently show that various peer metrics provide statistically significant predictive information for IBIT's future returns, volatility, trading volume, and market depth, even after controlling for IBIT's own lagged market conditions. This underscores the importance of considering peer network dynamics in understanding and potentially forecasting the behavior of individual assets like IBIT.

[Table 7 about here.]

V. Data

Building on the prototype retail trading network analyses of Bitcoin ETF, I then extend my analyses to common stocks in the US equity market. I build a comprehensive dataset that integrates twelve distinct data sources. My research sample spans January 2016 through December 2024, encompassing 18,724,340 security-day observations for 14,861 unique securities across 2,264 trading days. This period captures the transformation of retail trading through commission-free platforms, the rise of social media coordination, and several episodes of extreme retail-driven price movements.

A. NASDAQ Retail Trading Activity Tracker (RTAT)

The NASDAQ Retail Trading Activity Tracker serves as my primary measure of retail trading intensity. RTAT is the daily market share of the retail trading volume of a ticker. The data covers the full sample period from January 2016 through December 2024, with 89.8% coverage (16,822,630 security-days) across 14,745 unique securities. The RTAT indicator captures the strength of daily retail trading activity of each ticker detrened by total retail trading volume in the market. It aligns with my research goal to study correlations of retail trading activity between stocks and perform peer network analyses based on that.

B. CRSP Daily Stock Data

The Center for Research in Security Prices (CRSP) provides the foundational market data for the entire sample period (January 2016-December 2024), including daily prices, returns, trading volumes, and shares outstanding for all common stocks (share codes 10 and 11) traded on NYSE, AMEX, and NASDAQ (exchange codes 1, 2, and 3). I calculate market capitalization as price times shares outstanding, dollar volume as price times share volume.

C. *TAQ Intraday Indicators*

Intraday Indicators aggregates millisecond-level NYSE Trade and Quote (TAQ) data to provide essential microstructure measures for the full sample period. This database achieves a 99.1% match rate with CRSP observations, covering 14,853 unique securities. Key measures include time-weighted quoted spreads, volume-weighted effective spreads, intraday volatility based on quotes, price impact measures, and variance ratios for measuring price efficiency. In addition, WRDS provides retail trading volume classification following Boehmer, Jones, Zhang, and Zhang (2021), which classifies retail trades based on sub-penny price improvement typically offered by wholesale market makers. Intraday Indicators cover the whole sample from January 2016 to December 2024.

D. *Robinhood and DriveWealth Trading Data*

Platform-specific retail trading data comes from Robinhood and DriveWealth (RHDW) for the period November 1, 2021, through December 31, 2024. I estimate RHDW trading activity using the latency-based methodology of Bartlett, McCrary, and O'Hara (2023) and Bartlett, McCrary, and O'Hara (2024), identifying retail trades by counting one-share off-exchange trades from Trade Reporting Facilities with according latency filters based on TAQ data from NYSE. The methodology exploits the characteristic execution latency of retail orders routed through wholesale market makers. I refine the algorithm to account for a latency distribution shift observed in mid-2024, reducing the lower bound threshold from 20ms to 10ms. This adjustment maintains identification accuracy, with the estimated RHDW trades achieving 99% R^2 when validated against FINRA OTC Transparency data. The final dataset covers 5,964,279 security-days (79.1% of observations during the restricted period) across 11,903 unique securities. This data is available from November 1, 2021 to December 31, 2024. I provide the details of RHDW fractional shares estimation in Appendix C.

E. Off-Exchange One-share Trades

I compute the daily off-exchange one-share trades data from NYSE TAQ data for the whole sample period from January 2016 to December 2024. This indicator is the main variable in Da, Fang, and Lin (2024) to measure fractional trades. My estimates of off-exchange one-share trades cover 87.3% of full CRSP observations (16,338,740 security-days) across 14,735 securities. This data is particularly relevant for retail trading analysis as wholesale market makers internalize the vast majority of retail marketable orders. Single-share trades executed off-exchange serve as a strong proxy for retail activity, as these odd-lot transactions are predominantly originated by individual investors.

F. Robintrack

I obtain Robintrack data from Welch (2022) replication files with the number of Robinhood users holding each stock, covering period from May 2018 through August 2020, when Robinhood discontinued the API that enabled this tracking. The dataset has 1,732,658 security-days observations, representing 9.3% of total observations in my research sample.

G. RavenPack News Sentiment

To incorporate news sentiment into my analysis, I utilize data from RavenPack Analytics (RPA), a prominent provider of structured news analytics derived from a vast array of global news sources. RavenPack processes content from over 40,000 news sources, including premium newswires (such as Dow Jones Newswires), press releases, and numerous web publications. For this study, covering the period 2016-2024, I draw upon data from three RavenPack editions: the Dow Jones & PR Edition, which contains premium newswire and press release content; the Web Edition, comprising online news publications; and the Full Edition, which integrates all available sources. My analysis is based on a total of 1,060,702,146 news records processed from the Full Edition. Within this, 100,127,414 records originate from the Dow Jones Newswire and Press Release Edition, and the remaining 960,574,732 records are sourced from the Web Edition and other content

integrated into the Full Edition.

My study focuses on two key sentiment variables provided by RavenPack, generated through proprietary machine learning algorithms:

- **Event Sentiment Score (ESS)**: This score quantifies the directional sentiment of specific news events concerning an entity, ranging from -1.00 (most negative) to +1.00 (most positive). The algorithm is trained on over 6,700 categorized financial, economic, and geopolitical events, with scores calibrated using expert consensus on expected market impact. Positive ESS values indicate bullish sentiment, while negative values suggest bearish sentiment.

- **Composite Sentiment Score (CSS)**: This score also ranges from -1.00 to +1.00 and aims to provide a more robust overall sentiment signal for a news story. It is derived by combining five specialized RavenPack sentiment classifiers (PEQ, BEE, BMQ, BAM, and BCA) through an ensemble approach. The CSS was specifically trained using intraday price movements of large-cap stocks to capture market-validated sentiment strength.

To transform the high-frequency, granular news data from RavenPack into entity-day level analytics suitable for my study of daily retail trading dynamics, I implemented a systematic filtering and aggregation pipeline. The following criteria were applied to the news records:

1. Entity Relevance ≥ 75: This retains news items where the company in question is prominently mentioned.

2. Event Relevance ≥ 90: This focuses the analysis on primary news events within a story.

3. Event Similarity Days ≥ 1: This filter is designed to remove exact same-day duplicate news events for the same entity and event type, while still preserving information from recurring corporate events (e.g., distinct announcements on consecutive days or similar events occurring with at least a one-day gap). This 1-day threshold aligns

with the daily frequency of retail trading decisions, as retail investors often react to news as it appears, and this approach allows for the inclusion of important recurring events like quarterly earnings or analyst updates that might elicit different market reactions based on prevailing conditions.

4. Exclusion of Neutral Sentiment: Records with an Event Sentiment Score (ESS) of exactly 0 are excluded to concentrate the analysis on news items with explicitly directional sentiment signals.

Following these filtering steps, the aggregation process computed daily statistics for each entity. For this study, the final security-day `ess_full` and `css_full` metrics used in regressions are the simple average of the sentiment scores from the filtered news events for stock i on day t. Event counts were also computed to measure daily news intensity for each entity. This methodological approach ensures that news meeting the relevance and novelty criteria contributes to the daily sentiment measures.

The coverage of these RavenPack editions varies across securities, reflecting differences in media attention, with larger and more visible firms typically receiving more frequent news coverage. For instance, in my dataset, the Dow Jones Newswire public edition covers approximately 7.1% of security-days across 7,832 securities; the Web edition covers 14.5% of security-days across 7,940 securities; and the Full edition provides the broadest coverage at 17.7% of security-days across 7,981 securities.

H. Reddit WallStreetBets (WSB) Mentions

I acquire the mentions data of Reddit r/WallStreetBets (WSB) from Quiver Quantitative, covering period from December 1, 2021 to December 31, 2024. The data captures 229,904 security-day mentions across 5,614 unique securities, representing 1.2% coverage of my full research sample. The WSB mentions exhibit extreme concentration—the top 10 most-discussed stocks account for 48.1% of all mentions. Beyond raw mention counts, the data includes sentiment scores and frequencies of retail trading terminology (e.g., "YOLO," "diamond hands," "to the moon"), capturing the cultural dimensions of coordinated retail

trading.

I. OptionMetrics

Options market data from OptionMetrics spans the full sample period and includes at-the-money implied volatilities for call options (19.8% coverage) and puts (19.1% coverage), historical volatility (20.2% coverage), and volume-weighted implied volatility of calls (33.1% coverage), which is employed in Bartlett, McCrary, and O'Hara (2024). The data covers 7,568 unique securities with listed options and sufficient trading activity for reliable estimates. The data is available from January 4, 2016 to August 31, 2023, the latest date that OptionMetrics is available. In Section B, I provide details about the configuration for computing the option volatility metrics.

J. FINRA Short Volume

I collect daily short sale volume data directly from FINRA's publicly available Short Sale Volume Data files for the entire sample period, aggregating daily reports across all trading venues.[7] This data covers 93.3% of observations across 14,832 securities. The short volume ratio, calculated as short volume divided by total volume, helps identify stocks under short-selling pressure that may experience squeeze events popular among retail traders. I use "Consolidated TRF/ADF Daily Short Sale Volume Files," which combines the volume for trades in exchange-listed securities reported to the TRFs (FINRA/Nasdaq Chicago, FINRA/Nasdaq Carteret, FINRA/NYSE) and the ADF. I use the consolidated files on and after August 1st, 2018. Before that, I aggregate the data on my own. FINRA daily short volume is available for my whole sample period from January 4, 2016 to December 31, 2024.

[7]Data downloaded from https://www.finra.org/finra-data/browse-catalog/short-sale-volume-data, with daily files concatenated to create the comprehensive dataset.

K. Fama-French Factor Loadings

I estimate time-varying factor loadings throughout the sample period using rolling 252-day windows with the Fama-French five-factor model plus momentum using data from WRDS. The estimations cover 97.6% of observations across 14,594 securities, providing daily measures of market beta, size (SMB), value (HML), momentum (UMD), and profitability (RMW) loadings, along with idiosyncratic volatility and model R-squared values. This data is available for my full sample period from January 4, 2016 to December 31, 2024.

L. Compustat/Thomson Reuters/IBES Fundamentals

Quarterly accounting data from Compustat, available throughout the sample period, provides fundamental characteristics for 43.8% of observations across 5,889 securities. Using CRSP-Compustat Merged (CCM) linkages, I match each security-day to the most recent quarterly report. I calculate return on assets (ROA) and book-to-market ratio (BM) for my analyses. Thomson Reuters 13F filings provide institutional ownership percentages, while I/B/E/S supplies the number of analysts and earnings forecast dispersion. This data is available for my full sample period from January 4, 2016 to December 31, 2024.

M. Summary Statistics: Full Sample

Table VIII presents summary statistics for the main variables in our analysis. The sample covers 18,724,340 security-day observations from January 2016 through December 2024.

[Table 8 about here.]

Several patterns emerge from the summary statistics. Retail trading activity displays substantial right skewness across all measures. The RTAT activity measure has a median value of zero but reaches 0.0018 at the 99th percentile, indicating concentrated bursts of retail interest in specific stocks. The BJZZ retail volume measure shows similar concentration, with a median of $0.17 million compared to $57.57 million at the 99th percentile. This skewness characterizes all retail trading proxies in our sample, from

platform-specific measures (RHDW, Robintrack) to market-wide indicators (Off-Exchange 1-Share).

The cross-sectional variation in market microstructure variables reveals heterogeneous trading conditions. Quoted spreads range from 3 basis points at the 5th percentile to 731 basis points at the 99th percentile. Effective spreads exhibit similar dispersion, though at lower levels, ranging from 2 to 485 basis points across the same percentiles. The variance ratio, which measures deviations from random walk behavior, averages 1.03 with substantial variation around unity.

Our sample spans stocks with diverse characteristics. Market capitalization ranges from $6.89 million at the 5th percentile to $91.89 billion at the 99th percentile. Institutional ownership averages 60.15% but varies from 3.96% to 100% across the distribution. The number of analysts covering each stock ranges from 1 to 31, with a median of 5 analysts. This heterogeneity allows us to examine how firm characteristics influence retail trading patterns and their propagation through the market.

The retail trading proxies capture different aspects of retail investor behavior. RTAT provides the broadest coverage with 16.8 million observations, while platform-specific measures offer more granular data during critical periods. RHDW data covers approximately 6 million observations, while Robintrack data is available for 1.7 million observations during the period when the platform operated. The WSB mentions variable, capturing social media attention, is available for approximately 230,000 observations where stocks received at least one mention on the WallStreetBets subreddit.

N. Retail Trading Proxy Correlations

To validate the retail trading measures used in my analysis, I examine the correlations among several established proxies. The sample period is from January 4, 2016, to December 31, 2024, although the availability of each specific measure varies. Table IX presents the correlation matrix for the unscaled values of these proxies.

All measures exhibit positive and statistically significant correlations, indicating that they capture related dimensions of retail trading activity. My primary proxy, the NASDAQ

Retail Trading Activity Tracker (RTAT), shows a correlation of 0.880 with the BJZZ retail volume measure from Boehmer, Jones, Zhang, and Zhang (2021). While the two measures are highly correlated, RTAT's proprietary Nasdaq algorithm may capture a different or broader set of retail trades compared to the tick-test methodology of BJZZ.[8] RTAT also has high correlations with other microstructure-based measures, including off-exchange single-share trades (0.696) and odd-lot volume (0.651). Platform-specific measures, such as RHDW volume and Robintrack holdings, show more moderate correlations (0.596 and 0.473, respectively), consistent with their narrower platform-specific coverage.

[Table 9 about here.]

Table X normalizes each measure by its daily market-wide total. After this adjustment, the correlation between RTAT and the off-exchange proxy increases from 0.696 to 0.783, while the correlation with BJZZ remains stable at 0.883. This suggests that microstructure-based proxies align more closely with RTAT after accounting for aggregate market activity.

[Table 10 about here.]

I next examine RTAT's correlation with market characteristics. The correlations, shown in Table XI, reveal three main empirical patterns. First, retail trading activity is concentrated in liquid, visible stocks. RTAT correlates positively with dollar volume (0.381) and analyst coverage (0.238), and negatively with quoted spreads (-0.038).

Second, retail trading is more strongly associated with social media activity than with professional news sentiment. RTAT shows a correlation of 0.570 with WallStreetBets mentions, compared to correlations near zero with news sentiment scores (ESS: 0.011, CSS: -0.000).

Third, retail trading activity is positively correlated with short sale volume (0.536) and market beta (0.038). This indicates that higher retail interest is associated with stocks that have both high shorting activity and higher sensitivity to market-wide movements.

[8]Nasdaq representatives confirmed that RTAT uses consolidated market feeds and a proprietary algorithm to identify and classify retail transactions as buy or sell, effectively isolating specific retail flows for inclusion in the product.

[Table 11 about here.]

Finally, the relationship between RTAT and returns is informative. RTAT has a small positive contemporaneous correlation with returns (0.023), while its unconditional correlation with next-day returns is zero.

VI. Empirical Design

This section presents my empirical framework to examine for cross-stock return predictability based on retail trading network. I construct a sample of common stocks from major U.S. exchanges covering 2016-2024 and define peer relationships using lagged correlations in retail trading activity. The main explanatory variable is the average return of a stock's retail trading peers. Control variables include factors documented in the literature to predict returns or influence retail trading activity: market-based characteristics, firm fundamentals, information environment variables, sentiment indicators, and risk factors. I estimate the relationship between peer returns and focal stock returns using panel regressions with firm and time fixed effects and Fama-MacBeth cross-sectional regressions.

A. Sample Selection: Common Stocks

My empirical analysis utilizes a comprehensive dataset of U.S. equities spanning January 1, 2016, to December 31, 2024. The sample comprises common stocks (CRSP share codes 10 and 11) listed on NYSE, AMEX/NYSE American, or NASDAQ (CRSP exchange codes 1, 2, and 3). Following standard practice, I exclude penny stocks by requiring a minimum closing price of \$1.00 and winsorize all continuous variables daily at the 1st and 99th percentiles of their cross-sectional distributions. The primary data source for retail trading activity is Nasdaq's Retail Trading Activity Tracker (RTAT), which I merge with CRSP for price and volume data, Compustat for accounting variables, Thomson Reuters for institutional holdings, I/B/E/S for analyst coverage, RavenPack for sentiment scores, and FINRA for short sale volumes.

B. *Main Explanatory Variable*

Central to my analysis is the construction of a dynamic peer network based on retail trading patterns using Nasdaq's RTAT data. This measure represents a stock's daily retail trading dollar volume as a share of the total retail trading dollar volume across all U.S. equities. To ensure robust correlation estimates, I require each stock to have at least 120 days of RTAT observations with non-zero variance. I then identify peers through a lagged correlation structure that captures potential information flows: stock j qualifies as a peer of stock i if the Pearson correlation between $RTAT_{i,t}$ and $RTAT_{j,t-1}$ over the trailing 252 trading days equals or exceeds 0.4. This threshold balances signal detection with a low false positive rate, making it unlikely that spurious relationships are identified as peers.[9] Among qualifying candidates, I select up to 20 peers ranked by average market capitalization to focus on economically significant relationships.

My primary explanatory variable, $Peer_Avg_Return_{i,t}$, is the equal-weighted average daily return of all stocks identified as peers for a focal stock i on day t. This variable is also winsorized daily at the 1st and 99th percentiles.

C. *Control Variables*

To isolate the predictive power of peer returns for focal stock outcomes, I incorporate comprehensive controls organized into five categories. First, market-based characteristics capture recent performance and trading dynamics: the focal stock's own return on day t and cumulative returns from $t-5$ to $t-1$ control for short-term reversal and momentum effects; log-transformed volume proxies for liquidity and information flow; log price captures trading frictions documented in Da, Fang, and Lin (2024) and Bartlett, McCrary, and O'Hara (2024); and log market capitalization provides the standard size control (Fama and MacBeth, 1973; Fama and French, 1993). These variables appear throughout the retail trading literature (Ruan, 2019; Barber, Huang, Odean, and Schwarz, 2022; Eaton, Green, Roseman, and Wu, 2022).

[9] Simulation evidence indicates that a 0.4 correlation threshold yields a false positive rate below 5% under the null hypothesis of independent retail trading patterns.

Second, firm fundamentals include the log book-to-market ratio, distinguishing value from growth stocks, and return on assets (ROA), controlling for profitability effects on returns (Fama and French, 2015). Third, information environment variables—institutional ownership percentage and analyst coverage—proxy for information production and dissemination, affecting return dynamics (Ying, 2024; Ali and Hirshleifer, 2020). Fourth, sentiment indicators from RavenPack capture news-driven price pressure: event sentiment scores (ESS) and composite sentiment scores (CSS) aggregating Dow Jones, press releases, and web sources demonstrate predictive power for asset prices (Hafez and Gomez, 2019; Hafez, Matas Navarro, Gomez, and Kangrga, 2021; Hafez, Liu, and Matas Navarro, 2020). Fifth, trading activity and risk adjustments include log FINRA short volume as a bearish sentiment indicator (Barber, Huang, Odean, and Schwarz, 2022; Bartlett, McCrary, and O'Hara, 2024) and risk controls from rolling 252-day Fama-French-Carhart four-factor regressions:

$$R_{i,\tau} - R_{f,\tau} = \alpha_i + \beta_i^{MKT}(R_{m,\tau} - R_{f,\tau}) + \beta_i^{SMB} SMB_\tau + \beta_i^{HML} HML_\tau + \beta_i^{UMD} UMD_\tau + \epsilon_{i,\tau}.$$

I include the estimated alpha, market beta, factor loadings, and idiosyncratic volatility to control for systematic risk exposures (Fama and French, 1993; Carhart, 1997).

D. Econometric Specifications

I employ two complementary econometric methodologies. The baseline specification uses panel regressions with firm and time fixed effects:

$$Y_{i,t+1} = \alpha_i + \theta_t + \beta \cdot Peer_Avg_Return_{i,t} + \mathbf{X}'_{i,t}\boldsymbol{\gamma} + \epsilon_{i,t+1},$$

where $Y_{i,t+1}$ is the outcome variable (typically returns or abnormal returns), α_i and θ_t are firm and time fixed effects, and $\mathbf{X}_{i,t}$ is the control vector. Standard errors are clustered at the firm level. This specification leverages within-firm variation to identify the effect of peer returns while controlling for unobserved time-invariant firm characteristics and market-wide shocks.

To assess robustness and examine cross-sectional patterns, I implement the Fama and MacBeth (1973) two-stage procedure. The first stage estimates daily cross-sectional regressions:

$$Y_{i,t+1} = a_t + b_t \cdot Peer_Avg_Return_{i,t} + \mathbf{X}'_{i,t}\boldsymbol{c}_t + e_{i,t+1}.$$

The second stage computes time-series averages of the daily coefficients with Newey-West standard errors using five lags to account for autocorrelation (Newey and West, 1987). This approach evaluates the average cross-sectional relationship between peer returns and focal stock outcomes. Together, these methodologies provide complementary perspectives: fixed effects regressions identify within-firm temporal effects, while Fama-MacBeth regressions assess the consistency of cross-sectional relationships. Appendix B provides detailed variable definitions.

VII. Empirical Results: Cross-Predictability of Returns

This section presents the results from fixed-effect panel regressions and Fama-MacBeth regressions, which include various sets of control variables. The results identify the specific effects of peer average return on the return of the focal stock and establish the robustness of these findings across various econometric specifications.

A. Fixed-effect Regressions

Table XII presents the fixed effects panel regression results where the dependent variable is the focal stock's next-day return ($Return_{i,t+1}$).

The coefficient for $Peer_Avg_Return_{i,t}$ is consistently positive and highly statistically significant ($p<0.01$) across all six specifications. In the base model (Column 1), a one-unit increase in peer average return is associated with a 0.0144 increase in the focal stock's future return. This effect remains remarkably stable with the inclusion of various control sets, ranging from 0.0126 in the full model (Column 6) to 0.0144 in the base and sentiment models. This suggests a robust positive spillover effect from peer performance to the focal

firm's subsequent returns, even after controlling for a wide array of factors.

Among the base controls, $Log(Price)_{i,t}$ shows a consistent negative and significant relationship with future returns, suggesting that lower-priced stocks tend to have higher subsequent returns. Current $Return_{i,t}$ and the sum of lagged returns $Return_{i,[t-5,t-1]}$ both exhibit a significant negative association with $Return_{i,t+1}$, indicating return reversal patterns. $Log(Market_Cap)_{i,t}$ and $Log(Volume)_{i,t}$ show mixed or less economically significant effects in the base specification but become more pronounced in fuller models.

The addition of fundamental controls (Column 2) reveals that higher $ROA_{i,t}$ is positively associated with future returns (0.0054, p<0.01), while higher institutional ownership is negatively associated (-0.0006, p<0.05). These effects largely persist in the full model. The inclusion of these controls slightly attenuates the peer return coefficient to 0.0127.

Sentiment controls (Column 3), specifically $CSS_{i,t}$ (0.0042, p<0.01) and $ESS_{i,t}$ (0.0030, p<0.01), both show a positive and significant relationship with future returns, suggesting that positive sentiment predicts higher subsequent stock returns. The peer return coefficient remains at 0.0144.

The short interest control (Column 4), $Log(FINRA_Short_Vol.)_{i,t}$, is positively and significantly related to future returns (0.0004, p<0.01). This finding contrasts with Engelberg, Reed, and Ringgenberg (2012), who document a negative relation between short volume ratio and future returns using data from 2005 and 2007. The positive coefficient indicates that shorts are not profitable within a one-day horizon. For example, during January 2021, retail investors targeted heavily shorted stocks such as GameStop and AMC, particularly those with short interest exceeding outstanding shares, resulting in substantial positive returns. However, whether such retail-driven short squeeze dynamics represent the general mechanism underlying the positive short volume-return relation throughout the sample period remains inconclusive. The peer effect coefficient remains 0.0143 after controlling for short interest.

Risk controls (Column 5) show that $Alpha_FF4_{i,t}$ (0.0946, p<0.01) and $Beta_UMD_{i,t}$ (0.0002, p<0.01) are positively associated with future returns, while $IVOL_FF4_{i,t}$ (-0.0218, p<0.01) has a negative association. Here, $Alpha_FF4$ represents the daily alpha estimated

using the Fama-French-Carhart four-factor model with a rolling window of 252 days, while $IVOL_FF4$ is the idiosyncratic volatility based on the residuals from these daily four-factor regressions. The peer return coefficient is 0.0142 in this specification.

In the full model (Column 6), which includes all controls, the $Peer_Avg_Return_{i,t}$ coefficient remains positive and highly significant at 0.0126 (p<0.01). Many of the previously significant controls retain their sign and significance, though some magnitudes are altered. For example, $Log(Market_Cap)_{i,t}$ becomes significantly negative (-0.0006, p<0.01). The overall explanatory power, as indicated by R^2 (Within), increases from 0.002 in the base model to 0.003 in the full model, suggesting that while statistically significant, these factors explain a small portion of the within-firm variation in next-day returns.

[Table 12 about here.]

I provide the results using contemporaneous rtat activity to construct peer network in Table XIII. For the fixed-effects model with full control variables, the coefficient estimate of peer_avg_ret reduces from 0.0126 to 0.086 (decrease by 31.7%), and it is positive and significant. A potential explanation is using lagged correlation allows the peer network to be flexibly asymmetric and hence provide more predictive power.

[Table 13 about here.]

I provide the results of alternative specifications with/without entity and fixed effects in Table XIV. In this table, I also consider the adding RTAT and use adjusted overall R-Squared.

The coefficient estimate of RTAT is negative and significant, providing additional predictive power than price and volume (together is dollar volume) in the equation.

The Pooled OLS, Entity FE Only, Time FE Only, and Two way FE models have peer_avg_ret coefficient estimates of 0.0162, 0.0162, 0.0126, and 0.0126 accordingly, and they are all statistically significant at 1%. The R-Squared is 0.0013, 0.0028, 0.1922, and 0.1932, indicating that time fixed effects significantly add to explanatory power of model.

[Table 14 about here.]

B. *Fama-MacBeth Regressions*

The Fama-MacBeth estimates for the predictability of next-day stock returns ($Return_{i,t+1}$) are presented in Table XV. Across all specifications, the $Peer_Avg_Return_{i,t}$ demonstrates a consistently positive and highly statistically significant (p<0.01) coefficient. In the baseline model (Column 1), which includes only base controls ($Log(Market_Cap)$, *Price*, *Volume*, and past returns), the coefficient for $Peer_Avg_Return_{i,t}$ is 0.0110. This indicates that, on an average day, a one percentage point increase in the average return of a firm's peers is associated with an approximate 0.0110 percentage point increase in the focal firm's return on the following day. The average R^2 for this specification over the 2262 periods (days) is 0.0400, suggesting that these base controls and peer returns explain about 4.0% of the cross-sectional variation in next-day returns on average.

As I introduce additional sets of controls, the peer return coefficient remains robust. With the inclusion of fundamental variables in Column 2 (*ROA*, *Institutional_Ownership*, etc.), the peer return coefficient is 0.0098 (p<0.01), and the average R^2 increases to 0.0604 over 2203 periods. This suggests that while fundamentals have their own predictive power (e.g., *ROA* shows a coefficient of 0.0056, p<0.01), the peer effect is distinct. Adding sentiment controls (Column 3), such as $CSS_{i,t}$ (coefficient 0.0054, p<0.01) and $ESS_{i,t}$ (coefficient 0.0029, p<0.01), which are themselves significant positive predictors, the peer return coefficient remains strong at 0.0110, and the average R^2 is 0.0418. The inclusion of short interest (Column 4, $Log(FINRA_Short_Vol)_{i,t}$ with coefficient 0.0002, p<0.01) or detailed risk controls (Column 5, where variables like $Alpha_FF4_{i,t}$ and $IVOL_FF4_{i,t}$ show significant coefficients of 0.0601 and -0.0386 respectively) does not diminish the significance of peer returns, with coefficients of 0.0107 and 0.0100 respectively. The average R^2 in the risk model (Column 5) is notably higher at 0.0907.

In the full specification (Column 6), encompassing all control categories, the coefficient of $Peer_Avg_Return_{i,t}$ is 0.0089 (p<0.01). The average R^2 for this comprehensive model is 0.1068, indicating that about 10.7% of the daily cross-sectional variation in future returns can be explained. The consistent significance and positive sign of peer returns across these varied FMB specifications strongly support the hypothesis of positive return

spillovers in the cross-section.

[Table 15 about here.]

[Figure 18 about here.]

Figure 18 presents the time series of daily peer return coefficients that I derive from Fama-MacBeth regressions for each trading day from 2016 to 2024. These regressions predict a focal stock's next-day total return using the average return of its retail trading peers, alongside a comprehensive set of control variables. My analysis of these coefficients reveals several patterns.

First, I observe that the coefficient estimates for peer average return exhibit time variation yet fluctuate around a positive mean of 0.0089 (standard error 0.0005). Approximately 75% of these daily estimates are positive, which indicates a prevailing positive peer effect in the cross-section of stocks throughout the sample period.

Second, I identify a distinct pattern in the behavior of these coefficients during the 2020-2021 period, which coincides with the COVID-19 pandemic and increased retail trading activity. In these months, both the average magnitude and the daily volatility of the peer return coefficients increase. Daily coefficients frequently exceed 0.10—a value more than ten times the full-sample average—and also show greater day-to-day dispersion. This pattern suggests an intensification of peer group influences within retail trading networks during periods of heightened retail market participation and uncertainty.

Third, while the coefficients show some movement toward their mean after 2021, they generally remain at levels higher than those observed prior to 2020. This observation points to a potential change in the propagation dynamics of peer-related information through retail trading networks in the more recent market environment.

Finally, the positive nature of these coefficients persists across different market conditions, including the 2018 volatility spike, the 2020 pandemic-related market decline, and the 2022 bear market. This persistence indicates that the observed peer effects are not confined to bull market periods but are present across various market states. Given that these coefficients arise from regressions that include a comprehensive set of standard control

variables, these findings support the interpretation that retail investors systematically incorporate the performance of peer stocks into their trading decisions.

VIII. Economic Channels

Having established robust evidence of cross-predictability of returns in the US equity market across various specifications and controls, I now turn to understanding the economic mechanisms underlying these peer network effects. To develop testable hypotheses about these channels, I draw on institutional knowledge and documented patterns of retail investor behavior in the US equity market.

A fundamental characteristic of retail investors is their persistent role as net buyers of securities in the US equity market (Mackintosh, 2021, 2025). This consistent directional flow, combined with the documented coordination in retail trading networks, provides the foundation for understanding how peer effects generate cross-predictability. Recent literature has documented that retail investors can contribute meaningfully to market efficiency. Welch (2022) provide evidence for the "wisdom of crowds" in retail trading, showing that aggregated retail investor decisions contain valuable information for price discovery. Additionally, Eaton, Green, Roseman, and Wu (2022) demonstrate that sophisticated retail investors improve market quality, pointing to their potential role in liquidity provision.

This study examines general retail trading activity across the market, which may exhibit different patterns from platform-specific phenomena documented in recent literature. Barber, Huang, Odean, and Schwarz (2022) study attention-induced trading on Robinhood, a platform that caters to less experienced retail investors, and document return reversals as attention wanes. In contrast, market-wide retail trading networks may exhibit more persistent effects if they capture the behavior of more sophisticated investors. Similarly, Da, Fang, and Lin (2024) document asset bubble dynamics associated with fractional share trading, a specific innovation that may particularly attract new investors. The broader retail trading activity examined here may show different dynamics that reflect the aggregate behavior of retail investors across all platforms and experience levels.

Given retail investors' role as consistent net buyers and the potential for persistent peer effects in market-wide data, I analyze two complementary economic channels through which retail trading networks generate cross-predictability: price discovery and liquidity provision. The price discovery hypothesis posits that retail investors, through their peer networks, aggregate dispersed information that helps incorporate new information into prices. The liquidity provision hypothesis suggests that coordinated retail trading supplies systematic liquidity to the market. These hypotheses are not mutually exclusive—retail networks may simultaneously discover prices and provide liquidity, earning compensation for both economic functions.

A. Price Discovery of Retail Investors

The price discovery hypothesis posits that retail investors capture subtle economic links between securities and learn information from their peer networks that helps incorporate new information into prices. Given their role as persistent net buyers, retail investors may be particularly effective at price discovery from the buy side. The liquidity provision hypothesis suggests that coordinated retail trading supplies liquidity to the market, potentially reducing transaction costs and facilitating price efficiency.

In this section, I focus on testing the price discovery hypothesis. Retail investors may have valuable local information or behavioral insights that, when aggregated through their trading networks, contribute to more efficient price formation. The peer network structure I document may serve as an information aggregation mechanism, where trading patterns in peer stocks reveal information relevant to focal stocks.

Based on the price discovery hypothesis, I develop several testable predictions:

A.1. Prediction 1: Asymmetric Effects on Positive Peer Returns

Given that retail investors are predominantly buyers, the price discovery effects should be stronger when peer returns are positive. I test this prediction by including an interaction term PeerAvgReturn $\times$ $\mathbf{1}$(PeerAvgReturn > 0) in the Fama-MacBeth regression.

Evidence: Column (2) of Table XVI shows that the interaction coefficient is 0.0213

(t-stat = 12.91), while the base effect for negative peer returns is only 0.0031. This indicates that the peer effect is nearly seven times stronger when peer returns are positive, consistent with retail investors' net buyer position facilitating price discovery primarily on the positive side.

A.2. Prediction 2: Information Gap Amplifies Peer Effects

The price discovery hypothesis predicts that effects should increase with the magnitude of the information gap between peer stocks and the focal stock, measured by $(PeerAvgReturn - Return)$. When peer stocks contain more positive information than the focal stock, retail networks should facilitate stronger information transmission.

Evidence: Column (3) of Table XVI confirms this prediction. The interaction term PeerAvgReturn × (PeerAvgReturn − Return) has a coefficient of 0.0851 (t-stat = 8.90), demonstrating that larger information gaps between peers and focal stocks lead to stronger cross-predictability.

A.3. Prediction 3: "Buy the Dip" Behavior

"Buy the Dip", which emerged as a new investment mantra in the retail investor crowd, now captures the interest of mainstream financial media.[10] According to this "Buy the Dip" pattern of retail investors, the peer effect should be particularly strong when the focal stock experiences a significant drop while its peer group maintains a higher average return. I test whether the coefficient of PeerAvgReturn × $\mathbf{1}$(Return < 5th percentile) × (PeerAvgReturn − Return) is positive.

Evidence: Column (4) of Table XVI provides striking evidence for coordinated Buy-the-Dip behavior. The triple interaction coefficient is 0.6553 (t-stat = 11.81), indicating that peer effects are amplified when stocks experience extreme negative returns while their peers perform better. This massive amplification suggests retail networks coordinate to identify and exploit temporary mispricings. This result also serves as evidence for liquidity

[10] See, for example, Financial Times: https://www.ft.com/content/9d037d36-f697-4ca2-b0be-c793ccb0a3dd

provision of retail investors for stocks in the network that experience extraordinary negative returns.

A.4. Prediction 4: Complementarity with Institutional Information Sources

If retail investors provide genuine price discovery, their effects should be stronger in weaker information environments where traditional information sources are limited. I test whether the coefficients of PeerAvgReturn × InstOwnership and PeerAvgReturn × NumAnalysts are negative.

Evidence: Columns (5) and (6) of Table XVI support this prediction. The interaction with institutional ownership is -0.0301 (t-stat = -18.22), and the interaction with analyst coverage is -0.0008 (t-stat = -13.72). The peer effect is approximately three times stronger for stocks with minimal institutional ownership compared to those with complete institutional ownership, suggesting retail investors play a complementary price discovery role to institutions and analysts.

A.5. Prediction 5: Gradual Information Diffusion

Due to the gradual process of information diffusion, the effect on open-to-close returns should be a fraction of the effect on close-to-close returns, reflecting partial information incorporation during trading hours. This is related to the notion of slow moving capital in Duffie (2010).

Evidence: Table XVII Column (2) shows that the peer effect on open-to-close returns is 0.0055 (t-stat = 13.73), approximately 47% of the close-to-close effect of 0.0118. This proportion is consistent with information being gradually incorporated during the trading day.

A.6. Prediction 6: Improved Price Efficiency

Higher peer average return should be associated with improved information efficiency, measured by variance ratio closer to one. Since the variance ratio variable represents the absolute difference from one, the coefficient should be negative.

Evidence: Column (3) of Table XVII shows a coefficient of -0.0264 (t-stat = -12.79) on peer average return when predicting variance ratio, confirming that positive peer information improves price efficiency by reducing deviations from random walk behavior.

A.7. Prediction 7: Options Market Spillovers

As positive information transmits through peer networks, implied volatility of call options should increase with peer average return. This prediction is consistent with retail investors' documented participation in options markets Bryzgalova, Pavlova, and Sikorskaya (2023) and Eaton, Green, Roseman, and Wu (2024).

Evidence: Columns (4) and (5) of Table XVII show positive coefficients for both at-the-money implied volatility (0.0174, t-stat = 3.33) and volume-weighted implied volatility (0.0281, t-stat = 3.91), indicating that positive peer information increases option prices as retail investors use options to express their bullish views.

A.8. Prediction 8: Permanent Price Impact

The price discovery hypothesis predicts persistent effects without return reversal, distinguishing it from attention-induced trading in Barber, Huang, Odean, and Schwarz (2022) which exhibits significant reversals. This can be interpreted through the lens of price impact literature, which decomposes trade-induced price movements into permanent components reflecting new information and transitory components from temporary liquidity effects (Kyle, 1985; Glosten and Milgrom, 1985; Hasbrouck, 1991; O'Hara, 1995; Hasbrouck, 2007).

Evidence: Table XVIII provides strong evidence for permanent price impact. The peer effect remains significant at all horizons, with coefficients of 0.0026 at t+2, gradually increasing to 0.0414 for cumulative returns from t+51 to t+126. The absence of sign reversal at any horizon, even at the six-month horizon, confirms that retail trading networks deliver genuine price discovery rather than temporary price pressure.

A.9. Summary

In summary, I find cohesive and consistent evidence supporting the price discovery role of retail investors, particularly from the buy side. All eight predictions derived from the price discovery hypothesis are strongly supported by the data. The effects are economically large, statistically robust, and persistent across time horizons. The evidence suggests that retail trading networks serve as an effective information aggregation mechanism that complements traditional institutional channels of price discovery.

[Table 16 about here.]

[Table 17 about here.]

[Table 18 about here.]

B. Liquidity Provision of Retail Investors

I examine whether retail trading networks contribute to market liquidity. The foundation for this question is that retail investors are persistent net buyers in US equity market (Mackintosh, 2021, 2025). I hypothesize that the peer network effects I document amplify this directional flow into a systematic form of liquidity provision, predicting that peer-correlated trading increases market depth, particularly on the bid side. My approach extends the work of Eaton, Green, Roseman, and Wu (2022) on sophisticated retail by focusing on a network-driven mechanism. Furthermore, by using the market-wide RTAT dataset, my analysis is distinct from studies of platform-specific attention effects (Barber, Huang, Odean, and Schwarz, 2022). From this framework, I derive five testable predictions:

B.1. Prediction 1: Peer Returns Drive Trading Activity

Higher peer average returns should be associated with increased future trading activity in focal stocks, as retail investors coordinate their trading through peer networks. This effect should manifest in higher dollar volume and increased RTAT activity.

Evidence: Table XIX Column (1) strongly supports this prediction. The coefficient on peer average return for log dollar volume is 0.3446 (t-stat = 12.02), indicating that a

one percentage point increase in peer returns leads to approximately 0.345% higher dollar volume.

B.2. Prediction 2: Asymmetric Depth Provision on the Buy Side

Given that RTAT is consistently biased towards net buying as documented in Mackintosh (2021, 2025), higher peer average returns should increase bid depth more than offer depth. This asymmetry reflects retail investors' role as liquidity providers primarily on the buy side.

Evidence: Table XIX Columns (2)-(5) confirm this asymmetric effect. While peer average return increases overall market depth by 0.0958 (t-stat = 9.07), the decomposition reveals striking asymmetry: bid depth increases by 0.1170 (t-stat = 10.82) compared to only 0.0477 (t-stat = 4.18) for offer depth. The bid-offer difference coefficient of 0.0697 (t-stat = 10.36) directly captures this buy-side bias in liquidity provision.

B.3. Prediction 3: Informed Trading and Spread Widening

Given the order book imbalance derived from directional trades of retail investors biased towards buys, market makers face heightened adverse selection risk (Glosten and Milgrom, 1985). Such systematic order flow may signal the presence of informed trading, leading market makers to widen both quoted and effective spreads to compensate for the increased probability of transacting with better-informed agents (Kyle, 1985; Easley, López de Prado, and O'Hara, 2012; Chordia, Roll, and Subrahmanyam, 2002). This prediction is also consistent with the price discovery hypothesis of retail investors.

Evidence: Table XX shows that higher peer returns lead to wider spreads. The quoted spread increases by 12 basis points (t-stat = 9.64) and the effective spread by 10 basis points (t-stat = 12.38). These results confirm that the liquidity provided by retail investors carries information content, leading market makers to adjust spreads to protect against adverse selection.

B.4. Prediction 4: Stronger Effects in Less Liquid Stocks

The effects of peer returns should be stronger for stocks with lower baseline liquidity, as retail liquidity provision has greater marginal impact in less liquid markets. Interaction terms with spread measures should be positive (higher spreads indicate lower liquidity), while interactions with market capitalization and depth should be negative (larger size and depth indicate higher liquidity). These liquidity variables are commonly seen in market microstructure literature such as (Foley and Putniņš, 2016).

Evidence: Table XXI provides strong support for this prediction. The interaction between peer returns and quoted spread is 0.6559 (t-stat = 12.53), and with effective spread is 1.2409 (t-stat = 13.53), indicating that peer effects are amplified in stocks with wider spreads. Conversely, the interactions with log market cap (-0.0053, t-stat = -18.62) and log depth (-0.0048, t-stat = -13.10) are negative, confirming that peer effects diminish in more liquid stocks.

B.5. Prediction 5: Persistent Liquidity Effects

The liquidity provision effects should persist over time, supporting the role of retail investors as consistent liquidity providers. This persistence is inconsistent with the alternative mechanism of attention-induced trading in Barber, Huang, Odean, and Schwarz (2022), which would predict return and trading activity reversals.

Evidence: Table XXII demonstrates remarkable persistence in bid depth effects. The coefficient remains positive and significant from t+2 (0.1073, t-stat = 9.48) through t+126 (0.0604, t-stat = 5.13). While the magnitude declines over time, the effect never reverses sign, confirming that retail investors provide persistent rather than transitory liquidity. This persistence aligns with genuine liquidity provision rather than temporary attention-driven trading.

B.6. Summary

I find comprehensive evidence supporting the liquidity provision hypothesis of retail investors in the US equity markets for explaining the cross-predictability of returns. The

results demonstrate that retail trading networks facilitate systematic liquidity provision, particularly on the buy side, with effects that are stronger in less liquid stocks and persist over extended horizons.

Combining the results of the price discovery and liquidity provision hypotheses, I provide a market microstructure-based explanation for cross-predictability of returns in the retail trading network, grounded in the institutional understanding that retail investors are consistently net buyers in the US equity market. The peer network effects I document represent a mechanism through which retail investors both discover prices and provide liquidity, contributing to market efficiency while earning compensation for their liquidity provision and information production roles.

[Table 19 about here.]

[Table 20 about here.]

[Table 21 about here.]

[Table 22 about here.]

IX. Peer Effects Beyond Common Stocks

Motivated by the wide-reaching footprints of retail trading activity across asset types and exchanges in Figure 4 and Table I, I extend my analysis of cross-predictability beyond common stocks, to include all securities in the CRSP database. By defining the peer group using the full CRSP sample, I examine whether network-driven return predictability is a pervasive feature of modern financial markets or confined to a specific asset class. I employ daily Fama-MacBeth regressions to invesigate the cross-predictability of returns in this cross-section by various subsamples of asset types and exchanges.

A. Peer Effects by Asset Type

Table XXIII presents Fama-MacBeth regression results where the sample is segmented by asset type. The specification remains consistent across all subsamples, allowing for direct

comparison. The key independent variable, $Peer_Avg_Return_{i,t}$, captures the average return of an asset's retail-defined peers. The results indicate that positive and statistically significant peer effects on next-day returns are present across all major asset categories.

[Table 23 about here.]

The magnitude of the peer effect varies logically across asset types. The peer return coefficient is largest for ADRs (0.0212) and Foreign Stocks (0.0178). This pattern is consistent with retail investors placing greater weight on peer trading activity in markets characterized by higher information asymmetry. Despite representing only 9.1% of retail trading volume (Table I), these international securities generate the strongest network effects. The persistence of a significant peer effect in ETPs (0.0033, t-statistic = 10.61) indicates that retail coordination extends beyond individual security selection to broader asset allocation decisions.

The explanatory power of the model also varies. The average R^2 is highest for ETPs (23.3%) and REITs (12.2%), reflecting the more systematic cross-sectional structure of returns in these asset classes. In contrast, the low average R^2 for Common Stocks (2.8%) reflects the high degree of idiosyncratic volatility inherent in individual stock returns.

B. Peer Effects by Exchange

To investigate the role of market structure in mediating these effects, I next segment the analysis by primary trading exchange. Table XXIV presents the results, which reveal substantial heterogeneity in the peer effect across different trading venues.

[Table 24 about here.]

The peer return coefficient on NASDAQ (0.0153) is nearly double that of NYSE (0.0085), a difference that is both statistically and economically significant. This differential aligns with NASDAQ's growing share of retail activity, which increased from 35% to 45% over the sample period (Figure 5). NYSE Arca, which specializes in ETPs, exhibits a smaller but highly significant coefficient (0.0039). The high average R^2 for NYSE Arca (23.0%) combined with modest peer effects reflects the systematic nature of ETP returns.

A noticeable result is the null finding for Cboe. The peer effect coefficient is economically and statistically zero (0.0001, t-statistic = 0.10). Cboe BZX Exchange lists primarily volatility-linked products such as VXX, UVXY, and SVXY, whose returns are mechanically driven by VIX futures and daily rebalancing rules. This finding serves as a placebo test: in a market segment where returns are structurally disconnected from peer returns, the measured peer effect disappears. This strengthens the interpretation that the effects observed on other exchanges reflect genuine network phenomena rather than mechanical correlations.

C. Economic Significance and Market Implications

The estimated coefficients imply meaningful economic effects. For ADRs, a one-standard-deviation increase in peer group returns (approximately 2.5%) predicts a 5.3 basis point increase in next-day returns. The persistence of these effects across more than 2,200 daily cross-sections indicates that retail coordination is a structural feature of the market.

From a market efficiency perspective, these findings indicate that prices do not instantaneously incorporate the information contained in retail order flow. The persistence of predictable patterns in highly liquid markets like NASDAQ, which processes 45.3% of retail activity (Table I), points to limits to arbitrage. The heterogeneity across exchanges highlights the role of market microstructure in mediating the transmission of retail sentiment. The contrast between strong effects on NASDAQ and the null result on Cboe demonstrates that market design and asset characteristics are critical determinants of whether retail peer effects manifest. These insights are relevant for asset managers, regulators, and exchange operators seeking to understand the information environment of modern financial markets.

X. Conclusion

In this study, I introduce a novel method for constructing a retail trading network based on lead-lag correlations in Nasdaq's Retail Trading Activity Tracker (RTAT). The

methodology's validity is first established using the iShares Bitcoin Trust (IBIT) as a prototype. This initial analysis demonstrates the network approach's ability to uncover sensible economic linkages—such as a connection to a TIPS ETF, which provides market-based evidence of an inflation-hedging motive among retail investors.

With the method validated, the analysis proceeds to its primary focus on the U.S. common stock universe. My findings show that the average return of a stock's peers, as identified through this network, robustly predicts its future returns. This network predictability also extends to key microstructure outcomes. The results hold across firm-level fixed effects and cross-sectional Fama-MacBeth regressions, and they are robust to a comprehensive set of controls for market-based characteristics, firm fundamentals, sentiment, short volume, and risk factors.

To test the pervasiveness of this phenomenon, the analysis is further extended across the full CRSP universe. This confirms the peer effect is a broad market feature, present across multiple asset classes including common stocks, foreign stocks, REITs, ETPs, and American Depositary Receipts (ADRs). The effect also varies by market structure, with the peer return coefficient on NASDAQ-listed stocks being nearly double that of those on the NYSE. Notably, the peer effect is absent for volatility-linked products on the Cboe exchange.

This study provides a unified, microstructure-based explanation for this cross-predictability, identifying two complementary economic channels. The first channel is the **price discovery** of retail investors. Retail networks aggregate dispersed information, leading to permanent price impacts without reversal at horizons up to 126 trading days. Consistent with the "Buy the Dip" pattern of retail investors, I find the peer effect is magnified when two conditions are met simultaneously: a focal stock has a large negative return, and its peer group maintains a higher average return. Further, I find that the effect is stronger in weaker information environments measured by institutional holding and analyst coverage. The second channel is the **liquidity provision** of retail investors. I find that the retail trading networks facilitate systematic, buy-side liquidity that persistently increases market depth. In response, market makers widen spreads to mitigate the adverse selection risk

posed by this persistent, imbalanced order flow, which they interpret as informed trading.

This research demonstrates that the collective actions of retail investors form a significant component of the U.S. equity market's infrastructure. This network, comprising over 100 million accounts with assets comparable in value to the U.S. GDP, acts as a persistent source of capital and liquidity, contributing over $100 billion in net purchases annually while continuously processing market information.

The findings have direct implications for market participants and regulators. The identified retail trading network offers a novel signal for asset managers' portfolio construction and risk assessment. For electronic market makers, these network dynamics are critical for anticipating correlated order flow and managing adverse selection risk. For regulators, the results highlight the importance of monitoring the systemic effects of this retail investor network to maintain financial stability and market integrity.

REFERENCES

Ali, Usman, and David Hirshleifer, 2020, Shared analyst coverage: Unifying momentum spillover effects, *Journal of Financial Economics* 136, 649–675.

Anton, Miguel, and Christopher Polk, 2014, Connected stocks, *The Journal of Finance* 69, 1099–1127.

Barber, Brad M, Xing Huang, Philippe Jorion, Terrance Odean, and Christopher Schwarz, 2024, A (sub) penny for your thoughts: Tracking retail investor activity in taq, *The Journal of Finance* 79, 2403–2427.

Barber, Brad M, Xing Huang, Terrance Odean, and Christopher Schwarz, 2022, Attention-induced trading and returns: Evidence from robinhood users, *The Journal of Finance* 77, 3141–3190.

Bartlett, Robert P, Justin McCrary, and Maureen O'Hara, 2023, A fractional solution to a stock market mystery, *Available at SSRN 4167890*.

Bartlett, Robert P., Justin McCrary, and Maureen O'Hara, 2024, Tiny trades, big questions: Fractional shares, *Journal of Financial Economics* 157.

Battalio, Robert, Robert Jennings, Mehmet Saglam, and Jun Wu, 2022, Identifying market maker trades as "retail" from taq: No shortage of false negatives and false positives, *SSRN*.

Boehmer, Ekkehart, Charles M Jones, Xiaoyan Zhang, and Xinran Zhang, 2021, Tracking retail investor activity, *The Journal of Finance* 76, 2249–2305.

Breitung, Christian, and Sebastian Müller, 2025, Global business networks, *Journal of Financial Economics* 166, 104007.

Bryzgalova, Svetlana, Anna Pavlova, and Taisiya Sikorskaya, 2023, Retail trading in options and the rise of the big three wholesalers, *The Journal of Finance* 78, 3465–3514.

Budish, Eric, 2025, Trust at scale: The economic limits of cryptocurrencies and blockchains, *The Quarterly Journal of Economics* 140, 1–62.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *The Journal of finance* 52, 57–82.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2002, Order imbalance, liquidity, and market returns, *Journal of Financial Economics* 65, 111–130.

Cohen, Lauren, and Andrea Frazzini, 2008, Economic links and predictable returns, *Journal of Finance* 63, 1977–2011.

Cohen, Lauren, and Dong Lou, 2012, Complicated firms, *Journal of Financial Economics* 104, 383–400.

Cong, Lin William, Ye Li, and Neng Wang, 2021, Tokenomics: Dynamic adoption and valuation, *The Review of Financial Studies* 34, 1105–1155.

Da, Zhi, Vivian W Fang, and Wenwei Lin, 2024, Fractional trading, *The Review of Financial Studies* p. hhae012.

Duffie, Darrell, 2010, Presidential address: Asset price dynamics with slow-moving capital, *Journal of Finance* 65, 1237–1267.

Easley, David, and Jon Kleinberg, 2010, Networks, crowds, and markets, *Cambridge Books*.

Easley, David, Marcos López de Prado, Maureen O'Hara, and Zhibai Zhang, 2021, Microstructure in the machine age, *The Review of Financial Studies* 34, 3316–3363.

Easley, David, Marcos M López de Prado, and Maureen O'Hara, 2012, Flow toxicity and liquidity in a high-frequency world, *The Review of Financial Studies* 25, 1457–1493.

Easley, David, Maureen O'Hara, Songshan Yang, and Zhibai Zhang, 2024, Microstructure and market dynamics in crypto markets, *Available at SSRN 4814346*.

Eaton, Gregory W, T Clifton Green, Brian Roseman, and Yanbin Wu, 2024, Retail option traders and the implied volatility surface, *Available at SSRN 4104788*.

Eaton, Gregory W, T Clifton Green, Brian S Roseman, and Yanbin Wu, 2022, Retail trader sophistication and stock market quality: Evidence from brokerage outages, *Journal of Financial Economics* 146, 502–528.

Engelberg, Joseph E, Adam V Reed, and Matthew C Ringgenberg, 2012, How are shorts informed?: Short sellers, news, and information processing, *Journal of Financial Economics* 105, 260–278.

Even-Tov, Omri, Kimberlyn Munevar, Shimon Kogan, and Eric C. So, 2022, Fee the people: Retail investor behavior and trading commission fees, MIT Sloan Research Paper No. 6801-22. Available at SSRN: https://ssrn.com/abstract=4226044.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

——— , 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Feng, Jian, Xiaolin Huo, Xin Liu, Yifei Mao, and Hong Xiang, 2025, Economic links from bonds and cross-stock return predictability, *Journal of Financial Economics* 171, 104110.

Finke, Christian, and Florian Weigert, 2017, Does foreign information predict the returns of multinational firms worldwide?, *Review of Finance* 21, 2199–2248.

Foley, Sean, and Tālis J Putniņš, 2016, Should we be afraid of the dark? dark trading and market quality, *Journal of Financial Economics* 122, 456–481.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.

Hafez, Peter, and Francisco Gomez, 2019, Socially responsible investing: Combining esg ratings with news sentiment generates alpha, Discussion paper, RavenPack.

Hafez, Peter, Alan Liu, and Ricard Matas Navarro, 2020, Forex carry, value, & momentum factors boosted by sentiment, Discussion paper, RavenPack.

Hafez, Peter, Ricard Matas Navarro, Francisco Gomez, and Marko Kangrga, 2021, Three trading signals generated systematically from earnings conference calls, Available at SSRN: https://ssrn.com/abstract=4323558.

Hamilton, James D, 2020, *Time series analysis* (Princeton university press).

Hasbrouck, Joel, 1991, Measuring the information content of stock trades, *The Journal of Finance* 46, 179–207.

——— , 2007, *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading* (Oxford University Press: New York).

Hoberg, Gerard, and Gordon M Phillips, 2018, Text-based industry momentum, *Journal of Financial and Quantitative Analysis* 53, 2355–2388.

Hornback, Andrew M, and Robert E Whaley, 2025, Spot bitcoin etfs: The struggle was worth it, *Financial Analysts Journal* 81, 39–50.

Huang, Xia, 2015, Thinking outside the borders: Investors' underreaction to foreign operations information, *Review of Financial Studies* 28, 3109–3152.

John, Kose, Leonid Kogan, and Fahad Saleh, 2023, Smart contracts and decentralized finance, *Annual Review of Financial Economics* 15, 523–542.

John, Kose, Maureen O'Hara, and Fahad Saleh, 2022, Bitcoin and beyond, *Annual Review of Financial Economics* 14, 95–115.

Kogan, Shimon, Igor Makarov, Marina Niessner, and Antoinette Schoar, 2024, Are cryptos different? evidence from retail trading, *Journal of Financial Economics* 159, 103897.

Krause, David, 2024, The rise of spot cryptocurrency etfs: Implications for institutional investors, Available at SSRN: https://ssrn.com/abstract=4868157.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1335.

Lee, Charles MC, Stephen Teng Sun, Rongfei Wang, and Ran Zhang, 2019, Technological links and predictable returns, *Journal of Financial Economics* 132, 76–96.

Liang, Qiqi, Mohammad Najand, David D. Selover, and Licheng Sun, 2024, Retail trading around earnings announcements: Evidence from robinhood traders, Available at SSRN: https://ssrn.com/abstract=4294065.

Liang, Qiqi, and Licheng Sun, 2024, A closer look at the substitution effects between retail trading and national lotteries, *Finance Research Letters* Forthcoming.

Mackintosh, Phil, 2021, A new way to look at retail trading trends, *Nasdaq*.

——— , 2025, Retail trading growth in perspective, *Nasdaq*.

Mazur, Mieszko, and Efstathios Polyzos, 2025, Spot bitcoin etfs: The effect of fund flows on bitcoin price formation, *Journal of Alternative Investments* 27, 110.

Menzly, Lior, and Oguzhan Ozbas, 2010, Market segmentation and cross-predictability of returns, *Journal of Finance* 65, 1555–1580.

Moskowitz, Tobias J, and Mark Grinblatt, 1999, Do industries explain momentum?, *Journal of Finance* 54, 1249–1290.

Müller, Sebastian, 2019, Economic links and cross-predictability of stock returns: Evidence from characteristic-based 'styles', *Review of Finance* 23, 363–395.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

O'Hara, Maureen, 1995, *Market Microstructure Theory* (Blackwell Publishers: Cambridge, MA).

Parsons, Christopher A, Riccardo Sabbatucci, and Sheridan Titman, 2020, Geographic lead-lag effects, *Review of Financial Studies* 33, 4721–4770.

Ruan, Qihong, 2019, Amihud premium: The compensation for illiquidity or mispricing?, *Available at SSRN 4052995*.

Tsay, Ruey S, 2013, *Multivariate time series analysis: with R and financial applications* (John Wiley & Sons).

Welch, Ivo, 2022, The wisdom of the robinhood crowd, *The Journal of Finance* 77, 1489–1527.

Ying, Jie, 2024, Gradual information diffusion across commonly owned firms, *Journal of Financial Economics* 156, 103852.

Appendix

A. US Retail Brokerage Landscape 2016-2024

The heterogeneity in average client wealth across retail brokerage platforms reveals fundamental market segmentation that persists despite technological convergence. Figure 19 documents striking disparities in average account size: Fidelity maintains $385,200 per account in 2024, compared to Charles Schwab's $277,000, Interactive Brokers' $170,200, and Robinhood's $7,700. This 50-fold differential between traditional and new-generation platforms reflects distinct client demographics and service models rather than merely platform age or features. The temporal dynamics reveal divergent trajectories—while Fidelity's average account size increased 21.0% from 2016 to 2024, Interactive Brokers experienced a 22.9% decline despite 767% account growth (Table XXV), suggesting that democratization of sophisticated trading tools attracted smaller retail investors to the platform.

The 2022 market downturn provides a natural experiment for examining client behavior across wealth segments. Figure 19 shows universal declines in average account values: Fidelity dropped from $359,200 to $277,600 (-22.7%), Schwab from $246,800 to $208,900 (-15.4%), Interactive Brokers from $223,200 to $146,800 (-34.2%), and Robinhood from $4,300 to $2,700 (-37.2%). The inverse relationship between platform average wealth and drawdown severity suggests that wealthier clients either maintained positions through volatility or had more diversified portfolios, while smaller investors potentially crystallized losses or reduced exposure. The subsequent recovery patterns—with traditional brokers recovering to exceed pre-downturn levels by 2024 while Interactive Brokers and Robinhood lagged—further supports behavioral differences across wealth segments.

Detailed year-by-year analysis (Table XXVI) reveals platform-specific dynamics obscured by headline growth figures. Charles Schwab's average account size exhibits unusual volatility, dropping from $294,300 in 2019 to $225,800 in 2020 despite the TD Ameritrade acquisition, suggesting the absorbed accounts had lower average balances than Schwab's existing base. The metric's recovery to $277,000 by 2024 indicates either account con-

solidation or wealth accumulation among retained clients. Robinhood's trajectory from $2,800 in 2019 to $7,700 in 2024 represents a 175% increase, yet this growth primarily occurred during market recovery (2023-2024), implying that early adopters who survived the 2022 downturn accumulated assets while new user acquisition slowed.

Market composition analysis (Figure 20) illustrates how scale and average wealth create distinct competitive positions. Fidelity and Schwab occupy the upper-right quadrant with both high account counts (39.2 million and 36.5 million respectively) and substantial average assets, managing $15.1 trillion and $10.1 trillion in total. Interactive Brokers serves a niche of active traders—3.3 million accounts with moderate wealth ($170,200 average) totaling $568 billion. Robinhood's position—25.2 million accounts but only $193 billion in assets—suggests a fundamentally different business model focused on transaction volume and payment for order flow rather than asset-based fees. This segmentation appears stable: despite feature convergence across platforms, path dependence and switching costs maintain distinct client bases.

The persistence of these wealth disparities has profound implications for market microstructure and systemic risk. Platforms serving wealthier clients likely generate more stable fee revenue but face succession challenges as client demographics age. Conversely, platforms targeting younger, less wealthy investors must navigate the transition from transaction-based to asset-based revenue models as their client base matures. The documented behavioral differences during market stress—with wealth-weighted drawdowns ranging from 15% to 37%—suggest that market volatility disproportionately impacts platforms serving less affluent investors, potentially creating procyclical effects where market downturns reduce participation among price-sensitive retail traders precisely when their liquidity provision might be most valuable.

[Figure 19 about here.]

[Figure 20 about here.]

[Table 25 about here.]

[Table 26 about here.]

[Table 27 about here.]

B. Variable Definitions

I define all variables used in the regression analyses below. Unless otherwise specified, the sample period for variables is January 4, 2016, to December 31, 2024. These definitions largely follow standard conventions in the empirical finance literature.

B.1. A. Dependent Variables (Measured at $t+1$)

- $Return_{i,t+1}$: Total return of stock i on day $t+1$. Source: CRSP.
- $Open_to_Close_Return_{i,t+1}$: Intraday return of stock i on day $t+1$, calculated as:

$$Open_to_Close_Return_{i,t+1} = \frac{P^{close}_{i,t+1} - P^{open}_{i,t+1}}{P^{open}_{i,t+1}}$$

  where $P^{close}_{i,t+1}$ and $P^{open}_{i,t+1}$ are the closing and opening prices for stock i on day $t+1$, respectively. Source: CRSP.

- $Return^2_{i,t+1}$: Squared total return of stock i on day $t+1$, serving as a proxy for realized volatility. Source: CRSP.
- $Log(Intraday_Volatility_Q)_{i,t+1}$: Logarithm of quote-based intraday volatility during market hours for stock i on day $t+1$. The intraday volatility of stock i on day d during market hours, is calculated as $IVol_{i,d} = \frac{\sum_{t=1}^{N}(Ret_{i,t} - \overline{Ret}_{i,d})^2}{N-1}$, where N is the total number of NBBO observations for stock i on day d, $Ret_{i,t} = \ln\left(\frac{M_{i,t}}{M_{i,t-1}}\right)$ is the log return between consecutive observations with $M_{i,t} = \frac{B_{i,t}+A_{i,t}}{2}$ being the bid-ask midpoint at observation t (where $B_{i,t}$ and $A_{i,t}$ are the best bid and ask prices), and $\overline{Ret}_{i,d} = \frac{1}{N}\sum_{t=1}^{N} Ret_{i,t}$ is the mean return. For second-frequency data, when multiple NBBO updates occur within the same second, only the last update is used; for millisecond-frequency data, each NBBO update is treated as a separate event. This Intraday Volatility is also a dependent variable in Bartlett, McCrary, and O'Hara (2024). Source: Intraday Indicators based on NYSE TAQ.

- $IV_CATM_{i,t+1}$: Implied volatility of at-the-money call options for stock i on day $t+1$. This measure is constructed using call options with a time-to-expiration between 7 and 186 days, moneyness (stock price divided by strike price) between 0.95 and 1.05, and reported individual option implied volatilities between 0% and 200%. The resulting individual option implied volatilities that meet these criteria are then equal-weighted to produce the daily stock-level measure. The data covers the period January 4, 2016, to August 31, 2023. Source: OptionMetrics.

- $VWIV_Call_OM_{i,t+1}$: Volume-weighted implied volatility of all call options for stock i on day $t+1$ from OptionMetrics. This variable is also used in Bartlett, McCrary, and O'Hara (2024). The data covers the period January 4, 2016, to August 31, 2023. Source: OptionMetrics.

- $Quoted_Spread_{i,t+1}$: The time-weighted average percentage quoted spread for stock i on day $t+1$. Source: Intraday Indicators based on NYSE TAQ.

- $Effective_Spread_{i,t+1}$: The simple average percentage effective spread across all trades for stock i on day $t+1$. Source: Intraday Indicators based on NYSE TAQ.

- $Log(BidDepth)_{i,t+1}$: Logarithm of the time-weighted average dollar depth at the best bid quotes for stock i on day $t+1$. Source: Intraday Indicators based on NYSE TAQ.

- $Log(OfferDepth)_{i,t+1}$: Logarithm of the time-weighted average dollar depth at the best offer quotes for stock i on day $t+1$. Source: Intraday Indicators based on NYSE TAQ.

- $Log(Depth)_{i,t+1}$: Logarithm of the average dollar depth at the best bid and offer quotes for stock i on day $t+1$. Source: Intraday Indicators based on NYSE TAQ.

- $Log(Dollar_Volume)_{i,t+1}$: Logarithm of the total dollar trading volume for stock i on day $t+1$, where dollar volume is calculated as closing price multiplied by share volume. Source: CRSP.

- $Variance_Ratio_{i,t+1}$: Variance ratio for stock i on day $t+1$ during market hours, calculated as the absolute deviation from 1 of the ratio of 15-second return variance to three times the 5-second return variance:

$$Variance_Ratio_{i,t+1} = \left| \frac{\text{Var}(\text{Ret}_{15s,i,t+1})}{3 \times \text{Var}(\text{Ret}_{5s,i,t+1})} - 1 \right|$$

where $\text{Var}(\text{Ret}_{15s,i,t+1})$ is the variance of 15-second log returns and $\text{Var}(\text{Ret}_{5s,i,t+1})$ is the variance of 5-second log returns for stock i on day $t+1$. A lower value (closer to zero) indicates that the price process is closer to a random walk, suggesting higher pricing efficiency. Source: Intraday Indicators based on NYSE TAQ.

B.2. B. Main Independent Variable

- $Peer_Avg_Return_{i,t}$: Average return of peer stocks for focal stock i on day t. Peers are identified based on the correlation of daily RTAT activity between the focal stock and potential peer stocks over the prior 252 trading days. Stocks exhibiting a lagged correlation (focal stock's RTAT at t with peer's RTAT at $t-1$) greater than 0.4 and ranking within the top 20 by average market capitalization are classified as peers. This variable represents the simple arithmetic mean of these peers' returns on day t.

B.3. C. Control Variables (Measured at t for focal stock i)

- $Return_{i,t}$: Total return on day t. Source: CRSP.
- $Return_{i,[t-5,t-1]}$: Sum of total returns from day $t-5$ to $t-1$. Source: CRSP.
- $Log(Market_Cap)_{i,t}$: Logarithm of market capitalization (price times shares outstanding) at the end of day t. Source: CRSP.
- $Log(Price)_{i,t}$: Logarithm of the closing price on day t. Source: CRSP.
- $Log(Volume)_{i,t}$: Logarithm of share trading volume on day t. Source: CRSP.
- $Log(Book_to_Market)_{i,t}$: Logarithm of the book-to-market ratio. Book value is from the most recent quarterly financial statements available prior to day t. For

each trading date, this variable is matched with the accounting data from the last fiscal quarter ending before the current trading day ($q-1$). Source: Compustat.

- $Institutional_Ownership_{i,t}$: Institutional ownership percentage, derived from 13F filings. For each trading date, this variable is matched with the latest available 13F filing data prior to day t. Source: Thomson Reuters.

- $Num_Analysts_{i,t}$: Number of analysts covering the stock. For each trading date, this variable is matched with the latest available analyst coverage data prior to day t. Source: I/B/E/S.

- $ROA_{i,t}$: Return on assets, calculated using data from the most recent quarterly financial statements available prior to day t (matched as $q-1$). Source: Compustat.

- $ESS_{i,t}$: RavenPack Event Sentiment Score (ESS) for stock i on day t. The underlying RavenPack ESS is a granular score from -1.00 (negative) to +1.00 (positive) representing news sentiment for an entity. It is determined by systematically matching stories to over 6,700 categorized financial, economic, and geopolitical events. The score calibration considers expert consensus on expected short-term financial or economic impact and dynamically adjusts based on event-specific details like magnitudes, sentiment-conveying words, comparative values, and ratings disclosed in the story. For this study, the daily ESS for each stock is derived from processing 1,060,702,146 news records from the RavenPack 'Full Edition,' which integrates Dow Jones Newswire & Press Releases (100,127,414 records), web sources from over 40,000 news sources (960,574,732 records), and other content. My methodology involves applying filters for high entity relevance (score ≥ 75) and high event relevance (score ≥ 90, focusing on primary news events mentioned earliest and frequently, such as in headlines or initial paragraphs). To focus on new, impactful information while retaining significant recurring events, I require event similarity days ≥ 1 (indicating the news is not a duplicate from the same day) and exclude events with an original neutral sentiment (raw ESS = 0). The final stock-day $ESS_{i,t}$ metric used in regressions is a simple average of these filtered news events for stock i on day t. For

stock-days without news coverage, I assign a value of 0 in the regressions. The Full Edition provides coverage for 17.7% of stock-days across 7,981 securities. Source: RavenPack (processed).

- $CSS_{i,t}$: RavenPack Composite Sentiment Score (CSS) for stock i on day t. The underlying RavenPack CSS is a granular score from -1.00 to +1.00 for a news story, combining five specialized sentiment classifiers (PEQ - Global Equities, BEE - Earnings Evaluations, BMQ - Editorials & Commentary, BAM - Venture, Company, Mergers & Acquisitions, and BCA - Reports on Corporate Actions). Its direction reflects emotionally charged words/phrases and expert consensus on short-term share price impact, while its strength is empirically modeled on intraday stock price reactions of approximately 100 large-cap stocks. CSS typically hovers between -0.2 and +0.2, with more extreme values assigned when confidence in short-term signals is high. For this study, similar to the ESS, the daily CSS for each stock is derived from the same 1,060,702,146 news records from the RavenPack 'Full Edition' covering over 40,000 global news sources. I apply the same filtering criteria (entity relevance ≥ 75, event relevance ≥ 90, event similarity days ≥ 1, and exclusion of initially neutral sentiment if applicable to the underlying news items before aggregation) to process this large volume of news. The stock-day $CSS_{i,t}$ metric is a simple average of the CSS values associated with the filtered news events for stock i on day t. For stock-days without news coverage, I assign a value of 0 in the regressions. Source: RavenPack (processed).

- $Log(FINRA_Short_Vol)_{i,t}$: Logarithm of daily short sale volume on day t. Source: FINRA.

- **Factor Loadings and Risk Measures**: Estimated from rolling 252-day regressions on the Fama-French-Carhart four-factor model (market (MKT), size (SMB), value (HML), and momentum (UMD)). These include:

  - $Alpha_FF4_{i,t}$: Daily alpha (intercept) from the Fama-French-Carhart four-factor model regression.

  - $Beta_MKT_{i,t}$: Market factor loading.
  - $Beta_SMB_{i,t}$: Size (SMB) factor loading.
  - $Beta_HML_{i,t}$: Value (HML) factor loading.
  - $Beta_UMD_{i,t}$: Momentum (UMD) factor loading.

  Source: Estimated using CRSP returns and Fama-French factor data.

- $IVOL_FF4_{i,t}$: Idiosyncratic volatility based on the residuals from the rolling 252-day Fama-French-Carhart four-factor model regressions, calculated as the standard deviation of residuals. Source: Estimated using CRSP returns and Fama-French factor data.

C. RHDW Fractional Share Estimates

Bartlett, McCrary, and O'Hara (2023) and Bartlett, McCrary, and O'Hara (2024) (hereafter BMO) develop a latency-based methodology to identify fractional share trades executed by Robinhood and DriveWealth (hereafter RHDW) in Trade and Quote (TAQ) data. Their approach exploits systematic differences in execution latencies—the time between trade execution and reporting—to isolate retail fractional trades from the universe of off-exchange one-share transactions. Specifically, BMO identify Robinhood trades using latencies between 135-300 milliseconds on Nasdaq Trade Reporting Facilities (TRFs) and DriveWealth trades using latencies exceeding 20 milliseconds on the NYSE TRF. And after October 7, 2021, DriveWealth trades are also on Nasdaq TRF. Since the latency range of Robinhood (135-300ms) overlaps with that of Drivewealth (>20ms), the trades of Robinhood and Drivewealth are combined as RHDW on and after October 7, 2021. As institutional investors rarely engage in fractional share trading, these RHDW estimates provide a relatively pure measure of retail trading activity.

This paper extends and refines the BMO methodology through December 31, 2024. My analysis reveals that the original classification parameters become increasingly inaccurate during 2024, necessitating methodological adjustments to maintain measurement validity. Specifically, I document a previously unidentified change in RHDW's latency distribution

that renders the original 20-millisecond latency threshold obsolete for trades occurring after mid-2024.

C.1. Evidence of Classification Breakdown

Application of the original BMO parameters to 2024 data produces anomalous results. Figures 21 and 22 illustrate a sharp decline in estimated RHDW fractional trades beginning in mid-2024, followed by a partial recovery—a pattern inconsistent with actual retail trading activity. This divergence between estimated and disclosed trades suggests a structural change in the data-generating process rather than a genuine shift in trading behavior.

To diagnose the source of this breakdown, I conduct two tests. First, I examine whether changes in TRF usage explain the anomaly by re-estimating RHDW trades without TRF restrictions, including all off-exchange one-share trades exceeding the 20-millisecond latency threshold regardless of reporting facility. The persistence of the mid-2024 anomaly under this specification rules out TRF migration as the primary cause. Second, I analyze the latency distribution of RHDW trades over time, which reveals a downward shift in reporting latencies beginning in mid-2024.

C.2. Methodological Refinement

Based on this analysis, I implement a time-varying latency threshold for DriveWealth classification. The refined methodology maintains BMO's original parameters through January 1, 2024, then reduces the DriveWealth latency threshold from 20 to 10 milliseconds for subsequent periods. This adjustment reflects DriveWealth's apparent infrastructure improvements that reduced their characteristic reporting latencies while preserving the systematic differences that enable identification.

The refined classification algorithm proceeds as follows: For each trading day, I extract all off-exchange trades (Exchange = 'D') with volume equal to one share. I calculate reporting latency as the difference between TRF and participant timestamps in nanoseconds, converted to milliseconds. For dates before October 7, 2022, I classify trades as Robinhood if they exhibit latencies between 135-300 milliseconds on Nasdaq TRFs (B,

T, or Q) and as DriveWealth if latencies exceed 20 milliseconds on the NYSE TRF (N). Following DriveWealth's documented TRF transition on October 7, 2022, I combine both brokers' trades on Nasdaq TRFs, applying a 20-millisecond minimum latency through January 1, 2024, and 10 milliseconds thereafter.

C.3. Validation and Results

Figures 23 and 24 demonstrate that the refined methodology successfully eliminates the mid-2024 anomaly, producing estimates that closely track FINRA OTC Transparency disclosures throughout the extended sample period. To formally validate the approach, I aggregate daily estimates to weekly frequencies and regress disclosed RHDW trades from FINRA data on my classified estimates. Table XXVIII presents results for 31,312 stock-week observations from November 1, 2021, through December 31, 2024, restricted to periods with positive DriveWealth activity.

The validation regression yields a coefficient of 0.938 (s.e. = 0.0006) on estimated trades, nearly identical to BMO's reported coefficient of 0.93, with an R^2 of 0.986. These results confirm that the refined methodology maintains classification accuracy throughout the extended sample despite structural changes in trade reporting infrastructure. The slight underestimation (coefficient < 1) likely reflects fractional trades that lack the characteristic latency signatures, consistent with BMO's original findings.

C.4. Implementation Details

Researchers seeking to replicate these estimates should note several critical implementation details. First, the analysis requires TAQ millisecond data files containing participant and TRF timestamps, trade volume, exchange indicators, and sale conditions. Second, proper handling of market holidays is essential when mapping daily estimates to weekly FINRA disclosures—I adjust Monday holidays to Tuesday to align with FINRA's reporting conventions. Third, trades with special sale conditions (F, 4, C, Q, or M) should be excluded as they represent non-standard executions.

[Table 28 about here.]

[Figure 21 about here.]

[Figure 22 about here.]

[Figure 23 about here.]

[Figure 24 about here.]

List of Figures

1 Growth in Retail Brokerage Accounts, 2016-2024 . . . 79
2 Total Client Assets Under Administration (AUA), 2016-2024 . . . 80
3 Robinhood's "People Also Own" Feature for SOXL . . . 81
4 Daily RTAT Activity Composition by Share Code Category . . . 82
5 Daily RTAT Activity Composition by Exchange Code Category . . . 83
6 RTAT Activity Time Series for Selected Tickers. Panel A: Technology Leaders (TSLA, NVDA). Panel B: Index ETFs (SPY, QQQ). Panel C: Leveraged ETFs (TQQQ, SOXL). Panel D: Meme Stocks (GME, AMC). Source: Nasdaq RTAT dataset. . . . 84
7 RTAT vs. Dollar Volume Share for Technology Leaders . . . 85
8 RTAT vs. Dollar Volume Share for Index ETFs . . . 86
9 RTAT vs. Dollar Volume Share for Leveraged ETFs . . . 87
10 RTAT vs. Dollar Volume Share for Meme Stocks . . . 88
11 RTAT vs. Dollar Volume Share for ADRs and Foreign Listings . . . 89
12 Distribution of Lead-Lag Correlations and Peer Group Sizes. Left Panel: Distribution of lead-lag correlations between $RTAT_{i,t}$ and $RTAT_{j,t-1}$ for all stock pairs. The red dashed line indicates the 0.4 threshold. Right Panel: Distribution of the number of peers per firm, capped at 20. Source: Nasdaq RTAT dataset. . . . 90
13 Market Capitalization (MSTR vs. IBIT) and IBIT Price . . . 91
14 IBIT and MSTR Retail Trading Activity (RTAT) . . . 92
15 Peer Candidate RTAT Correlation Analysis . . . 93
16 Peer Selection: Correlation vs. Market Cap (Top 20 Highlighted) . . . 94
17 IBIT Retail Trading Peers by Group . . . 95
18 Time Series of Peer Return Coefficients from Fama-MacBeth Regressions . . . 96
19 Average Asset Size per Client Account, 2016-2024 . . . 97
20 Retail Brokerage Market Composition by Account Size and Client Assets, 2024 . . . 98
21 Tesla (TSLA): Combined Estimated vs Total Disclosed Fractional Trades (20ms Threshold) . . . 99
22 Apple (AAPL): Combined Estimated vs Total Disclosed Fractional Trades (20ms Threshold) . . . 100
23 Tesla (TSLA): Combined Estimated vs Total Disclosed Fractional Trades (Refined Methodology) . . . 101
24 Apple (AAPL): Combined Estimated vs Total Disclosed Fractional Trades (Refined Methodology) . . . 102

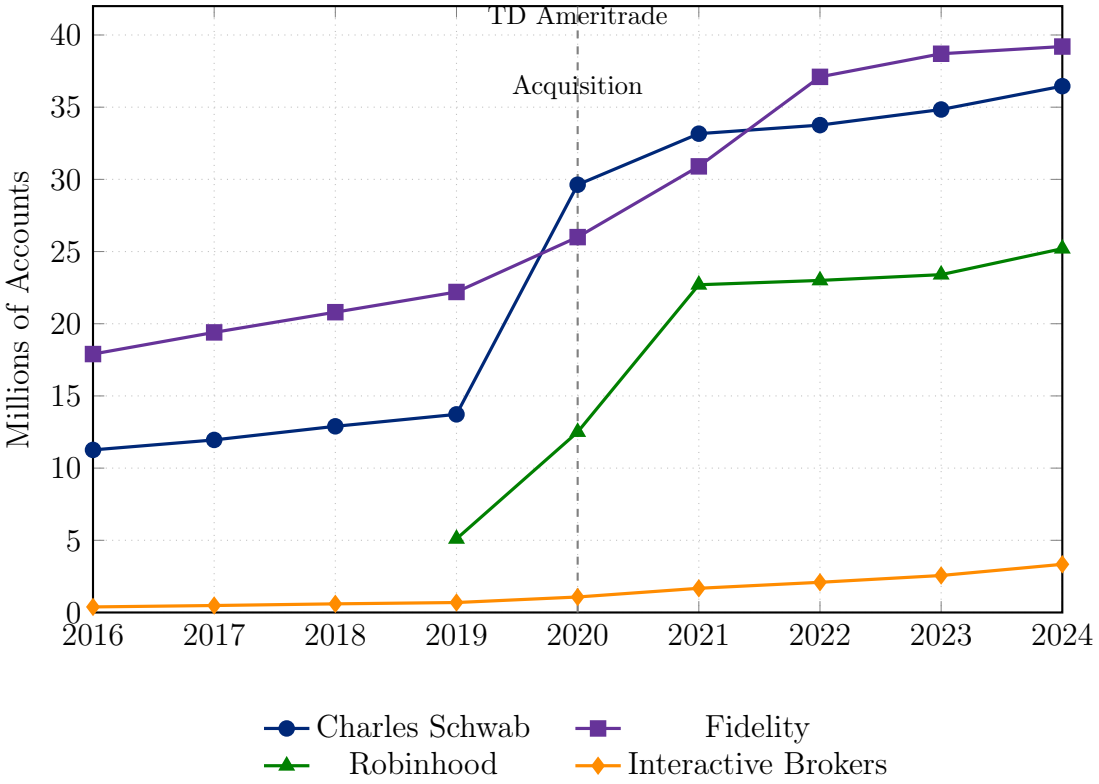


Figure 1. Growth in Retail Brokerage Accounts, 2016-2024

Notes: Charles Schwab reports Active Brokerage Accounts. The 2020 figure includes 14.5 million new brokerage accounts related to the acquisition of TD Ameritrade and 1.1 million new brokerage accounts related to the acquisition of the assets of USAA's Investment Management Company. Fidelity metrics refer to Total retail accounts for 2016-2023. In 2024, the definition in the Fidelity Annual Report changed slightly to 39.2 million unique customers engaging with Fidelity.com, NetBenefits.com, or mobile apps, including unique customers who engage on Fidelity.com, NetBenefits.com, the Fidelity Investments Mobile App, or the NetBenefits Mobile App. Robinhood reports Net Cumulative Funded Accounts (2019-2022) and Funded Customers (2023-2024). IBKR reports Client Accounts.

Sources: Charles Schwab annual reports (2016-2024); Fidelity shareholder update (2018), annual reports (2020-2024), and Q3 2021 Business Update; Robinhood S-1 filing (2019), SEC filings, and annual reports (2020-2024); Interactive Brokers annual reports (2016-2024). See Table XXVII for complete source documentation.

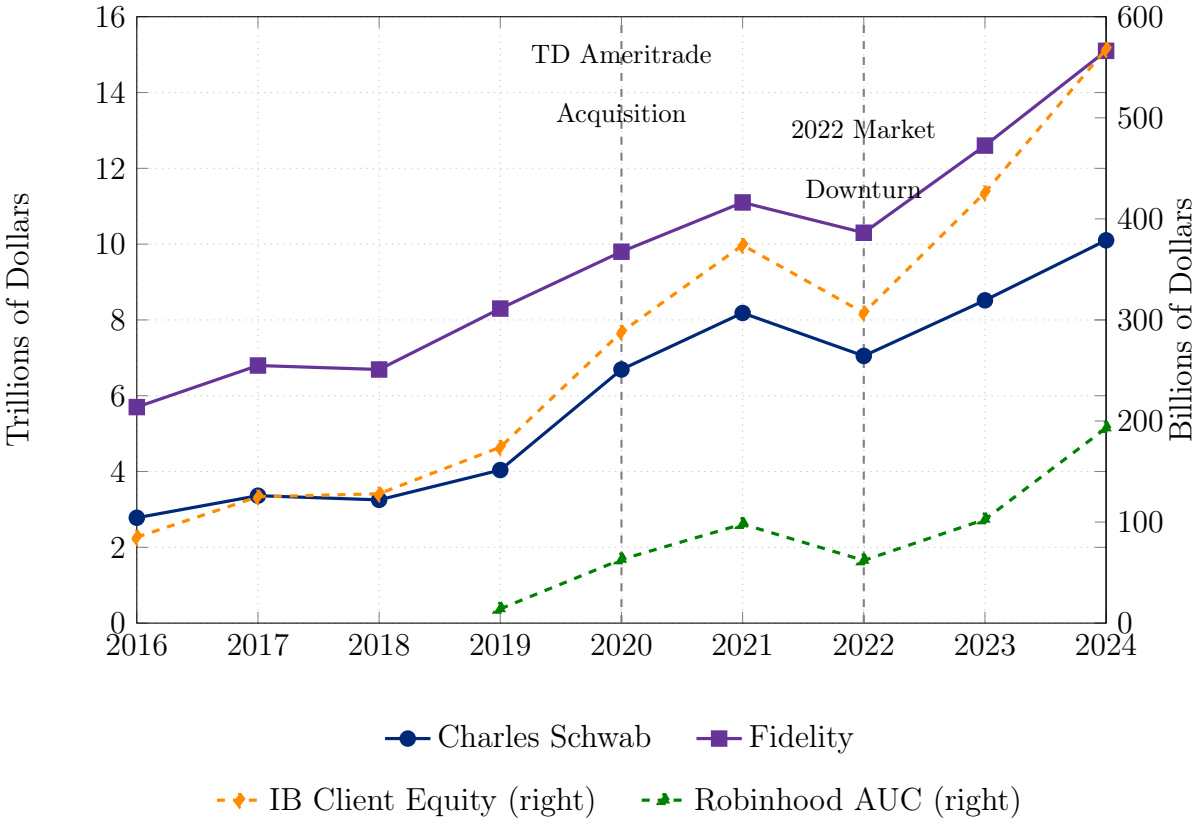


Figure 2. Total Client Assets Under Administration (AUA), 2016-2024
Notes: Charles Schwab and Fidelity report Assets Under Administration (AUA). IBKR reports Client Equity. Robinhood reports Assets Under Custody (AUC). Charles Schwab Client Assets for 2020 totaled $6,691.7 billion. The 2020 figure includes the impact of 14.5 million new brokerage accounts related to the acquisition of TD Ameritrade and 1.1 million new brokerage accounts related to the acquisition of the assets of USAA's Investment Management Company.
Sources: Same as Figure 1. See Table XXVII for complete source documentation.

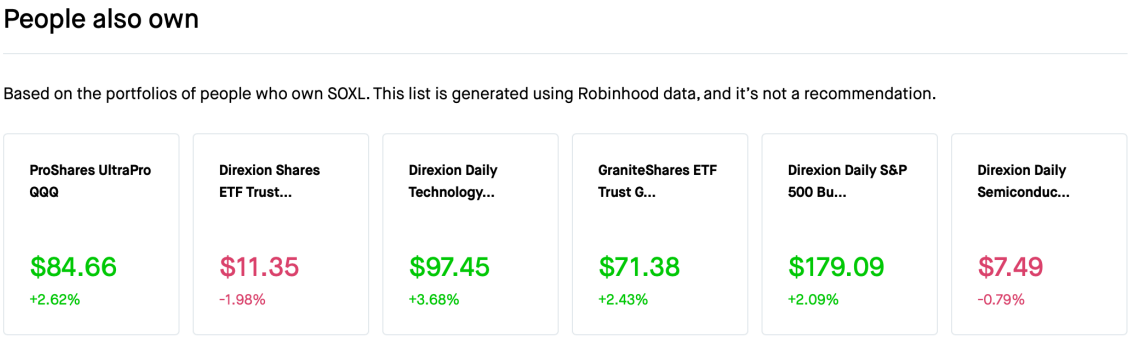


Figure 3. Robinhood's "People Also Own" Feature for SOXL

Note: This is the "People Also Own" section for SOXL from the Robinhood (https://robinhood.com/stocks/SOXL?source=search). SOXL provides 3x daily exposure to a modified market-cap-weighted index of 30 US-listed semiconductor companies. The listed name for SOXL is Direxion Daily Semiconductor Bull 3X Shares. Its peers in this section include TQQQ (ProShares UltraPro QQQ), TSLL (Direxion Shares ETF Trust Direxion Daily TSLA Bull 2X Shares), NVDL (GraniteShares ETF Trust GraniteShares 2x Long NVDA Daily ETF), TECL (Direxion Daily Technology Bull ETF), SOXS (Direxion Daily Semiconductor Bear 3X Shares), and TNA (Direxion Daily Small Cap Bull ETF). This list is generated by Robinhood using its customer data.

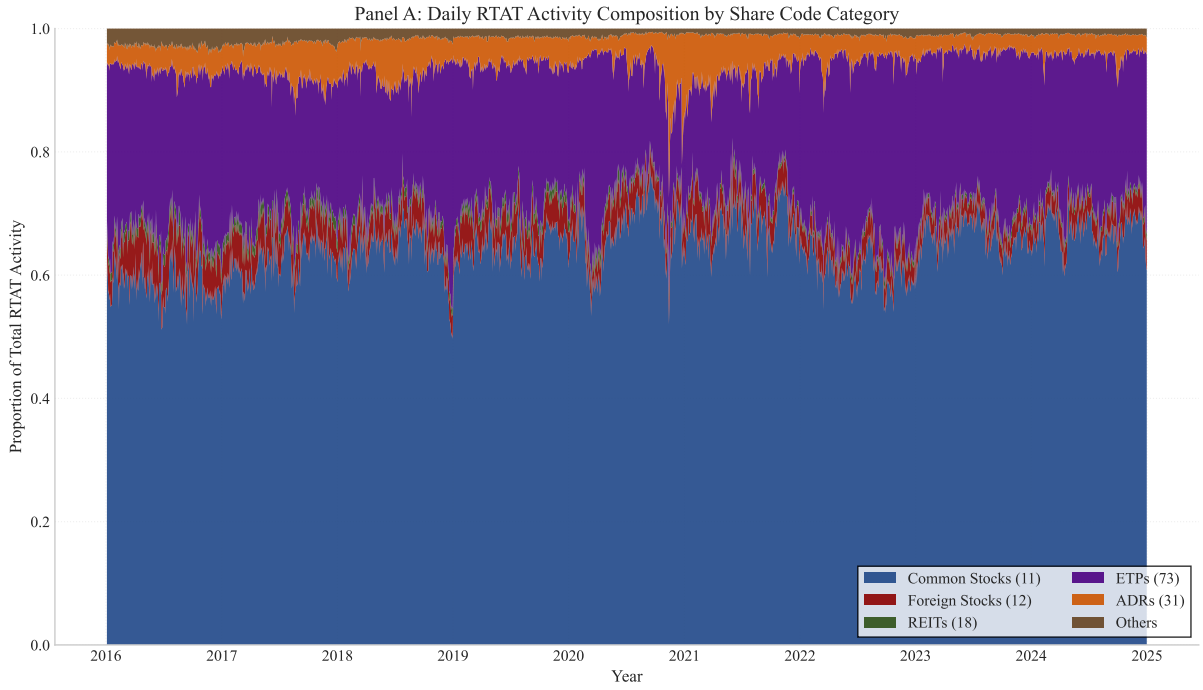


Figure 4. Daily RTAT Activity Composition by Share Code Category

Note: This figure displays the daily composition of retail trading activity across different share code categories from January 2016 to December 2024. Share codes are based on CRSP classifications: Common Stocks (SHRCD 11), Foreign Stocks (SHRCD 12), REITs (SHRCD 18), ETPs (SHRCD 73), and ADRs (SHRCD 31). The stacked area chart shows the proportion of total retail trading activity (measured in U.S. dollars) allocated to each category on a daily basis. Note that traditionally, CRSP distinguished between share codes 11 (standard, ordinary common stock) and 10 (a more general category for common shares that don't fit a more specific classification). In the current CRSP dataset, these categories have been consolidated into share code 11, which I use throughout this analysis. Bartlett, McCrary, and O'Hara (2024) also follow this practice of using only share code 11.

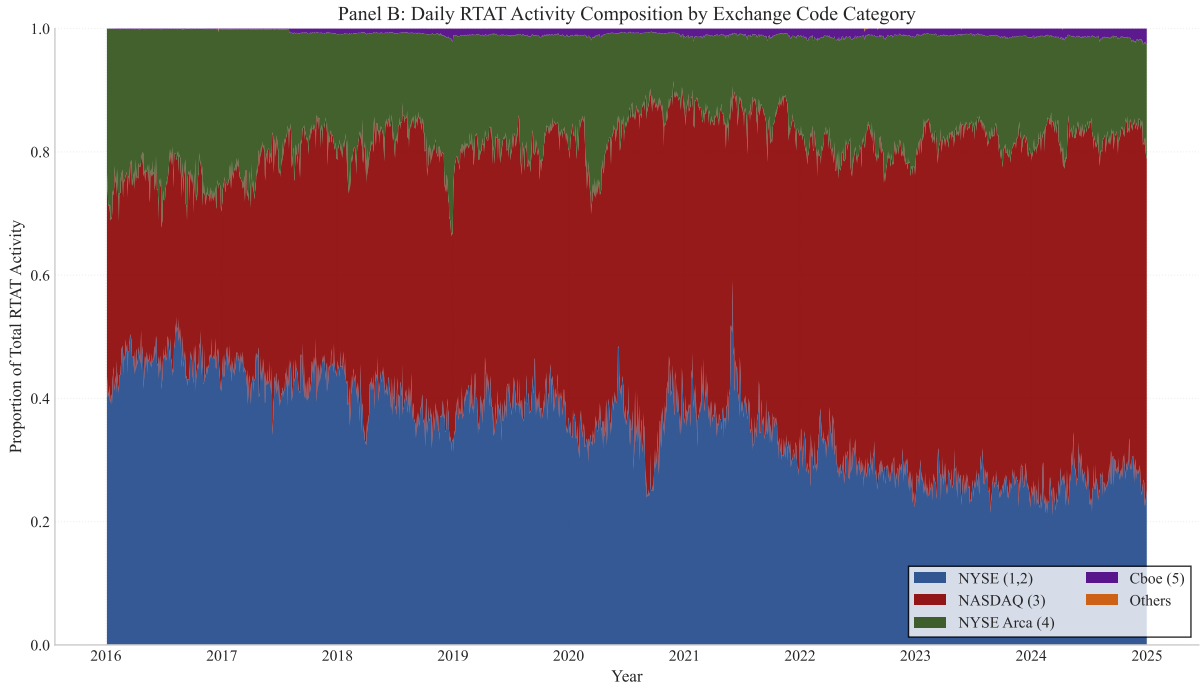


Figure 5. Daily RTAT Activity Composition by Exchange Code Category

Note: This figure presents the daily composition of retail trading activity across different exchanges from January 2016 to December 2024. Exchange codes are based on CRSP classifications: NYSE (exchcd 1,2), NASDAQ (exchcd 3), NYSE Arca (exchcd 4), and Cboe (exchcd 5). The stacked area chart displays the proportion of total retail trading activity (measured in U.S. dollars) executed on each exchange on a daily basis.

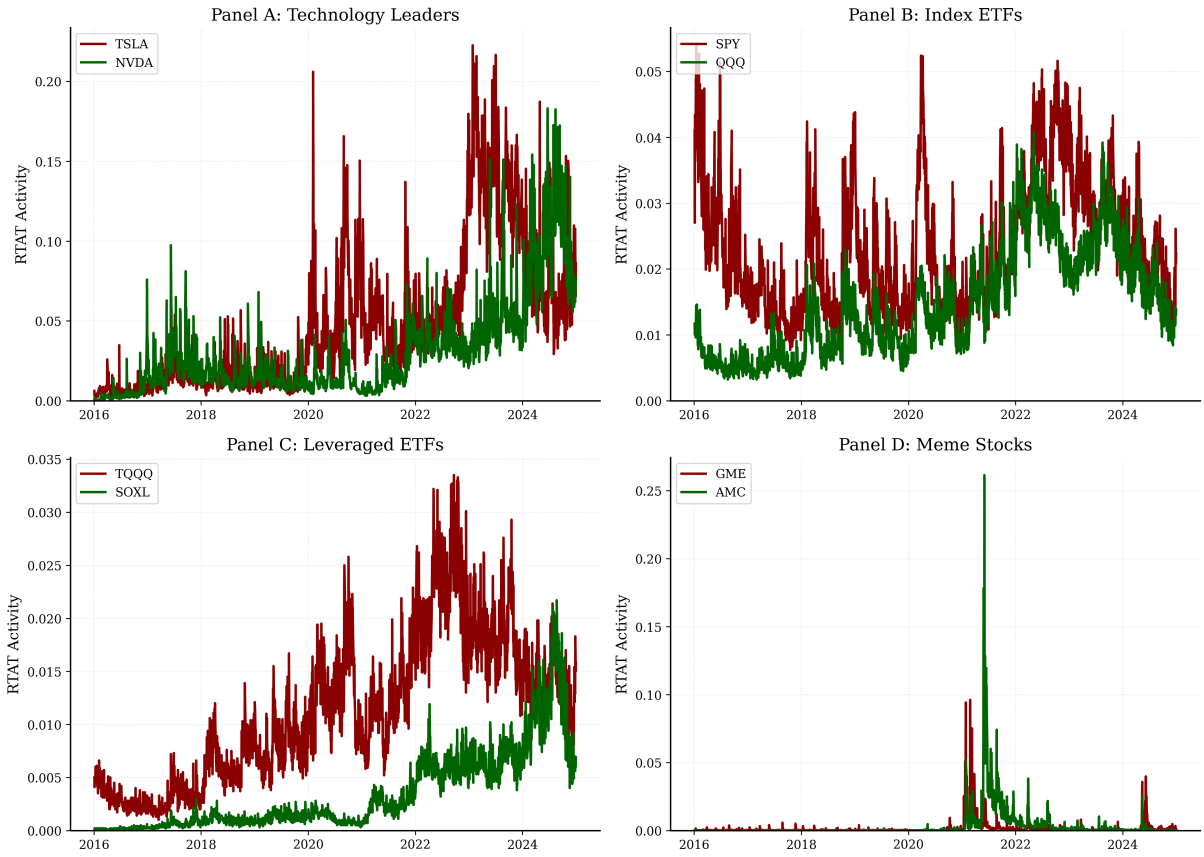
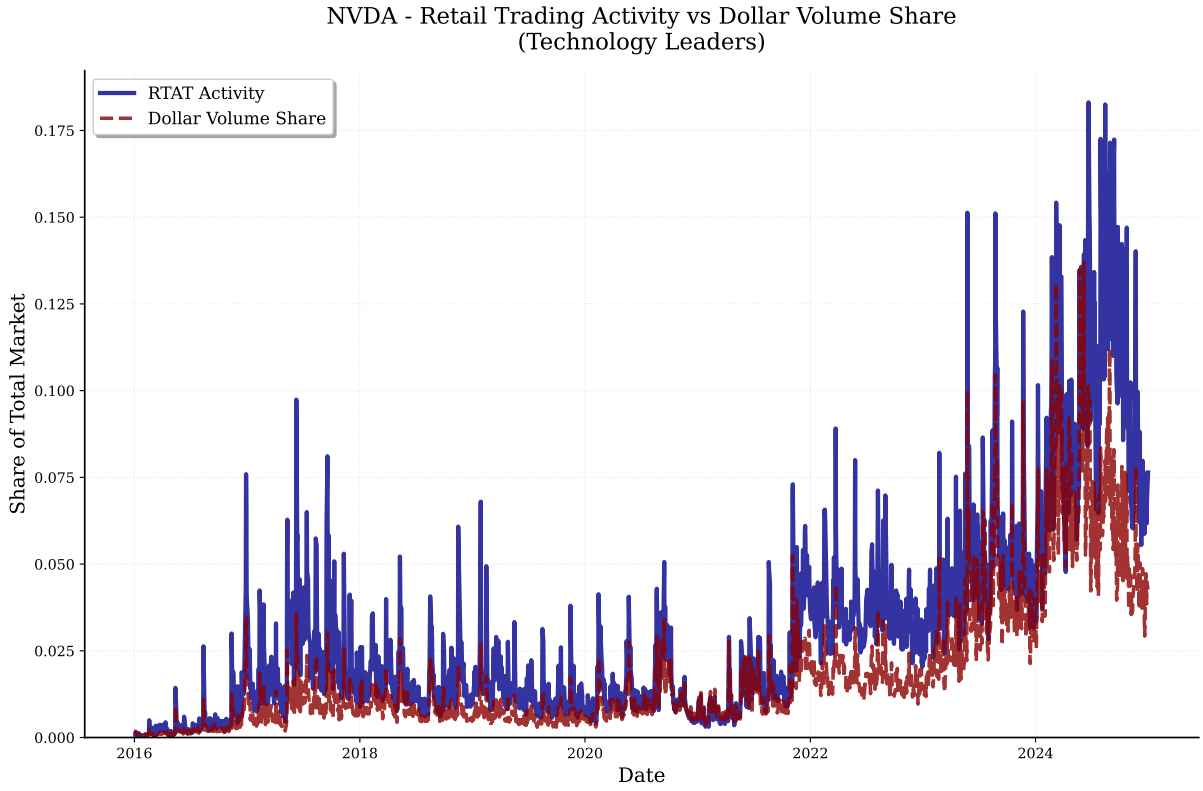
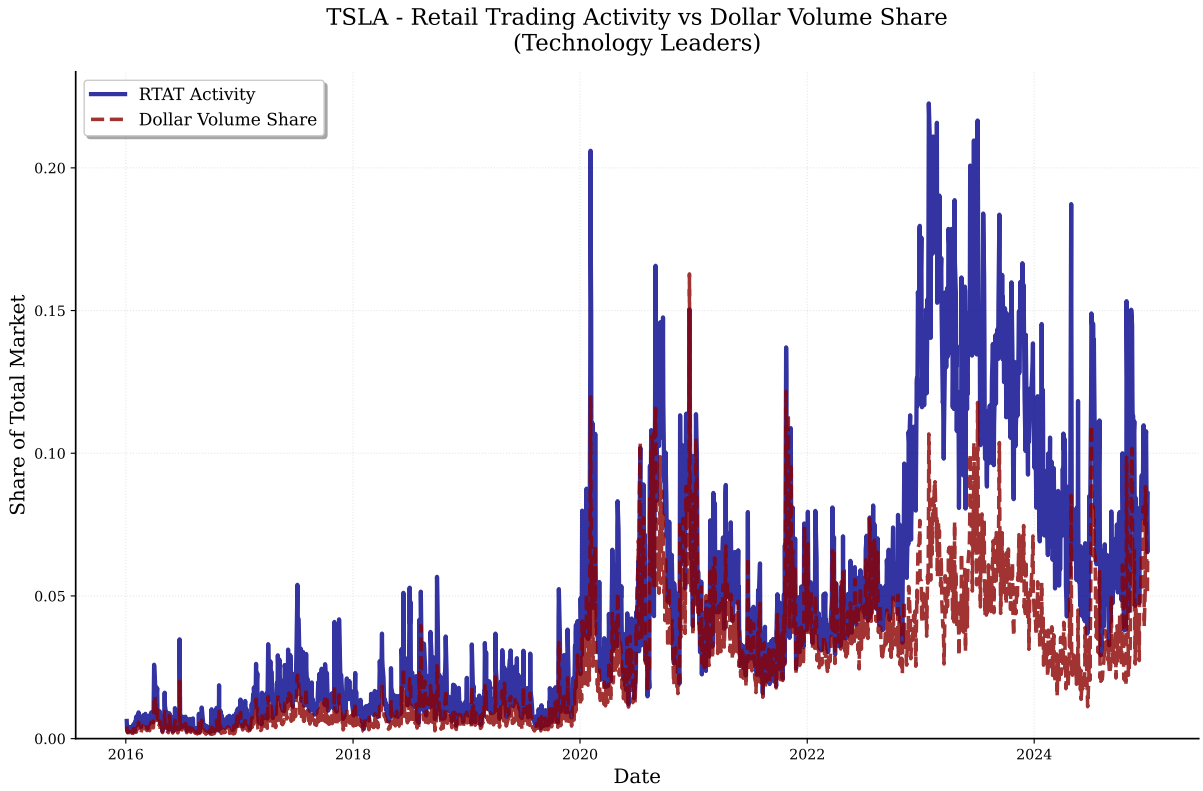


Figure 6. RTAT Activity Time Series for Selected Tickers. Panel A: Technology Leaders (TSLA, NVDA). Panel B: Index ETFs (SPY, QQQ). Panel C: Leveraged ETFs (TQQQ, SOXL). Panel D: Meme Stocks (GME, AMC). Source: Nasdaq RTAT dataset.

Figure 7. RTAT vs. Dollar Volume Share for Technology Leaders

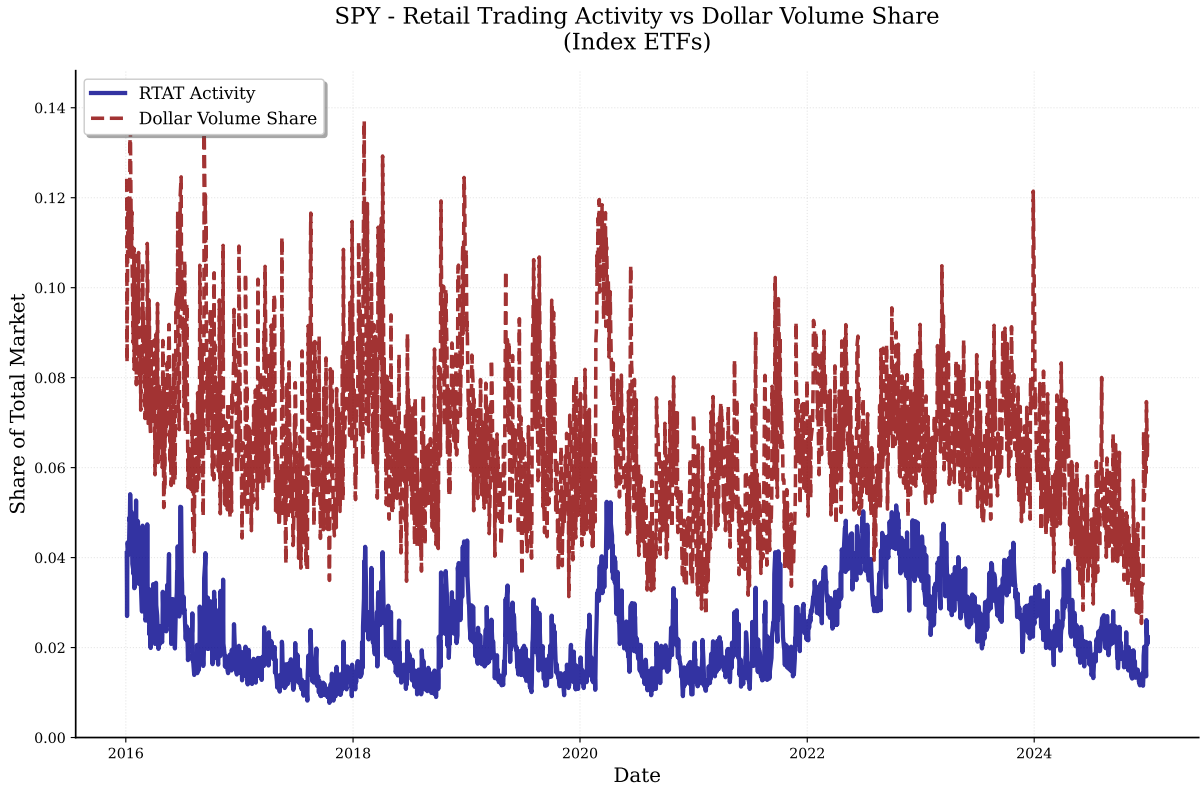


(a) NVDA

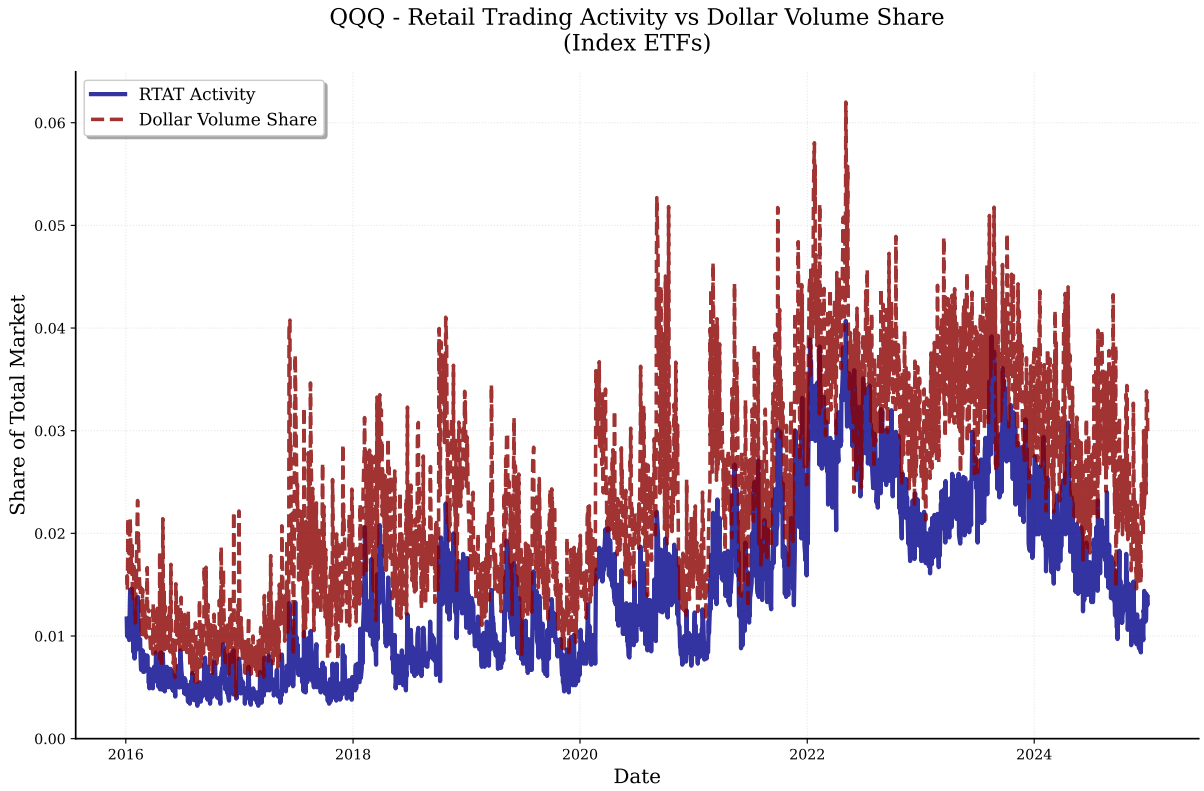


(b) TSLA

Figure 8. RTAT vs. Dollar Volume Share for Index ETFs

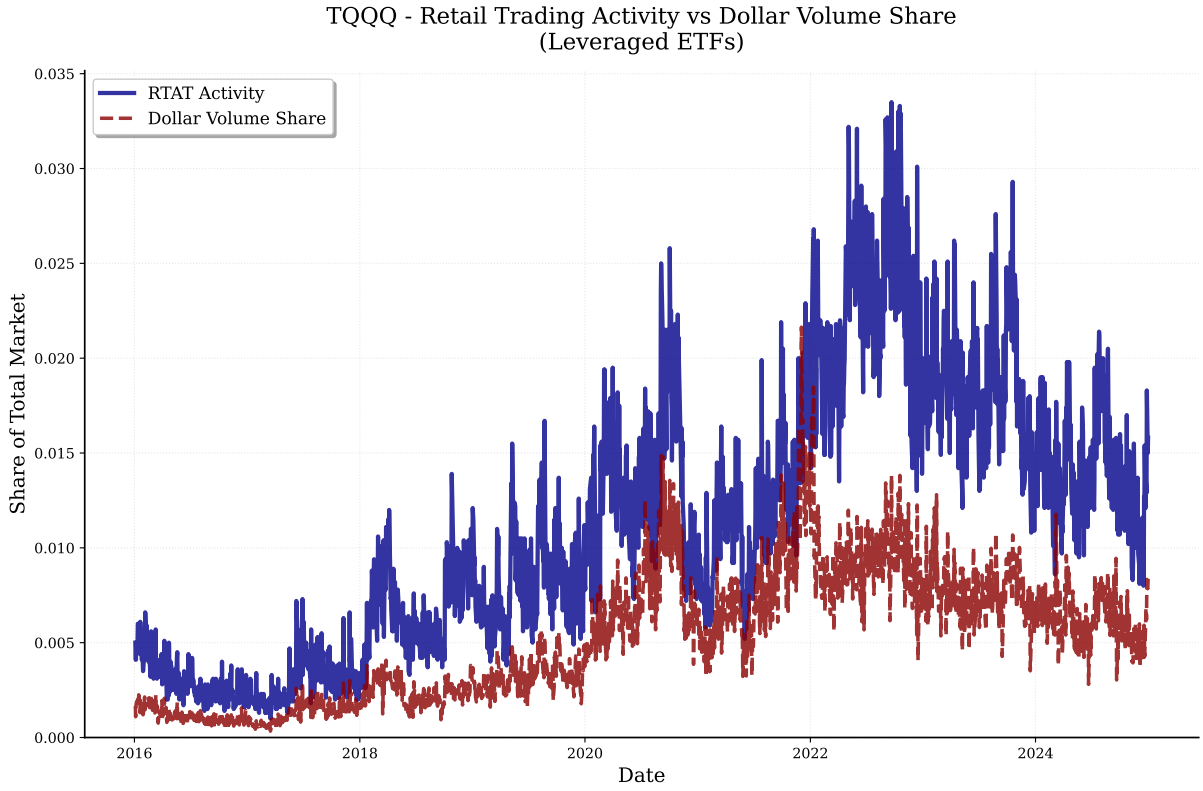


(a) SPY

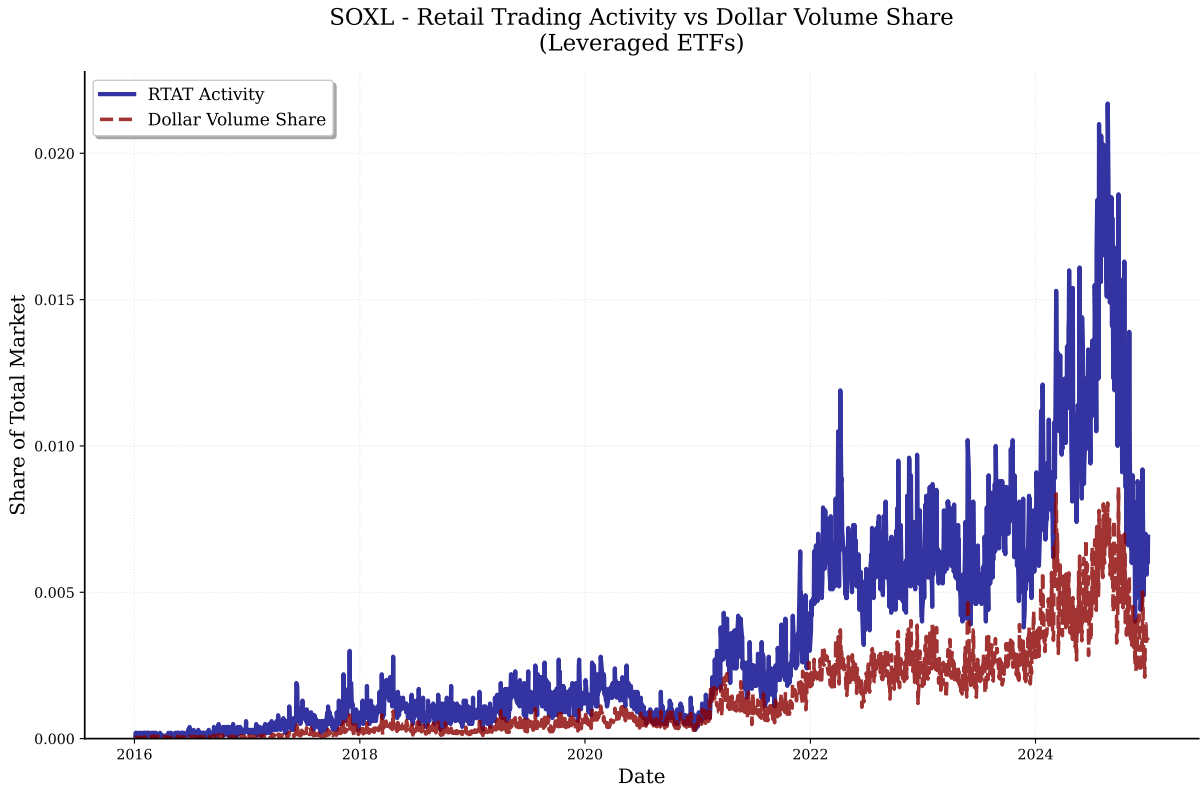


(b) QQQ

Figure 9. RTAT vs. Dollar Volume Share for Leveraged ETFs

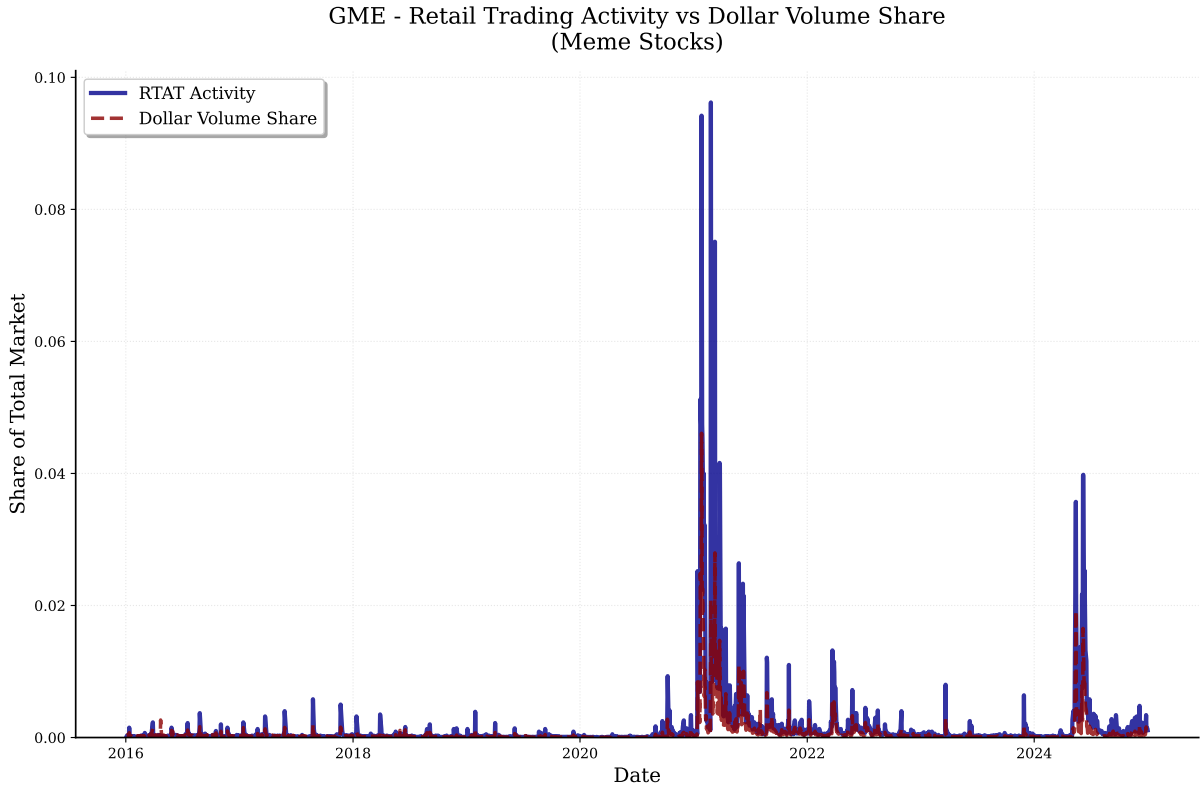


(a) TQQQ

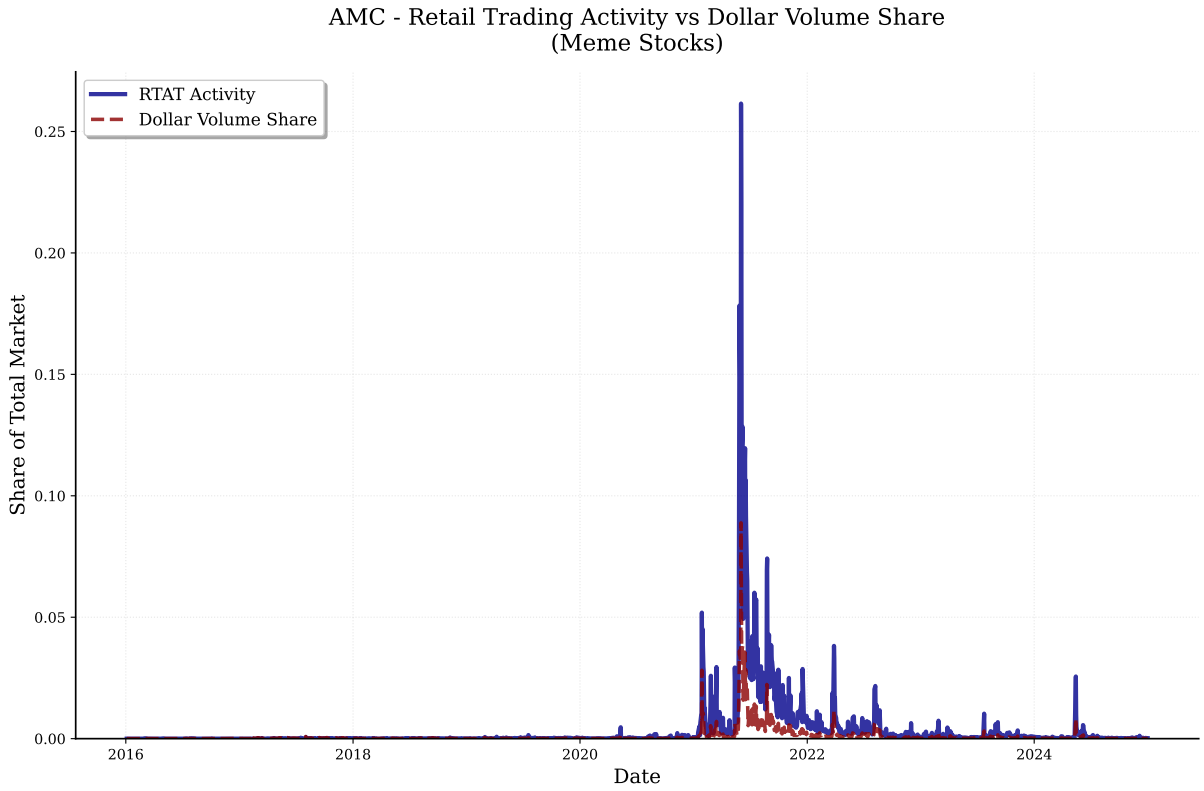


(b) SOXL

Figure 10. RTAT vs. Dollar Volume Share for Meme Stocks

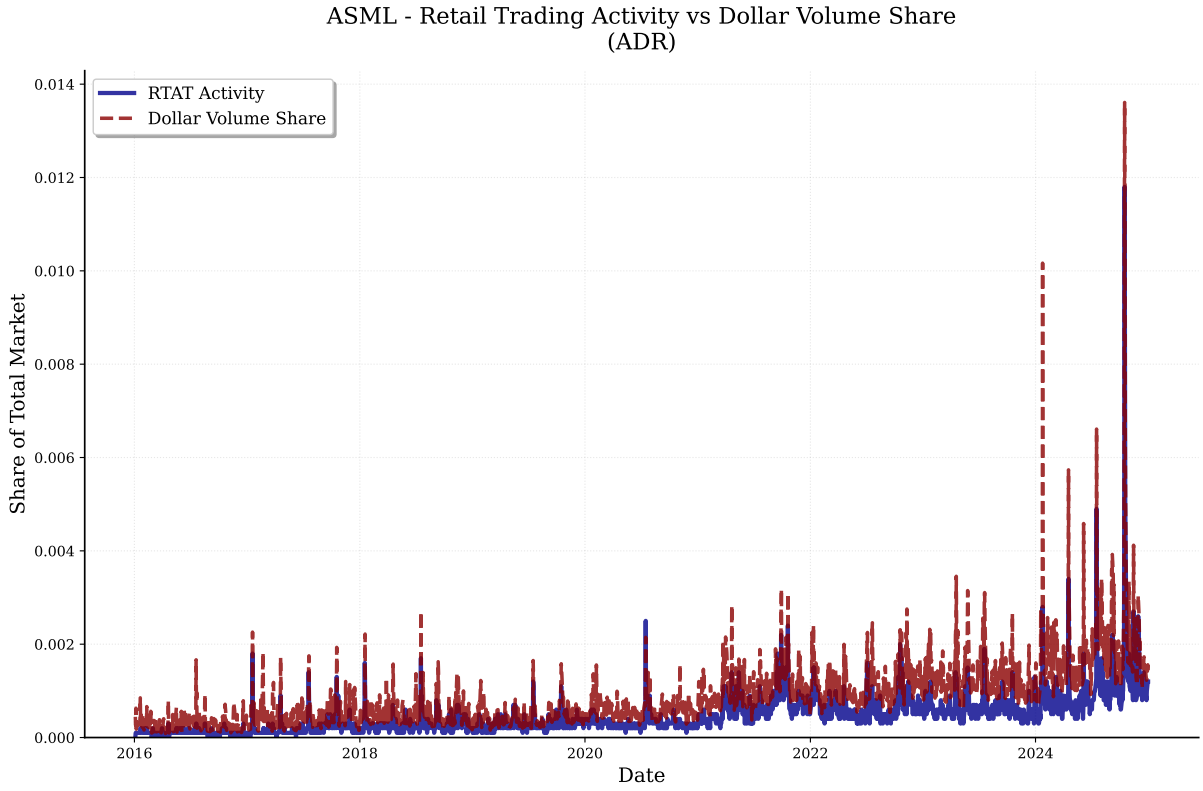


(a) GME

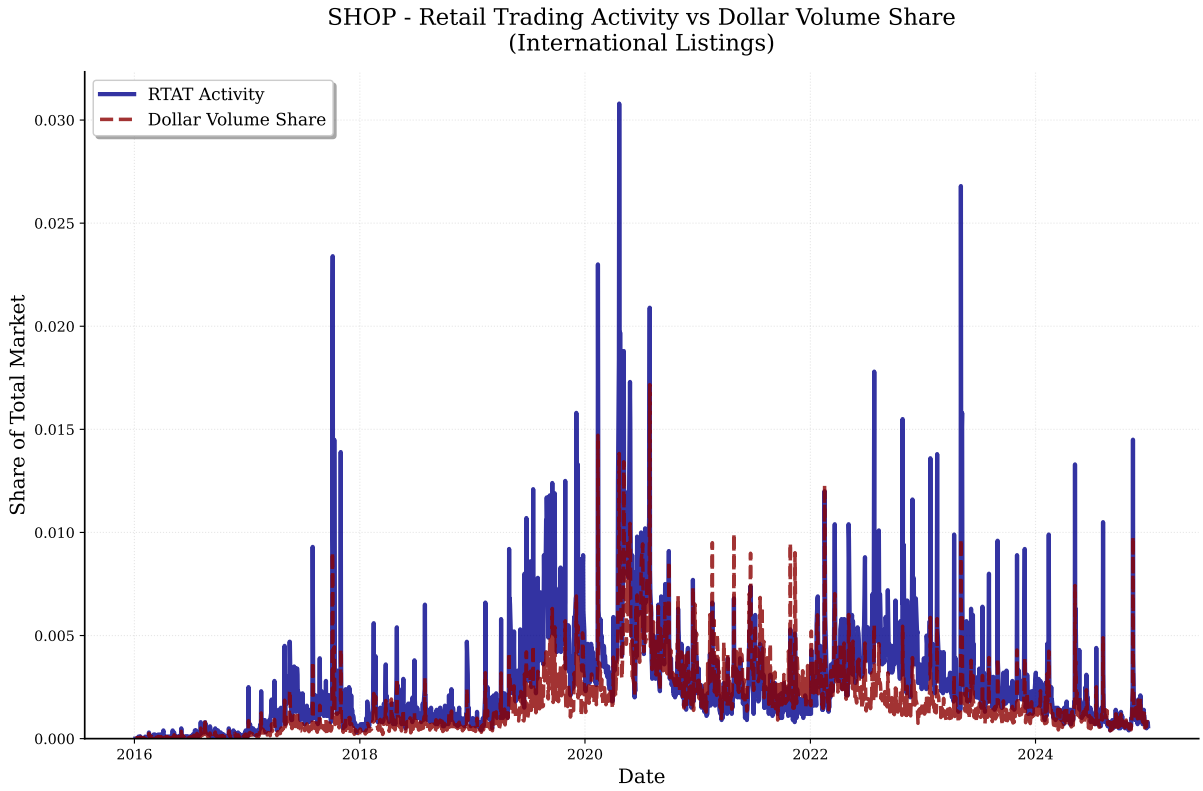


(b) AMC

Figure 11. RTAT vs. Dollar Volume Share for ADRs and Foreign Listings



(a) ASML (ADR)



(b) SHOP (Foreign Direct Listing)

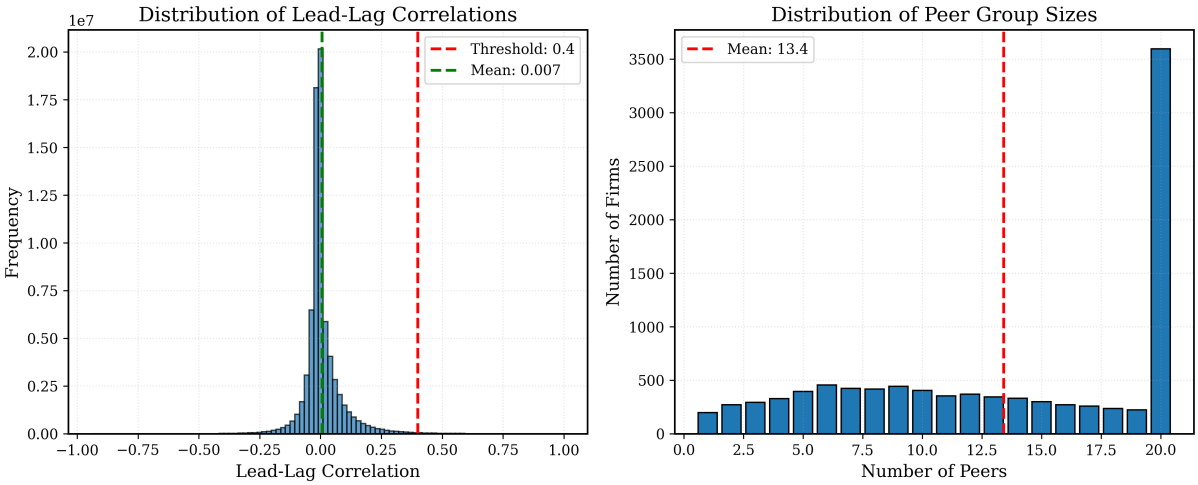


Figure 12. Distribution of Lead-Lag Correlations and Peer Group Sizes. Left Panel: Distribution of lead-lag correlations between $RTAT_{i,t}$ and $RTAT_{j,t-1}$ for all stock pairs. The red dashed line indicates the 0.4 threshold. Right Panel: Distribution of the number of peers per firm, capped at 20. Source: Nasdaq RTAT dataset.

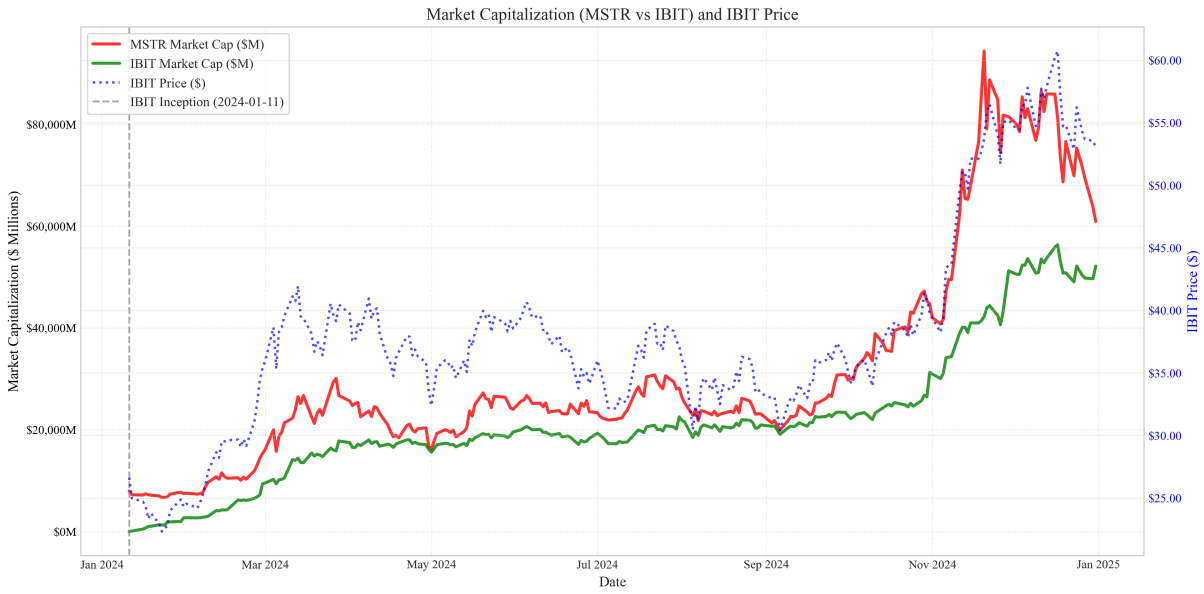


Figure 13. Market Capitalization (MSTR vs. IBIT) and IBIT Price

Notes: This figure displays the market capitalization of IBIT (green line, left axis) and MSTR (red line, left axis), along with IBIT's price (blue dotted line, right axis) from January 2024 to January 2025. IBIT's inception date is marked with a vertical dashed line. Both IBIT and MSTR market capitalizations show significant growth, with IBIT starting much lower and growing rapidly. Source: CRSP.

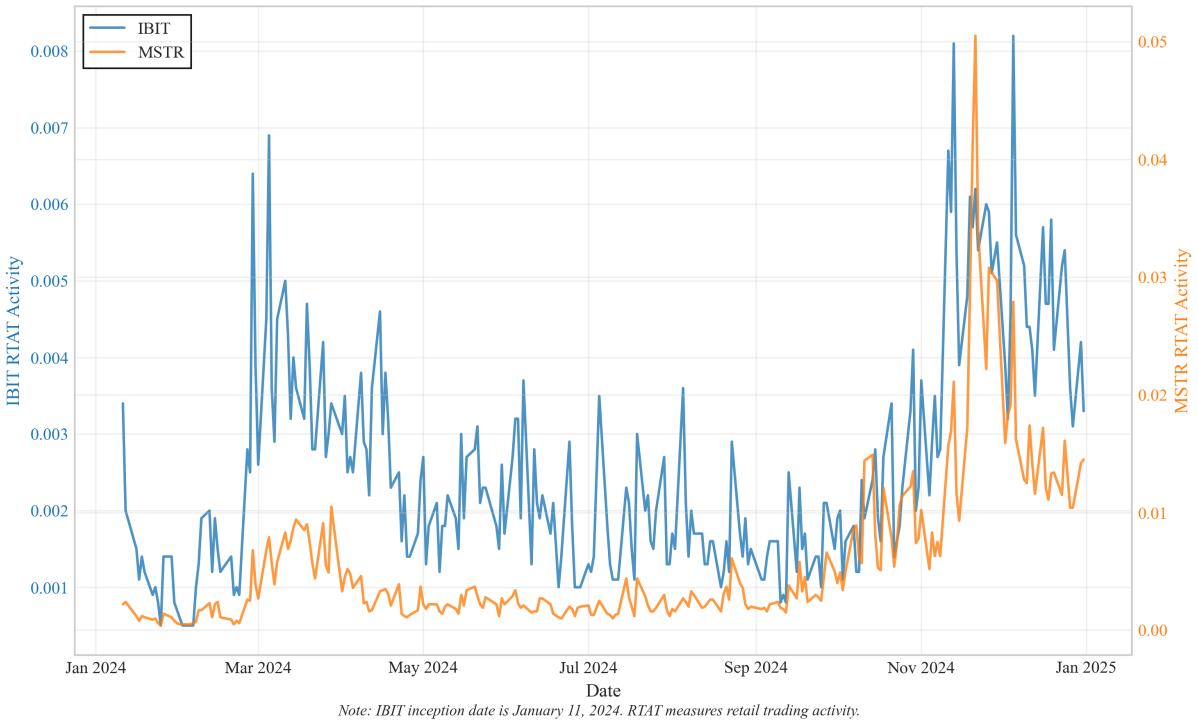


Figure 14. IBIT and MSTR Retail Trading Activity (RTAT)

Notes: This figure shows the time series of IBIT RTAT (blue line, left y-axis) and MSTR RTAT (orange line, right y-axis) from January 2024 to January 2025. Both series show volatility and some degree of comovement, with peaks in retail activity. IBIT inception date is January 11, 2024. RTAT measures retail trading activity. Source: Nasdaq RTAT dataset.

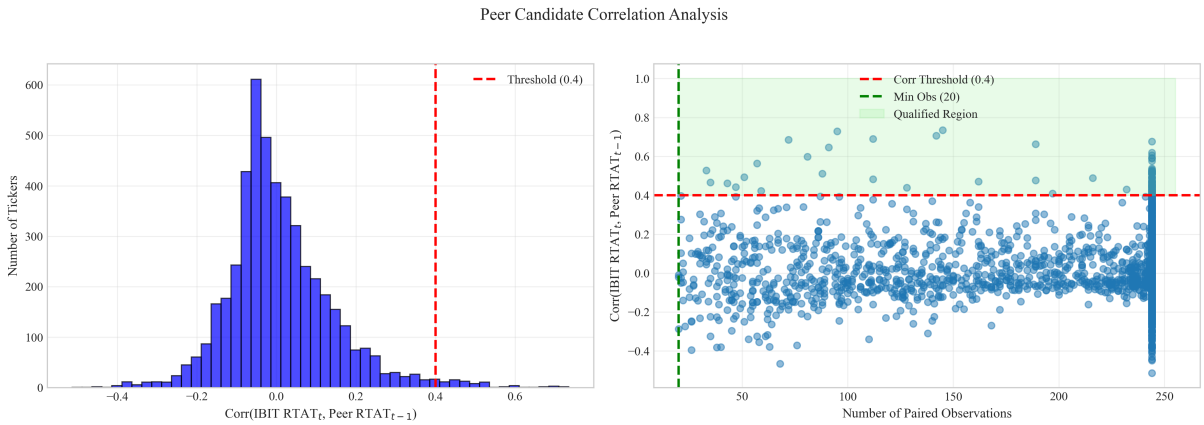


Figure 15. Peer Candidate RTAT Correlation Analysis

Notes: Two-panel figure. Left panel: Histogram of correlations Corr($RTAT_{IBIT,t}$, $RTAT_{Peer,t-1}$). A vertical red dashed line shows the 0.4 threshold. Right panel: Scatter plot of correlation versus number of paired observations. Horizontal red dashed line at correlation = 0.4, vertical green dashed line at minimum observations = 20. Green shaded area shows qualified region. Source: Nasdaq RTAT dataset.

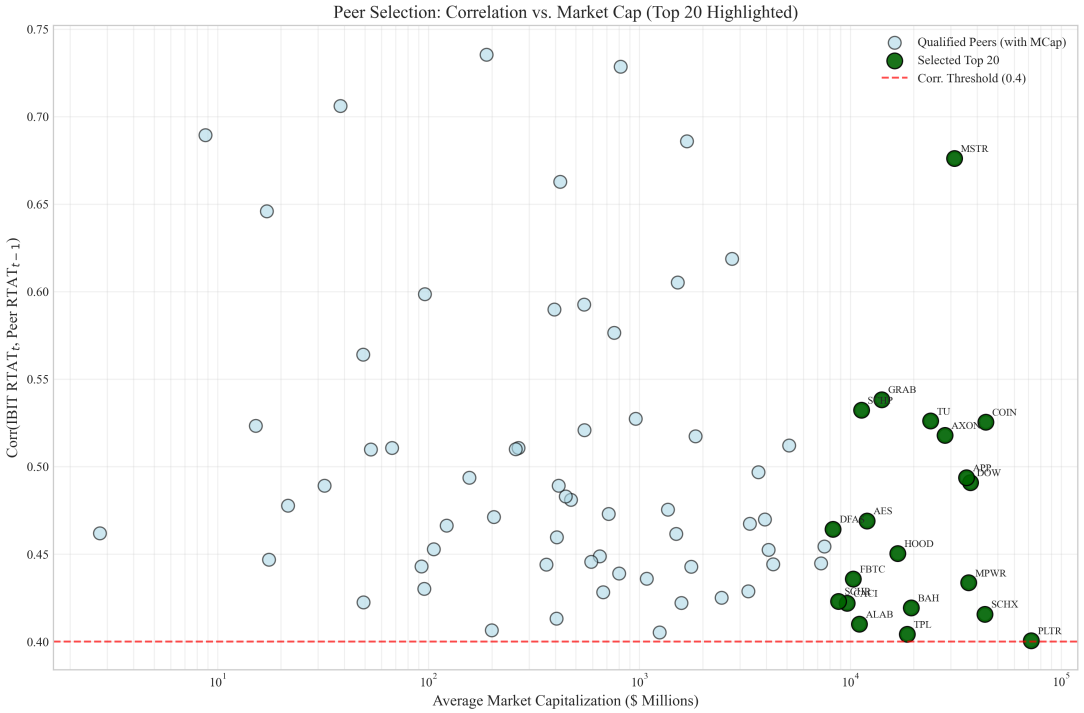


Figure 16. Peer Selection: Correlation vs. Market Cap (Top 20 Highlighted)

Notes: Scatter plot with average market capitalization ($ millions, log scale) on the x-axis and Corr($RTAT_{IBIT,t}$, $RTAT_{Peer,t-1}$) on the y-axis. Light blue circles are qualified peers. Dark green circles are the selected top 20. A horizontal red dashed line indicates the correlation threshold of 0.4. Source: CRSP and Nasdaq RTAT dataset.

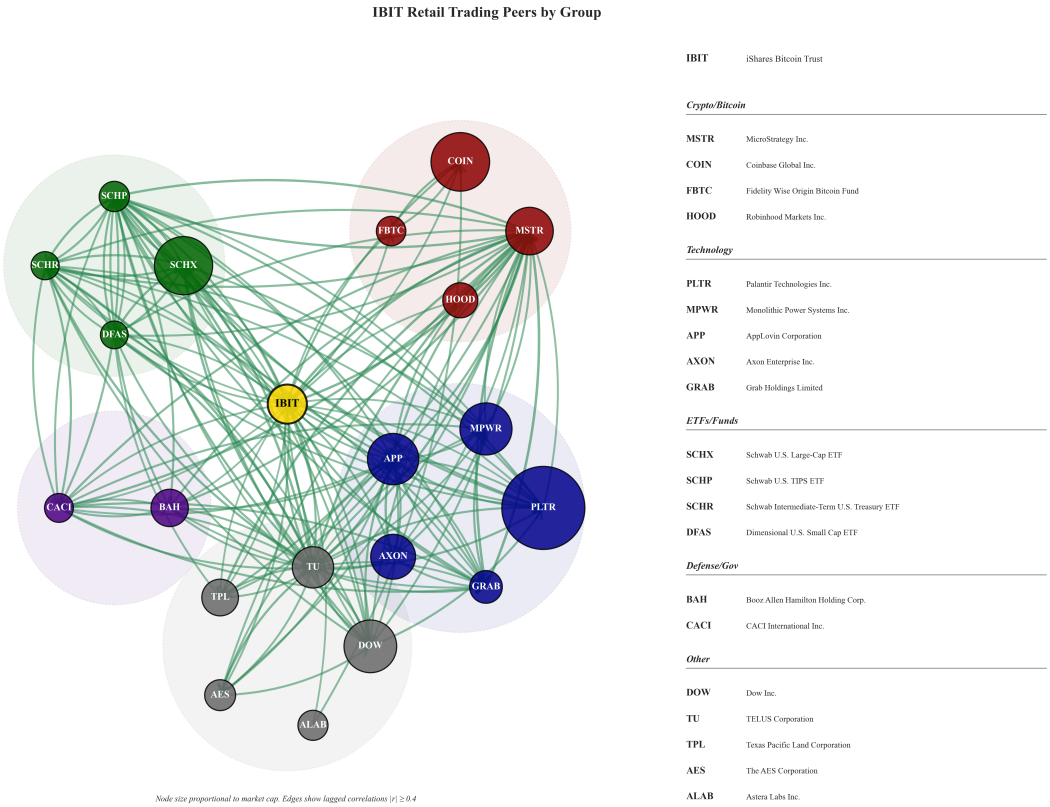


Figure 17. IBIT Retail Trading Peers by Group

Notes: Network graph showing IBIT (yellow node, center) and its 20 selected peers. Peers are color-coded by sector (Crypto/Bitcoin in red, Technology in dark blue, ETFs/Funds in light green, Defense/Gov in purple, Other in grey). Node size is proportional to market cap. Edges show lagged RTAT correlation > 0.4. Company names are listed next to tickers on the right, grouped by category. Source: CRSP and Nasdaq RTAT dataset.

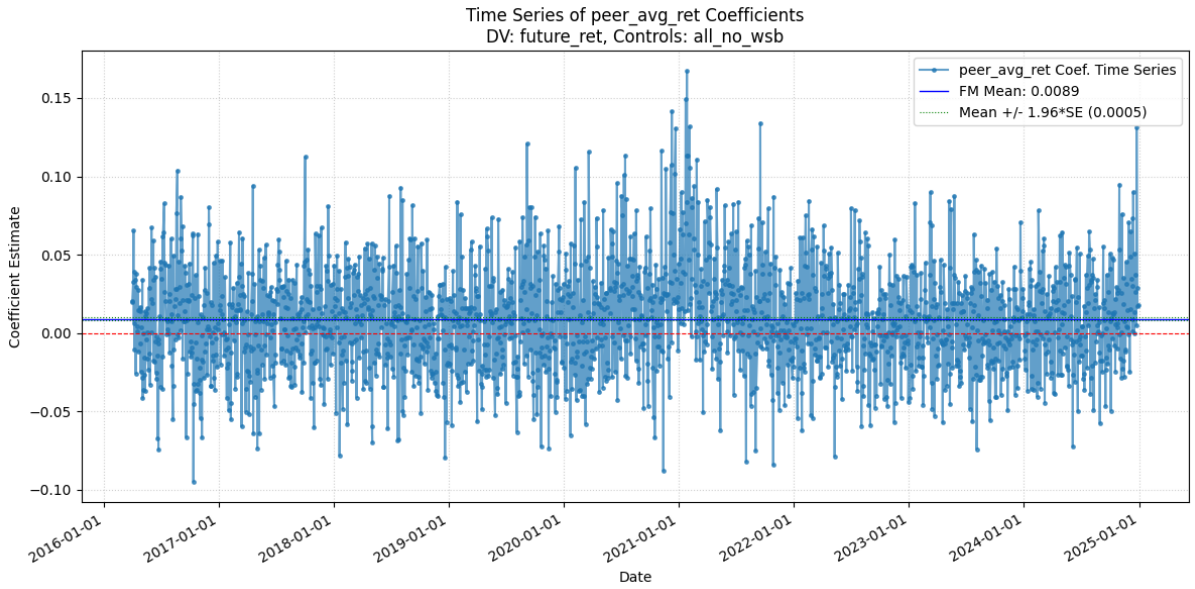


Figure 18. Time Series of Peer Return Coefficients from Fama-MacBeth Regressions

This figure plots the daily coefficient estimates on peer average returns ($PeerAvgRet_{i,t}$) from cross-sectional regressions. The dependent variable is the focal stock's next-day total return ($Return_{i,t+1}$). The regressions include firm characteristics, fundamentals, sentiment, short interest, and risk factors as controls. The sample covers each trading day from 2016 to 2024. The blue line shows the time series of daily coefficient estimates. The solid blue horizontal line indicates the time-series mean of the Fama-MacBeth coefficients (0.0089), and the dashed red line indicates zero. The green dotted lines represent the 95% confidence interval (mean $\pm$ 1.96 $\times$ standard error of 0.0005).

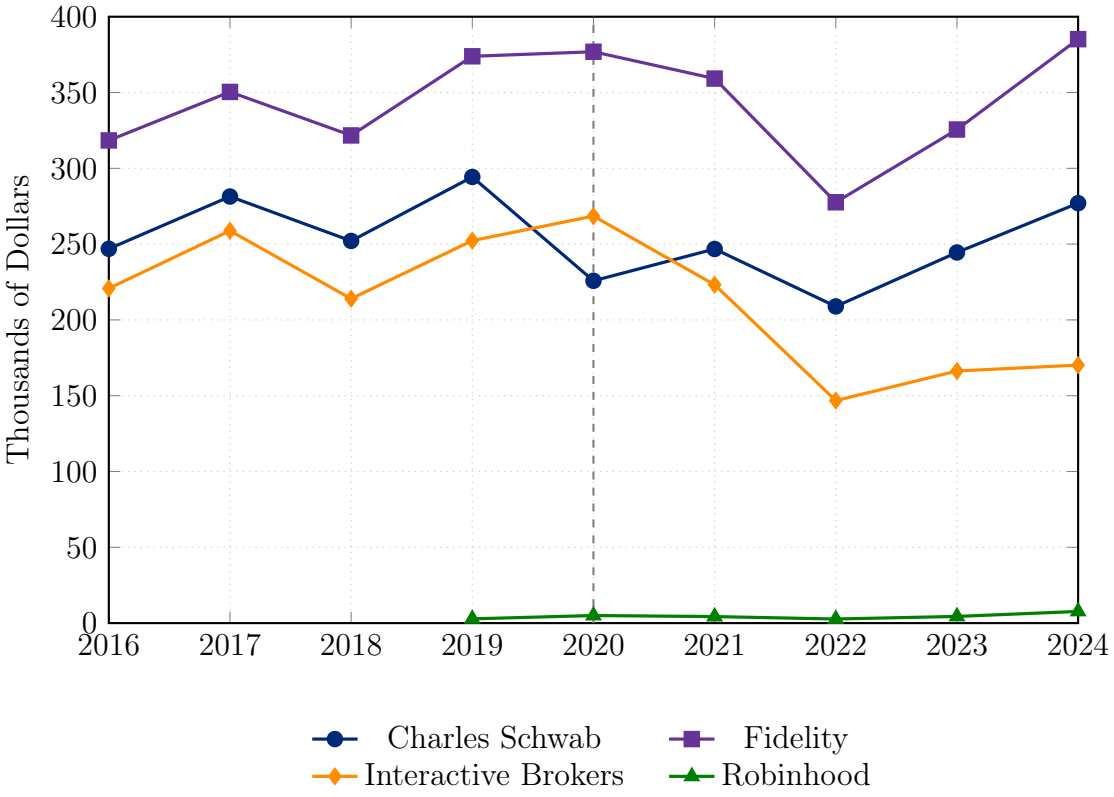


Figure 19. Average Asset Size per Client Account, 2016-2024

Notes: Average asset size per client account is calculated as total client assets divided by number of accounts. All values are in thousands of US dollars.

Calculation: Average Asset Size = Total Client Assets ÷ Number of Accounts.

Account Definitions: Charles Schwab: Active Brokerage Accounts. Fidelity: Total retail accounts (2016-2023), unique customers engaging with digital platforms (2024). Robinhood: Net Cumulative Funded Accounts (2019-2022), Funded Customers (2023-2024). IBKR: Client Accounts.

Asset Definitions: Charles Schwab and Fidelity: Assets Under Administration (AUA). IBKR: Client Equity. Robinhood: Assets Under Custody (AUC).

Sources: Calculated from data in company annual reports and SEC filings. See Table XXVII for complete source documentation.

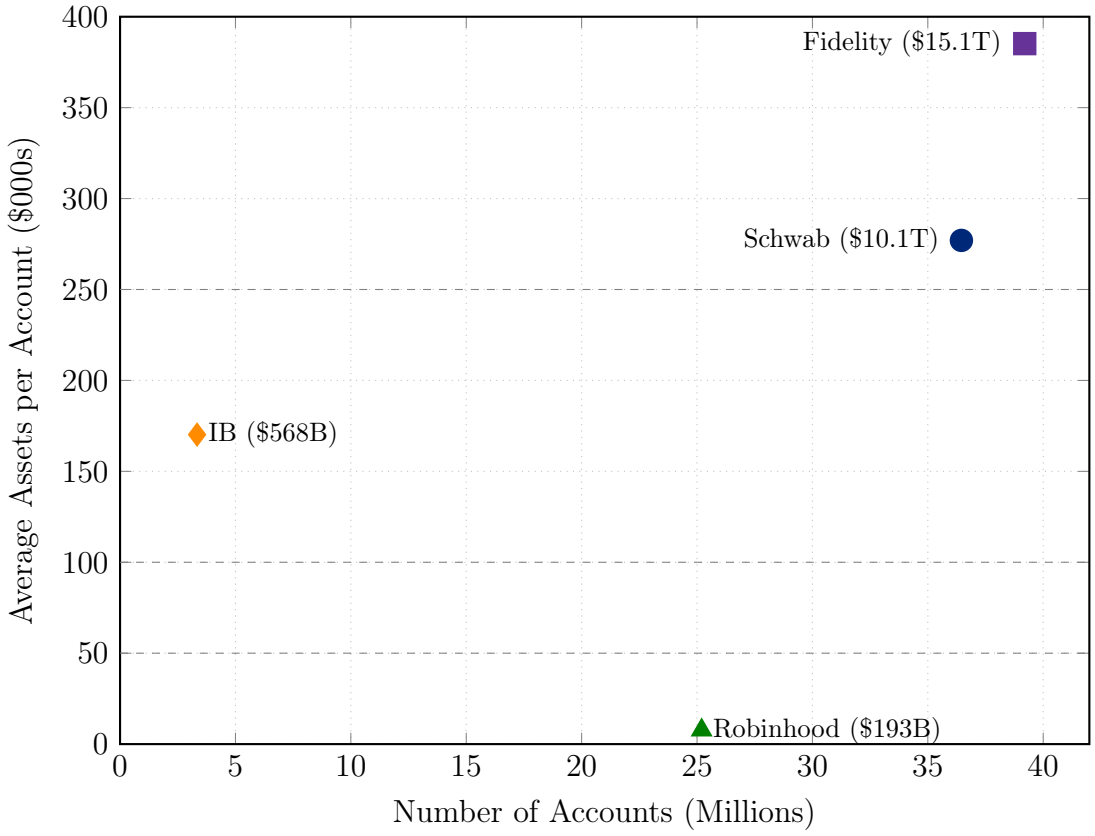


Figure 20. Retail Brokerage Market Composition by Account Size and Client Assets, 2024

Notes: Market structure as of year-end 2024. Each broker is positioned by number of accounts (x-axis) and average assets per account (y-axis), with total assets shown in parentheses.

2024 Data: Charles Schwab: 36.456 million Active Brokerage Accounts, Client Assets $10,101.3 billion. Fidelity: 39.2 million unique customers engaging with Fidelity.com, NetBenefits.com, or mobile apps, AUA $15.1 trillion. Robinhood: 25.2 million Funded Customers, AUC $193 billion. IBKR: 3.337 million accounts, Client equity $568 billion.

Sources: 2024 company earnings releases and SEC filings. See Table XXVII for complete source documentation.

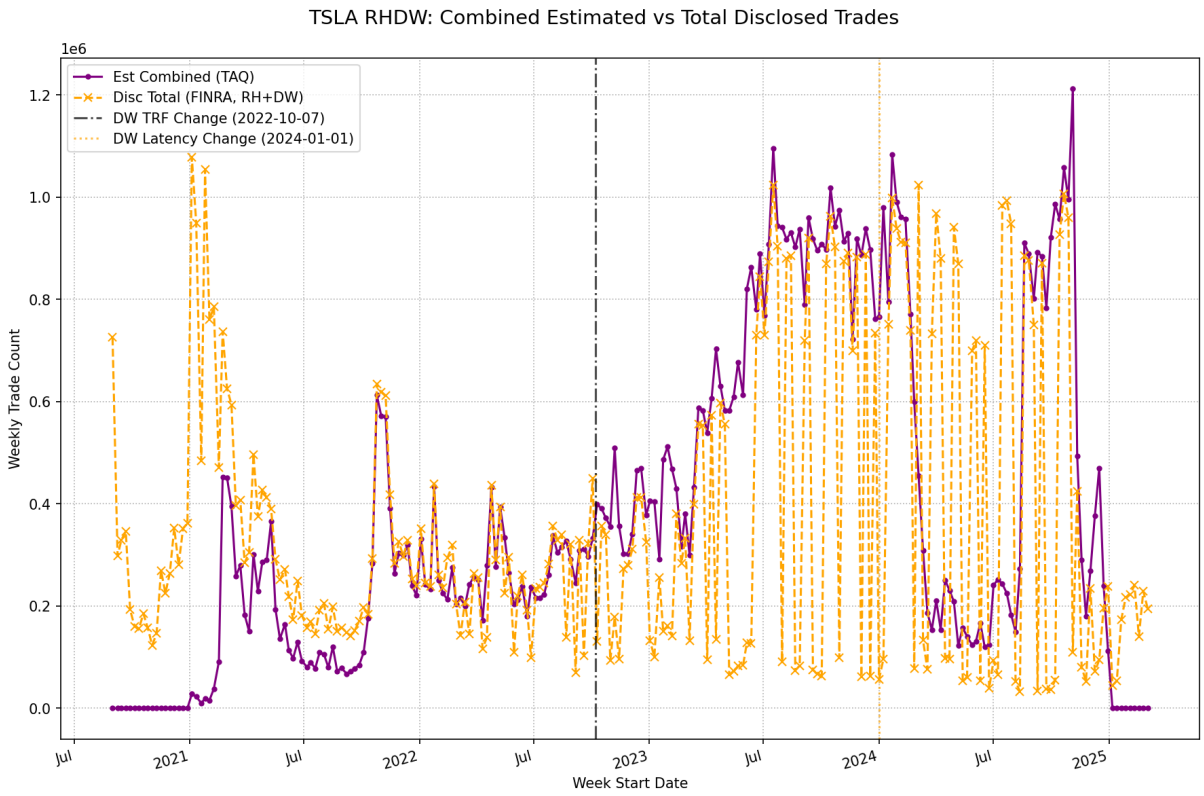


Figure 21. Tesla (TSLA): Combined Estimated vs Total Disclosed Fractional Trades (20ms Threshold)

Notes: Weekly RHDW fractional share trades for Tesla from January 2021 to December 2024. Purple line shows TAQ-based estimates using the original 20-millisecond latency threshold throughout the sample period. Orange dashed line shows total disclosed trades from FINRA OTC Transparency data. The divergence beginning mid-2024 indicates classification breakdown due to changes in DriveWealth's reporting latency.

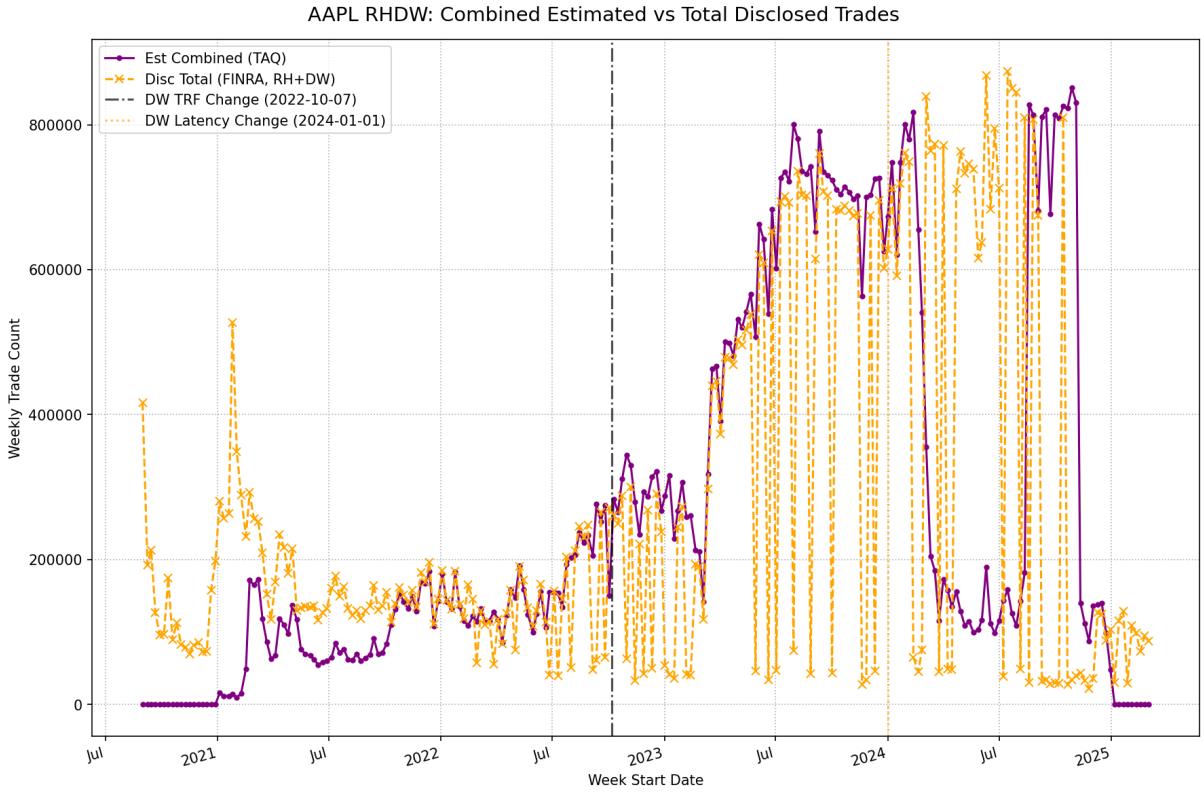


Figure 22. Apple (AAPL): Combined Estimated vs Total Disclosed Fractional Trades (20ms Threshold)

Notes: Weekly RHDW fractional share trades for Apple from January 2021 to December 2024. Purple line shows TAQ-based estimates using the original 20-millisecond latency threshold throughout the sample period. Orange dashed line shows total disclosed trades from FINRA OTC Transparency data. The divergence beginning mid-2024 indicates classification breakdown due to changes in DriveWealth's reporting latency.

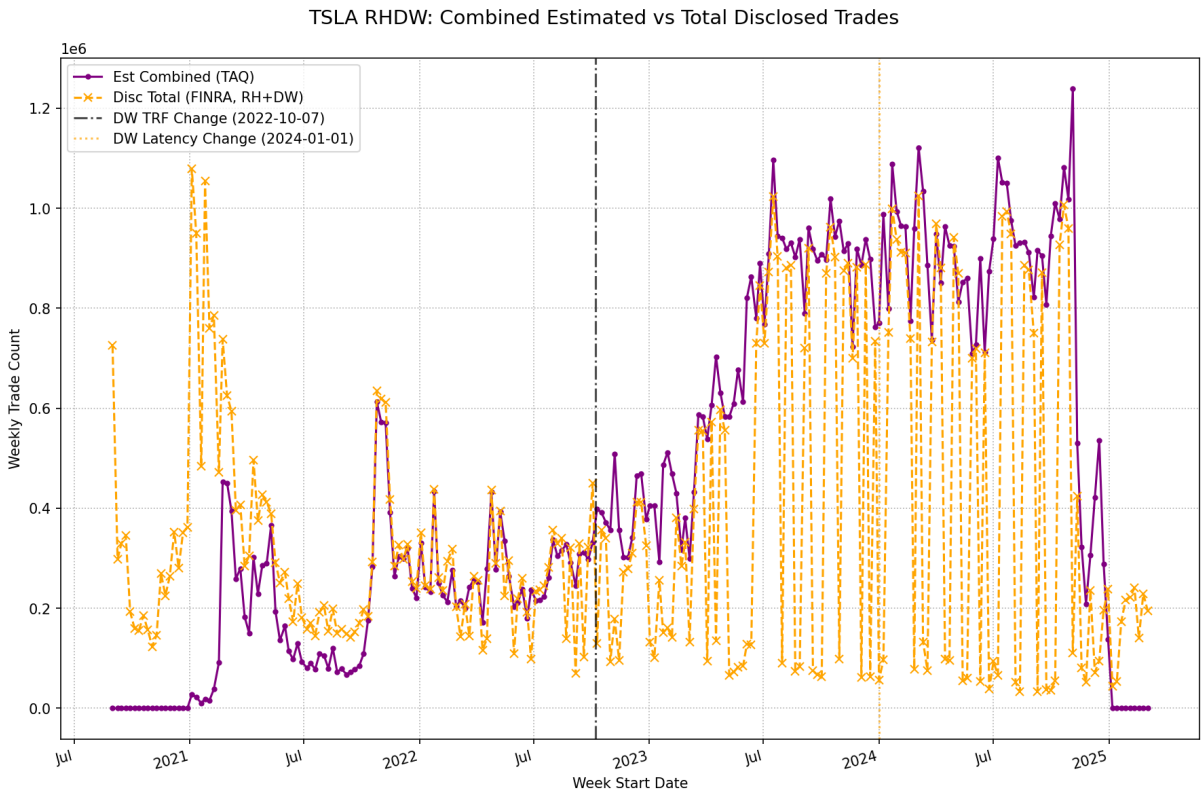


Figure 23. Tesla (TSLA): Combined Estimated vs Total Disclosed Fractional Trades (Refined Methodology)

Notes: Weekly RHDW fractional share trades for Tesla from January 2021 to December 2024. Purple line shows TAQ-based estimates using the refined methodology with time-varying latency thresholds (20ms through January 1 2024, 10ms thereafter). Orange dashed line shows total disclosed trades from FINRA OTC Transparency data. The close correspondence throughout the sample validates the methodological refinement.

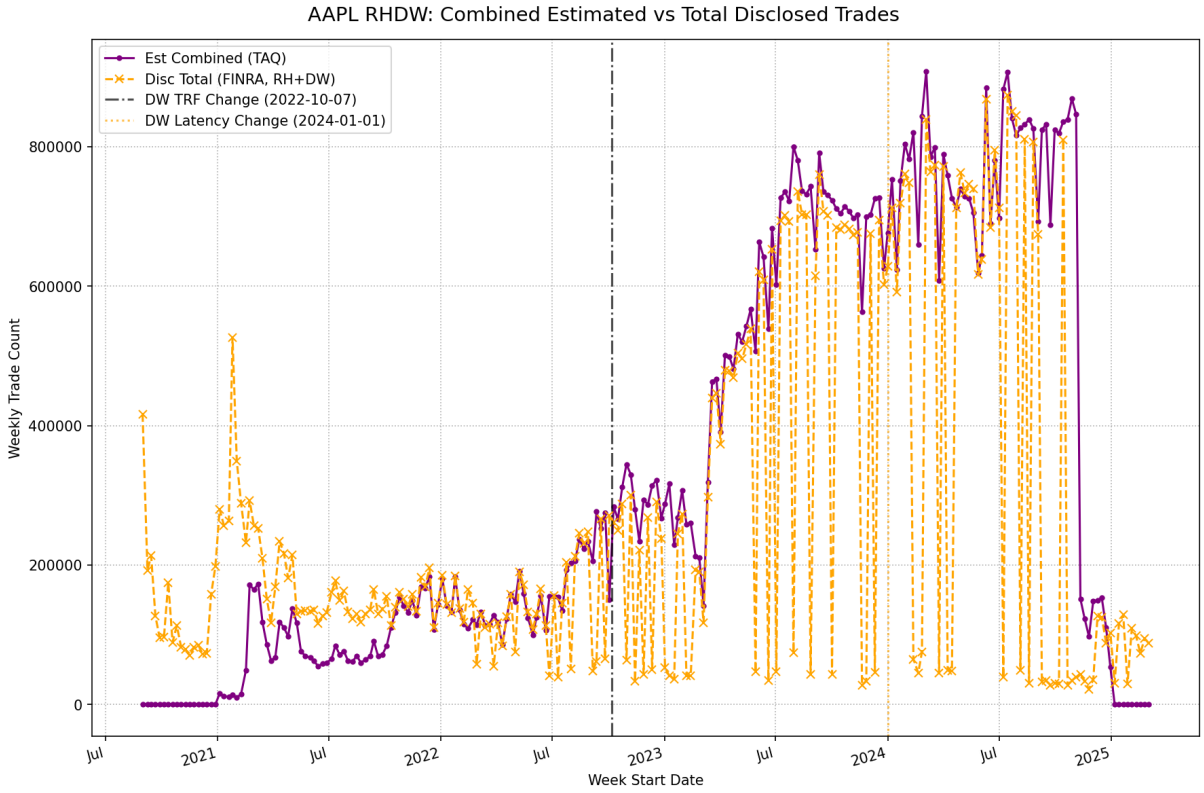


Figure 24. Apple (AAPL): Combined Estimated vs Total Disclosed Fractional Trades (Refined Methodology)

Notes: Weekly RHDW fractional share trades for Apple from January 2021 to December 2024. Purple line shows TAQ-based estimates using the refined methodology with time-varying latency thresholds (20ms through January 1 2024, 10ms thereafter). Orange dashed line shows total disclosed trades from FINRA OTC Transparency data. The close correspondence throughout the sample validates the methodological refinement.

List of Tables

I Average Retail Trading Activity Distribution across Asset Categories and Exchanges . . . 104
II Top-10 tickers by average retail activity, market share, and fractional-share activity, 2016–2024 . . . 105
III Logit Model of Edge Formation . . . 106
IV Robustness to Sample Size . . . 107
V Top 20 Selected Retail Trading Peers for IBIT . . . 108
VI Peer Network Effects on IBIT Return Predictability . . . 109
VII Peer Network Effects on IBIT Volatility, Trading Activity, and Market Quality 110
VIII Summary Statistics . . . 111
IX Correlation Matrix of Retail Trading Proxies . . . 112
X Correlation Matrix of Retail Trading Market Share Ratios . . . 113
XI Correlation Matrix of Retail Trading Activity and Market Variables . . . 114
XII Peer Returns and Focal Stock's Future Return: Fixed Effects . . . 115
XIII Peer Returns and Focal Stock's Future Return (Period-t RTAT Correlations) 116
XIV Peer Effects on Future Returns: Alternative Fixed Effects Specifications . . . 117
XV Peer Returns and Focal Stock's Future Return: Fama-MacBeth . . . 118
XVI Price Discovery: Cross-Sectional Variations . . . 119
XVII Price Discovery: Microstructure Outcomes . . . 120
XVIII Price Discovery: Persistence Across Time Horizons . . . 121
XIX Liquidity Provision: Trading Activity and Market Depth . . . 122
XX Liquidity Provision: Spread Effects . . . 123
XXI Liquidity Provision: Cross-Sectional Variations with Market Liquidity . . . 124
XXII Liquidity Provision: Persistence of Bid Depth Effects . . . 125
XXIII Peer Effects on Future Returns by Asset Type: Fama-MacBeth Regressions 126
XXIV Peer Effects on Future Returns by Exchange: Fama-MacBeth Regressions . 127
XXV Summary Statistics: Average Asset Size per Client, 2016-2024 . . . 128
XXVI Detailed Average Asset Size per Client by Year, 2016-2024 . . . 129
XXVII Detailed Source Documentation . . . 130
XXVIII Validation of RHDW Fractional Share Trade Classification . . . 131

Table I. Average Retail Trading Activity Distribution across Asset Categories and Exchanges

Share Code Category	Exchange Code					Row Total
	NYSE (1,2)	NASDAQ (3)	NYSE Arca (4)	Cboe (5)	Others	
Common Stocks (11)	0.2754	0.3687	0.0000	0.0001	0.0000	0.6442
Foreign Stocks (12)	0.0324	0.0122	0.0000	0.0000	0.0000	0.0446
REITs (18)	0.0099	0.0014	0.0000	0.0000	0.0000	0.0113
ETPs (73)	0.0000	0.0560	0.1740	0.0091	0.0000	0.2392
ADRs (31)	0.0324	0.0133	0.0000	0.0000	0.0000	0.0457
Others	0.0106	0.0014	0.0030	0.0000	0.0000	0.0150
Column Total	0.3608	0.4530	0.1770	0.0092	0.0000	1.0000

Notes: This table presents the average distribution of retail trading activity across different asset categories and exchanges from the RTAT dataset. Share code categories are based on CRSP classifications, where numbers in parentheses indicate the SHRCD values. Exchange codes represent the primary listing exchanges as defined in CRSP: NYSE (exchcd 1,2), NASDAQ (exchcd 3), NYSE Arca (exchcd 4), and Cboe (exchcd 5). Values represent proportions of total retail trading activity measured in U.S. dollars. The sample period spans from January 2016 to December 2024.

Table II. Top-10 tickers by average retail activity, market share, and fractional-share activity, 2016–2024

Rank	RTAT Activity		Dollar-Volume Share		RHDW Share	
	Ticker	%	Ticker	%	Ticker	%
1	TSLA	0.0472	SPY	0.0645	TSLA	0.1662
2	AAPL	0.0300	TSLA	0.0267	AMZN	0.0739
3	NVDA	0.0300	QQQ	0.0238	AAPL	0.0636
4	SPY	0.0234	AAPL	0.0202	NVDA	0.0631
5	AMD	0.0223	NVDA	0.0183	MSFT	0.0467
6	AMZN	0.0183	AMZN	0.0178	VOO	0.0346
7	FB	0.0155	MSFT	0.0115	SPY	0.0339
8	QQQ	0.0146	IWM	0.0110	NFLX	0.0207
9	TQQQ	0.0112	FB	0.0108	META	0.0201
10	META	0.0104	META	0.0092	QQQ	0.0192
Sum		0.2230		0.2139		0.5420

Notes: RTAT Activity is the fraction of retail-traded dollars (Nasdaq RTAT) for a ticker; Dollar-Volume Share is the fraction of total market dollar volume (CRSP); RHDW Share is the fraction of estimated fractional-share trading via Robinhood / DriveWealth. Averages are Jan 2016–Dec 2024 for RTAT and Dollar Volume, and Nov 2021–Dec 2024 for RHDW. FB and META denote Meta Platforms, Inc. before and after its 2021 name change.

Table III. Logit Model of Edge Formation

	Dependent Variable: $\text{Edge}_{j \to i}$		
	(1) Basic	(2) Trading	(3) Full
$\log(\text{MCap}_{\text{Source}}/\text{MCap}_{\text{Target}})$	0.0895***	0.0885***	0.0948***
	(0.0030)	(0.0063)	(0.0066)
1(Same Industry)			0.5143***
			(0.0207)
$\log(\text{Price})_{\text{Source}}$		-0.0584***	-0.0489***
		(0.0102)	(0.0107)
$\log(\text{Price})_{\text{Target}}$		0.0256**	0.0404***
		(0.0102)	(0.0108)
$\log(\text{Volume})_{\text{Source}}$		0.2095***	0.2011***
		(0.0069)	(0.0073)
$\log(\text{Volume})_{\text{Target}}$		0.1447***	0.1565***
		(0.0068)	(0.0074)
$\text{Volatility}_{\text{Source}}$			0.0625***
			(0.0146)
$\text{Volatility}_{\text{Target}}$			0.0025
			(0.0166)
Avg $\text{RTAT}_{\text{Source}}$			26.5959***
			(3.8278)
Avg $\text{RTAT}_{\text{Target}}$			2.9759
			(9.1779)
Constant	-6.6484***	-11.1817***	-11.4427***
	(0.0089)	(0.1059)	(0.1117)
Observations	9,965,031	9,965,031	9,965,031
Pseudo R^2	0.0044	0.0159	0.0190

Notes: This table reports coefficients from Logit regressions where the dependent variable is a binary indicator equal to 1 if a directed edge exists from source firm j to target firm i. Edge formation is based on $\text{Corr}(RTAT_{j,t-1}, RTAT_{i,t}) > 0.4$ and market cap ranking. The regression is run on a 10% random sample of all possible directed pairs. Volatility is the annualized standard deviation of daily returns. Avg RTAT is the time-series mean of the Nasdaq Retail Trading Activity Tracker. Standard errors are in parentheses. *, **, *** denote significance at 10%, 5%, 1% levels.

Table IV. Robustness to Sample Size

Sample	Observations	Coefficient on log(MCap Ratio)	Std. Error	Pseudo R^2
1%	996,503	0.0941***	(0.0215)	0.0206
5%	4,982,515	0.0882***	(0.0094)	0.0186
10%	9,965,031	0.0948***	(0.0066)	0.0190

Notes: This table reports the coefficient on the log market cap ratio from the full Logit specification (Model 3 in Table III) estimated on random subsamples of the data. Each subsample is drawn randomly from the universe of 99,650,306 possible directed pairs. Standard errors are in parentheses. *** denotes significance at the 1% level.

Table V. Top 20 Selected Retail Trading Peers for IBIT

Ticker	Full Name	Correlation	Market Cap ($M)
PLTR	Palantir Technologies Inc.	0.401	72,118
COIN	Coinbase Global Inc.	0.525	43,921
SCHX	Schwab U.S. Large-Cap ETF	0.416	43,506
DOW	Dow Inc.	0.491	37,107
MPWR	Monolithic Power Systems Inc.	0.434	36,432
APP	AppLovin Corporation	0.494	35,548
MSTR	MicroStrategy Incorporated	0.676	31,185
AXON	Axon Enterprise Inc.	0.518	28,102
TU	TELUS Corporation	0.526	23,995
BAH	Booz Allen Hamilton Holding Corp.	0.419	19,458
TPL	Texas Pacific Land Corporation	0.404	18,635
HOOD	Robinhood Markets Inc.	0.450	16,785
GRAB	Grab Holdings Limited	0.538	14,085
AES	The AES Corporation	0.469	12,011
SCHP	Schwab U.S. TIPS ETF	0.532	11,282
ALAB	Astera Labs Inc.	0.410	11,023
FBTC	Fidelity Wise Origin Bitcoin Fund	0.436	10,341
CACI	CACI International Inc	0.422	9,652
SCHR	Schwab Interm. U.S. Treasury ETF	0.423	8,775
DFAS	Dimensional U.S. Small Cap ETF	0.464	8,282

Notes: This table lists the top 20 retail trading peers for the iShares Bitcoin Trust (IBIT) ETF. Peers are identified based on a two-stage process. First, tickers are selected if their previous-day Retail Trading Activity Tracker (RTAT) activity has a Pearson correlation coefficient of 0.4 or greater with IBIT's current-day RTAT activity, ($RTAT_{IBIT,t}$, $RTAT_{Peer,t-1}$). A minimum of 20 paired daily observations from IBIT's launch (January 11, 2024) through the end of the analysis period (December 31, 2024) is required for correlation calculation. Second, from these qualifying tickers, the top 20 are chosen based on their average market capitalization over the same analysis period (January 11, 2024 - December 31, 2024). Market capitalization is presented in millions of U.S. dollars.

Table VI. Peer Network Effects on IBIT Return Predictability

Dependent Variable:	(1) $Return_{t+1}$	(2) $Return_{t+1}$	(3) $Return_{t+1}$	(4) OTC_Return_{t+1}
Peer Variables				
$Peer_Return_t$	0.4313**			
	(0.2069)			
$Peer_Return^2{}_t$		4.1298**		
		(1.6328)		
$Peer_DVol_t$ (log)			0.0230***	
			(0.0072)	
$Peer_RTAT_t$				7.7399***
				(2.6313)
Control Variables				
$Return_t$	−0.2203**	−0.1393**	−0.1285**	−0.0576
	(0.0890)	(0.0659)	(0.0636)	(0.0405)
$Log(Price)_t$	−0.0320**	−0.0298**	−0.0634***	−0.0297***
	(0.0141)	(0.0141)	(0.0168)	(0.0105)
$Log(Volume)_t$	0.0128**	0.0085	0.0020	−0.0003
	(0.0053)	(0.0055)	(0.0062)	(0.0034)
$Return_{[t-5,t-1]}$	−0.0155	−0.0185	−0.0330	−0.0239
	(0.0296)	(0.0295)	(0.0297)	(0.0187)
Constant	−0.1004	−0.0391	−0.2614***	0.1035*
	(0.0741)	(0.0757)	(0.0917)	(0.0549)
Observations	243	243	243	243
R^2	0.0546	0.0625	0.0763	0.0542
Adjusted R^2	0.0346	0.0427	0.0568	0.0343

Notes: This table reports OLS regression results. Standard errors are in parentheses. $Peer_Return^2$ represents the average of squared returns across peer firms. OTC_Return_{t+1} is $Open_to_Close_Return_{t+1}$. $Peer_DVol_t$ (log) is the log of peer average dollar volume. $Peer_RTAT_t$ is the peer average Nasdaq RTAT activity. See Appendix B for detailed variable definitions. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table VII. Peer Network Effects on IBIT Volatility, Trading Activity, and Market Quality

Dependent Variable:	(5) $Return^2_{t+1}$	(6) $Log(IVolQ)_{t+1}$	(7) $Log(DVol)_{t+1}$	(8) $Log(Depth)_{t+1}$
Peer Variables				
$Peer_Return^2{}_t$	0.4449***			
	(0.1086)			
$Peer_DVol_t$ (log)			0.1935**	
			(0.0783)	
$Peer_RTAT_t$		143.1771***		−88.4194**
		(49.5665)		(35.8499)
Control Variables				
$Return_t$	−0.0140***	−0.4222	−1.1138	0.4579
	(0.0044)	(0.7633)	(0.6874)	(0.5520)
$Log(Price)_t$	−0.0003	−1.4891***	1.4703***	0.0657
	(0.0009)	(0.1979)	(0.1815)	(0.1431)
$Log(Volume)_t$	0.0001	0.2144***	0.4084***	−0.1315***
	(0.0004)	(0.0645)	(0.0668)	(0.0466)
$Return_{[t-5,t-1]}$	−0.0003	0.5398	0.6484**	0.0922
	(0.0020)	(0.3515)	(0.3213)	(0.2542)
Constant	−0.0000	−20.0511***	4.6073***	14.2131***
	(0.0050)	(1.0346)	(0.9911)	(0.7483)
Observations	243	243	243	243
R^2	0.0884	0.2023	0.7107	0.1248
Adjusted R^2	0.0692	0.1855	0.7046	0.1063

Notes: This table reports OLS regression results. Standard errors are in parentheses. $Log(IVolQ)_{t+1}$ represents the log of quote-based intraday volatility during market hours. $Log(DVol)_{t+1}$ is the log of dollar volume. $Peer_Return^2$ represents the average of squared returns across peer firms. $Peer_DVol_t$ (log) is the log of peer average dollar volume. $Peer_RTAT_t$ is the peer average Nasdaq RTAT activity. See Appendix B for detailed variable definitions. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table VIII. Summary Statistics

Variable	N	Mean	Std	P5	P50	P95	P99
Panel A: Stock Characteristics							
Price ($)	18,644,893	85.23	4,436.28	0.71	21.84	127.25	296.18
Market Cap ($mil)	18,644,893	5,721.80	42,217.35	6.89	350.72	20,575.26	91,887.33
Dollar Volume ($mil)	18,644,886	52.99	538.20	0.00	1.73	186.37	763.81
Share Volume (000s)	18,644,891	1,125.06	6,450.52	0.25	121.89	4,274.18	16,603.27
Returns (%)	18,637,184	0.04	4.26	-9.17	0.00	9.06	10.16
Panel B: Retail Trading Measures							
RTAT Activity	16,822,630	0.0001	0.0013	0.0000	0.0000	0.0004	0.0018
BJZZ Retail Volume ($mil)	18,724,340	4.81	81.56	0.00	0.17	11.44	57.57
RHDW Count	5,964,279	193.10	2,589.81	1.00	18.00	439.00	2,385.00
RHDW Value ($000s)	5,964,279	3.98	101.23	0.00	0.04	4.09	35.74
Robintrack Holdings	1,732,658	3,167.14	19,989.79	16.00	301.00	8,818.00	58,002.43
Off-Exchange 1-Share	16,338,740	282.80	2,716.81	1.00	24.00	991.00	3,652.00
Panel C: Market Microstructure							
Quoted Spread (bps)	18,550,865	77.00	245.00	3.00	25.00	317.00	731.00
Effective Spread (bps)	18,425,817	50.00	195.00	2.00	15.00	203.00	485.00
Variance Ratio	18,402,599	1.03	0.58	0.44	0.95	1.81	2.65
Panel D: Short Interest and Sentiment							
Short Volume (000s)	17,466,512	212.43	1,730.28	0.12	22.73	712.08	3,097.02
Short Ratio	17,466,512	0.462	0.226	0.094	0.456	0.882	1.000
ESS Full	3,321,753	0.089	0.427	-0.580	0.180	0.680	0.950
CSS Full	5,441,982	0.015	0.082	-0.080	0.010	0.100	0.220
WSB Mentions	229,904	9.08	48.22	1.00	1.00	25.00	171.00
Panel E: Risk and Fundamentals							
Market Beta	18,277,665	0.748	0.636	-0.024	0.800	1.570	2.270
Idiosyncratic Vol (%)	18,277,665	2.38	2.99	0.21	1.56	6.84	12.33
Book Value ($mil)	8,210,481	2,893.22	16,358.40	-11.83	299.92	10,464.00	47,026.00
ROA (%)	8,197,999	-3.67	90.15	-22.96	0.23	4.50	10.67
Inst. Ownership (%)	7,776,215	60.15	31.66	3.96	68.46	100.00	100.00
Number of Analysts	7,731,603	7.11	7.05	1.00	5.00	22.00	31.00

This table reports summary statistics for the sample period January 2016 through December 2024. The sample includes 18,724,340 security-day observations. N denotes the number of non-missing observations. P5, P50, P95, and P99 denote the 5th, 50th, 95th, and 99th percentiles, respectively. Dollar Volume and Market Cap are in millions of dollars. RHDW Value is in thousands of dollars. Share Volume and Short Volume are in thousands of shares. Quoted Spread and Effective Spread are in basis points. Returns, Idiosyncratic Vol, and ROA are in percentage terms. Number of Analysts refers to the number of sell-side analysts covering the stock.

Table IX. Correlation Matrix of Retail Trading Proxies

	RTAT	RHDW	Odd-lot	Robintrack	BJZZ	Off-Exch
RTAT	1.000					
RHDW	0.596***	1.000				
Odd-lot	0.651***	0.807***	1.000			
Robintrack	0.473***		0.294***	1.000		
BJZZ	0.880***	0.634***	0.741***	0.450***	1.000	
Off-Exch	0.696***	0.765***	0.613***	0.623***	0.675***	1.000

This table presents Pearson correlations among retail trading proxies. RTAT is the NASDAQ Retail Trading Activity Tracker. RHDW is dollar volume estimates of fractional shares trading from Robinhood and DriveWealth (November 1, 2021 to December 31, 2024). Odd-lot is dollar volume from trades of less than 100 shares. Robintrack is the number of Robinhood investors holding each ticker (May 2, 2018 to August 13, 2020) based on replication files from Welch (2022). BJZZ is retail volume based on Boehmer, Jones, Zhang, and Zhang (2021). Off-Exch is the sum of off-exchange one-share trades from Da, Fang, and Lin (2024). The general sample period is January 4, 2016 to December 31, 2024. Significance levels are based on two-tailed t-tests: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table X. Correlation Matrix of Retail Trading Market Share Ratios

	RTAT	RHDW	Odd-lot	Robintrack	BJZZ	Off-Exch
RTAT	1.000					
RHDW	0.619***	1.000				
Odd-lot	0.681***	0.840***	1.000			
Robintrack	0.551***		0.318***	1.000		
BJZZ	0.883***	0.638***	0.714***	0.504***	1.000	
Off-Exch	0.783***	0.768***	0.671***	0.640***	0.661***	1.000

This table presents retail trading metrics correlations where each variable represents the daily ratio of a ticker's volume to the total market volume across all tickers. For example, RHDW share is the stock's RHDW volume divided by total RHDW volume across all stocks on that day. RTAT remains in its original normalized form. All other definitions follow Table IX. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table XI. Correlation Matrix of Retail Trading Activity and Market Variables

	RTAT	Return	Ret(t+1)	$ Volume	Q-Spread	Depth	Short Vol	ESS	CSS	WSB	Mkt Beta	Inst Own	Analysts
RTAT	1.000												
Return	0.023***	1.000											
Return(t+1)	-0.000	-0.030***	1.000										
Dollar Volume	0.381***	0.004***	0.001**	1.000									
Quoted Spread	-0.038***	-0.002***	0.003***	-0.026***	1.000								
Avg Depth	0.024***	0.000	-0.000	0.013***	-0.016***	1.000							
Short Volume	0.536***	0.144***	-0.005***	0.141***	-0.035***	0.037***	1.000						
ESS	0.011***	0.077***	0.028***	0.011***	0.016***	0.005***	0.014***	1.000					
CSS	-0.000	0.059***	0.010***	0.003***	-0.028***	0.002***	-0.001***	0.359***	1.000				
WSB	0.570***	0.009***	-0.001	0.569***	-0.035***	0.003	0.244***	0.017***	-0.015***	1.000			
Beta on Market	0.038***	0.005***	0.005***	0.015***	-0.091***	-0.028***	0.025***	-0.027***	0.000	0.044***	1.000		
Inst Own	0.010***	-0.000	-0.000	0.025***	-0.409***	-0.002***	-0.018***	-0.049***	0.037***	-0.042***	0.280***	1.000	
Number of Analyst	0.238***	0.001**	0.001**	0.211***	-0.294***	0.009***	0.116***	-0.004***	0.049***	0.160***	0.212***	0.426***	1.000

This table presents the correlation matrix between retail trading activity and market variables. RTAT is the NASDAQ Retail Trading Activity Tracker. Return(t+1) represents next-day returns. Dollar Volume is daily trading value. Q-Spread is the quoted spread in basis points. Depth is the average of bid and offer depth in dollars. Short Volume is daily short sale volume. ESS and CSS are RavenPack event and composite sentiment scores. WSB represents WallStreetBets mention counts. Mkt Beta is the market beta from factor model regressions. Inst Own is institutional ownership percentage. Analysts is analyst coverage. The sample period is January 1, 2016 to December 31, 2024. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table XII. Peer Returns and Focal Stock's Future Return: Fixed Effects

	Dependent Variable: $Return_{i,t+1}$					
	(1) Base	(2) Fundamental	(3) Sentiment	(4) Short Interest	(5) Risk	(6) Full
Peer Avg. Return$_{i,t}$	0.0144***	0.0127***	0.0144***	0.0143***	0.0142***	0.0126***
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
log(Market Cap)$_{i,t}$	-0.0000	-0.0004***	-0.0000	-0.0000	-0.0003***	-0.0006***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
log(Price)$_{i,t}$	-0.0019***	-0.0016***	-0.0020***	-0.0020***	-0.0020***	-0.0017***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
log(Volume)$_{i,t}$	-0.0001***	0.0001***	-0.0001***	-0.0005***	-0.0000	-0.0003***
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0001)
Return$_{i,t}$	-0.0211***	-0.0200***	-0.0231***	-0.0208***	-0.0225***	-0.0221***
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)
Return$_{i,[t-5,t-1]}$	-0.0065***	-0.0070***	-0.0070***	-0.0066***	-0.0067***	-0.0077***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Inst. Ownership (%)$_{i,t}$		-0.0006**				-0.0005**
		(0.0002)				(0.0002)
log(Book-to-Market)$_{i,t}$		0.0000				0.0001*
		(0.0000)				(0.0000)
Num. Analysts$_{i,t}$		-0.0000**				-0.0000*
		(0.0000)				(0.0000)
ROA$_{i,t}$		0.0054***				0.0050***
		(0.0004)				(0.0004)
CSS (DJ & Web)$_{i,t}$			0.0042***			0.0039***
			(0.0003)			(0.0003)
ESS (DJ & Web)$_{i,t}$			0.0030***			0.0030***
			(0.0001)			(0.0001)
log(FINRA Short Vol.)$_{i,t}$				0.0004***		0.0004***
				(0.0000)		(0.0000)
Alpha (FF4)$_{i,t}$					0.0946***	0.0659***
					(0.0130)	(0.0138)
Beta (HML)$_{i,t}$					-0.0000	-0.0000
					(0.0000)	(0.0001)
Beta (MKT)$_{i,t}$					0.0001*	0.0001*
					(0.0001)	(0.0001)
Beta (SMB)$_{i,t}$					-0.0001*	-0.0001
					(0.0000)	(0.0000)
Beta (UMD)$_{i,t}$					0.0002***	0.0001***
					(0.0001)	(0.0001)
IVOL (FF4)$_{i,t}$					-0.0218***	-0.0141***
					(0.0021)	(0.0023)
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,521,721	4,877,031	5,521,721	5,484,786	5,450,500	4,853,987
R^2 (Within)	0.002	0.002	0.002	0.002	0.002	0.003

Notes: I estimate $Return_{i,t+1} = \alpha + \beta_1 \text{PeerAvgReturn}_{i,t} + \sum_k \gamma_k X_{k,i,t} + \mu_i + \tau_t + \epsilon_{i,t}$ fixed effects panel regression. PeerAvgReturn$_{i,t}$ is the simple average return of stock i's peers on day t. Peers are defined based on retail trading activity comovement with correlation coefficient > 0.4 and top 20 market cap. Sample: 2016-2024. μ_i and τ_t are firm and time fixed effects. Column (1): basic controls (log market cap, price, volume, return$_t$, cumulative return$_{[t-5,t-1]}$). Column (2) adds fundamentals (institutional ownership, log book-to-market, analysts, ROA). Column (3) adds sentiment scores (CSS and ESS from Dow Jones and web sources using RavenPack). Column (4) adds log FINRA short volume. Column (5) adds risk measures (FF4 alpha, factor betas, IVOL from 252-day rolling window). Column (6): full specification. Clustered standard errors (firm level) in parentheses. *, **, *** denote significance at 10%, 5%, 1%.

Table XIII. Peer Returns and Focal Stock's Future Return (Period-t RTAT Correlations)

	Dependent Variable: $Return_{i,t+1}$					
	(1) Base	(2) Fundamental	(3) Sentiment	(4) Short Interest	(5) Risk	(6) Full
Peer Avg. Return$_{i,t}$	0.0094***	0.0086***	0.0095***	0.0095***	0.0092***	0.0086***
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
log(Market Cap)$_{i,t}$	-0.0001	-0.0005***	-0.0001	-0.0001	-0.0003***	-0.0006***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
log(Price)$_{i,t}$	-0.0019***	-0.0016***	-0.0020***	-0.0019***	-0.0019***	-0.0016***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
log(Volume)$_{i,t}$	-0.0001***	0.0002***	-0.0001***	-0.0005***	0.0000	-0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Return$_{i,t}$	-0.0209***	-0.0198***	-0.0229***	-0.0206***	-0.0223***	-0.0218***
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)
Return$_{i,[t-5,t-1]}$	-0.0064***	-0.0070***	-0.0069***	-0.0065***	-0.0067***	-0.0076***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Inst. Ownership (%)$_{i,t}$		-0.0006***				-0.0004*
		(0.0002)				(0.0002)
log(Book-to-Market)$_{i,t}$		-0.0000				0.0000
		(0.0000)				(0.0000)
Num. Analysts$_{i,t}$		-0.0000***				-0.0000*
		(0.0000)				(0.0000)
ROA$_{i,t}$		0.0057***				0.0054***
		(0.0004)				(0.0004)
CSS (DJ & Web)$_{i,t}$			0.0041***			0.0039***
			(0.0003)			(0.0003)
ESS (DJ & Web)$_{i,t}$			0.0030***			0.0029***
			(0.0001)			(0.0001)
log(FINRA Short Vol.)$_{i,t}$				0.0004***		0.0004***
				(0.0000)		(0.0000)
Alpha (FF4)$_{i,t}$					0.0941***	0.0604***
					(0.0129)	(0.0137)
Beta (HML)$_{i,t}$					0.0000	-0.0000
					(0.0000)	(0.0000)
Beta (MKT)$_{i,t}$					0.0001*	0.0001
					(0.0001)	(0.0001)
Beta (SMB)$_{i,t}$					-0.0001*	-0.0001
					(0.0000)	(0.0000)
Beta (UMD)$_{i,t}$					0.0001***	0.0001**
					(0.0000)	(0.0000)
IVOL (FF4)$_{i,t}$					-0.0212***	-0.0131***
					(0.0021)	(0.0023)
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,639,340	4,991,641	5,639,340	5,602,732	5,568,221	4,968,770
R^2 (Within)	0.002	0.002	0.002	0.002	0.002	0.003

Notes: I estimate $Return_{i,t+1} = \alpha + \beta_1 \text{PeerAvgReturn}_{i,t} + \sum_k \gamma_k X_{k,i,t} + \mu_i + \tau_t + \epsilon_{i,t}$ fixed effects panel regression. PeerAvgReturn$_{i,t}$ is the simple average return of stock i's peers on day t. Peers are defined based on retail trading activity comovement with correlation coefficient > 0.4 and top 20 market cap, using period-t correlations. Sample: 2016-2024. μ_i and τ_t are firm and time fixed effects. Column (1): basic controls (log market cap, price, volume, return$_t$, cumulative return$_{[t-5,t-1]}$). Column (2) adds fundamentals (institutional ownership, log book-to-market, analysts, ROA). Column (3) adds sentiment scores (CSS and ESS from Dow Jones and web sources using RavenPack). Column (4) adds log FINRA short volume. Column (5) adds risk measures (FF4 alpha, factor betas, IVOL from 252-day rolling window). Column (6): full specification. Clustered standard errors (firm level) in parentheses. *, **, *** denote significance at 10%, 5%, 1%.

Table XIV. Peer Effects on Future Returns: Alternative Fixed Effects Specifications

	Dependent Variable: $Return_{i,t+1}$			
	(1) Pooled OLS	(2) Entity FE Only	(3) Time FE Only	(4) Two-way FE
Peer Avg. Return$_{i,t}$	0.0162***	0.0162***	0.0126***	0.0126***
	(0.0006)	(0.0010)	(0.0007)	(0.0008)
RTAT Activity$_{i,t}$	-0.1347***	-0.4358***	-0.2138***	-0.5163***
	(0.0503)	(0.1189)	(0.0480)	(0.1147)
log(Market Cap)$_{i,t}$	-0.0000**	-0.0007***	-0.0000	-0.0005***
	(0.0000)	(0.0001)	(0.0000)	(0.0001)
log(Price)$_{i,t}$	-0.0004***	-0.0017***	-0.0003***	-0.0014***
	(0.0000)	(0.0001)	(0.0000)	(0.0001)
log(Volume)$_{i,t}$	-0.0005***	-0.0005***	-0.0003***	-0.0003***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Return$_{i,t}$	-0.0232***	-0.0235***	-0.0234***	-0.0239***
	(0.0005)	(0.0011)	(0.0008)	(0.0011)
Return$_{i,[t-5,t-1]}$	-0.0069***	-0.0072***	-0.0074***	-0.0078***
	(0.0002)	(0.0003)	(0.0003)	(0.0003)
Inst. Ownership (%)$_{i,t}$	0.0010***	0.0006***	0.0004***	-0.0005**
	(0.0001)	(0.0002)	(0.0001)	(0.0002)
log(Book-to-Market)$_{i,t}$	0.0001***	0.0002***	0.0000	0.0001**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Num. Analysts$_{i,t}$	0.0000***	0.0000**	0.0000***	-0.0000**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
ROA$_{i,t}$	0.0053***	0.0054***	0.0052***	0.0053***
	(0.0002)	(0.0005)	(0.0003)	(0.0004)
CSS (DJ & Web)$_{i,t}$	0.0046***	0.0051***	0.0034***	0.0038***
	(0.0004)	(0.0004)	(0.0003)	(0.0003)
ESS (DJ & Web)$_{i,t}$	0.0030***	0.0030***	0.0030***	0.0030***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
log(FINRA Short Vol.)$_{i,t}$	0.0004***	0.0005***	0.0003***	0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Alpha (FF4)$_{i,t}$	0.1627***	0.1495***	0.1079***	0.0838***
	(0.0086)	(0.0149)	(0.0105)	(0.0135)
Beta (HML)$_{i,t}$	0.0001***	0.0001	-0.0001**	-0.0000
	(0.0000)	(0.0001)	(0.0000)	(0.0000)
Beta (MKT)$_{i,t}$	0.0004***	0.0002***	0.0004***	0.0001**
	(0.0000)	(0.0001)	(0.0000)	(0.0001)
Beta (SMB)$_{i,t}$	0.0001**	-0.0000	0.0001***	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Beta (UMD)$_{i,t}$	-0.0000	0.0002***	-0.0001	0.0001*
	(0.0000)	(0.0001)	(0.0000)	(0.0000)
IVOL (FF4)$_{i,t}$	-0.0203***	0.0091***	-0.0296***	-0.0136***
	(0.0011)	(0.0024)	(0.0015)	(0.0022)
Firm Fixed Effects	No	Yes	No	Yes
Time Fixed Effects	No	No	Yes	Yes
Observations	4,826,880	4,826,880	4,826,880	4,826,880
Adj. R^2	0.001	0.003	0.192	0.193
R^2 (Within)	–	0.003	–	0.003

Notes: I estimate $Return_{i,t+1} = \alpha + \beta_1 \text{PeerAvgReturn}_{i,t} + \beta_2 \text{RTAT}_{i,t} + \sum_k \gamma_k X_{k,i,t} + \mu_i + \tau_t + \epsilon_{i,t}$ with alternative fixed effects specifications. PeerAvgReturn$_{i,t}$ is the simple average return of stock i's peers on day t. Peers are defined based on retail trading activity comovement with correlation coefficient > 0.4 and top 20 market cap. RTAT Activity$_{i,t}$ is the retail trading activity measure. Sample: 2016-2024. μ_i and τ_t are firm and time fixed effects, respectively. Column (1): pooled OLS with no fixed effects. Column (2): only firm fixed effects. Column (3): only time fixed effects. Column (4): both firm and time fixed effects. All specifications include the full set of controls: basic controls (log market cap, price, volume, return$_t$, cumulative return$_{[t-5,t-1]}$), fundamentals (institutional ownership, log book-to-market, analysts, ROA), sentiment scores (CSS and ESS from Dow Jones and web sources using RavenPack), log FINRA short volume, and risk measures (FF4 alpha, factor betas, IVOL from 252-day rolling window). Standard errors in columns (2) and (4) are clustered at the firm level; columns (1) and (3) use heteroskedasticity-robust standard errors. *, **, *** denote significance at 10%, 5%, 1%.

Table XV. Peer Returns and Focal Stock's Future Return: Fama-MacBeth

	Dependent Variable: $Return_{i,t+1}$					
	(1) Base	(2) Fundamental	(3) Sentiment	(4) Short Interest	(5) Risk	(6) Full
Peer Avg. Return$_{i,t}$	0.0110***	0.0098***	0.0110***	0.0107***	0.0100***	0.0089***
	(0.0006)	(0.0005)	(0.0006)	(0.0006)	(0.0005)	(0.0005)
log(Market Cap)$_{i,t}$	0.0005***	0.0001***	0.0005***	0.0005***	0.0002***	-0.0001**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
log(Price)$_{i,t}$	-0.0004***	-0.0003***	-0.0004***	-0.0005***	-0.0005***	-0.0003***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
log(Volume)$_{i,t}$	-0.0002***	-0.0000	-0.0002***	-0.0004***	-0.0001***	-0.0003***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Return$_{i,t}$	-0.0274***	-0.0288***	-0.0297***	-0.0268***	-0.0324***	-0.0331***
	(0.0010)	(0.0011)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Return$_{i,[t-5,t-1]}$	-0.0023***	-0.0030***	-0.0028***	-0.0024***	-0.0025***	-0.0034***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Inst. Ownership (%)$_{i,t}$		0.0004***				0.0002*
		(0.0001)				(0.0001)
log(Book-to-Market)$_{i,t}$		-0.0000				-0.0000***
		(0.0000)				(0.0000)
Num. Analysts$_{i,t}$		0.0000**				0.0000***
		(0.0000)				(0.0000)
ROA$_{i,t}$		0.0056***				0.0044***
		(0.0004)				(0.0003)
CSS (DJ & Web)$_{i,t}$			0.0054***			0.0046***
			(0.0002)			(0.0002)
ESS (DJ & Web)$_{i,t}$			0.0029***			0.0029***
			(0.0001)			(0.0001)
log(FINRA Short Vol.)$_{i,t}$				0.0002***		0.0003***
				(0.0000)		(0.0000)
Alpha (FF4)$_{i,t}$					0.0601***	0.0249**
					(0.0124)	(0.0124)
Beta (HML)$_{i,t}$					0.0000	0.0000
					(0.0001)	(0.0001)
Beta (MKT)$_{i,t}$					0.0004***	0.0003***
					(0.0001)	(0.0001)
Beta (SMB)$_{i,t}$					0.0001***	0.0001***
					(0.0000)	(0.0000)
Beta (UMD)$_{i,t}$					-0.0000	-0.0000
					(0.0001)	(0.0001)
IVOL (FF4)$_{i,t}$					-0.0386***	-0.0264***
					(0.0015)	(0.0014)
Constant	-0.0057***	-0.0013**	-0.0053***	-0.0056***	-0.0003	0.0023***
	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0003)	(0.0004)
Number of Periods	2,262	2,203	2,262	2,262	2,262	2,203
Avg. Obs. per Period	2,441	2,214	2,441	2,425	2,410	2,203
Observations	5,521,721	4,877,031	5,521,721	5,484,786	5,450,500	4,853,987
Avg. R^2	0.040	0.060	0.042	0.042	0.091	0.107

Notes: I estimate $Return_{i,t+1} = \alpha + \beta_1 \text{PeerAvgReturn}_{i,t} + \sum_k \gamma_k X_{k,i,t} + \epsilon_{i,t}$ using Fama-MacBeth regressions. PeerAvgReturn$_{i,t}$ is the simple average return of stock i's peers on day t. Peers are defined based on retail trading activity comovement with correlation coefficient > 0.4 and top 20 market cap. Sample: 2016-2024. Column (1): basic controls (log market cap, price, volume, return$_t$, cumulative return$_{[t-5,t-1]}$). Column (2) adds fundamentals (institutional ownership, log book-to-market, analysts, ROA). Column (3) adds sentiment scores (CSS and ESS from Dow Jones and web sources using RavenPack). Column (4) adds log FINRA short volume. Column (5) adds risk measures (FF4 alpha, factor betas, IVOL from 252-day rolling window). Column (6): full specification. Standard errors in parentheses. *, **, *** denote significance at 10%, 5%, 1%.

Table XVI. Price Discovery: Cross-Sectional Variations

	Dependent Variable: $Return_{i,t+1}$					
	(1) Base	(2) Positive Asymmetry	(3) Information Gap	(4) Buy the Dip	(5) Institutional Holdings	(6) Analyst Coverage
Peer Avg. $\text{Return}_{i,t}$	0.0118***	0.0031***	0.0127***	0.0129***	0.0308***	0.0165***
	(0.0006)	(0.0009)	(0.0006)	(0.0007)	(0.0015)	(0.0008)
$\mathbf{1}(\text{Peer Avg. Return}_{i,t} > 0)$		-0.0001**				
		(0.0000)				
Peer Avg. $\text{Return}_{i,t} \times \mathbf{1}(> 0)$		0.0213***				
		(0.0016)				
Peer Avg. $\text{Return}_{i,t}\times$ (Peer-Own)			0.0851***	0.0481***		
			(0.0096)	(0.0098)		
$\mathbf{1}(\text{Return}_{i,t} < 5\text{th percentile})$				-0.0005***		
				(0.0001)		
Peer Avg. $\text{Return}_{i,t} \times \mathbf{1}(< 5\text{th})$				-0.0277***		
				(0.0040)		
Peer Avg. $\text{Return}_{i,t} \times \mathbf{1}(< 5\text{th}) \times$ (Peer-Own)				0.6553***		
				(0.0555)		
Inst. Ownership $(\%)_{i,t}$					0.0008***	
					(0.0001)	
Peer Avg. $\text{Return}_{i,t}\times$ Inst. Own.					-0.0301***	
					(0.0017)	
Num. $\text{Analysts}_{i,t}$						-0.0000
						(0.0000)
Peer Avg. $\text{Return}_{i,t}\times$ Num. Analysts						-0.0008***
						(0.0001)
$\text{Return}_{i,t}$	-0.0219***	-0.0219***	-0.0226***	-0.0125***	-0.0220***	-0.0216***
	(0.0011)	(0.0011)	(0.0011)	(0.0014)	(0.0011)	(0.0011)
$\text{Return}_{i,[t-5,t-1]}$	-0.0022***	-0.0023***	-0.0023***	-0.0022***	-0.0025***	-0.0024***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
$\log(\text{Book-to-Market})_{i,t}$	0.0002***	0.0002***	0.0002***	0.0002***	0.0002***	0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\log(\text{Market Cap})_{i,t}$	0.0002***	0.0002***	0.0002***	0.0001***	0.0001***	0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Constant	-0.0001	-0.0002	-0.0001	-0.0001	0.0010*	-0.0003
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0007)
Number of Periods	2,203	2,203	2,203	2,203	2,203	2,203
Avg. Obs. per Period	2,392	2,392	2,392	2,392	2,297	2,303
Observations	5,270,599	5,270,599	5,270,599	5,270,599	5,060,785	5,072,437
Avg. R^2	0.033	0.034	0.034	0.038	0.039	0.039

Notes: This table reports Fama-MacBeth regression results testing cross-sectional variations in the price discovery hypothesis. The dependent variable is $Return_{i,t+1}$. $\text{PeerAvgReturn}_{i,t}$ is the equal-weighted average return of stock i's peers on day t, where peers are identified based on retail trading activity comovement (correlation > 0.4, top 20 by market cap). Column (1) shows the base specification. Column (2) tests asymmetric effects for positive peer returns. Column (3) examines information gap effects using the interaction with peer-own return difference. Column (4) tests "buy the dip" behavior when focal stock returns fall below the daily 5th percentile. Columns (5) and (6) test complementarity with institutional information sources. Standard errors in parentheses. *, **, *** denote significance at 10%, 5%, 1% levels.

Table XVII. Price Discovery: Microstructure Outcomes

	Dependent Variables				
	(1)	(2)	(3)	(4)	(5)
	$\text{Return}_{i,t+1}$	$\text{Open-Close}_{i,t+1}$	$\text{Var Ratio}_{i,t+1}$	$\text{IV CATM}_{i,t+1}$	$\text{IV VW}_{i,t+1}$
$\text{Peer Avg. Return}_{i,t}$	0.0118***	0.0055***	-0.0264***	0.0174***	0.0281***
	(0.0006)	(0.0004)	(0.0021)	(0.0052)	(0.0072)
$\text{Return}_{i,t}$	-0.0219***	-0.0089***	0.0031	-0.0639***	-0.1323***
	(0.0011)	(0.0010)	(0.0052)	(0.0188)	(0.0198)
$\text{Return}_{i,[t-5,t-1]}$	-0.0022***	0.0003	0.0191***	-0.0130	-0.0787***
	(0.0004)	(0.0004)	(0.0038)	(0.0166)	(0.0176)
$\log(\text{Book-to-Market})_{i,t}$	0.0002***	0.0002***	-0.0087***	-0.0284***	-0.0325***
	(0.0000)	(0.0000)	(0.0001)	(0.0006)	(0.0007)
$\log(\text{Market Cap})_{i,t}$	0.0002***	0.0003***	-0.0623***	-0.0760***	-0.1286***
	(0.0000)	(0.0000)	(0.0002)	(0.0005)	(0.0009)
Constant	-0.0001	-0.0041***	1.5754***	1.6814***	2.8739***
	(0.0006)	(0.0005)	(0.0050)	(0.0134)	(0.0212)
Number of Periods	2,203	2,203	2,203	1,868	1,868
Avg. Obs. per Period	2,392	2,392	2,390	1,296	1,510
Observations	5,270,599	5,270,635	5,265,258	2,421,172	2,820,297
Avg. R^2	0.033	0.031	0.579	0.397	0.457

Notes: This table reports Fama-MacBeth regression results for different microstructure outcomes. Column (1) replicates the base specification for comparison. Column (2) uses open-to-close returns as the dependent variable. Column (3) uses variance ratio (absolute deviation from 1) to measure price efficiency. Columns (4) and (5) use call option implied volatilities: at-the-money (CATM) and volume-weighted (VW). All specifications include the same control variables. The positive coefficient on implied volatilities and negative coefficient on variance ratio support the price discovery hypothesis. Standard errors in parentheses. *, **, *** denote significance at 10%, 5%, 1% levels.

Table XVIII. Price Discovery: Persistence Across Time Horizons

	Dependent Variable: Future Returns at Different Horizons							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\text{Return}_{i,t+2}$	$\text{Return}_{i,t+3}$	$\text{Return}_{i,t+4}$	$\text{Return}_{i,t+5}$	$\sum\text{Return}_{i,[t+6,t+10]}$	$\sum\text{Return}_{i,[t+11,t+20]}$	$\sum\text{Return}_{i,[t+21,t+50]}$	$\sum\text{Return}_{i,[t+51,t+126]}$
Peer Avg. $\text{Return}_{i,t}$	0.0026***	0.0017***	0.0006	0.0004	0.0063***	0.0098***	0.0278***	0.0414***
	(0.0004)	(0.0004)	(0.0008)	(0.0008)	(0.0009)	(0.0013)	(0.0023)	(0.0039)
$\text{Return}_{i,t}$	-0.0022***	-0.0045***	-0.0016	-0.0024	-0.0083***	-0.0057**	0.9750***	0.9905***
	(0.0008)	(0.0009)	(0.0016)	(0.0016)	(0.0021)	(0.0028)	(0.0045)	(0.0075)
$\text{Return}_{i,[t-5,t-1]}$	-0.0016***	-0.0005	0.0002	0.0006	-0.0009	0.0044*	0.9851***	0.9937***
	(0.0004)	(0.0004)	(0.0008)	(0.0008)	(0.0016)	(0.0023)	(0.0038)	(0.0073)
$\log(\text{Book-to-Market})_{i,t}$	0.0002***	0.0002***	0.0002***	0.0002***	0.0009***	0.0016***	0.0015***	-0.0036***
	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0005)	(0.0009)
$\log(\text{Market Cap})_{i,t}$	0.0001***	0.0001***	0.0001***	0.0001***	0.0007***	0.0014***	0.0036***	0.0107***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0002)	(0.0003)	(0.0005)
Constant	0.0002	-0.0000	-0.0001	-0.0002	-0.0006	-0.0019	-0.0436***	-0.2482***
	(0.0006)	(0.0005)	(0.0009)	(0.0008)	(0.0025)	(0.0045)	(0.0076)	(0.0137)
Number of Periods	2,202	2,201	2,200	2,199	2,198	2,193	2,183	2,138
Avg. Obs. per Period	2,392	2,391	2,390	2,389	2,388	2,383	2,369	2,312
Observations	5,266,437	5,262,281	5,258,128	5,253,979	5,249,127	5,226,448	5,170,779	4,942,408
Avg. R^2	0.031	0.031	0.030	0.029	0.035	0.036	0.224	0.118

Notes: This table reports Fama-MacBeth regression results examining the persistence of peer effects across different time horizons. Columns (1)-(4) show single-day future returns from $t+2$ to $t+5$. Columns (5)-(8) show cumulative returns over longer horizons: days $t+6$ to $t+10$, $t+11$ to $t+20$, $t+21$ to $t+50$, and $t+51$ to $t+126$. The persistent positive coefficients on peer average return across all horizons, with no sign reversal, indicate permanent price impact consistent with genuine price discovery rather than temporary price pressure. The coefficients for own returns approach 1 for longer horizons due to the cumulative nature of returns. Standard errors in parentheses. *, **, *** denote significance at 10%, 5%, 1% levels.

Table XIX. Liquidity Provision: Trading Activity and Market Depth

	Dependent Variables				
	(1) log(Dollar Volume$_{t+1}$)	(2) log(Avg Depth$_{t+1}$)	(3) log(Offer Depth$_{t+1}$)	(4) log(Bid Depth$_{t+1}$)	(5) log(Bid/ Offer$_{t+1}$)
Peer Avg. Return$_{i,t}$	0.3446***	0.0958***	0.0477***	0.1170***	0.0697***
	(0.0287)	(0.0106)	(0.0114)	(0.0108)	(0.0067)
Return$_{i,t}$	0.8683***	0.0435***	0.4437***	-0.3274***	-0.7710***
	(0.0642)	(0.0105)	(0.0158)	(0.0164)	(0.0249)
Return$_{i,[t-5,t-1]}$	0.1022**	0.1670***	0.3378***	-0.0079	-0.3457***
	(0.0444)	(0.0098)	(0.0108)	(0.0110)	(0.0117)
log(Book-to-Market)$_{i,t}$	-0.0399***	0.0113***	0.0103***	0.0141***	0.0037***
	(0.0020)	(0.0011)	(0.0012)	(0.0011)	(0.0004)
log(Market Cap)$_{i,t}$	1.0392***	0.3966***	0.4232***	0.4058***	-0.0174***
	(0.0025)	(0.0014)	(0.0015)	(0.0015)	(0.0006)
Constant	-6.5572***	0.9619***	0.3394***	0.7828***	0.4433***
	(0.0446)	(0.0354)	(0.0359)	(0.0339)	(0.0143)
Number of Periods	2,203	2,203	2,203	2,203	2,203
Avg. Obs. per Period	2,359	2,359	2,359	2,359	2,359
Observations	5,196,886	5,196,778	5,196,850	5,196,778	5,196,778
Avg. R^2	0.822	0.621	0.627	0.607	0.024

Notes: This table reports Fama-MacBeth regression results for trading activity and market depth outcomes. Column (1) examines trading activity response through dollar volume. Columns (2)-(5) decompose market depth effects, showing stronger impacts on bid depth than offer depth, consistent with retail investors' net buyer position. Column (5) directly measures the bid-offer asymmetry. All specifications include standard controls. Standard errors in parentheses. *, **, *** denote significance at 10%, 5%, 1% levels.

Table XX. Liquidity Provision: Spread Effects

	Dependent Variables	
	(1) Quoted Spread$_{t+1}$	(2) Effective Spread$_{t+1}$
Peer Avg. Return$_{i,t}$	0.0012***	0.0010***
	(0.0001)	(0.0001)
Return$_{i,t}$	0.0013***	0.0006***
	(0.0002)	(0.0001)
Return$_{i,[t-5,t-1]}$	-0.0006***	-0.0006***
	(0.0002)	(0.0001)
log(Book-to-Market)$_{i,t}$	-0.0003***	-0.0002***
	(0.0000)	(0.0000)
log(Market Cap)$_{i,t}$	-0.0029***	-0.0018***
	(0.0000)	(0.0000)
Constant	0.0617***	0.0391***
	(0.0004)	(0.0003)
Number of Periods	2,203	2,203
Avg. Obs. per Period	2,359	2,359
Observations	5,196,778	5,196,683
Avg. R^2	0.410	0.430

Notes: This table reports Fama-MacBeth regression results for spread responses to peer returns. Column (1) shows the effect on quoted spreads and Column (2) on effective spreads. The positive coefficients indicate that peer returns lead to wider spreads, consistent with market makers adjusting for adverse selection risk from informed retail order flow. Standard errors in parentheses. *, **, *** denote significance at 10%, 5%, 1% levels.

Table XXI. Liquidity Provision: Cross-Sectional Variations with Market Liquidity

	Dependent Variable: $Return_{i,t+1}$			
	(1) Quoted Spread Interaction	(2) Effective Spread Interaction	(3) Market Cap Interaction	(4) Depth Interaction
Peer Avg. $\text{Return}_{i,t}$	0.0075***	0.0073***	0.1209***	0.0547***
	(0.0006)	(0.0005)	(0.0062)	(0.0035)
Quoted Spread $(\%)_{i,t}$	-0.0026			
	(0.0030)			
Peer Avg. $\text{Return}_{i,t}\times$ Quoted Spread	0.6559***			
	(0.0523)			
Effective Spread $(\%)_{i,t}$		0.0037		
		(0.0048)		
Peer Avg. $\text{Return}_{i,t}\times$ Eff. Spread		1.2409***		
		(0.0917)		
Peer Avg. $\text{Return}_{i,t} \times \log(\text{MCap})$			-0.0053***	
			(0.0003)	
$\log(\text{Depth})_{i,t}$				-0.0000**
				(0.0000)
Peer Avg. $\text{Return}_{i,t} \times \log(\text{Depth})$				-0.0048***
				(0.0004)
$\text{Return}_{i,t}$	-0.0216***	-0.0214***	-0.0219***	-0.0219***
	(0.0011)	(0.0011)	(0.0011)	(0.0011)
$\text{Return}_{i,[t-5,t-1]}$	-0.0021***	-0.0021***	-0.0022***	-0.0022***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
$\log(\text{Book-to-Market})_{i,t}$	0.0002***	0.0002***	0.0002***	0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\log(\text{Market Cap})_{i,t}$	0.0001***	0.0002***	0.0002***	0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Constant	0.0000	-0.0004	-0.0003	-0.0003
	(0.0007)	(0.0007)	(0.0006)	(0.0006)
Number of Periods	2,203	2,203	2,203	2,203
Avg. Obs. per Period	2,359	2,359	2,359	2,359
Observations	5,196,742	5,196,646	5,196,849	5,196,742
Avg. R^2	0.039	0.040	0.034	0.036

Notes: This table reports Fama-MacBeth regression results testing how peer effects vary with stock liquidity characteristics. Columns (1)-(2) show that peer effects are stronger in stocks with wider spreads (lower liquidity). Columns (3)-(4) show that peer effects are weaker in stocks with larger market capitalization and greater depth (higher liquidity). These patterns confirm that retail liquidity provision has greater marginal impact in less liquid markets. Standard errors in parentheses. *, **, *** denote significance at 10%, 5%, 1% levels.

Table XXII. Liquidity Provision: Persistence of Bid Depth Effects

	Dependent Variable: log(Bid Depth) at Different Horizons							
	(1) $t+2$	(2) $t+3$	(3) $t+4$	(4) $t+5$	(5) $t+10$	(6) $t+20$	(7) $t+50$	(8) $t+126$
Peer Avg. Return$_{i,t}$	0.1073***	0.1097***	0.1034***	0.0945***	0.0658***	0.0868***	0.0764***	0.0604***
	(0.0113)	(0.0115)	(0.0211)	(0.0211)	(0.0210)	(0.0213)	(0.0112)	(0.0118)
Return$_{i,t}$	-0.0097	0.0086	0.0054	-0.0071	-0.0485**	-0.0393*	-0.0541***	-0.0550***
	(0.0145)	(0.0126)	(0.0216)	(0.0213)	(0.0205)	(0.0205)	(0.0101)	(0.0110)
Return$_{i,[t-5,t-1]}$	-0.0096	-0.0143	-0.0194**	-0.0258***	-0.0382***	-0.0283***	-0.0691***	-0.0427***
	(0.0105)	(0.0102)	(0.0097)	(0.0097)	(0.0095)	(0.0093)	(0.0089)	(0.0099)
log(Book-to-Market)$_{i,t}$	0.0142***	0.0143***	0.0144***	0.0145***	0.0149***	0.0155***	0.0153***	0.0119***
	(0.0011)	(0.0011)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0011)	(0.0012)
log(Market Cap)$_{i,t}$	0.4056***	0.4056***	0.4055***	0.4056***	0.4055***	0.4051***	0.4043***	0.4017***
	(0.0015)	(0.0015)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0015)	(0.0017)
Constant	0.7890***	0.7906***	0.7934***	0.7949***	0.8021***	0.8173***	0.8331***	0.8364***
	(0.0339)	(0.0340)	(0.0196)	(0.0196)	(0.0197)	(0.0198)	(0.0346)	(0.0380)
Number of Periods	2,202	2,201	2,200	2,199	2,194	2,184	2,154	2,078
Avg. Obs. per Period	2,358	2,357	2,357	2,356	2,352	2,344	2,319	2,253
Observations	5,192,678	5,188,585	5,184,487	5,180,391	5,159,906	5,118,748	4,995,036	4,681,040
Avg. R^2	0.607	0.606	0.606	0.606	0.604	0.601	0.594	0.580

Notes: This table reports Fama-MacBeth regression results examining the persistence of bid depth effects across different time horizons. The dependent variable is log bid depth at horizons from $t+2$ to $t+126$ trading days. The persistent positive coefficients on peer average return, with magnitudes declining gradually from 0.1073 to 0.0604 but never reversing sign, demonstrate that retail investors provide persistent rather than transitory liquidity. This pattern is inconsistent with attention-driven trading that would predict reversals. Standard errors in parentheses. *, **, *** denote significance at 10%, 5%, 1% levels.

Table XXIII. Peer Effects on Future Returns by Asset Type: Fama-MacBeth Regressions

	Dependent Variable: $Return_{i,t+1}$					
	All Assets	Common Stocks	Foreign Stocks	REITs	ETPs	ADRs
$Peer_Avg_Return_{i,t}$	0.0122***	0.0121***	0.0178***	0.0038***	0.0033***	0.0212***
	(18.81)	(16.56)	(13.50)	(3.48)	(10.61)	(14.43)
$Return_{i,t}$	-0.0269***	-0.0268***	-0.0391***	-0.0197***	-0.0093***	-0.0504***
	(-24.18)	(-24.95)	(-33.32)	(-8.27)	(-2.67)	(-41.87)
$Return_Lags_{i,t}$	-0.0016***	-0.0023***	-0.0005	-0.0040***	0.0005	-0.0030***
	(-2.93)	(-4.92)	(-0.91)	(-3.93)	(0.27)	(-4.74)
$Log(Market_Cap)_{i,t}$	0.0001***	0.0002***	0.0002***	0.0000**	-0.0000	0.0002***
	(11.80)	(10.47)	(8.89)	(1.95)	(-0.25)	(10.50)
Constant	-0.0022***	-0.0035***	-0.0032***	-0.0005	0.0003***	-0.0035***
	(-8.57)	(-7.82)	(-7.38)	(-1.12)	(3.45)	(-8.68)
Average R^2	0.031	0.028	0.039	0.122	0.233	0.046
Time Periods	2,262	2,262	2,262	2,262	2,262	2,262
Observations	12,003,303	6,078,171	1,059,927	294,697	2,913,411	674,163

Notes: This table reports Fama-MacBeth cross-sectional regression results for the effect of peer average returns on next-day returns across different asset types. The sample period is 2016-2024. $Return_Lags_{i,t}$ represents the sum of returns from t-1 to t-5. T-statistics based on Newey-West standard errors are reported in parentheses. Statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table XXIV. Peer Effects on Future Returns by Exchange: Fama-MacBeth Regressions

	Dependent Variable: $Return_{i,t+1}$				
	All Exchanges	NYSE	NASDAQ	NYSE Arca	Cboe
$Peer_Avg_Return_{i,t}$	0.0122***	0.0085***	0.0153***	0.0039***	0.0001
	(18.81)	(14.91)	(18.11)	(10.09)	(0.10)
$Return_{i,t}$	-0.0269***	-0.0201***	-0.0343***	-0.0071**	-0.0243***
	(-24.18)	(-15.06)	(-35.17)	(-2.04)	(-5.64)
$Return_Lags_{i,t}$	-0.0016***	-0.0022***	-0.0021***	-0.0010	0.0000
	(-2.93)	(-3.63)	(-4.76)	(-0.55)	(0.01)
$Log(Market_Cap)_{i,t}$	0.0001***	0.0001***	0.0002***	0.0000	-0.0000***
	(11.80)	(5.17)	(13.40)	(1.25)	(-2.80)
Constant	-0.0022***	-0.0009***	-0.0040***	0.0001**	0.0005***
	(-8.57)	(-2.93)	(-10.48)	(1.98)	(4.44)
Average R^2	0.031	0.039	0.024	0.230	0.284
Time Periods	2,262	2,262	2,262	2,262	2,262
Observations	12,003,303	4,331,723	5,236,814	2,035,067	399,448

Notes: This table reports Fama-MacBeth cross-sectional regression results for the effect of peer average returns on next-day returns across different exchanges. The sample period is 2016-2024. $Return_Lags_{i,t}$ represents the sum of returns from t-1 to t-5. T-statistics based on Newey-West standard errors are reported in parentheses. Statistical significance: *** p < 0.01, ** p < 0.05, * p < 0.10.

Table XXV. Summary Statistics: Average Asset Size per Client, 2016-2024

	Average Assets per Account ($000s)			Growth Metrics		
Broker	2016	2024	Mean	CAGR	Total Growth	Acct. Growth
Charles Schwab	246.9	277.0	253.8	1.5%	12.2%	224%
Fidelity	318.4	385.2	338.8	2.4%	21.0%	119%
Interactive Brokers	220.8	170.2	216.4	-3.2%	-22.9%	767%
Robinhood	2.8*	7.7	4.6	13.4%**	175%**	394%**

Notes: Average assets per account calculated as total client assets divided by number of accounts. CAGR = Compound Annual Growth Rate. *Robinhood 2016 data not available; 2019 data used as starting point. **Robinhood growth metrics calculated from 2019-2024. Account growth for Robinhood calculated from 2019 (5.10M Net Cumulative Funded Accounts) to 2024 (25.2M Funded Customers).

Definitions: Charles Schwab: Active Brokerage Accounts. Fidelity: Total retail accounts (2016-2023), unique customers engaging with digital platforms (2024). Robinhood: Net Cumulative Funded Accounts (2019-2022), Funded Customers (2023-2024). IBKR: Client Accounts. Assets Under Administration (AUA) for Schwab and Fidelity. Client Equity for IBKR. Assets Under Custody (AUC) for Robinhood.

Sources: Calculated from data in company annual reports and SEC filings. See Table XXVII for complete source documentation.

Table XXVI. Detailed Average Asset Size per Client by Year, 2016-2024

	Average Assets per Account (Thousands of Dollars)								
Broker	2016	2017	2018	2019	2020	2021	2022	2023	2024
Panel A: Traditional Brokers									
Charles Schwab	246.9	281.4	252.1	294.3	225.8	246.8	208.9	244.5	277.0
Total Assets ($B)	2,780	3,362	3,252	4,039	6,691.7	8,183.0	7,049.8	8,516.6	10,101.3
Active Brokerage Accts (000s)	11,261	11,952	12,895	13,723	29,629	33,165	33,758	34,838	36,456
Fidelity	318.4	350.4	321.7	373.9	376.9	359.2	277.6	325.6	385.2
Total AUA ($T)	5.700	6.797	6.692	8.3	9.8	11.1	10.3	12.6	15.1
Total retail accounts (M)	17.9	19.4	20.8	22.20	26.0	30.9	37.1	38.7	39.2
Panel B: Active Trader Platform									
Interactive Brokers	220.8	258.8	214.0	252.3	268.5	223.2	146.8	166.3	170.2
Client Equity ($B)	85	125	128	174.1	288	374	307	426	568
Accounts (000s)	385	483	598	690	1,073	1,676	2,091	2,562	3,337
Panel C: New Generation Platform									
Robinhood	–	–	–	2.8	5.0	4.3	2.7	4.4	7.7
AUC ($M)	–	–	–	14,136	63,000	98,000	62,000	102,600	193,000
Funded Accts/Cust (M)	–	–	–	5.10	12.5	22.7	23.0	23.4	25.2

Notes: Average asset size per client account is calculated as total client assets divided by number of accounts. All values are in thousands of US dollars. Charles Schwab's 2020 figure includes 14.5 million new brokerage accounts related to the acquisition of TD Ameritrade and 1.1 million new brokerage accounts related to the acquisition of the assets of USAA's Investment Management Company. Fidelity metrics refer to Total retail accounts for 2016-2023. In 2024, the definition in the Fidelity Annual Report changed slightly to 39.2 million unique customers engaging with Fidelity.com, NetBenefits.com, or mobile apps, including unique customers who engage on Fidelity.com, NetBenefits.com, the Fidelity Investments Mobile App, or the NetBenefits Mobile App.
Definitions: Charles Schwab: Active Brokerage Accounts. Fidelity: Total retail accounts (2016-2023), unique customers engaging with digital platforms (2024). Robinhood: Net Cumulative Funded Accounts (2019-2022), Funded Customers (2023-2024). IBKR: Client Accounts. Assets Under Administration (AUA) for Schwab and Fidelity. Client Equity for IBKR. Assets Under Custody (AUC) for Robinhood.
Sources: See Table XXVII for complete source documentation with URLs.

Table XXVII. Detailed Source Documentation

Company	Source Details and URLs
Charles Schwab	Annual reports (2016-2024). Sources: Company annual reports.
Fidelity	Sources: - 2016, 2017, 2018 data from 2018 shareholder update: https://www.mutualfundobserver.com/wp-content/uploads/2019/03/2018-Fidelity-Shareholder-Update.pdf - 2019 data inferred from 2020 annual report and growth rate: https://www.fidelity.com/bin-public/060_www_fidelity_com/documents/about-fidelity/Fidelity_Investments_2020_AnnualReportInfographic.pdf (also provides 2020 data) - 2021 Q3 Business Update: https://www.fidelity.com/about-fidelity/our-company/quarterly-updates/quarterlyupdates-q3-2021 - 2022 annual report: https://www.fidelity.com/bin-public/060_www_fidelity_com/documents/about-fidelity/2022-Fidelity-Investments-Annual-Report.pdf - 2023 annual report: https://www.fidelity.com/bin-public/060_www_fidelity_com/documents/about-fidelity/2023-Fidelity-Investments-Annual-Report.pdf - 2024 annual report: https://www.fidelity.com/bin-public/060_www_fidelity_com/documents/about-fidelity/2024-Fidelity-Investments-Annual-Report.pdf
Robinhood	Sources: Annual reports, SEC filings, official website, and the following specific data points: - 2019 data: https://investors.robinhood.com/static-files/31f4ce87-6047-481c-a903-6bf08fe009d3 - 2021 data: https://www.sec.gov/Archives/edgar/data/1783879/000178387922000022/robinhoodex991_12312021.htm - 2024 data: https://www.sec.gov/Archives/edgar/data/1783879/000178387925000045/q42024robinhoodexhibit991.htm
Interactive Brokers	Annual reports (2016-2024). Sources: Company annual reports.

Table XXVIII. Validation of RHDW Fractional Share Trade Classification

	FINRA OTC Weekly Disclosed Trades	
	Coefficient	Standard Error
Weekly Estimated Combined Trades	0.9378***	(0.0006)
Constant	−2314.233***	(39.029)
Observations	31,312	
R^2	0.9862	

Notes: This table reports the regression of weekly RHDW fractional share trades disclosed in FINRA OTC Transparency data on estimated weekly RHDW fractional share trades. The sample includes stock-weeks from November 1, 2021 to December 31, 2024 with positive DriveWealth trading activity. Standard errors are heteroskedasticity-robust. *** $p < 0.01$.