

7459859-retail-trading-and-the-varying-predictive-power-of-aggregate-short-interest

AVTORJI: Shiran Froymovich, Shera Gong, Charles Wasley, Jason J. Xiao

POVZETEK: Raziskava je identificirala, da je skupna brezplačna zanimanje (short interest) ena od najinformativnejših prediktorjev prihodnjih skupnih povratkov trga. No, to informacijsko vrednost skupnega brezplačnega zanimanja odvisna je od sestave tržnih delavcev, ki ga omenjajo, ki se je signifikantno spremenila v zadnjih desetletjih. Raziskovalci nameravajo, da je predikativni moč skupnega brezplačnega zanimanja padel v času in da ta padec je negativno povezan z narastom delovanja brezplačnih tržnikov. Dodatne analize pokažejo, da delovanje brezplačnih tržnikov učinkovito vpliva skozi dva mehanizma: narastajočo delo brezplačnih tržnikov pri brezplačnem zanimanju in večja rizika zatiska brezplačnega zanimanja. Naši rezultati podčrtajo pomembnost sestave tržnikov pri ustreznosti skupnega brezplačnega zanimanja kot prediktorja skupnih povratkov trga.

METODA: Raziskovalci uporabljajo mesečne podatke od leta 2012 do 2023. Definirajo predikativni moč skupnega brezplačnega zanimanja (β_{SII}) kot koeficient v regresiji, kjer je skupno brezplačno zanimanje (SII) prediktor skupnih mesečnih povratkov S&P 500 za naslednjih 12 mesecev. Raziskovalci preučujejo, kako se predikativni moč skupnega brezplačnega zanimanja vpliva delovanje brezplačnih tržnikov, ki je definirano kot delo mesečnega tržnega obdobja, ki mu je pripadajoče brezplačno zanimanje. Podatki kontrolirajo mrežno likvidnost, makroekonomske nesigurnosti, volatilitnost trga in skupno ekonomsko dejavnost.

KLJUČNE UGOTOVITVE:

- Predikativni moč skupnega brezplačnega zanimanja je padel v času.
- Padek predikativni moči je negativno povezan z narastom delovanja brezplačnih tržnikov.
- Delovanje brezplačnih tržnikov vpliva skozi dva mehanizma: narastajočo delo brezplačnih tržnikov pri brezplačnem zanimanju in večja rizika zatiska brezplačnega zanimanja.

POMEN ZA TRGOVANJE: Raziskava podpira pomembnost sestave tržnikov pri ustreznosti skupnega brezplačnega zanimanja kot prediktorja skupnih povratkov trga. Trgovci lahko uporabljajo te rezultate za boljše razumevanje in predviđanje tržnih potez, sicer pa je pomembno razumeti, da se predikativni moč skupnega brezplačnega zanimanja lahko spremeni glede na spremembe sestave tržnikov.

Retail Trading and the Varying Predictive Power of Aggregate Short Interest

Shiran Froymovich
School of Management
Binghamton University, SUNY
svakninfroym@binghamton.edu

Shera Gong
School of Management
Binghamton University, SUNY
shera.gong@binghamton.edu

Charles Wasley[†]
Simon Business School
University of Rochester
charles.wasley@simon.rochester.edu

Jason J. Xiao
School of Management
Binghamton University, SUNY
jxiao3@binghamton.edu

September 2026

Abstract: Prior research identifies aggregate short interest as one of the strongest predictors of aggregate stock market returns. Yet, the informativeness of aggregate short interest ultimately depends on the composition of market participants whose trading activity it reflects, a population that has changed significantly over the past decade. We hypothesize and find that the predictive ability of aggregate short interest has declined significantly over time and that this decline is negatively associated with the concurrent rise in retail investor participation. Additional analyses suggest that retail investor participation affects predictability through two mechanisms: a growing share of retail short selling activity and a greater risk of short squeeze activity. Collectively, our findings highlight the importance of investor composition in shaping the information content of short selling activity and its ability to predict aggregate stock market returns.

Keywords: aggregate short interest; short selling; short squeeze; retail trading; institutional investors; aggregate stock returns

JEL classifications: E30; E44; G14; M41; O16

[†] Corresponding author. We gratefully acknowledge financial support from Binghamton University, SUNY, and the Simon School at the University of Rochester.

1. Introduction

Aggregate short interest is widely regarded as one of the most informative signals of future stock returns. For example, prior research finds that aggregate short interest predicts future aggregate stock returns and contains information about future economic fundamentals, an indication that it reflects the trading activity of informed short sellers (Lynch, Nikolic, Yan, and Yu, 2014; Rapach, Ringgenberg, and Zhou, 2016). Ultimately, the informativeness of aggregate short interest depends on the composition of market participants whose trading activity it aggregates. We hypothesize that as the composition of investors in the market changes, so will the amount and type of information embedded in aggregate short interest. Specifically, we investigate whether changes in investor composition, particularly the growth of retail investor participation, explains variability in aggregate short interest's ability to predict future aggregate stock market returns (hereafter, simply aggregate returns).

Two economic aspects of financial capital markets motivate our research question. First, the SEC's efforts to expand short selling disclosure underscore regulatory interest in short selling activity and its role in capital markets. Specifically, in 2023, the SEC adopted Rule 13f-2, expanding reporting requirements for institutional investment managers' short positions and short selling activity.¹ By examining how investor composition relates to the informativeness of aggregate short interest, our study can provide regulators with insights into the conditions under which short interest serves as an informative signal of future market returns. Second, retail investor participation has increased substantially over time (Boehmer, Jones, Zhang, and Zhang, 2021). Understanding how this shift in investor composition affects the information conveyed by aggregate short interest is important because short interest serves as a market wide signal of future

¹ Compliance with Rule 13f-2 and Form SHO has been deferred until January 2, 2028.

returns and economic fundamentals (Lynch et al., 2014; Rapach et al., 2016). Goyal, Welch, and Zafirov (2024) find that aggregate short interest continues to predict aggregate returns, but its predictive ability weakens in more recent years. Their finding leaves open whether changes in investor composition contribute to this decline. We examine whether and how greater retail participation helps explain time-series variation in the predictive ability of aggregate short interest.

Ex ante, it is unclear whether greater retail investor participation strengthens or weakens the predictive ability of aggregate short interest. On one hand, greater retail participation may strengthen predictability by increasing informed short selling. Prior research finds that retail investors possess value-relevant information, with evidence of retail order flow predicting future stock returns and firm news, and retail short selling predicting negative future stock returns and generating economically significant trading profits (Kelly and Tetlock 2013; 2017). Greater participation by informed retail short sellers may therefore increase the information embedded in aggregate short interest. In addition, technological innovations have also expanded access to information, analytical tools, and low-cost trading platforms, potentially enabling a broader set of retail investors to acquire information and trade on it (Boehmer et al., 2021; Farrell, Green, Jame, and Markov, 2022; Hvide, Meling, Mogstad, and Vestad, 2024). To the extent that these advances facilitate informed retail short selling, they may further strengthen aggregate short interest's predictive ability.

On the other hand, greater retail participation may weaken aggregate short interest's predictive ability through two mechanisms: retail short selling and short squeeze risk. First, through the retail short selling channel, greater retail participation may increase short selling driven by sentiment or attention rather than information about fundamentals. Prior research documents sentiment- and attention-driven retail trading (Kumar and Lee, 2006; Barber, Huang, Odean, and

Schwarz, 2022; Michels, 2024), as well as the presence of both informed and noise retail traders (Eaton, Green, Roseman, and Wu, 2022). To the extent that such noise trading extends to short positions, increased retail short selling may weaken the information content of aggregate short interest. Second, through a short squeeze risk channel, greater retail participation may increase short squeeze risk and thus the expected costs of short selling, reducing informed short sellers' incentives to produce information and trade on it (Schultz, 2024; Stice-Lawrence, Wong, and Zhao, 2025). Reduced informed short selling may, in turn, weaken the predictive ability of aggregate short interest. Taken together, whether greater retail investor participation strengthens or weakens the predictive ability of aggregate short interest is ultimately an empirical question.

We examine these competing predictions using monthly data from 2012 through 2023. We start by constructing β_{SII} , our proxy for aggregate short interest's predictive ability, as the coefficient on Rapach et al.'s (2016) aggregate short interest index (*SII*) in a regression of S&P 500 monthly excess returns averaged over the subsequent 12 months on *SII*. We re-estimate this regression each month using an expanding estimation window, starting in January 2007, to obtain a monthly series of β_{SII} .² We document a substantial decline in aggregate short interest predictability over our sample period. Four complementary tests of coefficient stability reject the hypothesis that the relation between aggregate short interest and future market returns is constant.

To examine whether greater retail participation helps explain variation in the predictive ability of aggregate short interest, we regress β_{SII} on *Retail%_Volume*, the fraction of monthly trading volume attributable to retail investors (identified using Boehmer et al.'s (2021) sub-penny methodology). Controlling for market liquidity, macroeconomic uncertainty, market volatility, and

² The estimation window starts in January 2007 because we require at least 60 monthly observations to estimate β_{SII} (see Goyal et al., 2024). The January 2012 estimate of β_{SII} uses *SII* observations from January 2007 through December 2011 and the corresponding subsequent 12-month aggregate return.

aggregate economic activity, we find a negative and significant coefficient on *Retail%_Volume*. This finding holds using alternative measures of retail participation and return-forecasting horizons, as well as weighted least squares estimation and a block bootstrap that incorporates uncertainty in the estimated β_{SII} . Consistent with the regression evidence, time series plots show that the declining predictive ability of aggregate short interest coincides with a sharp increase in retail participation.

Next, we investigate two mechanisms designed to explain the negative relation between retail participation and predictive power of aggregate short interest. First, to examine a retail short selling channel, we regress β_{SII} on the fraction of short selling volume attributable to retail investors along with controls. We find a negative and significant relation, consistent with the prediction that increased retail short selling dilutes the information embedded in aggregate short interest. Second, to examine a short squeeze risk channel, we regress β_{SII} on a measure of aggregate short squeeze risk along with controls. We find that predictability of aggregate short interest is significantly weaker during periods when short squeezes are more prevalent. This finding is consistent with the prediction that greater short squeeze risk raises the expected costs of short selling, discouraging informed short sellers from trading on their information.

To assess the relative importance of these mechanisms, we conduct a path analysis. For each channel, we decompose the total relation between retail participation and predictive ability of aggregate short interest into a direct component and an indirect component. Both indirect effects are negative and statistically significant. The retail short selling channel accounts for approximately half of the total relation, whereas the short squeeze risk channel accounts for approximately 14% to 21% of the total relation. These results suggest that both channels contribute

to the negative relation, with the retail short-selling channel comprising a larger portion of the effect.

Our study contributes to several literatures. First, we contribute to the literature on aggregate return predictability by providing evidence on why one of the most influential aggregate return predictors identified in prior research has become less informative over time (Lynch et al., 2014; Rapach et al., 2016; Goyal et al., 2024). While previous studies document the overall predictive ability of aggregate short interest, little is known about how and why it changes over time. We find that the weakening of aggregate short interest's predictive power for future aggregate returns is strongly associated with the rise of retail participation, an effect that helps explain the deterioration of an important market-wide information signal.

Second, we contribute to the short interest literature by showing that the information content of aggregate short interest depends on the market environment (composition) in which short sellers operate (Boehmer, Jones, and Zhang, 2008; Engelberg, Reed, and Ringgenberg, 2012; Lynch et al., 2014; Rapach et al., 2016). Prior research attributes the predictive ability of aggregate short interest to informed short sellers' ability to acquire, process, and trade on information about future economic fundamentals (Lynch et al., 2014; Rapach et al., 2016). Our findings suggest that the informativeness of aggregate short interest varies with changes in the composition of short sellers and the incentives they face.

Finally, we contribute to the growing literature on retail investors by documenting that the rise of retail participation has market-wide implications beyond those documented for individual stock pricing (Kumar and Lee, 2006; Kelley and Tetlock, 2013; Kelley and Tetlock, 2017; Boehmer et al., 2021). Our evidence indicates that retail participation affects the amount and type of information embedded in aggregate short interest, which in turn influences the predictive

content of this widely-followed future market return indicator. More broadly, our findings indicate that changes in investor composition can alter the extent to which information produced by sophisticated traders is reflected in market signals.

The remainder of the paper is organized as follows. Section 2 reviews related literature and develops our hypothesis. Section 3 describes the data and research design. Sections 4 and 5 discuss the results of the primary and additional analyses, respectively. Finally, Section 6 concludes.

2. Literature review and hypothesis development

A large literature concludes that short sellers are informed market participants whose trades reflect value-relevant information about future firm performance. For example, prior studies find that heavily shorted stocks subsequently earn lower returns, that institutional short sellers possess valuable information, and that short interest activity predicts negative corporate events, including disappointing earnings announcements and analyst downgrades (Desai, Ramesh, Thiagarajan, and Balachandran, 2002; Christophe, Ferri, and Angel, 2004; Boehmer et al., 2008; Diether, Lee, and Werner, 2009; Christophe, Ferri, and Hsieh, 2010). Consistent with the inference that short sellers are sophisticated information intermediaries, Engelberg et al. (2012) find that short-sellers profit from superior processing of publicly available information, while Boehmer and Wu (2013) find that short selling contributes to price discovery, facilitating the incorporation of information into stock prices. Collectively, this evidence suggests that short sellers possess informational advantages and process information in ways that allow their trading activity to predict future stock returns and corporate outcomes.

Evidence of the informational role of short sellers at the firm level naturally raises the question of whether their collective trading activity contains information about future market-wide outcomes. Early evidence in Lamont and Stein (2004) suggests that aggregate short interest reflects

broad market valuations and the ability of arbitrageurs to respond to market-wide mispricing. Building on this idea, Lynch et al. (2014) show that aggregate short selling exhibits substantial commonality across stocks and predicts future market returns. They also provide evidence that short sellers possess information about future aggregate earnings, macroeconomic conditions, and investor sentiment. Rapach et al. (2016) provide perhaps the strongest evidence on the predictive ability of aggregate short interest by showing that it outperforms numerous commonly-used return predictors both in- and out-of-sample. In specific, their findings indicate that aggregate short interest predicts future aggregate returns primarily through a cash flow channel, consistent with short sellers anticipating future aggregate economic fundamentals. Together, these studies suggest that aggregate short interest contains information about future aggregate returns because short sellers possess information about future aggregate cash flows and economic conditions that market prices do not yet fully reflect.

While prior research attributes the predictive ability of aggregate short interest to the informational advantages of short sellers, the composition of market participants has changed substantially over recent decades. One of the most notable changes has been the rapid growth of retail investor participation. This development raises the question of how the informativeness of aggregate short interest for aggregate returns varies over time due to an increasing portion of trading activity originating from retail investors and the pressures they may place on short sellers.

On one hand, the predictive ability of aggregate short interest for future aggregate returns may increase over time with retail investing. While retail investors are sometimes viewed as less sophisticated market participants, recent evidence suggests that at least some retail traders possess valuable information. For example, Kelley and Tetlock (2013) show that retail order flow predicts future stock returns and future firm news, including earnings surprises, while Kelley and Tetlock

(2017) find that retail short selling predicts negative future stock returns and generates economically significant trading profits. The rapid growth in retail investor participation coincides with technological advancements that have expanded investors' access to information, analytical tools, and low-cost trading platforms (Boehmer et al., 2021; Farrell et al., 2022; Hvide et al., 2024). These innovations not only facilitate greater retail investor participation, but also reduce information acquisition and processing costs, potentially enabling a broader set of investors to acquire, analyze, and act on value-relevant information. Consistent with the notion that information-processing ability contributes to short sellers' informational advantage, Engelberg et al. (2012) find that a substantial portion of short sellers' trading profits arises from their ability to analyze publicly available information. Technological developments that enable a growing number of investors, including retail short sellers, to trade on value-relevant information suggest that aggregate short interest may incorporate more information about future economic fundamentals, thereby increasing its ability to predict future aggregate returns.

On the other hand, alternative arguments can be made that would imply that the predictive ability of aggregate short interest for future aggregate returns could weaken over time. First, although Kelley and Tetlock (2017) document that retail short selling predicts negative future returns, that predictability is concentrated among smaller firms and stocks with substantial retail trading activity. Consequently, information generated by retail short sellers may have limited implications for future aggregate returns if such firms represent a relatively small portion of the overall market. In addition, other studies find that retail trading includes both informed and noise trading (Eaton et al., 2022), with noise trading often reflecting investor sentiment or attention (Kumar and Lee, 2006; Michels, 2024). Thus, to the extent that increased retail participation

introduces trading that is less informative about future aggregate economic fundamentals, the signal-to-noise ratio of aggregate short interest may decline over time.

Finally, increased retail participation in financial capital markets can alter the incentives facing informed short sellers. For example, Schultz (2024) shows that short squeezes impose economically significant costs on short sellers by forcing them to close positions prematurely, incur trading costs, and forgo future profits. These costs are particularly pronounced when short selling demand is high and the supply of lendable shares is limited. In addition, Stice-Lawrence et al. (2025) find that retail investor activity is positively associated with the likelihood of short squeezes following activist short selling campaigns. As a result, the growth of retail participation may increase the risks and costs associated with maintaining short positions. By increasing the likelihood and costs of short squeezes, greater retail participation could serve to reduce the expected profitability of informed short selling strategies and thereby weaken sophisticated investors' incentives to produce information and engage in arbitrage activity. This mechanism is particularly important because prior research concludes that informed short sellers devote substantial resources to identifying and trading against overvaluation, despite the presence of short-sale constraints (Ljungqvist and Qian, 2016). In summary, to the extent that aggregate short interest derives its predictive power from informed traders' ability to acquire, process, and trade on information about future economic fundamentals (Lynch et al., 2014; Rapach et al., 2016), a reduction in information production and arbitrage activity could weaken its informational content.

Collectively, the literature provides competing predictions regarding how greater retail participation affects the predictive ability of aggregate short interest. On the one hand, increased retail participation and technological innovations may strengthen its information content by expanding access to information and analytical tools and lowering information processing costs.

On the other hand, greater retail participation may introduce less informative short selling and increase short squeeze risk, reducing the expected profitability of informed short selling and discouraging information production. Whether greater retail participation strengthens or weakens the predictive ability of aggregate short interest is ultimately an empirical question. Accordingly, we test the following hypothesis (stated in null form):

H₀: Retail investor participation is unrelated to the predictive ability of aggregate short interest for future market returns.

3. Research design and data

3.1. Research design and variable measurement

We investigate changes in the aggregate return predictability of an aggregate short interest index (*SII*) using a two-stage procedure. We first estimate a monthly series of *SII* predictability coefficients. We then examine whether time-series variation in those coefficients is associated with changes in retail trading activity.

First stage estimation is based on the following forecasting regression:

$$Return_{t,h} = \alpha_t + \beta_{SII,t,h} SII_t + \epsilon_{t,h}, (Eq. 1)$$

where t denotes the month in which the forecast is formed and h denotes the return forecasting horizon (i.e., holding period). The dependent variable, $Return_{t,h}$, is the monthly log excess return on the S&P 500 averaged over the subsequent h months. The primary independent variable, SII_t , is the aggregate short interest index measured in month t , which is detrended and standardized to have unit standard deviation (see Rapach et al., 2016). The coefficient of interest, $\beta_{SII,t,h}$, measures *SII*'s return predictability.³

³ Our main analyses use $h = 12$ months as the return forecasting horizon (i.e., holding period). As in Rapach et al. (2016), we also estimate Eq.1 for $h = 1, 3$, and 6 as alternative forecasting horizons. As discussed in Section 5, our main results are robust under these alternative specifications.

Following Rapach et al. (2016), we estimate an alternative regression to control for commonly used aggregate return predictors. Specifically, in Eq. 2 we augment Eq. (1) with the first three principal components, extracted from the 14 return predictors examined by Welch and Goyal (2008), to estimate the following model:

$$Return_{t,h} = \alpha_t + \beta_{SII|PC,t,h}SII_t + \sum_{j=1}^3 \gamma_{j,t,h}PC_{j,t} + \epsilon_{t,h}, (Eq. 2)$$

where $PC_{j,t}$ denotes the j -th principal component.⁴ The principal-component approach is a parsimonious way to control for the information contained in the broader set of established return predictors. As in Rapach et al. (2016), we use this specification to assess whether SII contains predictive information beyond that captured collectively by other return predictors. In Eq. 2, $\beta_{SII|PC,t,h}$ measures the incremental relation between SII and future aggregate returns after controlling for the information contained in the other predictors.

In the first stage, Eq. (1) and (2) are estimated recursively using an expanding estimation window that begins in January 2007. Specifically, each month, we add that month's (new) observation to the estimation sample and re-estimate the forecasting regression. This procedure yields a monthly time-series of SII predictability coefficients, $\beta_{SII,t,h}$ and $\beta_{SII|PC,t,h}$. For readability, we refer to these measures as β_{SII} and $\beta_{SII|PC}$, respectively, omitting the subscripts t and h since the relevant month (t) and forecasting horizon (h) are clear from the context. To account for heteroskedasticity and serial correlation induced by overlapping return observations, we compute standard errors using the Newey and West (1987) estimator with 12 lags.

⁴ The 14 return predictors are the dividend-price ratio (DP); dividend yield (DY); earnings-price ratio (EP); dividend-payout ratio (DE); stock return volatility (RVOL); book-to-market ratio (BM); net equity expansion (NTIS); Treasury bill rate (TBL); long-term government bond yield (LTY); long-term government bond return (LTR); term spread (TMS); default yield spread (DFY); default return spread (DFR); and inflation (INFL).

In the second stage, we investigate whether the time-series variation in the return predictability of SII is associated with changes in retail trading activity. To do so, we estimate the following monthly regression:

$$\beta_{SII,t,h}(\beta_{SII|PC,t,h}) = \eta_h + \theta_h Retail\%_Volume_t + \gamma' Controls_t + \mu_{t,h} \text{ (Eq. 3)}$$

where t denotes the month and h denotes the return forecasting horizon ($h = 12$ in our main tests). The dependent variables, $\beta_{SII,t,h}$ or $\beta_{SII|PC,t,h}$, are the estimated SII predictability coefficients from Eqs. (1) and (2), respectively. The primary independent variable of interest, $Retail\%_Volume$, is aggregate retail participation in the market, which is measured as the fraction of monthly trading volume attributable to retail investors. Retail trades are identified using the sub-penny method of Boehmer et al. (2021). To investigate alternative dimensions of retail trading activity, in additional analyses we replace $Retail\%_Volume$ with the fraction of monthly retail dollar trading volume ($Retail\%_Dollar Volume$) or the fraction of monthly retail trade count ($Retail\%_Trade Count$). Estimation of second-stage models includes a vector of market-level control variables that capture market liquidity and macroeconomic conditions. These variables consist of Amihud (2002)'s illiquidity measure, Jurado, Ludvigson, and Ng (2015)'s macroeconomic uncertainty index, the CBOE Volatility Index (VIX), and the Chicago Fed National Activity Index (CFNAI). Detailed variable definitions are provided in Appendix A.

We utilize three alternative estimation procedures to enhance the reliability of our inferences. First, the error terms in Eq. (3) may exhibit heteroskedasticity and serial correlation, rendering Ordinary Least Squares (OLS) standard errors and t -statistics unreliable. To account for these concerns, we use Newey and West (1987) heteroskedasticity- and autocorrelation-consistent standard errors in an alternative estimation of Eq. (3). Second, since the number of observations in the first-stage estimation window increases each month, the precision with which each monthly

model estimates $\beta_{SII}(\beta_{SII|PC})$ varies over time. Estimates from earlier in the sample period are based on fewer observations and are therefore less precise, whereas estimates from later in the sample period are based on more observations and are more precise. This variation in precision may render OLS estimation inefficient (Saxonhouse, 1976). To account for such differences, Saxonhouse (1976) recommends weighted least squares (WLS) estimation, which assigns greater weight to more precisely estimated observations. As an alternative specification, we estimate Eq. (3) using WLS by weighting each monthly $\beta_{SII}(\beta_{SII|PC})$ observation by the inverse of its squared first-stage standard error. The WLS specification places greater weight on the more precisely estimated $\beta_{SII}(\beta_{SII|PC})$ observations of later periods, thereby theoretically improving the efficiency of the second-stage coefficient estimates. Finally, conventional second-stage standard errors may not fully reflect the estimation uncertainty introduced in the first stage, potentially overstating the statistical significance of the second-stage estimates (Pagan, 1984; Murphy and Topel, 1985; Chen, Hribar, and Melessa, 2018). To address this, Chen et al. (2018) recommend bootstrapping the complete two-stage estimation. To do that, we implement a block bootstrap that re-estimates both stages, including the first-stage estimation of $\beta_{SII}(\beta_{SII|PC})$ and the WLS second-stage regression, for 500 replications. This procedure incorporates first-stage estimation uncertainty into second-stage statistical inference.

3.2. Data and sample selection

Our sample period spans from January 2012 to December 2023, yielding 144 monthly observations. The sample begins in January 2012 for two reasons. First, the sub-penny method used to identify retail trading activity can be implemented only beginning in January 2007 (Boehmer et al., 2021). We therefore begin the first-stage estimation window in that month so that it coincides with the period over which retail trading activity can be measured. This alignment

ensures that the estimated predictability coefficients reflect the same market structure regime in which retail participation is observed. Second, following Goyal et al. (2024), we require a minimum estimation period of 60 monthly observations to obtain a reliable estimate of the SII predictive ability coefficient, β_{SII} . Together, these two constraints result in January 2012 being the first month for which we can estimate β_{SII} .

We obtain the monthly SII and monthly market excess returns from the Goyal and Welch Data Library.⁵ Retail trading measures are constructed from the TAQ database, while retail short selling measures are constructed by combining FINRA Reg SHO short-sale data with TAQ data (Boehmer and Song, 2020). We obtain market-level control variables from several sources. Market liquidity is constructed using CRSP data. We obtain the remaining variables directly from public databases: Jurado et al.'s (2015) macroeconomic uncertainty indices from the authors' website;⁶ VIX from the Federal Reserve Bank of St. Louis's FRED database; and CFNAI from the Federal Reserve Bank of Chicago.

4. Results

4.1. Descriptive evidence

We begin by providing descriptive evidence that documents changes in SII , its return predictability, and retail trading activity over our January 2012 to December 2023 sample period. Figure 1, Panel A, plots the monthly time-series of the aggregate short interest variable, SII , (detrended and standardized to have unit standard deviation) which reveals a substantial decline over the sample period. Panel B plots the corresponding SII predictive ability coefficient, β_{SII} , estimated using a 12-months-ahead return forecasting horizon (holding period), which reveals a similarly pronounced decline. The annual summary statistics in Table 1 corroborate these patterns.

⁵ Available at: <https://sites.google.com/view/agoyal145>.

⁶ Available at: <https://www.sydneyludvigson.com/macro-and-financial-uncertainty-indexes>.

In specific, Panel A shows a general decline in SII , while Panel B shows that the annual mean of β_{SII} decreases from 1.205 in 2012 to 0.705 in 2023. Consistent with a significant decline, the Spearman Rank correlation between calendar year and the annual mean of β_{SII} is -0.958, indicating a strong negative trend over time. While descriptive, these findings provide initial evidence consistent with the hypothesis that the predictive ability of aggregate short interest has changed over time.

Turning to retail trading activity, Table 1, Panel C, reveals a substantial increase in retail trading activity over the sample period. For example, the annual mean of $Retail\%_Volume$ increases from 6.8% in 2012 to 16.1% in 2023, indicating that the fraction of aggregate trading volume attributable to retail investors more than doubles over our sample period. Alternative measures of retail trading activity based on retail dollar volume and trade count exhibit similar increases. Consistent with a significant increase, the Spearman Rank correlations between calendar year and the annual means of the three retail trading measures are positive and large in magnitude, ranging from 0.923 to 0.993. The increase in these measures is consistent with prior research documenting the growing importance of retail trading in U.S. equity markets over the past decade (Boehmer et al., 2021; Yin and Zou, 2026). In summary, the findings reported in Figure 1 and Table 1 document two pronounced trends over our sample period: a decline in the aggregate return predictability of SII and an increase in retail trading activity.

To provide some evidence on the relation between these two trends, Figure 2 plots β_{SII} along with $Retail\%_Volume$. The plot illustrates that the decline in the SII predictive ability coefficients coincides with the increase in retail trading activity. Table 2 provides complementary quantitative evidence based on correlations between the β_{SII} and the three retail trading measures. The Spearman Rank and Pearson Product Moment correlations between β_{SII} and $Retail\%_Volume$

are both negative and large in magnitude at -0.919, as are the negative correlations for the retail dollar volume and trade count measures. Altogether, these results provide preliminary evidence of a negative relation between the aggregate return predictability of *SII* and retail trading activity.

4.2. Tests of the intertemporal stability of *SII*'s predictive ability

To more clearly assess the determinants of the decline in β_{SII} shown in Figure 1, we perform regression tests of β_{SII} 's intertemporal stability. We enhance the reliability of our inferences by utilizing four complementary tests based on Andrews (1993), Bai and Perron (1998), Hansen (2000), and Elliott and Müller (2006). Prior research uses these methods to evaluate the hypothesis that the coefficient of interest is constant over time ($H_0: \beta_{SII,t} = \beta$ for all t in our setting) where each models an alternative form of coefficient instability (see, Rapach and Wohar, 2006; Paye and Timmermann, 2006). Table 3 reports the results.

First, Table 3 reports results based on the Elliott and Müller (2006) *J*-test, which tests for general parameter instability without specifying the timing or form of the change. This test rejects the null of coefficient stability ($J = -8.61$; p -value = 0.042). The second specification is based on the Andrews (1993) SupF test, which tests for a structural break at an unknown point in time. This test also rejects the null of coefficient stability (SupF = 25.79; p -value = 0.010). The third test is based on the Bai and Perron (1998) UDMax test, which tests for multiple structural breaks in the coefficient. This test continues to reject the null of coefficient stability (UDMax = 26.88; p -value = 0.010). The fourth and final test is based on the Hansen (2000) heteroskedasticity-robust SupF test, which tests for a structural break allowing for heteroskedasticity. Consistent with the other three tests, this test also rejects the null of coefficient stability (SupF = 158.50; p -value = 0.003). In summary, all four tests reject the null hypothesis at the 0.05 level or better. Moreover, all four lead to the same inference that the predictive ability of *SII* for future aggregate returns varies over

time, or alternatively stated, changes over our sample period. These findings complement and extend Rapach et al. (2016), who show a generally stable coefficient for a time period (1973 to 2014) predating our sample.

4.3. The role of retail trading on SII 's predictive ability for future aggregate returns

Table 4 reports the results from estimating Eq. (3) under the alternative specifications previously described. In Panel A, Col. (1) examines the relation between retail trading activity and β_{SII} for our baseline specification based on Newey and West (1987) heteroskedasticity- and autocorrelation-consistent standard errors. The coefficient on *Retail%_Volume* is negative and significant (coef. = -4.614 ; HAC t -stat. = -11.233 ; p -value < 0.001), indicating that greater retail trading activity is associated with weaker return predictability for SII . To help put this finding into perspective, Col. (2) examines the explanatory power only for the control variables: market liquidity, macroeconomic uncertainty, market volatility, and aggregate economic activity (i.e., absent the retail trading activity variable). The results reveal that these variables alone explain a substantial portion of the variation in β_{SII} (adjusted $R^2 = 0.642$). More importantly, however, when *Retail%_Volume* is included with these controls in Col. (3), its coefficient remains negative and significant (coef. = -4.236 ; HAC t -stat. = -6.624 ; p -value < 0.001), evidence that broader market and macroeconomic factors do not subsume the relation between retail trading activity and SII 's predictive ability.

Cols. (4) and (5) of Panel A report results based on the two other estimation procedures described above. Specifically, Col. (4) re-estimates Eq. (3) using WLS, which assigns greater weight to more precisely estimated values of the monthly SII coefficients. In Col. (4), the coefficient on *Retail%_Volume* remains negative and significant (coef. = -4.065 ; HAC t -stat. = -6.663 ; p -value < 0.001). Finally, Col. (5) reports the results based on the bootstrap procedure,

which addresses the precision embedded in the monthly β_{SII} coefficients. The resulting coefficient is -4.065 and significant ($z\text{-stat.} = -3.640$; $p\text{-value} < 0.001$). In summary, the negative relation between retail trading activity and SII 's predictive ability is not sensitive to the specification of the underlying regression model.

Panel B extends the analysis to capture SII 's predictive ability using $\beta_{SII|PC}$, which controls for the principal components in the first-stage return forecasting regression. By absorbing variation in returns attributable to common return predictors, this specification mitigates concerns that the estimated SII coefficient simply reflects correlated sources of return predictability. The results largely mirror those in Panel A, with Col. (1) reporting a coefficient on $Retail\%_Volume$ of -3.733 (HAC $t\text{-stat.} = -9.036$; $p\text{-value} < 0.001$). Even after including the market and macroeconomic controls in Col. (3), the coefficient remains negative and significant (coef. = -3.954 ; HAC $t\text{-stat.} = -5.107$; $p\text{-value} < 0.001$). Consistent results obtain as well when using the WLS specification (coef. = -3.588 ; HAC $t\text{-stat.} = -5.236$; $p\text{-value} < 0.001$) and the bootstrap procedure (coef. = -3.588 ; $z\text{-stat.} = -3.086$; $p\text{-value} = 0.002$) in Cols. (4) and (5), respectively.

Overall, the results in Table 4 provide consistent evidence that greater retail trading activity is associated with weaker return predictability for SII , and that relation does not change after controlling for market and macroeconomic conditions, accounting for other sources of return predictability through a principal component analysis, weighting observations by the precision of the estimated coefficients, and after addressing uncertainty in the first stage estimation by using a bootstrap procedure. In the next section we evaluate potential mechanisms to explain how the increasing participation of retail investors in the equity market contributes to the decline of SII 's predictive ability over our sample period.

5. Additional analyses

5.1. Mechanism 1: The rise of retail short selling

The first mechanism that we examine is whether the rise in retail trading changes the composition of short sellers and, as a consequence, the information embedded in aggregate short interest. This mechanism is motivated by prior research which attributes the predictive ability of aggregate short interest to sophisticated short sellers who possess information about future aggregate fundamentals and aggregate returns (Lynch et al., 2014; Rapach et al., 2016). That said, it is also reasonable to expect that the information embedded in short selling activity may change as the composition of short sellers changes. In particular, although retail short sellers may be informed, the predictive ability of their trading concentrates among small firms and stocks heavily traded by other retail investors (Kelley and Tetlock, 2017). Under this scenario, an increase in the share of short selling attributable to retail investors would shift the information embedded in aggregate short interest away from market-wide fundamentals and towards firm-specific information. This line of reasoning predicts a negative relation between retail short selling volume and the predictive ability of SII for future aggregate returns.

We start by providing visual evidence of this mechanism in Figure 3 which illustrates that retail short selling increases substantially over the sample period while β_{SII} declines over that same period. Supporting Figure 3's visual evidence, Table 2 reports Spearman and Pearson correlations between *Retail Short%_Volume* and β_{SII} of -0.913 and -0.909 , respectively. Using alternative measures of retail short selling activity based on dollar volume and trade count produces similarly large negative correlations as well.

Moving beyond univariate analysis, Table 5 reports the results from estimating a modified version of Eq. (3) using *Retail Short%_Volume*, the fraction of aggregate monthly trading volume

attributable to retail short selling, as the explanatory variable. We scale retail short selling volume by total market trading volume rather than by total short selling because total short selling volume is not observable. *Retail Short%_Volume* therefore measures the intensity of retail short selling activity in the market and, by construction, reflects a portion of the broader growth in retail investor participation over time documented in Section 4.⁷

Panel A of Table 5 examines the relation between retail short selling and β_{SII} . Col. (1)'s results reveal a significant negative coefficient on *Retail Short%_Volume* (coef. = -25.742 ; HAC t -stat. = -10.767 ; p -value < 0.001), indicating that *SII* is less predictive of aggregate returns when retail short selling accounts for a greater share of aggregate trading volume. To contextualize Col. (1)'s results, Col. (2) examines the explanatory power of only the control variables (market liquidity, macroeconomic uncertainty, market volatility, and aggregate economic activity) in the absence of the retail short selling variable. Col. (2)'s results indicate that these variables alone explain a substantial portion of the variation in β_{SII} (adjusted $R^2 = 0.642$). Turning to Col. (3), which includes *Retail Short%_Volume* alongside the controls, its coefficient remains negative and significant (coef. = -22.161 ; HAC t -stat. = -6.695 ; p -value < 0.001), evidence that the control variables that capture broad market and macroeconomic conditions do not subsume the relation between retail short selling and *SII*'s predictive ability. Cols. (4) and (5) report results for the alternative WLS and Bootstrap estimation procedures. Both specifications continue to yield negative and significant coefficients on *Retail Short%_Volume*. The collective results indicate that our inference about the negative relation is consistent across alternative model specifications.

Panel B extends the analysis in Panel A to using $\beta_{SII|PC}$ as the measure of *SII*'s predictive power. The results are entirely consistent with those in Panel A. For example, in Col. (1), the

⁷ We revisit this point in Section 5.3 where we estimate the relation between *Retail Short%_Volume* and β_{SII} conditional on *Retail%_Volume* to further isolate the incremental role of retail short-selling.

coefficient on *Retail Short%_Volume* is -20.983 (HAC t -stat. = -9.135 ; p -value < 0.001). Moreover, even after including the market and macroeconomic controls in Col. (3), the coefficient remains negative and significant (coef. = -20.816 ; HAC t -stat. = -4.985 ; p -value < 0.001). This finding also holds under the WLS specification in Col. (4) (coef. = -18.892 ; HAC t -stat. = -5.073 ; p -value < 0.001) and under the bootstrap procedure in Col. (5) (z -stat. = -2.973 ; p -value = 0.003). Overall, the results in Table 5 continue to provide consistent evidence that greater retail participation in short selling is associated with a weaker return predictability of *SII*. Thereby supporting Mechanism 1, namely, that as retail short selling accounts for a greater share of aggregate trading volume, the information embedded in aggregate short interest becomes less informative about future aggregate returns, which in turn, contributes to the decline in the return-predictive ability of *SII*.

5.2. Mechanism 2: Increased short squeeze risk

The second mechanism that we examine is whether the rise in retail trading increases short squeeze risk and, as a consequence, discourages informed short selling. The growth of retail trading, particularly through zero-commission trading platforms and social media, increases retail investors' ability to trade collectively, potentially increasing the risks faced by short sellers (Allen, Haas, Nowak, Pirovano, and Tengulov, 2025a). Consistent with this, Stice-Lawrence et al. (2025) document that positive online media sentiment can trigger short squeezes. Short squeeze risk can prove particularly costly to informed short sellers because it exposes them to the possibility of being forced to close their positions before their information is incorporated into market prices. Thus, greater short squeeze risk can reduce the expected profitability of informed short selling and disincentivize sophisticated short sellers from trading on their information. As a result, greater

short squeeze risk may reduce the amount of market-wide information reflected in SII , and thereby weaken its predictive power for aggregate returns.

We start by providing visual evidence of this mechanism in Figure 4. The figure plots *Short Squeeze* over our sample period and illustrates that short squeezes tend to become more prevalent coincident with the decline in β_{SII} . Supporting Figure 4's visual evidence, Table 2 reports Spearman and Pearson correlations between *Short Squeeze* and β_{SII} that are negative and large in magnitude: -0.595 and -0.625 , respectively.

Panel A of Table 6 formally examines this mechanism by using *Short Squeeze* as an explanatory variable where β_{SII} is the dependent variable. Col. (1) reveals that the coefficient on *Short Squeeze* is negative and significant (coef. = -7.460 ; HAC t -stat. = -4.906 ; p -value < 0.001), evidence that SII is less predictive of aggregate returns during periods of greater short squeeze activity. Cols. (2) and (3) show that while the control variables (market liquidity, macroeconomic uncertainty, market volatility, and aggregate economic activity) explain a substantial portion of the variation in β_{SII} in the absence of *Short Squeeze* (adjusted $R^2 = 0.642$), the *Short Squeeze* coefficient remains negative and significant (coef. = -4.973 ; HAC t -stat. = -4.724 ; p -value < 0.001) even after including the control variables alongside it. This finding continues to hold under the WLS specification in Col. (4) and the bootstrap procedure in Col. (5), where the coefficient on *Short Squeeze* remains negative and significant. Thus, our inference about the negative relation is robust across alternative model specifications.

Panel B continues the analysis of Panel A using $\beta_{SII/PC}$. As before, the results remain consistent with those in Panel A. In particular, in Col. (1), the coefficient on *Short Squeeze* is -6.017 (HAC t -stat. = -5.236 ; p -value < 0.001). More importantly, even after including the market and macroeconomic control variables in Col. (3), the coefficient remains negative and significant

(coef. = -4.206 ; HAC t -stat. = -3.897 ; p -value < 0.001). As with all prior results reported above, the WLS specification in Col. (4) (coef. = -4.130 ; HAC t -stat. = -3.393 ; p -value = 0.001) and the bootstrap procedure in Col. (5) (z -stat. = -2.162 ; p -value = 0.031) both yield consistent inferences.

Altogether, the results in Table 6 provide consistent evidence that greater short squeeze activity is associated with a weaker return predictability for *SII*. These findings are consistent with a proposed short squeeze mechanism, in that greater short squeeze risk appears to discourage informed short sellers from taking or maintaining short positions. In turn, their diminished short selling participation reduces the market-wide information incorporated into aggregate short interest and, as a consequence, weakens *SII*'s predictive ability for future aggregate returns.

5.3. Path analysis

We use path analysis to provide evidence on the relative importance of the two mechanisms through which retail trading may associate with the decline in *SII*'s predictive power. While Tables 5 and 6 provide evidence separately consistent with each of the two mechanisms (changes in the composition of short sellers and increased short squeeze risk), they do not indicate how the mechanisms interact or the proportion of the relation between retail trading and *SII*'s predictive ability that operates through each channel. This motivates us to use path analysis to decompose the total effect of *Retail%_Volume* into a direct effect and an indirect effect through each channel, allowing us to quantify and compare the relative importance of the two channels.

Table 7 reports the results of the path analysis where the relation between *Retail%_Volume* and *Retail Short%_Volume* (or *Short Squeeze*) is denoted as path a while the relation between *Retail Short%_Volume* (or *Short Squeeze*) and *SII*'s predictive ability, conditional on *Retail%_Volume*, is denoted as path b . The product $a \times b$ captures the indirect effect of

Retail%_Volume on *SII*'s predictive ability operating through the respective channel, while c and c' represent the total and direct effects, respectively.

Turning to the results in Col. (1), which examines the retail short selling channel, the estimated path a is positive and significant ($a = 0.171$; $t\text{-stat.} = 8.72$; $p\text{-value} < 0.001$). This indicates that a 10-percentage-point increase in *Retail%_Volume* is associated with a 1.71-percentage-point increase in *Retail Short%_Volume*. Path b is negative and significant ($b = -12.096$; $t\text{-stat.} = -2.83$; $p\text{-value} = 0.005$), which indicates that, conditional on overall retail trading, a one-percentage-point increase in *Retail Short%_Volume* is associated with a 0.121 decrease in β_{SII} . Combining the two paths yields a negative and significant indirect effect ($a \times b = -2.064$; $t\text{-stat.} = -2.69$; $p\text{-value} = 0.007$), which accounts for 48.7 percent of the total relation between retail trading and *SII*'s predictive ability. More precisely, nearly half of the relation between retail trading and *SII*'s predictive ability derives from the retail short-selling channel.

Turning to Col. (2), which examines the short squeeze channel, the positive path a ($a = 0.397$; $t\text{-stat.} = 4.59$; $p\text{-value} < 0.001$) indicates that greater retail trading is associated with greater short squeeze activity, while the negative path b ($b = -2.187$; $t\text{-stat.} = -2.96$; $p\text{-value} = 0.003$) indicates that greater short squeeze activity is associated with weaker *SII* predictive power, conditional on overall retail trading. The resulting indirect effect is negative and significant ($a \times b = -0.868$; $t\text{-stat.} = -2.49$; $p\text{-value} = 0.013$) and accounts for 20.5 percent of the total relation. Cols. (3) and (4) yield similar results using $\beta_{SII|PC}$ where the indirect effect through *Retail Short%_Volume* is -2.031 ($t\text{-stat.} = -2.18$; $p\text{-value} = 0.029$) and accounts for 51.4 percent of the total relation, whereas the indirect effect through *Short Squeeze* is -0.553 ($t\text{-stat.} = -1.67$; $p\text{-value} = 0.096$), accounting for only 14.0 percent of the total relation.

In summary, the path analysis provides evidence consistent with both channels, although the retail short selling channel accounts for approximately one half of the relation between retail trading and *SII* predictability compared to only between 14 and 21 percent for the short squeeze channel. As such, for the two channels we examine, the evidence indicates that changes in the composition of short sellers act as the quantitatively stronger mechanism.

5.4. Robustness tests

We conduct a series of tests to evaluate the robustness of our results and inferences. These tests can be categorized into three groups: 1) alternative return forecasting horizons (holding periods); 2) alternative measures of retail trading and retail short selling activity; and 3) whether the results are driven by the specification of our two-stage estimation procedure. For brevity, Table 8 reports only the coefficient of interest and its associated test-statistic for each alternative robustness specification.

In the first set of robustness tests, we examine whether our results depend on the 12-months-ahead return forecasting horizon (return holding period) we use in the primary analyses. To do so, we re-estimate the models in Table 4, Panel A using one-, three-, and six-months-ahead return horizons to construct β_{SII} . Across all three alternative return horizons, the coefficient on *Retail%_Volume* remains negative and significant. Specifically, as reported in Table 8, coefficients of -2.844 ($t\text{-stat.} = -7.675$; $p\text{-value} < 0.001$), -5.557 ($t\text{-stat.} = -5.734$; $p\text{-value} < 0.001$), and -7.167 ($t\text{-stat.} = -10.840$; $p\text{-value} < 0.001$) for the one-, three-, and six-months-ahead return horizons, respectively. This evidence indicates that the negative relation between retail trading activity and *SII*'s return predictability is robust to alternative return horizons.

In the second set of robustness tests, we examine whether our findings are sensitive to the measurement of retail trading activity. To do so, we replace *Retail%_Volume* with alternative

measures based on retail dollar volume and trade count. *Retail%_Dollar Volume* captures retail participation based on the dollar value of trades and therefore places greater weight on larger trades. In comparison, *Retail%_Trade Count* captures the frequency of retail transactions and is less influenced by trade size. The coefficients on both alternative measures remain negative and significant (*Retail%_Dollar Volume*: coef. = -5.432 , t -stat. = -3.849 , p -value < 0.001 ; *Retail%_Trade Count*: coef. = -7.315 , t -stat. = -3.729 , p -value = 0.011). We conduct analogous tests for the retail short selling channel examined in Table 5 by replacing *Retail Short%_Volume* with measures based on retail short selling dollar volume and trade count. The coefficients continue to remain negative and significant (*Retail Short%_Dollar Volume*: coef. = -29.600 , t -stat. = -4.789 , p -value < 0.001 ; *Retail Short%_Trade Count*: coef. = -54.307 , t -stat. = -4.404 , p -value = 0.004). Thus, our main results and inferences are not sensitive to how retail trading and retail short selling are measured.

In our final set of robustness tests, we examine whether the relation documented in Table 4 is driven by the two-stage procedure used to estimate β_{SII} . To do so, rather than first estimating β_{SII} and then relating that estimated coefficient to retail trading activity, we instead estimate a single-stage return regression that interacts *SII* directly with *Retail%_Volume*. The coefficient on $SII \times Retail\%_Volume$ is negative and significant (coef. = -13.548 ; t -stat. = -2.970 ; p -value = 0.003), evidence that the predictive relation between *SII* and future market returns weakens as retail trading activity increases. This result corroborates Table 4's findings without relying on an estimated *SII* predictive ability coefficient as the dependent variable in our two-stage model.

Overall, the results in Table 8 indicate that our main findings and inferences are robust to alternative return forecasting horizons, alternative measures of retail trading and retail short selling activity, and a single-stage model specification. As such, these tests serve to mitigate concerns that

the change in *SII*'s predictive ability and its association with rising retail participation is attributable to any of our research design choices.

6. Conclusion

Motivated by prior research documenting aggregate short interest as one of the strongest predictors of aggregate stock returns (Rapach et al., 2016; Goyal et al., 2024), we examine whether that predictive ability has changed over time, particularly over the past decade. We find that both aggregate short interest and its predictive ability for future aggregate returns have significantly declined from 2012 to 2023. Given that reduction in predictive ability, we investigated whether changes in the composition of investors in the financial markets, specifically the growing participation of retail investors over time, explains that reduction, at least in part. Our results provide consistent evidence that retail trading activity exhibits a significant negative relation with the predictive ability of aggregate short interest for market returns. Additional results reveal that retail short selling and the incidence of short squeezes constitute two mechanisms for the negative relation we document.

Our study makes the following contributions. First, the effects we document can help inform regulators about the changing landscape of retail trading and how it impacts the informativeness of short-selling activity. Such effects are especially relevant given the dramatic rise in retail trading following the COVID-19 pandemic and the growth of commission-free trading platforms such as Robinhood (see, e.g., Ozik, Sadka, and Shen, 2021; Barber et al., 2022; Welch, 2022). Since an increase in the presence and activity of new retail investors can introduce added costs and risks to short selling by informed investors, the shifting investor composition has the potential to alter the cost-benefit tradeoffs of imposing further regulatory requirements on high-volume short sellers, such as those affected by SEC Rule 13f-2. Second, our study extends and

complements prior research (Rapach et al., 2016), documenting the predictive ability of aggregate short interest, while at the same time documenting that increased retail trading diminishes that predictive ability. The evidence we report suggests that *who* takes part in short selling, not necessarily the volume of short selling activity, plays a prominent role in explaining the decline in the predictive ability of aggregate short interest for aggregate returns. In this spirit, one avenue for future research would be to explore whether the introduction of new investors in the post-pandemic period warrants separate classification from traditional retail trading definitions and, in turn, how that classification bears on the continued evolution of short selling in general, and on its predictive ability for returns at the aggregate and firm levels.

References

- Allen, F., Haas, M., Nowak, E., Pirovano, M., & Tengulov, A. (2025a). Squeezing shorts through social media platforms. *Management Science*, 1-30.
- Allen, F., Haas, M., Pirovano, M., & Tengulov, A. (2025b). How prevalent are short squeezes? Evidence from the US and Europe. *Journal of Banking and Finance*, 176, 1-15.
- Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5(1), 31-56.
- Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4), 821-856.
- Atmaz, A., Basak, S., & Ruan, F. (2024). Dynamic equilibrium with costly short selling and lending market. *The Review of Financial Studies*, 37(2), 444-506.
- Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1), 47-78.
- Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593-1636.
- Barber, B. M., Huang, X., Odean, T., & Schwarz, C. (2022). Attention-induced trading and returns: Evidence from Robinhood users. *The Journal of Finance*, 77(6), 3141-3190.
- Boehmer, E., Jones, C. M., & Zhang, X. (2008). Which shorts are informed? *The Journal of Finance*, 63(2), 491-527.
- Boehmer, E., Jones, C. M., Zhang, X., & Zhang, X. (2021). Tracking retail investor activity. *The Journal of Finance*, 76(5), 2249-2305.
- Boehmer, E., & Song, W. (2020). Smart retail traders, short sellers, and stock returns [Working paper]. SSRN.
- Boehmer, E., & Wu, J. (2013). Short selling and the price discovery process. *The Review of Financial Studies*, 26(2), 287-322.
- Chen, W., Hribar, P., & Melessa, S. (2018). Incorrect inferences when using residuals as dependent variables. *Journal of Accounting Research*, 56(3), 751-796.
- Christophe, S. E., Ferri, M. G., & Angel, J. J. (2004). Short selling prior to earnings announcements. *The Journal of Finance*, 59(4), 1845-1875.
- Christophe, S. E., Ferri, M. G., & Hsieh, J. (2010). Informed trading before analyst downgrades: Evidence from short sellers. *Journal of Financial Economics*, 95(1), 85-106.

- Diether, K. B., Lee, K.-H., & Werner, I. M. (2009). Short-sale strategies and return predictability. *The Review of Financial Studies*, 22(2), 575-607.
- Eaton, G. W., Green, T. C., Roseman, B. S., & Wu, Y. (2022). Retail trader sophistication and stock market quality: Evidence from brokerage outages. *Journal of Financial Economics*, 146(2), 502-528.
- Elliott, G., & Müller, U. K. (2006). Efficient tests for general persistent time variation in regression coefficients. *The Review of Economic Studies*, 73(4), 907-940.
- Engelberg, J. E., Reed, A. V., & Ringgenberg, M. C. (2012). How are shorts informed? Short sellers, news, and information processing. *Journal of Financial Economics*, 105(2), 260-278.
- Farrell, M., Green, T. C., Jame, R., & Markov, S. (2022). The democratization of investment research and the informativeness of retail investor trading. *Journal of Financial Economics*, 145(2), 616-641.
- Goyal, A., Welch, I., & Zafirov, A. (2024). A comprehensive 2022 look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 37(11), 3490-3557.
- Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68(3), 575-603.
- Hvide, H. K., Meling, T. G., Mogstad, M., & Vestad, O. L. (2024). Broadband internet and the stock market investments of individual investors. *The Journal of Finance*, 79(3), 2163-2194.
- Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. *American Economic Review*, 105(3), 1177-1216.
- Kelley, E. K., & Tetlock, P. C. (2013). How wise are crowds? Insights from retail orders and stock returns. *The Journal of Finance*, 68(3), 1229-1265.
- Kelley, E. K., & Tetlock, P. C. (2017). Retail short selling and stock prices. *The Review of Financial Studies*, 30(3), 801-834.
- Kumar, A., & Lee, C. M. C. (2006). Retail investor sentiment and return comovements. *The Journal of Finance*, 61(5), 2451-2486.
- Lamont, O. A., & Stein, J. C. (2004). Aggregate short interest and market valuations. *American Economic Review*, 94(2), 29-32.
- Ljungqvist, A., & Qian, W. (2016). How constraining are limits to arbitrage? *The Review of Financial Studies*, 29(8), 1975-2028.

- Lynch, A., Nikolic, B., Yan, X. S., & Yu, H. (2014). Aggregate short selling, commonality, and stock market returns. *Journal of Financial Markets*, 17, 199-229.
- Michels, J. (2024). Retail investor trade and the pricing of earnings. *Review of Accounting Studies*, 29, 1-36.
- Murphy, K. M., & Topel, R. H. (1985). Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics*, 3(4), 370-379.
- Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703-708.
- Ozik, G., Sadka, R., & Shen, S. (2021). Flattening the illiquidity curve: Retail trading during the COVID-19 lockdown. *Journal of Financial and Quantitative Analysis*, 56(7), 2356-2388.
- Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, 25(1), 221-247.
- Paye, B. S., & Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, 13(3), 274-315.
- Rapach, D. E., Ringgenberg, M. C., & Zhou, G. (2016). Short interest and aggregate stock returns. *Journal of Financial Economics*, 121(1), 46-65.
- Rapach, D. E., & Wohar, M. E. (2006). Structural breaks and predictive regression models of aggregate US stock returns. *Journal of Financial Econometrics*, 4(2), 238-274.
- Saxonhouse, G. R. (1976). Estimated parameters as dependent variables. *The American Economic Review*, 66(1), 178-183.
- Schultz, P. (2024). Short squeezes and their consequences. *Journal of Financial and Quantitative Analysis*, 59(1), 68-96.
- Stice-Lawrence, L., Wong, Y. T. F., & Zhao, W. (2025). Short squeezes after short selling attacks. *Journal of Accounting Research*, 63(3), 1187-1236.
- Welch, I. (2022). The wisdom of the Robinhood crowd. *The Journal of Finance*, 77(3), 1489-1527.
- Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), 1455-1508.
- Yin, X., & Zou, D. (2026). Buying the dip in retail trading [Working paper]. SSRN 6796758.

Appendix A Detailed Variable Definitions

Variable Name	Detailed Variable Definition	Data Source
Stage 1 Regression: $Return_{t,h} = \alpha_t + \beta_{SII,t,h}SII_t + \varepsilon_{t,h}$ for $t = 1, \dots, T - h$ $Return_{t,h} = \alpha_t + \beta_{SII PC,t,h}SII_t + \sum_{j=1}^3 \gamma_{j,t,h}PC_{j,t} + \varepsilon_{t,h}$ for $t = 1, \dots, T - h$		
$Return_{t, h=1/3/6/12}$	Monthly log excess return on the S&P 500 averaged over the next $h = 1, 3, 6$ and 12 months, respectively.	Goyal and Welch Data Library
SII	Aggregate short interest index. Detrended and standardized to have unit standard deviation (Rapach, Ringgenberg, and Zhou, 2016).	Goyal and Welch Data Library
$PC_{j,t}$ ($j = 1, 2, 3$)	First three principal components extracted from 14 aggregate return predictors (Welch and Goyal, 2008): the dividend-price ratio (DP), dividend yield (DY), earnings-price ratio (EP), dividend-payout ratio (DE), stock return volatility (RVOL), book-to-market ratio (BM), net equity expansion (NTIS), Treasury bill rate (TBL), long-term government bond yield (LTY), long-term government bond return (LTR), term spread (TMS), default yield spread (DFY), default return spread (DFR), and inflation (INFL). The predictors are standardized prior to extraction. The three principal components are included in Eq. (2) to control for established return predictors.	Goyal and Welch Data Library
Stage 2 Regression: $\beta_{SII,t,h}(\beta_{SII PC,t,h}) = \eta_h + \theta_h Retail\%_Volume_t + \gamma'Controls_t + \mu_{t,h}$		

Dependent Variables:

$\beta_{SII, h=1,3,6,12}$	Monthly aggregate short interest (SII) return-predictability coefficient estimated recursively from regressions of h -month-ahead market excess returns on SII where the number of observations (months) in the estimation period increases by one as each new month is added over the sample period.
$\beta_{SII PC, h=1,3,6,12}$	Monthly aggregate short interest (SII) return-predictability coefficient estimated recursively from regressions of h -month-ahead market excess returns on SII after controlling for the first three principal components of the Welch and Goyal (2008) return predictors as specified in Eq. (2) and where the number of observations (months) in the estimation period increases by one as each new month is added over the sample period. $\beta_{SII PC}$ measures the incremental predictive ability of SII after conditioning on the information contained in established return predictors.

Independent Variables:

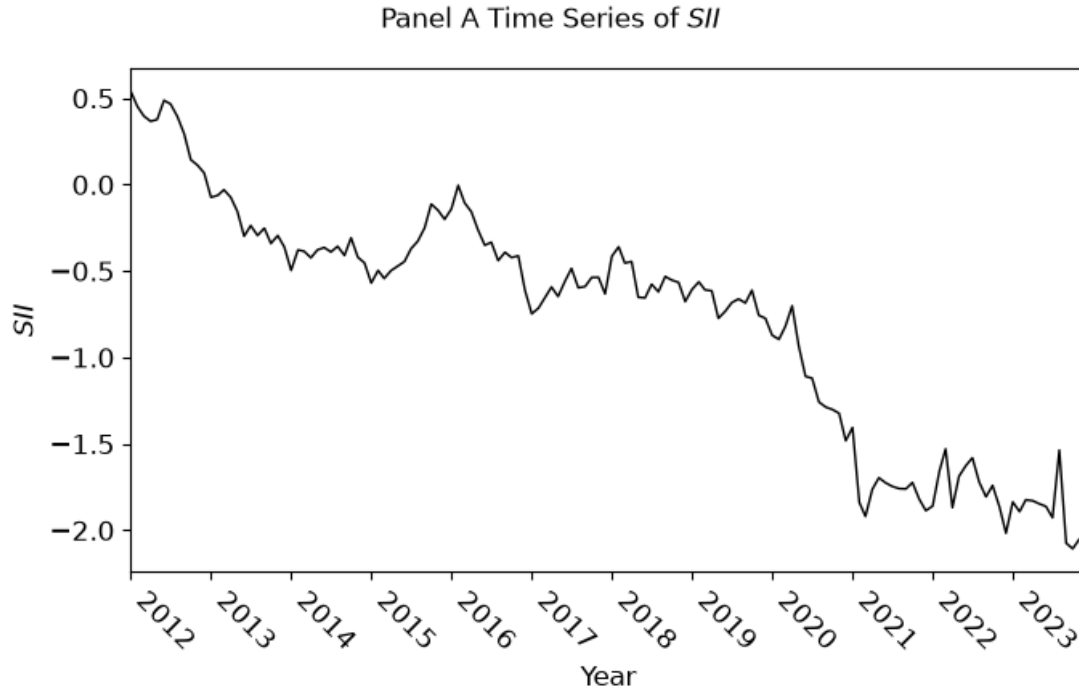
<i>Retail%_Volume</i>	Fraction of aggregate monthly trading volume attributable to retail investors computed as monthly retail trading volume divided by total market trading volume, aggregated across all stocks and trading days within the month. Retail trades are identified using the sub-penny method of Boehmer, Jones, Zhang, and Zhang (2021).	TAQ (WRDS Intraday Indicators)
<i>Retail%_Dollar Volume</i>	Fraction of aggregate monthly dollar trading volume attributable to retail investors computed as monthly retail dollar trading volume divided by total market dollar trading volume, aggregated across all stocks and trading days within the month. Retail trades are identified using the sub-penny method of Boehmer, Jones, Zhang, and Zhang (2021).	TAQ (WRDS Intraday Indicators)
<i>Retail%_Trade Count</i>	Fraction of aggregate monthly trade count attributable to retail investors computed as the monthly number of retail trades divided by the total number of market trades, aggregated across all stocks and trading days within the month. Retail trades are identified using the sub-penny method of Boehmer, Jones, Zhang, and Zhang (2021).	TAQ (WRDS Intraday Indicators)
<i>Retail Short%_Volume</i>	Fraction of aggregate monthly trading volume attributable to retail short selling computed as monthly retail short selling volume divided by total market trading volume, aggregated across all stocks and trading days within the month. Retail short selling is identified following Boehmer and Song (2020).	FINRA; TAQ
<i>Retail Short%_Dollar Volume</i>	Fraction of aggregate monthly dollar trading volume attributable to retail short selling computed as monthly retail short selling dollar volume divided by total market dollar trading volume, aggregated across all stocks and trading days within the month. Retail short selling is identified following Boehmer and Song (2020).	FINRA; TAQ
<i>Retail Short%_Trade Count</i>	Fraction of aggregate monthly trade count attributable to retail short selling computed as the monthly number of retail short selling trades divided by the total number of market trades, aggregated across all stocks and trading days within the month. Retail short selling is identified following Boehmer and Song (2020).	FINRA; TAQ
<i>Short Squeeze</i>	Fraction of CRSP-listed common stocks experiencing at least one trading-day securities lending squeeze during the month. A trading-day securities lending squeeze is identified following Schultz (2024) which is when a stock's lendable share supply is less than the previous trading day's shares on loan.	Markit Securities Finance; CRSP

Control Variables:

<i>Market Liquidity</i>	Natural logarithm of the aggregate monthly Amihud (2002) illiquidity measure computed as the sum of firm level Amihud illiquidity measures across all CRSP-listed common stocks.	CRSP
<i>Macro Uncertainty</i>	Monthly macroeconomic uncertainty index from Jurado, Ludvigson, and Ng (2015).	Jurado-Ludvigson-Ng Data Library
<i>VIX</i>	Monthly average of the CBOE Volatility Index.	FRED
<i>CFNAI</i>	Monthly Chicago Fed National Activity Index.	Federal Reserve Bank of Chicago

Figure 1
Time Series Behavior of Aggregate Short Interest Index (SII) and the Coefficient (β_{SII}) for its Predictive Ability for Future Aggregate Returns

The figure plots the time series of the aggregate short interest index (SII) and the corresponding return predictability coefficient (β_{SII}) that measures its predictive ability for future aggregate returns. The sample period is January 2012 through December 2023. Panel A plots the monthly SII . Panel B plots the monthly estimates of β_{SII} which are estimated recursively using a 12-month return holding period for a forecasting regression of market excess returns on SII where the length of the estimation period increases by one month as each new month is added to the regression over the January 2012 to December 2023 sample period. Variables are defined in Appendix A.



Panel B Time Series of β_{SII}

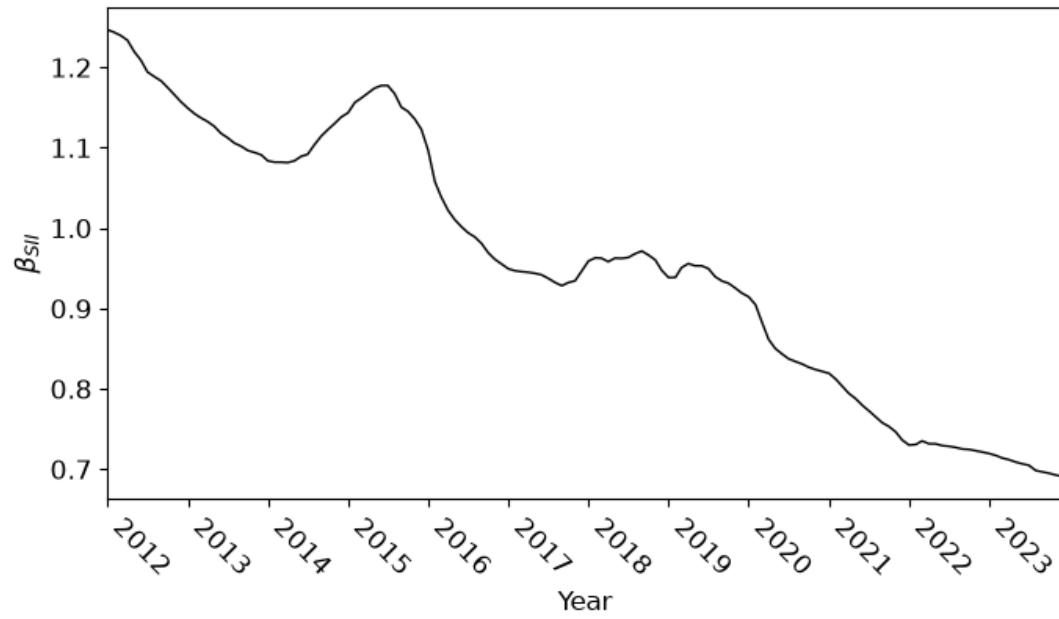


Figure 2
Time Series Behavior of Aggregate Short Interest Index (SII) Predictive Ability Coefficient (β_{SII}) and Retail Trading Activity

The figure plots the monthly aggregate short interest index (SII) predictive ability coefficient, β_{SII} (solid line, left axis) along with retail trading volume (dashed line, right axis). The sample period is January 2012 through December 2023. Month-by-month estimates of β_{SII} are estimated recursively using a 12-month return holding period for a forecasting regression of market excess returns on SII where the length of the estimation period increases by one month as each new month is added to the regression over the January 2012 to December 2023 sample period. Retail trading volume share ($Retail\%_Volume$) is measured as the proportion of total market trading volume attributable to retail investors. Variables are defined in Appendix A.



Figure 3
Time Series Behavior of Aggregate Short Interest Index (SII) Predictive Ability Coefficient (β_{SII}) and Retail Short Selling Activity

The figure plots the monthly aggregate short interest index (SII) predictive ability coefficient (β_{SII} , solid line, left axis) along with retail short selling volume (dashed line, right axis). The sample period is January 2012 through December 2023. Month-by-month estimates of β_{SII} are estimated recursively using a 12-month return holding period for a forecasting regression of market excess returns on SII where the length of the estimation period increases by one month as each new month is added to the regression over the January 2012 to December 2023 sample period. Retail short selling volume share (*Retail Short%_Volume*) is measured as the fraction of aggregate monthly trading volume attributable to retail short selling. Variables are defined in Appendix A.

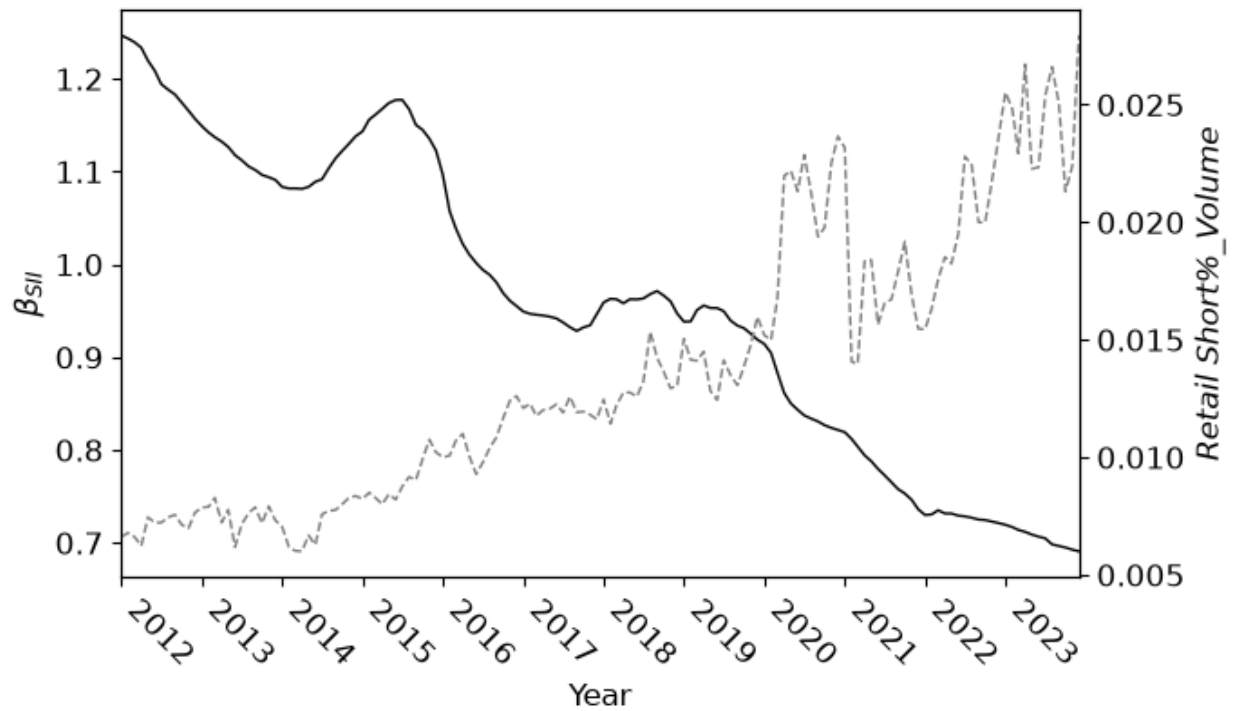


Figure 4
Time Series Behavior of Aggregate Short Interest Index (*SII*) Predictability Coefficient (β_{SII}) and Short Squeezes

The figure plots the monthly aggregate short interest index (*SII*) predictive ability coefficient (β_{SII} , solid line, left axis) along with the monthly short squeeze measure (dashed line, right axis). The sample period is January 2012 through December 2023. Month-by-month estimates of β_{SII} are estimated recursively using a 12-month return holding period for a forecasting regression of market excess returns on *SII* where the length of the estimation period increases by one month as each new month is added to the regression over the January 2012 to December 2023 sample period. The monthly short squeeze measure is defined as the fraction of CRSP-listed common stocks experiencing at least one trading-day securities lending squeeze in a given month. Variables are defined in Appendix A.

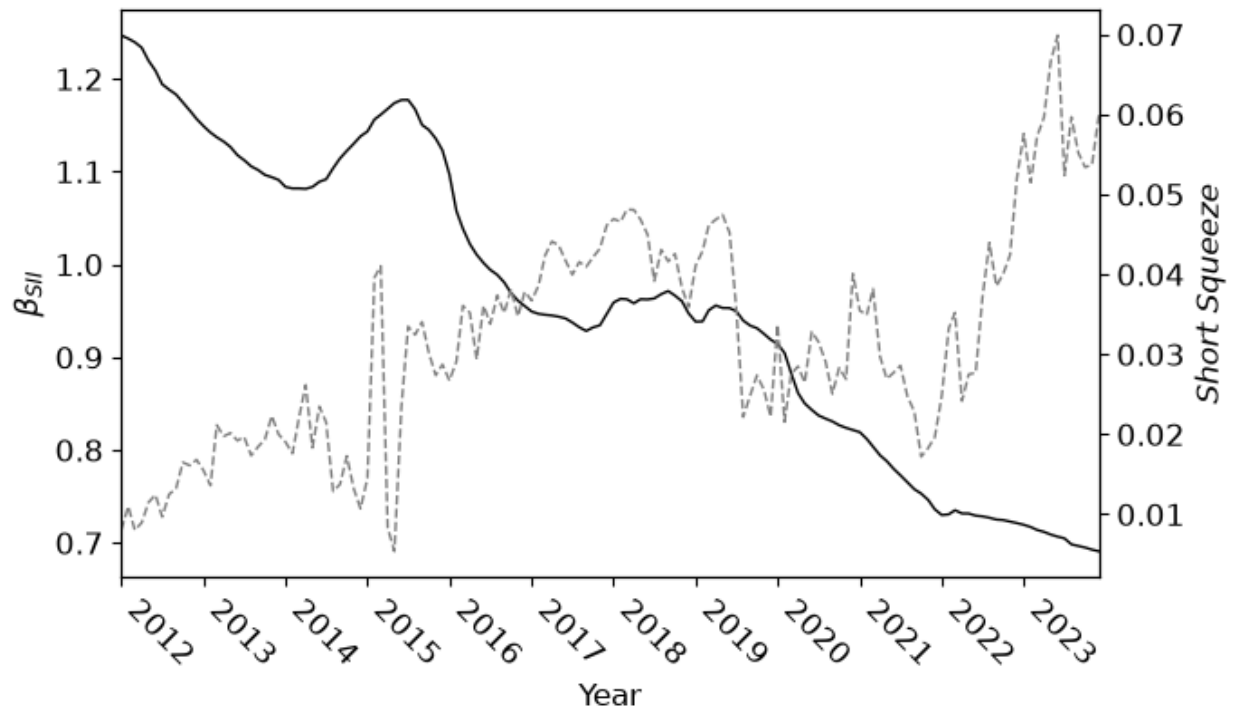


Table 1
Summary Descriptive Statistics

Annual summary statistics for aggregate short interest index (*SII*); the monthly *SII* return predictive ability coefficient (β_{SII}) and retail trading activity over the January 2012 – December 2023 sample period. Panel A reports the annual mean, standard deviation, lower quartile, median and upper quartiles for the monthly *SII* values. Panel B reports corresponding annual summary statistics for monthly estimates of β_{SII} , which is estimated recursively using a 12-month holding period return forecasting regression of market excess returns on *SII*. The bottom row of Panel B reports the Spearman Rank correlation between the calendar year and the annual mean of β_{SII} . Panel C reports the annual mean of three retail trading measures: retail volume share (*Retail% Volume*), retail dollar-volume share (*Retail% Dollar Volume*), and retail trade-count share (*Retail% Trade Count*). Numbers in parentheses in Panel C are standard deviations. The bottom row of Panel C reports the Spearman Rank correlation between the calendar year and the annual mean for each retail trading measure. Variables are defined in Appendix A.

Panel A: Annual Summary Statistics for Aggregate Short Interest Index (<i>SII</i>)					
Year	Mean	Std. Dev.	Lower Quartile	Median	Upper Quartile
2012	0.341	0.156	0.256	0.385	0.455
2013	-0.206	0.120	-0.296	-0.244	-0.074
2014	-0.397	0.048	-0.421	-0.388	-0.374
2015	-0.370	0.159	-0.497	-0.408	-0.238
2016	-0.302	0.173	-0.415	-0.342	-0.153
2017	-0.608	0.076	-0.650	-0.595	-0.554
2018	-0.542	0.103	-0.628	-0.559	-0.452
2019	-0.672	0.073	-0.739	-0.672	-0.610
2020	-1.092	0.244	-1.291	-1.115	-0.889
2021	-1.754	0.129	-1.825	-1.759	-1.725
2022	-1.746	0.141	-1.860	-1.731	-1.649
2023	-1.906	0.161	-2.056	-1.877	-1.834

Panel B: Annual Statistics of the <i>SII</i> Predictive Ability Coefficient (β_{SII})					
Year	Mean	Std. Dev.	Lower Quartile	Median	Upper Quartile
2012	1.205	0.032	1.181	1.202	1.235
2013	1.117	0.020	1.100	1.115	1.134
2014	1.100	0.021	1.083	1.090	1.116
2015	1.157	0.017	1.145	1.159	1.169
2016	1.006	0.042	0.978	0.998	1.026
2017	0.940	0.007	0.934	0.943	0.946
2018	0.962	0.006	0.960	0.963	0.964
2019	0.941	0.012	0.933	0.939	0.951
2020	0.853	0.032	0.830	0.840	0.867
2021	0.777	0.027	0.757	0.776	0.797
2022	0.728	0.004	0.725	0.729	0.731
2023	0.705	0.010	0.696	0.706	0.712
Spearman Rank correlation (<i>Year</i> , β_{SII})		-0.958			

Table 1 Cont'd

Panel C: Proportion of Retail Trading Activity (Retail%)			
Year	<i>Retail%_Volume</i>	<i>Retail%_Dollar Volume</i>	<i>Retail%_Trade Count</i>
2012	6.80%	5.30%	3.50%
	(0.005)	(0.004)	(0.002)
2013	7.20%	5.00%	3.40%
	(0.006)	(0.006)	(0.004)
2014	7.60%	5.90%	3.50%
	(0.008)	(0.014)	(0.005)
2015	8.10%	7.30%	4.20%
	(0.004)	(0.003)	(0.002)
2016	9.20%	7.90%	4.70%
	(0.007)	(0.005)	(0.004)
2017	9.60%	8.20%	5.00%
	(0.002)	(0.003)	(0.003)
2018	9.70%	8.60%	5.40%
	(0.005)	(0.004)	(0.004)
2019	9.90%	8.70%	5.40%
	(0.007)	(0.005)	(0.005)
2020	13.70%	10.80%	7.90%
	(0.019)	(0.009)	(0.010)
2021	14.60%	9.60%	8.20%
	(0.014)	(0.006)	(0.013)
2022	14.10%	9.90%	6.60%
	(0.014)	(0.008)	(0.007)
2023	16.10%	11.00%	7.40%
	(0.010)	(0.004)	(0.003)
Spearman Rank correlation (Year, Retail%)	0.993	0.972	0.923

Table 2
Spearman Rank and Pearson Product Moment Correlations between Monthly SII Predictive Ability Coefficients (β_{SII}) and the Explanatory Variables Used in the Empirical Tests

The table reports Spearman Rank and Pearson Product Moment correlations between monthly SII predictive ability coefficients (β_{SII}) and the explanatory variables used in the empirical tests. Month-by-month estimates of β_{SII} are estimated recursively using a 12-month return holding period for a forecasting regression of market excess returns on SII where the length of the estimation period increases by one month as each new month is added to the regression over the January 2012 to December 2023 sample period. Bold indicates significance at the 0.05 level or lower. Variables are defined in Appendix A.

	Spearman	Pearson
	β_{SII}	
Independent Variables:		
<i>Retail%_Volume</i>	-0.919	-0.919
<i>Retail%_Dollar Volume</i>	-0.888	-0.860
<i>Retail%_Trade Count</i>	-0.867	-0.846
<i>Retail Short%_Volume</i>	-0.913	-0.909
<i>Retail Short%_Dollar Volume</i>	-0.859	-0.869
<i>Retail Short%_Trade Count</i>	-0.867	-0.840
<i>Short Squeeze</i>	-0.595	-0.625
Control Variables:		
<i>Market Liquidity</i>	0.138	0.150
<i>EPU</i>	-0.466	-0.331
<i>Macro Uncertainty</i>	-0.241	-0.677
<i>Financial Uncertainty</i>	-0.602	-0.685
<i>VIX</i>	-0.241	-0.279
<i>CFNAI</i>	-0.136	-0.023

Table 3
Tests for Intertemporal Stability in the Aggregate Short Interest Index (SII) Predictability Coefficient (β_{SII})

The table reports four tests of the intertemporal stability of β_{SII} , the predictive ability coefficient on the short-interest index variable (SII). Reported test statistics include the J -test of Elliott and Müller (2006), the SupF test of Andrews (1993), the UDMax test of Bai and Perron (1998), and Hansen's (2000) heteroskedasticity-robust SupF test based on HAC covariance estimation. Following Paye and Timmermann (2006) and Rapach and Wohar (2006), the trimming parameter for the SupF tests is set to 15 percent of the sample size. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels for a two-tailed test, respectively. Variables are defined in Appendix A.

Test	Statistic for β_{SII}
J -Stat	-8.61**
SupF	25.79**
UDMax	26.88**
Hansen SupF	158.50**

Table 4
The Relation between Aggregate Short Interest Index (*SII*) Predictive Ability Coefficient (β_{SII}) and Retail Trading Activity

The table reports results for the relation between the aggregate short interest index (*SII*) predictive ability coefficient (β_{SII}) and retail trading activity. The sample consists of monthly observations from January 2012 - December 2023 (N = 144). In Panel A, the dependent variable is β_{SII} , which is estimated recursively using a 12-month return holding period for a forecasting regression of market excess returns on *SII* where the length of the estimation period increases by one month as each new month is added to the regression over the January 2012 to December 2023 sample period. In Panel B, the dependent variable ($\beta_{SII|PC}$) is the analogous regression coefficient after controlling for a set of principal components in the first stage return forecasting regression. The independent variable of interest is *Retail%_Volume*, the proportion of market trading volume attributable to retail investors. Cols. (1) and (2) report univariate OLS regressions of the dependent variable on *Retail%_Volume* alone and on the control variables alone, respectively, where the controls are *Market Liquidity*, the *Macro Uncertainty* index, the *VIX*, and the Chicago Fed National Activity Index (*CFNAI*). Col. (3) reports the OLS regression including both *Retail%_Volume* and the full set of controls. Col. (4) reports WLS estimates of the Col. (3) specification, weighting observations by the precision of the estimated dependent variable in each month. Col. (5) reports the WLS specification based on a bootstrap obtained by re-estimating the complete two-stage procedure in each iteration of the bootstrap. Parentheses below coefficient estimates in Cols. (1) – (4) report Newey-West HAC *t*-statistics while Col. (5) reports bootstrap *z*-statistics. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels for a two-tailed test, respectively. Variables are defined in Appendix A.

Panel A: Regression of the Aggregate Short Interest Index (<i>SII</i>) Predictive Ability Coefficient (β_{SII}) on Retail Trading Activity					
Dependent Variable:	β_{SII}				
	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	WLS	WLS with Bootstrap
<i>Retail%_Volume</i>	−4.614*** (−11.233)		−4.236*** (−6.624)	−4.065*** (−6.663)	−4.065*** (−3.640)
<i>Market Liquidity</i>		0.049*** (4.232)	0.019*** (2.625)	0.022*** (3.020)	0.022 (0.999)
<i>Macro Uncertainty</i>		−1.492*** (−3.343)	0.059 (0.245)	0.037 (0.168)	0.037 (0.067)
<i>VIX</i>		12.764** (1.987)	−0.676 (−0.179)	−1.055 (−0.310)	−1.055 (−0.107)
<i>CFNAI</i>		0.000 (0.116)	0.002 (1.495)	0.002 (1.419)	0.002 (0.211)
Newey-West (HAC) SE	Yes	Yes	Yes	Yes	N/A
Number of Monthly Obs.	144	144	144	144	144
Adj. R ²	0.843	0.642	0.857	0.850	0.850

Table 4 Cont'd

Panel B: Regression of Aggregate Short Interest Index (<i>SII</i>) Predictive Ability Coefficient after Controlling for Principal Components ($\beta_{SII PC}$) on Retail Trading Activity					
Dependent Variable:	$\beta_{SII PC}$				
	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	WLS	WLS with Bootstrap
<i>Retail%_Volume</i>	-3.733*** (-9.036)		-3.954*** (-5.107)	-3.588*** (-5.236)	-3.588*** (-3.086)
<i>Market Liquidity</i>		0.036*** (3.521)	0.009 (1.280)	0.015** (2.138)	0.015 (0.607)
<i>Macro Uncertainty</i>		-1.169*** (-3.017)	0.278 (0.985)	0.177 (0.792)	0.177 (0.313)
<i>VIX</i>		10.705** (1.988)	-1.842 (-0.464)	-0.963 (-0.282)	-0.963 (-0.089)
<i>CFNAI</i>		0.000 (-0.147)	0.001 (0.945)	0.002 (1.063)	0.002 (0.139)
Newey-West (HAC) SE	Yes	Yes	Yes	Yes	N/A
Number of Monthly Obs.	144	144	144	144	144
Adj. R ²	0.774	0.517	0.779	0.766	0.766

Table 5
Tests of Mechanism 1: Retail Short Selling

The table reports results for the relation between the aggregate short interest index (SII) predictive ability coefficient (β_{SII}) and retail short selling activity. The sample consists of monthly observations from January 2012 - December 2023 ($N = 144$). In Panel A, the dependent variable is β_{SII} , which is estimated recursively using a 12-month return holding period for a forecasting regression of market excess returns on SII where the length of the estimation period increases by one month as each new month is added to the regression over the January 2012 to December 2023 sample period. In Panel B, the dependent variable ($\beta_{SII|PC}$) is the analogous regression coefficient after controlling for a set of principal components in the first stage return forecasting regression. The independent variable of interest is *Retail Short% Volume*, measured as the fraction of aggregate monthly trading volume attributable to retail short selling. Cols. (1) and (2) report univariate OLS regressions of the dependent variable on *Retail Short% Volume* alone and on the control variables alone, respectively, where the controls are *Market Liquidity*, the *Macro Uncertainty* index, the *VIX*, and the Chicago Fed National Activity Index (*CFNAI*). Col. (3) reports the OLS regression including both *Retail Short% Volume* and all control variables. Col. (4) reports WLS estimates of the Col. (3) specification, weighting observations by the precision of the estimated dependent variable in each month. Col. (5) reports the WLS specification based on a bootstrap obtained by re-estimating the complete two-stage procedure in each iteration of the bootstrap. Parentheses below coefficient estimates in Cols. (1) – (4) report Newey-West HAC t -statistics while Col. (5) reports bootstrap z -statistics. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels for a two-tailed test, respectively. Variables are defined in Appendix A.

Panel A: Regression of the Aggregate Short Interest Index (SII) Predictive Ability Coefficient (β_{SII}) on Retail Short Selling					
Dependent Variable:	β_{SII}				
	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	WLS	WLS with Bootstrap
<i>Retail Short% Volume</i>	−25.742*** (−10.767)		−22.161*** (−6.695)	−21.258*** (−6.734)	−21.258*** (−3.740)
<i>Market Liquidity</i>		0.049*** (4.232)	0.027*** (3.171)	0.031*** (3.191)	0.031 (1.446)
<i>Macro Uncertainty</i>		−1.492*** (−3.343)	−0.047 (−0.145)	−0.067 (−0.217)	−0.067 (−0.131)
<i>VIX</i>		12.764** (1.987)	0.648 (0.182)	0.113 (0.035)	0.113 (0.012)
<i>CFNAI</i>		0.000 (0.116)	−0.001 (−0.249)	−0.001 (−0.256)	−0.001 (−0.057)
Newey-West (HAC) SE	Yes	Yes	Yes	Yes	N/A
Number of Monthly Obs.	144	144	144	144	144
Adj. R^2	0.826	0.642	0.859	0.853	0.853

Table 5 Cont'd

Panel B: Regression of Aggregate Short Interest Index (<i>SII</i>) Predictive Ability Coefficient after Controlling for Principal Components ($\beta_{SII/PC}$) on Retail Short Selling					
Dependent Variable:	$\beta_{SII/PC}$				
	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	WLS	WLS with Bootstrap
<i>Retail Short% Volume</i>	−20.983*** (−9.135)		−20.816*** (−4.985)	−18.892*** (−5.073)	−18.892*** (−2.973)
<i>Market Liquidity</i>		0.036*** (3.521)	0.015** (2.040)	0.024** (2.376)	0.024 (0.996)
<i>Macro Uncertainty</i>		−1.169*** (−3.017)	0.188 (0.555)	0.088 (0.296)	0.088 (0.160)
<i>VIX</i>		10.705** (1.988)	−0.675 (−0.187)	−0.009 (−0.003)	−0.009 (−0.001)
<i>CFNAI</i>		0.000 (−0.147)	−0.001 (−0.645)	−0.001 (−0.398)	−0.001 (−0.082)
Newey-West (HAC) SE	Yes	Yes	Yes	Yes	N/A
Number of Monthly Obs.	144	144	144	144	144
Adj. R ²	0.769	0.517	0.785	0.769	0.769

Table 6
Tests of Mechanism 2: Short Squeeze Risk

The table reports results for the relation between the aggregate short interest index (SII) predictive ability coefficient (β_{SII}) and short squeeze risk. The sample consists of monthly observations from January 2012 - December 2023 (N = 144). In Panel A, the dependent variable is β_{SII} , which is estimated recursively using a 12-month return holding period for a forecasting regression of market excess returns on SII where the length of the estimation period increases by one month as each new month is added to the regression over the January 2012 to December 2023 sample period. In Panel B, the dependent variable ($\beta_{SII|PC}$) is the analogous regression coefficient after controlling for a set of principal components in the first stage return forecasting regression. The independent variable of interest is *Short Squeeze*, measured as the fraction of CRSP-listed common stocks experiencing at least one trading-day securities lending squeeze in a given month. Cols. (1) and (2) report univariate OLS regressions of the dependent variable on *Short Squeeze* alone and on the control variables alone, respectively, where the controls are *Market Liquidity*, the *Macro Uncertainty* index, the *VIX*, and the Chicago Fed National Activity Index (*CFNAI*). Col. (3) reports the OLS regression including both *Short Squeeze* and all control variables. Col. (4) reports WLS estimates of the Col. (3) specification, weighting observations by the precision of the estimated dependent variable in each month. Col. (5) reports the WLS specification based on a bootstrap obtained by re-estimating the complete two-stage procedure in each iteration of the bootstrap. Parentheses below coefficient estimates in Cols. (1) – (4) report Newey-West HAC t -statistics while Col. (5) reports bootstrap z -statistics. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels for a two-tailed test, respectively. Variables are defined in Appendix A.

Panel A: Regression of the Aggregate Short Interest Index (SII) Predictive Ability Coefficient (β_{SII}) on Short Squeeze Risk					
Dependent Variable:	β_{SII}				
	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	WLS	WLS with Bootstrap
<i>Short Squeeze</i>	-7.460*** (-4.906)		-4.973*** (-4.724)	-4.858*** (-4.218)	-4.858** (-2.573)
<i>Market Liquidity</i>		0.049*** (4.232)	0.030*** (2.888)	0.037*** (2.827)	0.037* (1.789)
<i>Macro Uncertainty</i>		-1.492*** (-3.343)	-1.301*** (-4.297)	-1.284*** (-4.187)	-1.284*** (-3.110)
<i>VIX</i>		12.764** (1.987)	6.241** (2.133)	6.315** (2.263)	6.315 (0.672)
<i>CFNAI</i>		0.000 (0.116)	-0.002 (-0.680)	-0.002 (-0.570)	-0.002 (-0.155)
Newey-West (HAC) SE	Yes	Yes	Yes	Yes	N/A
Number of Monthly Obs.	144	144	144	144	144
Adj. R ²	0.387	0.642	0.783	0.762	0.762

Table 6 Cont'd

Panel B: Regression of Aggregate Short Interest Index (<i>SII</i>) Predictive Ability Coefficient after Controlling for Principal Components ($\beta_{SII/PC}$) on Short Squeeze Risk					
Dependent Variable:	$\beta_{SII/PC}$				
	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	WLS	WLS with Bootstrap
<i>Short Squeeze</i>	-6.017*** (-5.236)		-4.206*** (-3.897)	-4.130*** (-3.393)	-4.130** (-2.162)
<i>Market Liquidity</i>		0.036*** (3.521)	0.021* (1.883)	0.031** (2.189)	0.031 (1.349)
<i>Macro Uncertainty</i>		-1.169*** (-3.017)	-1.008*** (-3.802)	-0.995*** (-3.486)	-0.995** (-2.337)
<i>VIX</i>		10.705** (1.988)	5.187* (1.790)	5.716** (2.230)	5.716 (0.558)
<i>CFNAI</i>		0.000 (-0.147)	-0.003 (-0.905)	-0.002 (-0.596)	-0.002 (-0.145)
Newey-West (HAC) SE	Yes	Yes	Yes	Yes	N/A
Number of Monthly Obs.	144	144	144	144	144
Adj. R ²	0.352	0.517	0.658	0.645	0.645

Table 7
Path Analysis for Aggregate Short Interest Predictive Ability Mechanisms

The table reports a path analysis that decomposes the total effect of retail trading activity (*Retail%_Volume*) on the return predictability of aggregate short interest (*SII*) into direct and indirect effects. *Retail%_Volume* measures the proportion of market trading volume attributable to retail investors. The mediating variables are *Retail Short%_Volume*, which measures the fraction of aggregate monthly trading volume attributable to retail short selling, and *Short Squeeze*, which measures the fraction of CRSP-listed common stocks experiencing at least one trading-day securities lending squeeze in the month. The direct effect captures the relation between *Retail%_Volume* and *SII*'s predictive ability after controlling for the respective mediator, while the indirect effect captures the portion of the relation operating through that mediator. Cols. (1) and (2) use β_{SII} as the measure of *SII*'s predictive ability, while Cols. (3) and (4) use $\beta_{SII/PC}$. Cols. (1) and (3) examine *Retail Short%_Volume* as the mediator, while Cols. (2) and (4) examine *Short Squeeze* as the mediator. Parentheses report Newey-West HAC *t*-statistics. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels for a two-tailed test, respectively. Variables are defined in Appendix A.

	Dependent Variable: β_{SII}		$\beta_{SII/PC}$	
	Mediator: <i>Retail Short%_Volume</i>	<i>Short Squeeze</i>	<i>Retail Short%_Volume</i>	<i>Short Squeeze</i>
	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	OLS
c Total effect (<i>Retail%_Volume</i> $\rightarrow$ β_{SII})	-4.236***	-4.236***	-3.954***	-3.954***
	(-6.62)	(-6.62)	(-5.11)	(-5.11)
c' Direct effect (<i>Retail%_Volume</i> $\rightarrow$ β_{SII} Mediator)	-2.172***	-3.368***	-1.923***	-3.402***
	(-3.30)	(-5.03)	(-2.91)	(-4.04)
a <i>Retail%_Volume</i> $\rightarrow$ Mediator	0.171***	0.397***	0.171***	0.397***
	(8.72)	(4.59)	(8.72)	(4.59)
b Mediator $\rightarrow$ β_{SII} <i>Retail%_Volume</i>	-12.096***	-2.187***	-11.904**	-1.393*
	(-2.83)	(-2.96)	(-2.26)	(-1.79)
a*b Indirect effect	-2.064***	-0.868**	-2.031**	-0.553*
	(-2.69)	(-2.49)	(-2.18)	(-1.67)
Proportion mediated	48.70%	20.50%	51.40%	14.00%
Newey-West (HAC) SE	Yes	Yes	Yes	Yes
Number of Monthly Obs.	144	144	144	144

Table 8
Results of Robustness Tests

The table reports the results of robustness tests of the relation between the aggregate short interest index (*SII*) predictive ability coefficient and retail trading activity, where features of the baseline specifications reported in Tables 4 and 5 are modified as described below. Each row lists an alternative specification and provides a brief description of the modification (the concern it addresses), the estimated coefficient on the variable of interest in the robustness test, and the resulting *t*- or *z*-test statistic. All robustness tests include the same control variables as the main tests. The first three robustness tests re-estimate the models in Table 4, Panel A using alternative return forecasting horizons (holding periods) of 1, 3, and 6 months, in place of the 12-month horizon used in the main tests. The next four robustness tests use alternative measures of retail trading activity and retail short selling activity that are based on dollar volume and trade count rather than share volume. The final robustness test (last row of the table) estimates a single-stage return forecasting regression that interacts *SII* with *Retail%_Volume* as an alternative approach to test whether the predictive relation between *SII* and future aggregate returns varies with retail trading activity. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels for a two-tailed test, respectively. Variables are defined in Appendix A.

Specification	Description of Robustness Test	Rationale for Robustness Test	coeff.	t or z-stat
<i>h</i> = 1 month	Re-estimate Table 4 Panel A's regressions using a 1-month-ahead return horizon (holding period).	Ensure the <i>Retail%_Volume</i> relation is not an artifact of the 12-month return horizon (holding period) used in the main tests.	−2.844***	(−7.675)
<i>h</i> = 3 months	Re-estimate Table 4 Panel A's regressions using a 3-month-ahead return horizon (holding period).	Same as the prior row.	−5.557***	(−5.734)
<i>h</i> = 6 months	Re-estimate Table 4 Panel A's regressions using a 6-month-ahead return horizon (holding period).	Same as the prior row.	−7.167***	(−10.840)
<i>Retail%_Dollar Volume</i>	Re-estimate Table 4 Panel A's regressions using <i>Retail%_Dollar Volume</i> as independent variable.	Ensure that the results are not sensitive to how retail trading activity is measured. The use of dollar volume weights larger trades more heavily than the share-volume measure in the main tests.	−5.432***	(−3.849)
<i>Retail%_Trade Count</i>	Re-estimate Table 4 Panel A's regressions using <i>Retail%_Trade Count</i> as independent variable.	Ensure that the results are not driven by a small number of large retail trades. The use of trade count instead captures the breadth of retail participation.	−7.315***	(−3.729)
<i>Retail Short%_Dollar Volume</i>	Re-estimate Table 5's regressions using <i>Retail Short%_Dollar Volume</i> as the independent variable.	Ensure that the results are not sensitive to how retail short-selling activity is measured.	−29.600***	(−4.789)
<i>Retail Short%_Trade Count</i>	Re-estimate Table 5's regressions using <i>Retail Short%_Trade Count</i> as the independent variable.	Ensure that the results are not sensitive to how retail short-selling activity is measured.	−54.307***	(−4.404)

<i>SII * Retail%_Volume</i> <i>interaction</i>	Estimate the coefficient (θ_1) on the interaction term in the following regression: $Return_{t,h} = \alpha_t + \theta_1 SII * Retail\%_Volume + \beta_2 * SII + \beta_3 * Retail\%_Volume + Controls_t + \varepsilon_t,$ where h (return holding period) = 12 months and the controls variables are <i>Market Liquidity</i>, <i>Macro Uncertainty</i>, <i>VIX</i>, <i>CFNAI</i>.	Ensure that the negative effect of <i>Retail%_Volume</i> is not an artifact of the two-stage estimation procedure used in the main tests by instead using a single-stage regression with an interaction term.	-13.548***	(-2.970)
---	--	--	------------	----------
