Retail Trading Intensity and the Overnight-Intraday Return Gap*

Yongkil Ahn† Alfred Qi Fan‡ Don Noh§ Stella Park¶

February 20, 2024

Abstract

In most stock markets, average overnight (close-to-open) returns are high while intraday (open-to-close) returns are low, even negative. We show that retail investors' trading intensity is a major explanatory variable for this "overnight-intraday return gap" by using data from Korea where exhaustive investor type-level trade flows are available. Our IV approach exploits retail investors' tendency to more actively trade stocks with low per-share prices. We attribute this relationship to retail demand for daytime stock market exposure: They flow into high volatility stocks near open and flow out near close. This intensifies during times of high funding liquidity and consumer sentiment.

* This research was supported by the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 grant. We thank Motohiro Yogo, Wei Xiong, Michael Junho Lee, Jingi Ha, and Rongchen Li for feedback and suggestions. We also thank participants of the Princeton Finance Student Workshop, HKUST Finance Brown Bag, SMU Finance Brown Bag for helpful comments.

† Seoul National University of Science and Technology (yongkil.ahn@seoultech.ac.kr)

‡ Singapore Management University (qi.fan.2020@pbs.smu.edu.sg)

§ Hong Kong University of Science and Technology (noh@ust.hk)

¶ Singapore Management University (stellapark@smu.edu.sg)

1 Introduction

Both traders and academics alike have noticed a large gap between overnight and intraday returns (Branch and Ma 2006; Kelly and Clark 2011). Overnight returns (close-to-open) tend to be much higher than intraday returns (open-to-close) in both developed and emerging stock markets.[1] Figure 1 replicates this pattern for the Korean stock market, the setting of this paper. Once we decompose daily returns into their overnight and intraday components and compound them separately, we can easily notice the "overnight-intraday return gap."

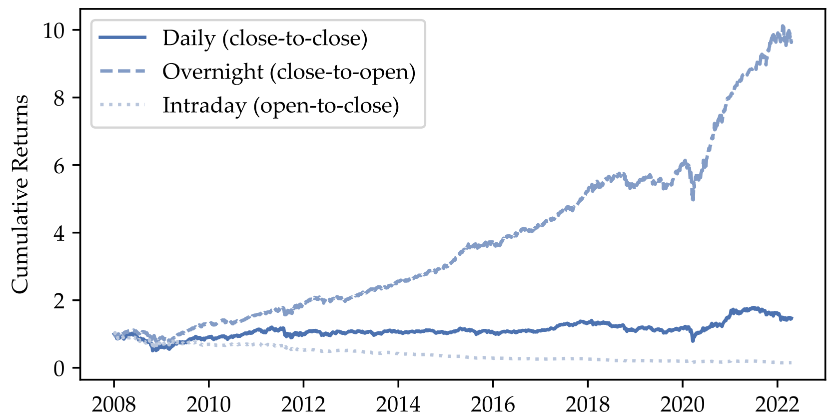


Figure 1. Cumulative Overnight and Intraday Returns

This figure shows cumulative overnight (close-to-open), intraday (open-to-close), and daily (close-to-close) returns of the Korea Composite Stock Price Index (KOSPI). The exact definition of each is given in Section 2.2.1.

In this paper, we find a positive relationship between this overnight-intraday return gap and retail investors' trading intensity. We also show that retail trading intensity has significant predictive power for the cross-section of return gaps in the following period. We proxy for retail intensity by retail investors' share of total trading volume, which can be computed precisely and severally for individual stocks thanks to the unique advantage of the Korean stock market data: Korea Exchange (KRX) reports the investor types (e.g., retail investors, foreign institutions, domestic pension funds, etc.) of the buyers and sellers. In our main empirical test, at the stock-month level a 1pp (cross-sectional standard deviation of retail volume share is around 20%) increase in retail volume share is associated with an approximately 0.3% higher monthly return gap in the following month.[2] While a similar relationship has been posited before (Berkman et al. 2012), an accurate and exhaustive measure of retail trading volume for such a long time-series has not been readily available for the US market. Also, previous studies have not established a causal relationship between these two quantities, nor investigate the sources of the persistent

[1] See Figure A1 for global evidence.

[2] The same results hold at the stock-day level, but with much lower R^2 and proportionally smaller magnitude.

diurnal patterns of retail flow.

The relationship between the overnight-intraday return gap and retail trading intensity is robust, but concerns about omitted variables or simultaneity remain. The return gap may be more pronounced for small, illiquid stocks which retail investors tend to be more attracted to. One may also worry about the fact that price and volume are jointly determined. To address this, we look for an instrumental variable (IV) that is correlated with retail trading activity but unrelated to fundamentals or sentiment. Firms often use the rationale that per-share prices are too high for individual investors to trade their shares to justify stock split decisions. Studies have also found the effect of stock splits on attracting trading activity (Easley et al. 2001).[3] At the same time, it is difficult to find a reason that the nominal per-share price should carry economic information. Based on these reasons, we use the log number of shares outstanding, after controlling for firm size, as an IV for retail volume shares. We show that it is a strong instrument (with a partial F-stat much larger than 10) and that our previous findings remain intact after instrumenting for retail volume share.

We also supplement these findings with a difference-in-differences (DID) design that exploits the shock to aggregate retail investor trading intensity caused by COVID-19. Likely due to the lack of entertainment under stay-at-home orders and/or to the generous stimulus packages, a retail trading frenzy ensued in the months following March 2020 (Ozik et al. 2021). The Korean stock market experienced a similar surge in retail trading, as shown in Figure 10. Consistent with our prediction, the overnight-intraday return gap also widens significantly in the post-pandemic period. But more importantly, we go from this increase in aggregate retail intensity to cross-sectional variation in retail volume shares by using the fact that *shares* are capped at 100%. Suppose there are only two stocks: a retail-dominant stock with 99% retail volume share and another stock with 50% retail volume share. An exogenous increase to general retail trading activity arrives and retail investors allocate comparable volume to these two stocks. The first stock will experience a negligible increase in retail volume share compared to the second stock because it was already close to the bound 100%. A more detailed example is discussed in Section 3.4 and Figure 11 shows this pattern empirically. Using the stocks with low (high) initial retail volume share as the treated (control) group, we find results consistent with our previous findings.

A natural question to ask now is: Why should more intense retail investor activity lead to higher overnight returns and lower intraday returns? We find suggestive evidence that this is driven by retail investors' demand for daytime stock market exposure. The aggregate pattern shown in Figure 2 foreshadows our related findings. Using the investor type identifiers, we can compute investor type-level normalized net buy (buy volume minus sell volume, normalized by shares outstanding) at intraday frequencies. Once we compute the average, across stock-days, of the investor type-level net buys within different 30-minute bins, we see that retail investors

[3] Also see Kumar and Lee (2006) and Green and Hwang (2009) for effect of price levels on trading.

consistently net buy stocks near market open and net sell near close. Stocks with high retail volume shares tend to have more investors who behave in such a way and have relatively less institutional volume to provide liquidity. Figure 5 shows why prices move in the direction of retail flows although retail and institutional net buy must sum to zero by market clearing – within the same stock-time. By comparing the average buy and sell prices of retail investors to those of institutional investors within the same stock-time interval, we find that retail investors are liquidity takers near market open. In other words, they buy high and sell low near market open.

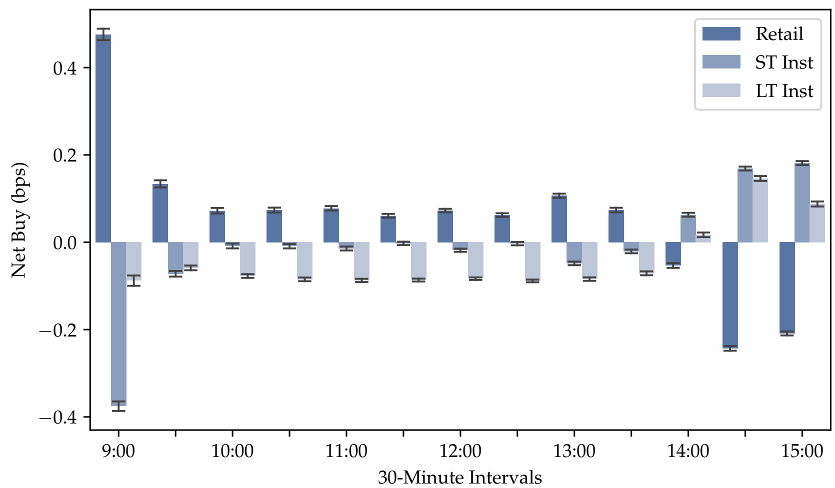


Figure 2. Intraday Net Buy Patterns

This figure shows the pooled average values for normalized net buys by different investor types during different times of the day. Net buys are normalized by shares outstanding. Each observation is at the stock-day-interval-investor level and each time interval is 30 minutes long. 95% confidence bands are constructed using bootstrapping with 1,000 bootstrap samples.

We also posit that this demand may be driven by sensation- or entertainment-seeking (Grinblatt and Keloharju 2009). The following findings corroborate this narrative. First, in the morning retail investors more strongly net buy stocks with high past volatility and in the afternoon they more strongly net sell such stocks.[4] Second, this pattern is stronger during times with ample funding liquidity and high consumer sentiment. Third, account-level data from a small Korean brokerage reveal that mobile phone traders are more prone to such behavior and that users tend to log in to the brokerage applications much more often when they have open positions. Taken together, retail investors appear to hold "exciting" stocks when they can monitor them in real-time and they do so when they are less financially constrained (Heimer and Imas 2022).

[4] We find similar results using skewness or maximum daily return in the past 20 trading days. All three variables are known to be significantly correlated (Bali et al. 2011).

Our first contribution to the study of the overnight-intraday return gap is that we bring exhaustive and accurate retail flow data over an almost 15-year period. The retail investor flag is accurate since it comes directly from KRX; retail flow is also exhaustive because KRX is the sole stock exchange in Korea on which all trade flows clear. While a recent innovation made by Boehmer et al. (2021) allows researchers to impute retail flows from the US trade and quotes (TAQ) data, these imputed flows remain a proxy for the exhaustive retail activity. Barber et al. (2022) also raise questions about the accuracy and comprehensiveness of the methodology.

Our second contribution is that we propose new identification strategies discussed earlier. Most previous studies (e.g., Berkman et al. 2012; Hendershott et al. 2020; Bogousslavsky 2021) explore the cross-sectional relationship between overnight returns and characteristics, such as market beta or proxies of retail attention, but stop short of establishing a causal relationship.

Lastly, we offer a new source of persistent intraday flow patterns of a certain "clientele," that is the retail investors. While there are discussions about the existence of repeated flows by heterogeneous groups of investors (Heston et al. 2010; Berkman et al. 2012; Lou et al. 2019), the sources of the flows are often only vaguely discussed. We find evidence that retail demand for real-time stock market exposure could be a source of intraday retail flows near open, and that these flows are aggressive and large enough to affect intraday return dynamics.

Additionally, our results provide a basis to believe that overnight-intraday return reversals may be a proxy for underlying – and in most markets, unobservable – retail activity. For example, Aboody et al. (2018) argues that overnight returns can be used as a measure of firm-specific sentiment; Akbas et al. (2022) relates the frequency of overnight-intraday return reversals to prominence of noise traders. Our findings show how degree opening price inflation can be a manifestation of retail (or noise trader) intensity. Also, the curious issue of "negative risk premium" in Hendershott et al. (2020) can be rationalized because we see that someone (retail investors) is eager to hold riskier assets near the open. At least near open, retail investors are the marginal investors (Barber et al. 2009).

The main novelty of this paper is not identifying the overnight-intraday return gap, as it has been already widely studied. Kelly and Clark (2011) compare Sharpe ratios of overnight and intraday returns of 5 common stock ETFs. They find that risk-adjusted returns are significantly higher during non-trading hours. Branch and Ma (2015) further show that high overnight returns tend to be followed by low intraday returns. While they speculate on several possible explanations, such as market maker trading patterns or retail investor activity near open, they do not investigate this further.

Berkman et al. (2012) exemplifies the "price impact" explanation: The paper points to the fact that retail investors are much more active around the opening hours and attributes the return gap to opening price inflation caused by attention-induced buying by retail investors. Lou et al. (2019) identifies a more general "clientele effect." The authors show that high overnight (intraday) returns during a given month predicts high overnight (intraday) returns in the following month.

This is to be expected if a certain group of investors trades during a certain hour of the day, and if they do so in the same direction over several days or weeks. In an older study, Heston et al. (2010) also find evidence of investors having a predictable demand for immediacy at certain times of the day. In a more recent study, Jones et al. (2022) attribute this pattern to price pressure by extrapolative trading by investors. They find that high intraday returns lead to high morning order imbalances and high overnight returns in the following day, which is expected in presence of unsophisticated and short-constrained investors. Lu et al. (2023) provide a more refined model with "fast and slow arbitrageurs," which can explain the diurnal return pattern even with a price pressure that does not revert throughout the day.

A related view is the "overnight risk" explanation that is articulated in Bogousslavsky (2021). He finds that mispricing corrects in the morning and worsens at the end of the day and attributes this to capital-constrained arbitrageurs closing their position at close. Similar to the view put forward by Savor and Wilson (2014), which shows that days with macroeconomic announcements pose inherently different risks, perhaps overnight periods are also distinct from intraday periods due to extreme illiquidity.[5] Consistent with this idea, Hendershott et al. (2020) shows that variation in market beta explains the cross-section of overnight returns well, notwithstanding the well-documented fact that CAPM performs poorly in explaining the cross-section of close-to-close returns (Fama and French 2004).

Other more recent papers also investigate overnight returns, but focus on different aspects. Akbas et al. (2022) test the relationship between the number of daytime reversals – positive overnight returns followed by negative intraday returns – and subsequent monthly returns. They find that more reversals lead to higher returns in the following month. Boyarchenko et al. (2023) compute market index futures returns for each hour of the trading day and find large positive returns during opening hours of European markets (2-3 a.m. EST). They argue that this is due to overnight resolution of order imbalances from end of previous U.S. trading day.

2 Data and Stylized Facts

2.1 Data

2.1.1 The KRX Data

We use 1-minute frequency price data, purchased from the KRX, for common stocks in Korea. Our price data covers the period from January 2009 to May 2021. There are approximately 2,400 stocks listed on the KRX as of 2021, but many of them are small and illiquid compared to the average listed US company. For this reason, we use a subset of 901 stocks (1) that traded every day throughout our sample period, (2) whose average market capitalization larger than 50 KRW

[5] For a theoretical basis for this day-night distinction, we can also look to Hong and Wang (2000) who explicitly model the regular closure of markets.

billion, and (3) whose average daily trading value is larger than KRW 500 million. Once we further restrict to stocks with Compustat data available, we arrive at 812 stocks. The last step was added to enable cross-checking of book values provided by the KRX and Compustat.

An important advantage of the KRX data is that the KRX provides identifiers for the investor type at the individual trade level through their trade and quote (TAQ) data. The provided investor types include: (1) retail investors, (2) foreign institutions, (3) proprietary traders, (4) asset managers, (5) pension funds, (6) banks, (7) insurance companies, and (8) private equity. All investor types except retail investors and foreign institutions refer to domestic institution types. At a more reasonable price, KRX also provides 1-minute frequency trade flows aggregated at the investor type level. We use the 1-minute frequency trade flow data over the period from January 2009 to May 2021.

KRX also provides various daily statistics such as short volume, short interest, book-to-market ratio, dividend yield, and shares outstanding on its website. For regulatory reasons, foreign investors' holdings shares are also reported at a daily frequency, but holdings of other investor types are not directly reported. These firm characteristics, such as book-to-market and shares outstanding, are obtained directly from the KRX website.

2.1.2 Brokerage Data

One limitation of the KRX data is that all retail investors are aggregated into one group. In other words, it is not an *account-level* data. To ask more detailed questions related to heterogeneity within retail investors, we obtain data from a mid-sized discount brokerage in Korea that has anonymized account identifiers.[6] The data is very similar to the one used by Barber and Odean (2000). There are 50,000 randomly chosen traders in our dataset and the dataset covers the period from January 2015 to December 2018.

The "trade file" includes all stock transactions by customers who trade through this brokerage firm. All usual information such as order quantity, order prices, order type (e.g., market or limit orders), executed quantity, executed price, and order time are provided in this file. In addition, the brokerage firm provides the order media (e.g., mobile app, desktop app, or in person) through which each order was submitted. This is useful information in distinguishing more dedicated retail traders from casual retail traders because dedicated traders tend to heavily utilize the desktop Home Trading System (HTS). The "balance file" includes all stock holdings by customers at the end of the day. The "log-in file" provides the time and media through which each customer logs onto the brokerage's platform.

[6] Trading volume from this brokerage firm covers approximately 5% of total retail volume on an average day.

2.1.3 Other Data Sources

We also obtain variables that proxy for aggregate funding liquidity and consumer sentiment. We proxy for the first with aggregate retail (uninvested) deposits, across all brokerage firms, provided by the Korea Financial Investment Association (KOFIA). KOFIA reports these deposits at a daily frequency. We use the Composite Consumer Sentiment Index (CCSI) reported by the Bank of Korea (BOK) to proxy for consumer sentiment. CCSI is constructed from a monthly survey of 2,500 Korean households.

2.2 Variable Construction and Stylized Facts

2.2.1 Overnight and Intraday Returns

The overnight return, or the close-to-open (CTO) return, and the intraday return, or the open-to-close (OTC) return, for stock i on day d are computed as follows:

$$\mathrm{CTO}_{id} = \frac{P_{id}^{\mathrm{open}} - P_{id-1}^{\mathrm{close}}}{P_{id-1}^{\mathrm{close}}} \tag{1}$$

$$\mathrm{OTC}_{id} = \frac{P_{id}^{\mathrm{close}} - P_{id}^{\mathrm{open}}}{P_{id}^{\mathrm{open}}}. \tag{2}$$

The usual daily return, or the close-to-close (CTC) return, can be recovered through the following:

$$1 + \mathrm{CTC}_{id} = (1 + \mathrm{CTO}_{id}) \times (1 + \mathrm{OTC}_{id}) = \frac{P_{id}^{\mathrm{close}}}{P_{id-1}^{\mathrm{close}}}. \tag{3}$$

This decomposition is as in Berkman et al. (2012) and Figure 3 demonstrates this breakdown graphically.

There are two institutional details worth mentioning. The first is that, as in many other stock exchanges, opening and closing prices are determined through opening and closing auctions in Korea. These auctions aggregate buy and sell orders during the 30 minutes (10 minutes) prior to market open (close), and determine the opening (closing) uniform auction prices. In our main analyses we use the value-weighted average prices (VWAPs) during the first and last 30 minutes of trading to compute the overnight and intraday returns. If we use the opening and closing auction prices instead, the results are unchanged. The second detail is that the closing time of the Korean stock market was extended by 30 minutes in August 2016, namely from 3:00 p.m. to 3:30 p.m.

Figure 1 shows the cumulative returns, ignoring transaction costs, for the CTO and OTC returns separately. For each day d, equal weighted overnight and intraday returns are computed by averaging across the different stocks. Figure 1 plots the cumulative returns defined as the

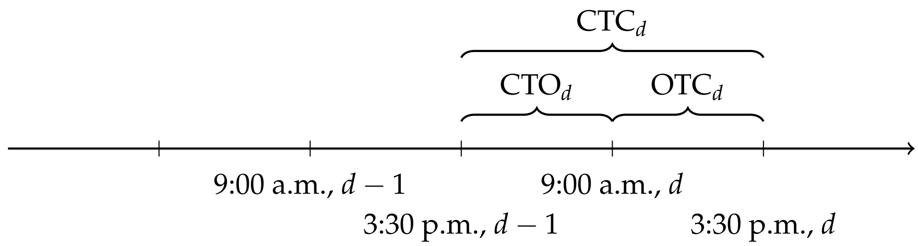


Figure 3. Overnight and Intraday Returns

This figure shows how overnight and intraday returns are defined. Each day's return (close-to-close) are decomposed into overnight return (close-to-open) and intrday return (open-to-close). The Korean stock market opens at 9:00 a.m. each day and closes at 3:30 p.m. During periods before August 1st of 2016, the closing time was 3:00 p.m.

following against time:

$$\prod_d (1 + \mathrm{CTO}_d) \quad \text{and} \quad \prod_d (1 + \mathrm{OTC}_d) \tag{4}$$

where the returns are computed from the opening and closing levels of the KOSPI. We can see that average overnight returns are much higher than average intraday returns. Also, this divergence becomes even starker after the COVID-19 crash of March 2020.

Later in our analyses, we aggregate observations to the stock-month level, as in Lou et al. (2019). We compute the *overnight-intraday return gap* of stock i during month t as the ratio between the cumulative CTO returns and the cumulative OTC returns minus 1:

$$\text{Overnight Ret}_{it} = \prod_{d \in \text{month } t} (1 + \mathrm{CTO}_{id}) - 1 \tag{5}$$

$$\text{Intraday Ret}_{it} = \prod_{d \in \text{month } t} (1 + \mathrm{OTC}_{id}) - 1 \tag{6}$$

$$\text{“Return Gap”} = \mathrm{RG}_{it} = \frac{1 + \text{Overnight Ret}_{it}}{1 + \text{Intraday Ret}_{it}} - 1. \tag{7}$$

For daily frequencies, we simply take the arithmetic difference between the overnight and intraday returns.

2.2.2 Intraday Trade Patterns by Investor Type

An important advantage of the KRX data is that KRX provides the investor type identifiers. For each stock i and investor type j, trades are aggregated at the 30-minute interval level, starting from 9:00 a.m. Suppose we index these intervals by t, then we can write:

$$\mathrm{BV}_{ijt} = \text{\# shares of stock } i \text{ bought by investor type } j \text{ during interval } t \tag{8}$$

$$\mathrm{SV}_{ijt} = \text{\# shares of stock } i \text{ sold by investor type } j \text{ during interval } t. \tag{9}$$

Then, *net buy* and normalized net buy are computed as follows:

$$\mathrm{NB}_{ijt} = \mathrm{BV}_{ijt} - \mathrm{SV}_{ijt} \tag{10}$$

$$\mathrm{nb}_{ijt} = \frac{\mathrm{NB}_{ijt}}{\text{\# shares outstanding}_{it}}. \tag{11}$$

This normalization allows us to compare volume and net buy across different stocks and periods.

Unconditionally, we should expect to see zero net buys for any randomly chosen (i, j, t). This is because markets have to clear during any interval $t \to t + 1$:

$$\sum_j \mathrm{NB}_{ijt} = 0 \text{ and } \sum_j \mathrm{nb}_{ijt} = 0 \text{ for any stock } i \text{ and time } t. \tag{12}$$

Figure 2 plots the average net buy by retail investors, short-term institutions (foreign institutions and domestic proprietary traders), and long-term institutions (other institutions) at different times of the day. We make this grouping because some institutional types, such as private equity and insurance companies, play too small a role in sub-daily frequencies. We compute the pooled averages—across stocks and days—of the net buys nb_{ijt} at different 30-minute intervals. Figure 2 plots these averages for retail investors and short-term institutions. We see that retail investors are net buyers in the morning and net sellers in the afternoon.

We repeat the same procedure with (normalized) volume by investor types. Investor type-level volume is similarly defined as with net buys:

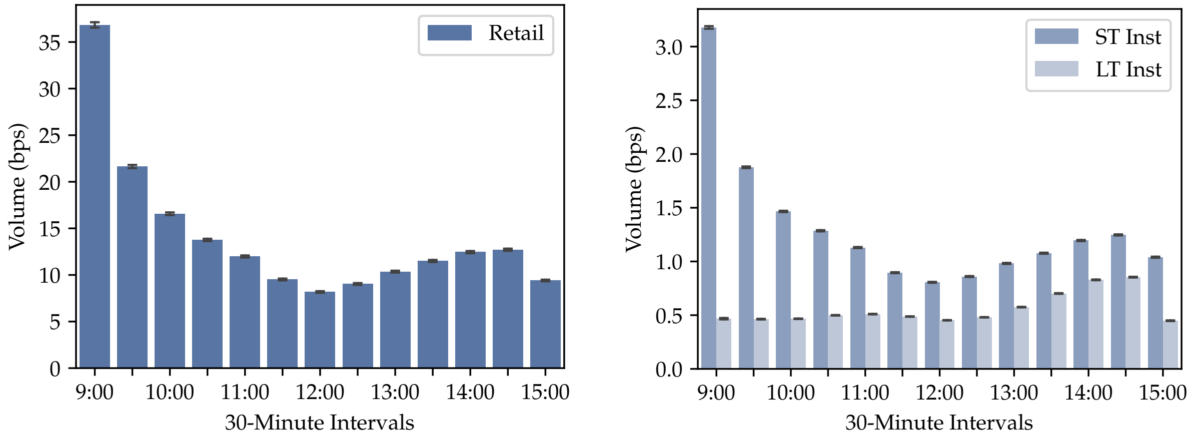
$$\mathrm{VLM}_{ijt} = \frac{\mathrm{BV}_{ijt} + \mathrm{SV}_{ijt}}{2} \tag{13}$$

$$\mathrm{vlm}_{ijt} = \frac{\mathrm{VLM}_{ijt}}{\text{\# shares outstanding}_{it}}. \tag{14}$$

The results are displayed in Figure 4. While volume is higher in the morning for both retail and short-term institutions, long-term institutions trade more in the afternoon. This pattern is consistent with short-term institutions providing liquidity to retail investors, while long-term institutions avoiding the turbulent mornings.

Using such investor type-level volume information, we can compute the *retail volume share* at various time intervals. This is simply the proportion of retail volume relative to the total volume for any stock i and time interval t:

$$\text{“Retail Share”} = \mathrm{RS}_{it} = \frac{\text{Retail Buy Volume}_{it} + \text{Retail Sell Volume}_{it}}{\text{Total Buy Volume}_{it} + \text{Total Sell Volume}_{it}}. \tag{15}$$



(a) Retail Volume at Different Times

(b) Institutional Volume at Different Times

Figure 4. Intraday Volume Patterns

This figure shows the pooled average values for normalized volume by different investor types during different times of the day. Volumes are normalized by shares outstanding. Each observation is at the stock-day-interval-investor level and each time interval is 30 minutes long. 95% confidence bands are constructed using bootstrapping with 1,000 bootstrap samples.

2.2.3 Morning and Afternoon "Aggressors"

Another exercise we can conduct is to compute average *relative buy/sell prices* by investor types. With these 1-minute frequency trade flow data, we can compute the average buy and sell prices for investor type j type as follows:

$$\text{Avg Buy (Sell) Price}_{ijt} = \frac{\text{Buy (Sell) Amount in KRW}_{ijt}}{\text{Buy (Sell) Volume}_{ijt}}. \tag{16}$$

Because the volume is in number of shares, above gives the average price per share within the interval t. This quantity is analogous to the usual value-weighted average price (VWAP).

To make fair comparisons, we keep 30-minute intervals during which both retail and institutional investors have nonzero buy and sell trades. Then, for each i and t, we take the average retail buy (sell) price, divide it by the average buy (sell) price for institutions, take the pooled average across stocks and days, and subtract one. This produces average relative price deviation for each time of the day. Figure 5 plots the relative price deviation against times of the day.

We see that retail investors buy (sell) at higher (lower) prices in the morning relative to institutions. In other words, they are demanding more immediacy close to the open. This pattern flips later in the day.

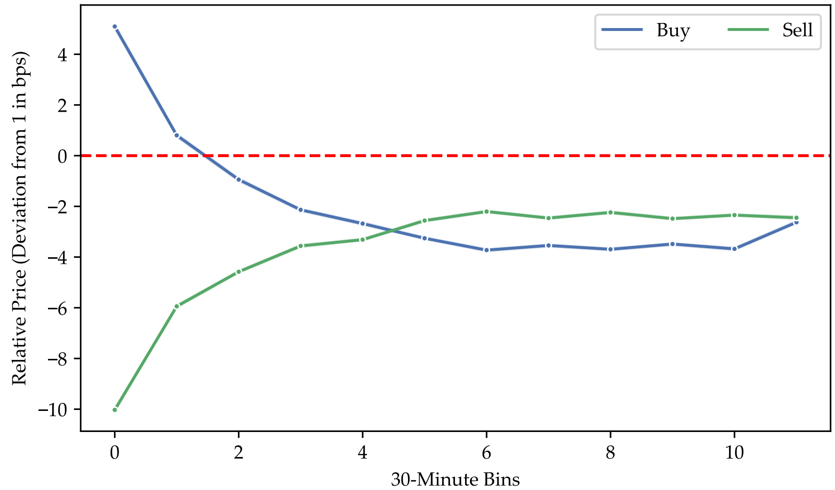


Figure 5. Relative Retail Trade Prices

This figure shows the pooled average values for the retail investors' buy and sell price deviations relative to those of the institutional investors. The deviations are expressed in basis points. Each observation is at the stock-day-interval-investor level and each time interval is 30 minutes long. 95% confidence bands are constructed using bootstrapping with 1,000 bootstrap samples.

2.2.4 Summary Statistics

We present the summary statistics for the return gap, firm characteristics, and retail shares in Table 1. One can see that an average firm in our sample has a market capitalization of roughly USD 2 billion, which is small relative to a listed US firm. The average monthly turnover is around 19%, which is high compared to NASDAQ's monthly turnover of around 5% as of 2023. The average retail volume is also high at around 67%. This quantity is difficult to compute exactly for the US, but estimates range between 5% and 25% for the US and appear to be on an increasing trajectory.[7]

What is also important to keep in mind is the cross-sectional correlation between the trade flow-related variables and firm characteristics. Admittedly, they have high correlations that are often consistent with one's priors about penny stocks vs blue chip stocks. Stocks with high retail trading volume share tend to be small, illiquid stocks with high turnover. We also see that retail volume share and monthly turnover has a high contemporaneous correlation with the monthly overnight-intraday return gap.

[7] See, for example, this article.

Table 1. Summary Statistics of Stock-Month Observations

	Mean	SD	25th	50th	75th
Cum. Overnight Ret (%)	4.65	11.52	-1.67	2.80	8.50
Cum. Intraday Ret (%)	-2.49	11.03	-9.07	-3.04	3.35
Monthly Ret Gap (%)	7.14	17.36	-2.89	5.93	15.46
Log Market Cap	5.37	1.53	4.26	5.01	6.13
Market Beta	0.88	0.37	0.63	0.88	1.13
Book-to-Market	0.96	0.83	0.45	0.77	1.25
Amihud	0.37	4.27	0.01	0.02	0.08
Foreign Ownership (%)	8.63	11.97	1.15	3.24	11.56
Monthly Turnover	0.40	0.78	0.08	0.16	0.37
Retail Vlm Share (%)	78.51	20.37	68.65	87.53	93.39

This table reports the summary statistics for the subset of 812 chosen stocks. Each observation is at the stock-month level and variables are winsorized at 1%. Cumulative overnight returns, intraday returns, and return gaps are computed according to equations (4) and (5). Amihud measure is computed using absolute returns in percentages and trading volume in KRW billions. Market capitalization is expressed in KRW billions and then log-transformed.

Table 2. Correlation Matrix of the Main Variables

	Ret Gap	Rtl Share	Turnover	Log Mkt Cap	Foreign Own	B/M
Retail Vlm Share	0.25					
Turnover	0.31	0.32				
Log Market Cap	-0.12	-0.80	-0.19			
Foreign Ownership	-0.12	-0.67	-0.18	0.65		
Book-to-Market	-0.07	-0.06	-0.13	-0.09	-0.01	
Amihud	-0.00	0.29	-0.01	-0.42	-0.22	0.15

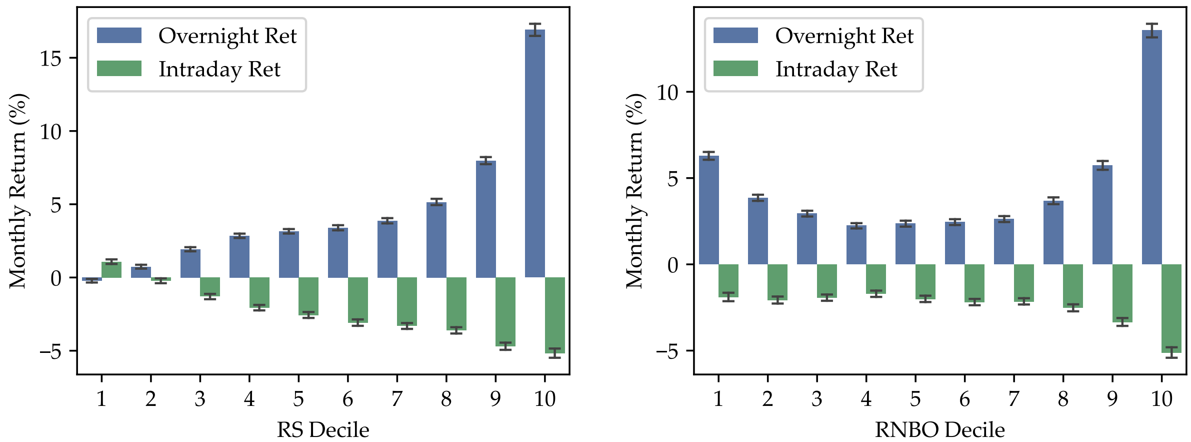
This table reports the time-series average of the cross-sectional correlation matrices of each month. First, the cross-sectional correlations among the variables for a given month are computed. Then, these matrices are averaged across all months.

3 Empirical Setup and Results

3.1 Portfolio Sorts

We start by examining the contemporaneous relationship between the monthly return gap and other variables. We do not make causal claims here, as prices and volume-related variables, such as investor type-level volume, are jointly determined. For example, retail flows not only determine the morning prices, they also may be reacting to overnight returns (e.g., due to salience). However, the patterns are still informative about the potential mechanisms behind the relationships. Here, we aggregate variables at the stock-month level, but results are similar when we use daily variables instead.

Figure 6 plots the monthly overnight and intraday returns as defined in Equation (5) of equal



(a) Sort on RS at Open

(b) Sort on RNB at Open

Figure 6. Contemporaneous Monthly Overnight and Intraday Portfolio Returns

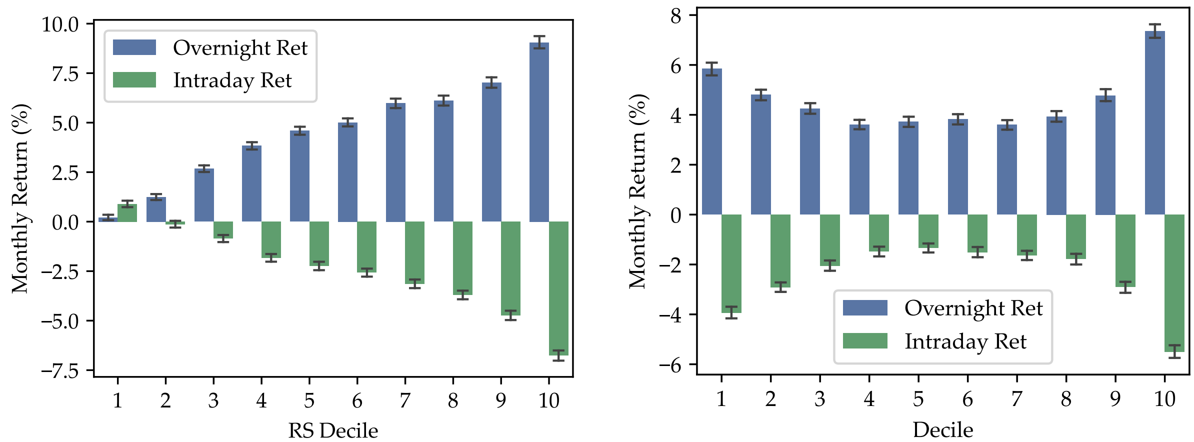
This figure shows the average monthly overnight and intraday returns for equal weighted decile portfolios. Stocks are assigned to decile portfolios based on their monthly retail volume share and morning retail net buy in the first and second panels respectively. Daily morning net buy is computed by summing up all retail net buys that occurs during the first 30 minutes of a trading day. Monthly morning net buy is computed by summing up all morning retail net buy for a given month. Returns for the same month are used. 95% confidence bands are constructed using bootstrapping with 1,000 bootstrap samples.

weighted decile portfolios. Each month, stocks are sorted into decile portfolios based on either the retail volume share (RS) in that month or normalized retail net buy (RNB) near the open (first 30 minutes) in that month. Returns are computed for the *same* month.

We see that intraday returns decrease monotonically with higher retail volume share and higher retail net buy near open. On the other hand, overnight returns increase monotonically with higher retail volume share, but is higher for *extreme* levels of contemporaneous opening retail net buy. This is somewhat at odds with a simple "retail price pressure" narrative: Morning retail flows, which are on average net positive, push prices higher in the morning. However, one can still rationalize this pattern with a more general "clientele effect" as in Lou et al. (2019). Because retail net buy is simply the negative of institutional net buy, lower deciles are including stocks that experience high institutional demand. For this reason, the intermediate deciles include stocks that do not have a strong buying pressure from either side.

Next, we repeat the same exercise, but use returns in the *subsequent* month.[8] The results are presented in Figure 7. Qualitative patterns are similar. We see that retail shares predict the cross-section of both overnight returns and intraday returns. We also see that the magnitude is very large: Top RS decile portfolio has a return gap that is more than 16% larger than that of the bottom RS decile portfolio.

[8] Empirically, both of these retail intensity-related variables are quite persistent.



(a) Sort on RS at Open

(b) Sort on RNB at Open

Figure 7. Forward Monthly Overnight and Intraday Portfolio Returns

This figure shows the average monthly overnight and intraday returns for equal weighted decile portfolios. Stocks are assigned to decile portfolios based on their monthly retail volume share and morning retail net buy in the first and second panels respectively. Daily morning net buy is computed by summing up all retail net buys that occurs during the first 30 minutes of a trading day. Monthly morning net buy is computed by summing up all morning retail net buy for a given month. Returns for the subsequent month are used. 95% confidence bands are constructed using bootstrapping with 1,000 bootstrap samples.

In Tables 3 and 4, we present the same results from above together with t-statistics based on Newey-West robust standard errors with 6-month lags.

3.2 Panel Regressions

We next examine the contemporaneous relationship in a panel regression setup. We start with a naive specification with stock-month-level controls and month fixed effects. The regression specification is as follows:

$$\mathrm{RG}_{it} = \beta \cdot \mathrm{RS}_{it} + \Gamma' \mathbf{X}_{it} + \gamma_t + \varepsilon_{it} \tag{17}$$

where RG_{it} is the overnight-intraday return gap for stock i during month t and RS_{it} is the retail volume share for stock i during month t, and γ_t is the month fixed effect. Controls $\mathbf{X}_{it}$ include size, book-to-market, Amihud measure, and foreign ownership. The exact definitions are provided in Table **??**.

Coefficient estimates are shown in Table 5. These results simply demonstrate simple correlations as we do not instrument for any of the flow-related variables. Coefficient estimates for β, shown in the fourth row, are positive and significant for columns (2) and (6). The estimate in column (6) indicates that a 1pp increase in retail volume share is associated with a 0.33pp increase in the monthly return gap. We also see that this comes from both the overnight and intraday legs.

Table 3. Monthly Returns for Deciles Portfolios Sorted on Retail Share

	Low	2	3	4	5	6	7	8	9	H	H-L
Overnight	-0.05	1.19	2.75	3.91	4.61	5.40	5.96	6.49	7.24	9.00	9.06
t-stat	(-0.20)	(3.91)	(7.40)	(9.31)	(9.68)	(11.47)	(11.78)	(10.81)	(12.60)	(16.65)	(14.77)
Intraday	1.08	-0.05	-1.02	-2.04	-2.50	-2.85	-3.23	-3.59	-4.63	-6.40	-7.48
t-stat	(4.93)	(-0.25)	(-3.76)	(-6.64)	(-7.48)	(-8.25)	(-8.78)	(-10.26)	(-11.72)	(-13.66)	(-16.41)
Ret Gap	-1.13	1.24	3.77	5.96	7.11	8.25	9.20	10.08	11.87	15.40	16.53
t-stat	(-3.30)	(4.07)	(7.94)	(10.81)	(11.13)	(12.63)	(13.43)	(13.71)	(14.60)	(19.56)	(17.74)

This table reports the monthly forward overnight returns, intraday returns, and return gaps for the decile portfolios constructed as in Figure 7a. *t*-statistics are computed using Newey-West robust standard errors with 6-month lags and are shown in parentheses.

Table 4. Monthly Returns for Deciles Portfolios Sorted on Retail Net Buy Near Open

	Low	2	3	4	5	6	7	8	9	H	H-L
Overnight	5.93	4.89	4.34	3.71	3.78	3.93	3.67	3.97	4.81	7.38	1.45
t-stat	(12.30)	(10.45)	(9.45)	(8.42)	(7.64)	(8.56)	(9.47)	(13.07)	(12.75)	(17.61)	(4.54)
Intraday	-3.96	-2.93	-2.07	-1.50	-1.35	-1.54	-1.66	-1.77	-2.89	-5.48	-1.52
t-stat	(-11.18)	(-8.70)	(-7.10)	(-5.04)	(-4.20)	(-4.89)	(-6.63)	(-6.12)	(-9.87)	(-15.97)	(-5.34)
Ret Gap	9.89	7.82	6.41	5.21	5.13	5.48	5.33	5.75	7.70	12.86	2.97
t-stat	(14.10)	(12.54)	(11.58)	(9.02)	(7.69)	(8.57)	(10.78)	(15.85)	(16.46)	(23.39)	(5.69)

This table reports the monthly forward overnight returns, intraday returns, and return gaps for the decile portfolios constructed as in Figure 7b. *t*-statistics are computed using Newey-West robust standard errors with 6-month lags and are shown in parentheses.

Table 5. Contemporaneous Panel Regression Results

	Overnight Ret		Intraday Ret		Return Gap	
	(1)	(2)	(3)	(4)	(5)	(6)
Rtl Turnover	0.06***		-0.01***		0.07***	
	(0.00)		(0.00)		(0.00)	
Inst Turnover	-0.07*		0.24***		-0.32***	
	(0.04)		(0.04)		(0.07)	
RNB Near Open	1.92***		-1.78***		3.69***	
	(0.41)		(0.22)		(0.43)	
Retail Vlm Share		0.21***		-0.16***		0.37***
		(0.01)		(0.01)		(0.01)
Log Mkt Cap	-0.47***	0.85***	0.21***	-1.03***	-0.68***	1.88***
	(0.07)	(0.10)	(0.07)	(0.07)	(0.11)	(0.11)
Beta	1.63***	0.36	-2.49***	-1.01***	4.12***	1.37***
	(0.19)	(0.22)	(0.22)	(0.20)	(0.27)	(0.28)
B/M	0.01	-0.06	0.64***	0.35***	-0.64***	-0.41***
	(0.05)	(0.06)	(0.09)	(0.07)	(0.10)	(0.08)
Amihud	0.01**	0.01**	-0.00	-0.01*	0.02	0.02**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Foreign Ownership	0.62	5.76***	3.30***	-1.70***	-2.67***	7.46***
	(0.45)	(0.43)	(0.49)	(0.47)	(0.68)	(0.61)
Month FE	✓	✓	✓	✓	✓	✓
Observations	113784	113784	113784	113784	113784	113784
Overall R^2	0.22	0.07	0.02	0.04	0.14	0.09
Within R^2	0.18	0.04	0.00	0.00	0.09	0.03

This table reports the panel regression results using specification (17). Standard errors are clustered at the month level and shown in parentheses. *, **, and *** denote 10%, 5%, and 1% statistical significance respectively.

What is also interesting is the opposite signs on the retail and institutional turnovers (or vlm_{ijt}) in the same month. This relationship survives after controlling for directional flow (retail net buy near open) in the morning. We should expect such an opposite relationship if institutional volume tends to provide liquidity to the aggressive retail flows in the morning.

Consistent with Berkman et al. (2012) and Lu et al. (2023), high contemporaneous retail net buying near the open is associated with higher overnight returns. One may be concerned that "contemporaneous" is only loosely true here. During a given month, retail net buy near open may be positive on average, but may be negative particularly on days with low overnight returns. We show in Table A2 that this is not the case by repeating a similar regression at daily frequencies. However, we should also recall that the portfolio sorts showed a non-monotonic relationship between opening retail net buy and overnight returns in Section 3.1.

A somewhat surprising result is that the coefficient on size is positive in column (6). This appears at first counterintuitive given that Berkman et al. (2012) finds smaller stocks to have

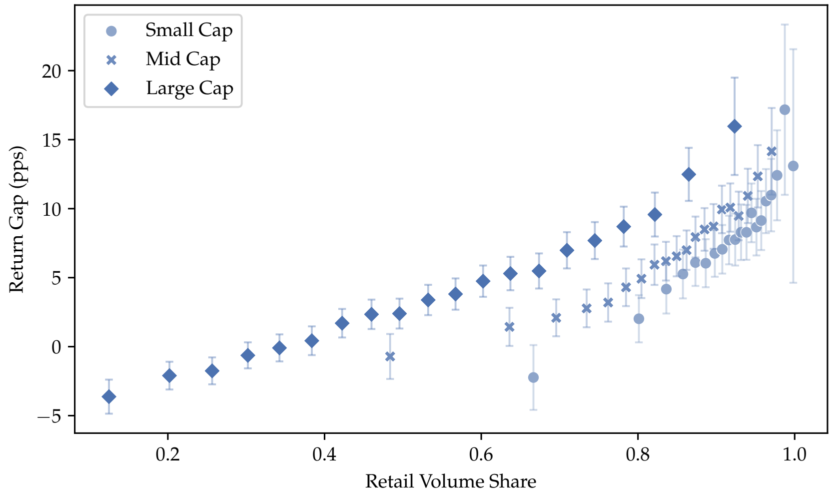


Figure 8. Binscatter of Return Gap against Volume Share

This figure shows a binscatter of the monthly return gap against the retail volume shares. Each observation is at the stock-month level. Stocks within each size tercile are assigned to 20 bins, and the bins are created so that each bin contains the same number of observations. 95% confidence bands are constructed using bootstrapping with 1,000 bootstrap samples.

higher overnight returns. The pattern is due to the fact that retail volume share is the main explanatory variable and size is (highly) negatively correlated with this variable. Figure 8 is revealing in this respect. In Figure 8, stock-month observations are sorted on the retail volume share and are assigned into 20 bins of equal sizes. A binscatter plot is done using this binning, and this processed is repeated separately for different size terciles. We see that the return gap increases with the retail volume share within each size tercile, but also that smaller stocks have a high retail volume share in aggregate.

Next, we repeat the regression based on specification (17), but use the return variables in the next month as the outcome variables:

$$\mathrm{RG}_{it+1} = \beta \cdot \mathrm{RS}_{it} + \Gamma' \mathbf{X}_{it} + \gamma_t + \varepsilon_{it+1}. \tag{18}$$

Table 6 presents the results. Possibly due to the persistence of the flow-related variables, the qualitative patterns of the coefficients remain unchanged from the contemporaneous regressions. High retail trading volume share in the current month predicts higher (lower) overnight (intraday) returns in the following month. The estimate in column (6) shows that a 1pp increase in retail volume share is associated with a 0.31pp increase in the monthly return gap. Results are similar if we use Fama-MacBeth regressions instead (see Table A4).

Table 6. Predictive Panel Regression Results

	Overnight Ret		Intraday Ret		Return Gap	
	(1)	(2)	(3)	(4)	(5)	(6)
Rtl Turnover	0.02***		-0.03***		0.05***	
	(0.00)		(0.00)		(0.00)	
Inst Turnover	-0.05**		0.04*		-0.10**	
	(0.03)		(0.03)		(0.04)	
RNB Near Open	-0.11		0.20*		-0.31	
	(0.18)		(0.12)		(0.22)	
Retail Vlm Share		0.15***		-0.16***		0.31***
		(0.01)		(0.01)		(0.01)
Stock Controls	✓	✓	✓	✓	✓	✓
Month FE	✓	✓	✓	✓	✓	✓
Observations	113784	113784	113784	113784	113784	113784
Overall R^2	0.05	0.05	0.05	0.04	0.08	0.07
Within R^2	0.02	0.01	0.02	0.01	0.03	0.01

This table reports the panel regression results using specification (18). Standard errors are clustered at the month level and shown in parentheses. *, **, and *** denote 10%, 5%, and 1% statistical significance respectively.

3.3 Number of Shares Outstanding as an IV

While the results in Table 6 are from predictive regressions, so that the return gap and retail share are not simultaneously determined, they may still be subject to omitted variables bias. Consider, for example, hard-to-value stocks. These stocks may attract a lot of retail trading, which persists, (Laarits and Sammon 2022) and also experience large price swings that lead to large return gaps. Thus, we need an exogenous source of variation in retail shares.

We suggest a way to use the number of outstanding shares as an instrument for retail activity. The per-share price of a stock should not carry economic meaning. At the same time, various firms have gone through stock splits in hopes of attracting more active trading by retail investors. The usual rationale is that a small retail investor will have difficulties trading a stock with a large per share price, exceeding hundreds of dollars, in absence of fractional shares trading.[9] This means that after controlling for market capitalization, higher number of outstanding shares (i.e., lower per-share price) should induce more retail trading.

The second panel of Figure 9 shows a binscatter plot of retail volume share against the log number of outstanding shares. We see that without controlling for market capitalization, there is no meaningful relationship between number of outstanding shares and retail activity. The third panel of Figure 9 repeats the binscatter plot with residuals of retail volume, after controlling for log market capitalization, on the y-axis instead. We take two different monthly cross-sections,

[9] During our sample period, fractional trading was not readily available for Korean retail investors.

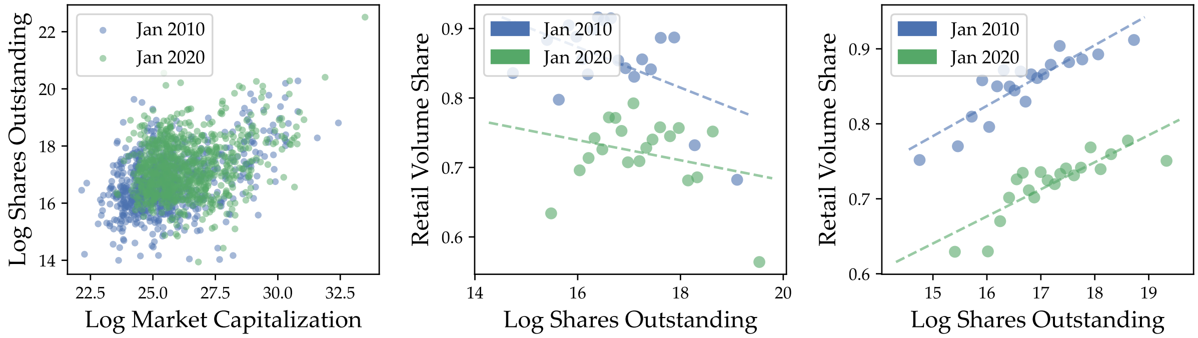


Figure 9. Binscatter of Retail Volume Share Against Log Shares Outstanding

The first panel is a scatter plot of log outstanding shares against log market capitalization. The second and third panels are binscatter plots between retail volume share and log outstanding shares. 20 bins of log share outstanding are used for both panels. The third panel uses the residuals from regressing retail volume share on a constant and log market capitalization on the y-axis. Two cross-sections, for the month January 2010 and the month January 2020, are used for all the panels.

one from January 2010 and another from January 2020, and use 20 bins of log shares outstanding. We can see that monthly retail volume share increases monotonically with log average shares outstanding in the same month as expected.

In practice, we need enough variation in number of shares outstanding across stocks with similar market capitalizations. In the first panel of Figure 9, we provide a scatter plot of all stocks in our dataset, where we put the stocks' log market capitalization on the x-axis and their log number of outstanding shares on the y-axis. Here, we can see that there is substantial variation in log outstanding shares for a given level of log market capitalization.

More specifically, we take the average end-of-the-day outstanding shares within a month, and then take logs. There are a few stock-month observations during which stock splits or large share buybacks occur. These observations will experience large changes in shares outstanding. For simplicity, we omit stock-month observations that include a daily change of 10% or larger in outstanding shares.

The relevance condition is satisfied. The instrument is strong with partial F-stats exceeding 1,000. It is worth noting that the instrument is valid only conditional on the log market capitalization. The exclusion restriction requires that the number of outstanding shares is not correlated with fundamentals, news, or sentiment, after controlling for other explanatory variables. In theory, number of outstanding shares should not carry an economic meaning.[10]

Table 7 presents the results from specification (18) after instrumenting for RS with log number of outstanding shares. OLS coefficients are provided for reference.[11] We see that the qualitative

[10] A lingering concern may be that our approach is equivalent to using "per-share price" as an instrument. This means that the design may be somehow affected by momentum or reversals.

[11] The slight difference in estimates compared to Table 6 is due to omitted stock-months during which stock splits occurred.

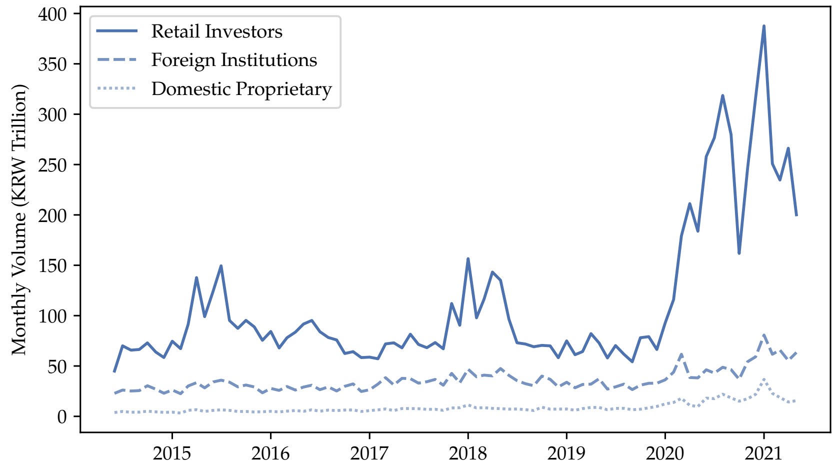


Figure 10. Monthly Volume by Investor Type

This figure shows the monthly trading volume of different investor types in terms of KRW trillions. The investor types considered are retail investors, foreign institutions, and domestic proprietary trading firms.

results are the same for all three outcome variables. We also see that the 2SLS coefficients are larger. One possibility is the following: Arbitrageurs (both institutional and retail) understand this phenomenon and chase naive retail investors in an attempt to take advantage of the return gap. Then, without an IV the OLS coefficient may be biased downward because of the shrinking return gap.

3.4 DID Using the Pandemic-Induced Retail Frenzy

In addition to the proposed IV, we also devise a DID regression to exploit the pandemic-induced retail trading frenzy. Similar to retail investors in the US (Ozik et al. 2021), possibly due to stimulus checks and stay-at-home orders, retail investors in Korea started trading in the stock market much more actively. This jump in retail volume is easily noticeable in Figure 10.

Suppose there are two stocks (A and B), and two agents (the representative retail investor and the representative institution). Also, suppose that the retail (institutional) investor allocates α ($\hat{\alpha}$) share of volume to asset A. If retail and institutional volume are V^r and V^i respectively, the retail volume share of assets A and B are as the following:

$$\mathrm{RS}_A = \frac{\alpha V^r}{\alpha V^r + \hat{\alpha} V^i} \qquad \mathrm{RS}_B = \frac{(1-\alpha)V^r}{(1-\alpha)V^r + (1-\hat{\alpha})V^i}. \tag{19}$$

Consider a scenario where αV^r is large relative to $\hat{\alpha} V^i$ so that $\mathrm{RS}_A \approx 1$. In other words, A is a retail-dominated stock. When an exogenous shock to V^r arrives, such as the retail frenzy following the COVID-19 stock market crash, we can see that RS_A will not increase significantly due to the fact that *shares* are capped at 1. On the other hand, RS_B increases more significantly,

Table 7. IV Regression Results

	Retail Share	Overnight Ret		Intraday Ret		Return Gap	
	1st Stage	OLS	2SLS	OLS	2SLS	OLS	2SLS
Retail Vlm Share		0.148***	0.226***	-0.159***	-0.206***	0.307***	0.432***
		(0.003)	(0.010)	(0.003)	(0.009)	(0.004)	(0.014)
Log Shares Oustanding	4.093***						
	(0.040)						
Stock Controls	✓	✓	✓	✓	✓	✓	✓
Month FE	✓	✓	✓	✓	✓	✓	✓
Observations	112184	112184	112184	112184	112184	112184	112184

This table reports the 2SLS regression results using specification (18) after instrumenting retail volume shares with log outstanding shares. Heteroskedasticity-robust standard errors are shown in parentheses. *, **, and *** denote 10%, 5%, and 1% statistical significance respectively.

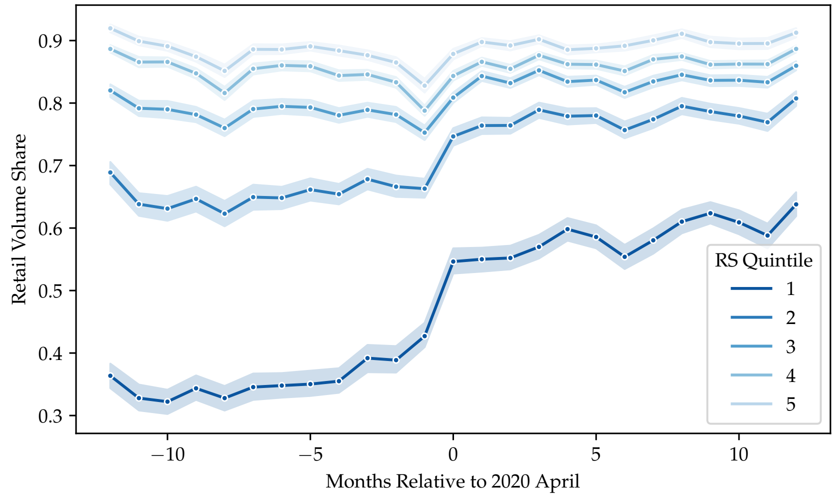


Figure 11. Retail Shares Over Time by Initial RS Quintiles

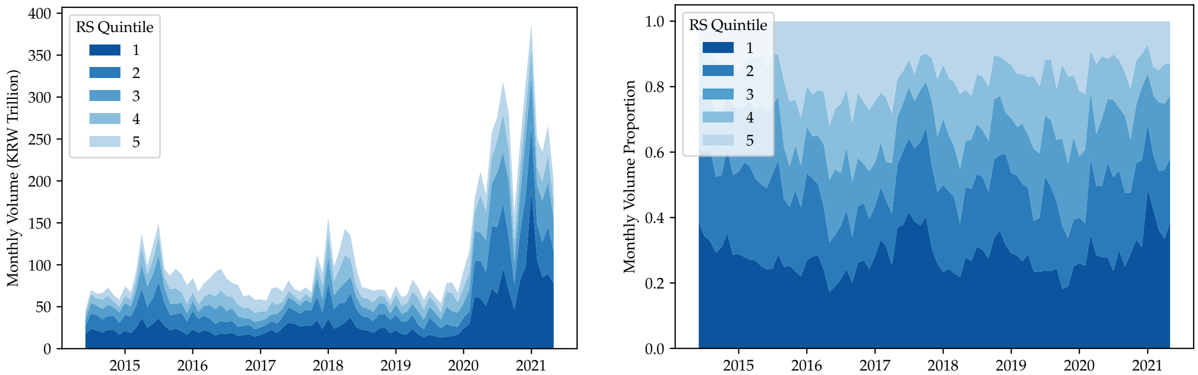
This figure shows the average retail investor volume shares for stocks within each initial retail share quintile. Stocks are first assigned into quintile portfolios based on their average retail volume share during the years 2014 to 2018. Then, average monthly retail volume shares are computed for each quintile group and plotted. 95% confidence bands are constructed using bootstrapping with 1,000 bootstrap samples.

assuming $\text{RS}_B \ll 1$.

Figure 11 demonstrates this idea empirically. Here, we compute each stock's average daily retail volume share during the years 2014 to 2018. Then, we sort stocks into quintiles based on this initial retail volume shares. We next compute monthly retail volume shares for each stock in the months around the COVID-19 shock. The average retail volume share for different quintiles are plotted in Figure 11. We see that the 5th quintile is always close to 1, while the 1st quintile experiences a significant increase around the COVID-19 shock. When the plausibly exogenous shock to aggregate retail trading intensity arrived with the pandemic only stocks whose retail volume shares were sufficiently away from 1 experienced significant increases in the retail volume share.

The implicit assumption behind the argument related to equation (19) is that the allocation of retail volume is somewhat stable over time. Figure 12 provides some support for this assumption. We see that even after the large jump in total retail volume shown in Figure 12a, the proportion of trading volume flowing to different initial retail share quintiles shown in Figure 12b is quite stable.

The initial retail shares are computed using the months between June 2014 and December 2018. Then, the *control* group is defined as the stocks in the highest initial retail share quintile. The period from January 2019 to December 2019 is defined as the *pre* period and the period from April 2020 to March 2021 is defined as the *post* period. The first quarter of 2020 is dropped. The



(a) Volume by RS Quintile

(b) Volume Proportion by RS Quintile

Figure 12. Monthly Retail Volume by RS Quintile

This figure shows the total monthly retail investor volume for each RS quintile in terms of KRW trillions and in terms of proportions. Stocks are assigned into quintile portfolios based on their current month's retail volume share and retail trading volume is summed within each quintile.

DID specification is as follows:

$$\text{RG}_{it} = \gamma_1 \cdot \text{Treated}_{it} + \gamma_2 \cdot \text{Post}_{it} + \beta \cdot \text{Treated}_{it} \times \text{Post}_{it} + \Gamma' \mathbf{X}_{it} + \alpha_i + \gamma_t + \varepsilon_{it} \tag{20}$$

where Treated_{it} and Post_{it} are defined as discussed above. Controls $\mathbf{X}_{it}$ include size, book-to-market, Amihud measure, and foreign ownership. Stock fixed effects α_i and month fixed effects γ_t are included. In specifications with stock and month fixed effects, the treated and post dummies are absorbed.

The validity of our DID strategy relies on the assumption that overnight and intraday returns for the treatment and control groups would have trended similarly without the arrival of the pandemic. One major concern is that retail volume shares are not randomly assigned. Table 2 shows that this quantity is significantly correlated with other variables. While we cannot resolve this issue with our current design, we show in an event study plot that there are no noticeable pre-trends in the control and treatment groups. In Figure A4, we plot the cross-sectional average of the monthly return gaps for the treatment and control groups separately.

Results from specification (20) are shown in Table 8. The coefficient estimates of interest are β's which are displayed in the third row. We see that the estimates are positive and significant in columns (1) and (3). Also, the estimates for γ_1 and γ_2 displayed in first and second rows are also consistent with our priors. As the treatment groups consist of stocks with lower initial retail volume share, we expect the estimates for γ_1 to be negative. Also, because retail trading increased in general following the outbreak of the pandemic, we expect the estimates for γ_2 to be positive.

However, we take the DID results with a grain of salt and only provide them as a supplement

Table 8. DID Regression Results

	Return Gap			Overnight	Intraday
	(1)	(2)	(3)	(4)	(5)
Treated	-5.03***	-5.44***			
	(0.91)	(1.14)			
Post	-0.09				
	(0.55)				
Treated × Post	3.86***	4.23**	4.14**	1.28	-1.87*
	(0.79)	(1.66)	(1.73)	(1.26)	(1.09)
Stock Controls	✓	✓	✓	✓	✓
Stock FE	–	–	✓	✓	✓
Month FE	–	✓	✓	✓	✓
Observations	8543	8543	8543	8543	8543
Within R^2	0.04	0.04	0.06	0.02	0.05

This table reports the DID regression results using specification (20). In columns (3), (4), and (5), standard errors are two-way clustered at the stock and month level. *, **, and *** denote 10%, 5%, and 1% statistical significance respectively.

to our IV results. While we believe that the pandemic-induced increase in retail activity is an exogenous shock, we admit that there are still issues with making a strong causal claim. First, we must truly believe that it is the *retail volume share* that drives the return gap, rather than any general measure of *retail activity*. This is because our strategy also relies on the fact that shares are mechanically capped at 1.

Another issue with this specification is the difficulty of interpretation. As opposed to finding an instrument for retail volume share, this artificial treatment-control setup deters us from drawing a quantitative interpretation at a per unit-basis. We can only comment that the treatment group experienced a larger average differential increase in return gap relative to the change in return gap for the control group. As a back-of-the-envelope calculation, we see that the estimate for beta β is approximately 4%. When the shock arrived, retail shares for the treated group increase by approximately 20%. A naive calculation leads to an estimate of around 0.2% increase in monthly return per a 1pp increase retail share, which is comparable to the panel regression estimates.

4 Mechanisms

4.1 Retail Demand for Daytime Stock Market Exposure

Figure 2 and Figure 5 together show how prices may move up together with the retail net buy in the morning and move back down in the afternoon. Given that volatility and illiquidity follow a U-shaped pattern throughout the day (Hong and Wang 2000), retail investors may be knowingly

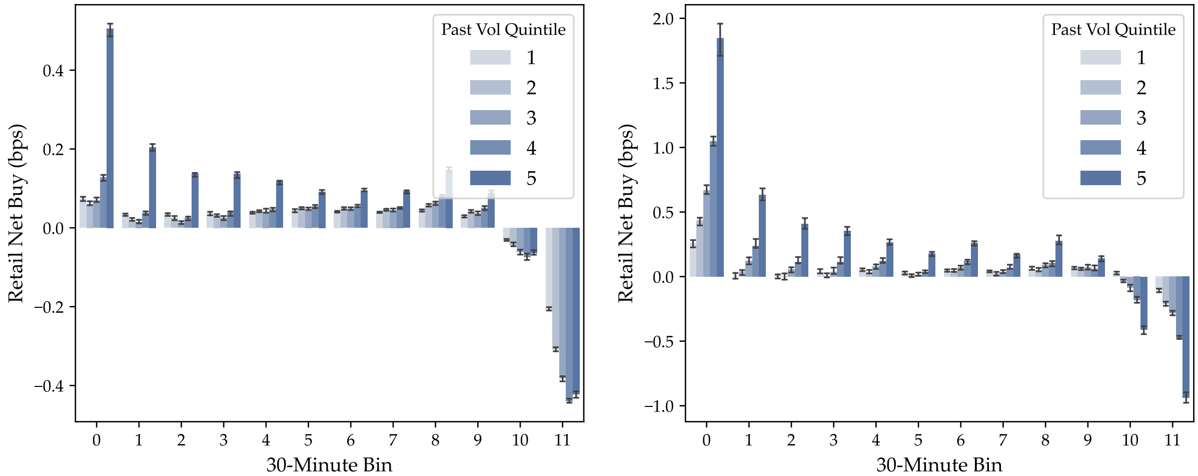


Figure 13. Intraday Retail Net Buy by Past Volatility Quintiles

This figure shows the average retail net buy during different 30-minute bins of the trading day. Averages are taken across stock-days within the same past volatility quintiles. Retail net buy is normalized by shares outstanding and is expressed in basis points. The left panel shows results for the pre-pandemic period and the right panel shows those for the post-pandemic period. 95% confidence bands are constructed using bootstrapping with 1,000 bootstrap samples.

flowing into the stock market in times of high volatility.

This volatility-chasing by retail investors can also be found in the cross-section. For each stock-day observation, we compute the volatility of daily returns during the past 20 trading days. Also, for each stock-day, retail net buy for different 30-minute intervals are computed. Then, 30-minute retail net buys are averaged across stock-days within the past volatility quintiles. This exercise is conducted for the pre-pandemic (left panel), up to March 2020, and post-pandemic (right panel) periods separately. Figure 13 plots these average retail net buys at different 30-minute bins. We see that retail investors more heavily net buy stocks with high past volatility in the morning, while they net sell those stocks in the afternoon. We also see that this pattern becomes stronger—notice the larger magnitude—during the post-pandemic period.

The inflow into volatile stocks may not be surprising given retail investors' preference for lottery-like stocks (Bali et al. 2011).[12] However, the ensuing outflow in the afternoon is more surprising. This suggests that retail investors want this exposure only during the day.

If this is driven by the desire to "enjoy the ride" of a volatile stock market, we expect a stronger pattern among investors who mainly trade through the mobile trading system (MTS) as opposed to the desktop home trading system (HTS). Mobile apps may be more suited for treating the stock market as a diversion, and MTS users tend to be less sophisticated than dedicated HTS users. Our brokerage data also specifies the order medium so that we can aggregate the MTS and HTS net flows separately. Using this dataset, we run a panel OLS regression with day fixed

[12] See Figure A5 for a sort on past max returns instead. Skewness, volatility, and max returns are all known to be correlated, and we can construct similar figures from any of these three variables.

Table 9. Retail Net Buy Panel Regression Results

	Morning RNB		Afternoon RNB	
	(1) MTS	(2) HTS	(3) MTS	(4) HTS
Past Vol	0.983***	-0.750***	-0.323***	0.488***
	(0.110)	(0.208)	(0.068)	(0.155)
Day FE	✓	✓	✓	✓
Observations	674288	674288	674288	674288
Overall R^2	0.001	0.000	0.000	0.000
Within R^2	0.000	0.000	0.000	0.000

This table reports the diff-in-diff regression results using specification (20). Standard errors are clustered at the day level. *, **, and *** denote 10%, 5%, and 1% statistical significance respectively.

effects:

$$\text{RNB}_{it}^{m,h} = \beta \cdot \text{Past Vol}_{it} + \Gamma' \mathbf{X}_{it} + \gamma_t + \varepsilon_{it} \tag{21}$$

We run separate regressions for $m \in \{\text{MTS}, \text{HTS}\}$ net buy for the $h \in \{\text{first 1hr}, \text{last 1hr}\}$ of the trading day, which results in four cases. Controls $\mathbf{X}_{it}$ include size, book-to-market, Amihud measure, and foreign ownership for stock i. Table 9 shows the coefficient estimates. We can see that, on average, MTS users chase (shun) volatility in the morning (afternoon), while HTS users take the opposite side of these trades.[13] The extremely low R^2 is due to the fact that we only have trades from a sample of 50,000 traders in our brokerage dataset.

Relatedly, we also see that MTS users log in more, relative to their own average log-in frequencies, on days when they have positive positions. This is also consistent with users coming back to check the movement of their positions' values throughout the day. Table 10 shows the coefficients from the following specification:

$$\begin{aligned} \ln(1 + \text{\# Log-In})_{ijd} &= \alpha_i + \delta_j + \gamma_d \\ &+ \beta_1 \cdot \mathbf{1}\{\text{Balance}_{ijd} > 0\} + \beta_2 \cdot \mathbf{1}\{\text{Balance}_{ijd} > 0\} \times \mathbf{1}\{j = \text{MTS}\} \end{aligned} \tag{22}$$

User i, whose main medium is j, is said to have a positive balance if she has any positive stocks holdings as of 9:30 a.m. of day d. The log-ins on day d are counted only if they happen after 9:30 a.m. of day d. This precludes the counting of day d log-ins in order to make trades that lead up to a positive balance on day d. We see that on days with positive balance, users log in more frequently, and also that this pattern is more prevalent among MTS users. A back-of-the-envelope calculation, which is very approximate due to fixed effects, for the magnitude of the effect is the following: Given an unconditional log-in frequency of 4 and the $\ln(1 + y)$ transformation, estimates of around 0.5 suggests approximately 3 more daily log-ins happening.

Finally, we find that this pattern is stronger during times with abundant liquidity and high

[13] Individuals whose main trading medium is the HTS tend to be more dedicated traders who are more sophisticated.

Table 10. Log-In Panel Regression Results

	ln (1 + # Log-In)		
	(1)	(2)	(3)
Positive Balance	0.898***	0.464***	0.537***
	(0.006)	(0.000)	(0.008)
MTS		-0.006***	
		(0.000)	
Pos. Bal. × MTS		0.390***	0.584***
		(0.000)	(0.011)
User FE	✓	–	✓
Observations	43734942	43734942	43734942
Overall R^2	0.111	0.109	0.122
Within R^2	0.111	0.116	0.122

This table reports the panel regression results using specification (**??**). In columns (1) and (3), standard errors are clustered at the user level and are shown in parentheses. *, **, and *** denote 10%, 5%, and 1% statistical significance respectively.

consumer sentiment. Here we repeat the regression as in specification (21) using aggregate retail net buy as the outcome variable. We also add the aggregate retail brokerage deposits and consumer sentiment on the RHS. We run separate regressions for the $h \in \{\text{first 1hr}, \text{last 1hr}\}$ of the trading day.

$$\text{RNB}_{it}^{h} = \Gamma' \mathbf{X}_{it} + \beta \cdot \text{Past Vol}_{it} + \gamma_t + \varepsilon_{it} \tag{23}$$

Results are shown in Table 11. The interaction terms are of main interest. We see that during times of abundant liquidity and high consumer sentiment, morning (afternoon) retail net buy flow into (out of) more volatile stocks. This suggests that retail investors are forced to tame such behavior when they are more financially constrained (Heimer and Imas 2022).

4.2 Other Mechanisms and Previous Studies

The positive relationship between retail trading intensity and the return gap is more apparently consistent with studies such as Berkman et al. (2012), Lou et al. (2019), or Jones et al. (2022). However, our paper does not rule out other relevant sources of intraday retail patterns or the return gap. In fact, the same relationship is implied by mechanisms that are suggested by various other papers.

For example, Lu et al. (2023) offer an explanation based on heteroegeous liquidity providers. If there are fast arbitrageurs who can better identify uninformed trades, they will offer a higher price when there uninformed buy orders are prevalent. Later on, slow arbitrageurs who have lower inventory costs offer better prices and there is a reversion of prices even though the order imbalance stays buy-heavy throughout the day. Even under this explanation, the authors'

Table 11. Retail Net Buy Panel Regression Results with Sentiment

	Morning RNB		Afternoon RNB	
	(1)	(2)	(3)	(4)
Past Vol	0.220***	0.234***	-0.109***	-0.090***
	(0.011)	(0.011)	(0.008)	(0.008)
Deposits	0.021***		-0.000	
	(0.002)		(0.001)	
Sentiment	0.006**		-0.003**	
	(0.003)		(0.002)	
Vol $\times$ Deposit	0.009***	0.010***	-0.006***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)
Vol $\times$ Sentiment	0.004***	0.004***	-0.002**	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Day FE	–	–	✓	✓
Observations	2173985	2173985	2175895	2175895
Overall R^2	0.007	0.005	0.004	0.003
Within R^2	0.006	0.005	0.005	0.004

This table reports the panel regression results using specification (23). In columns (3) and (4), standard errors are clustered at the day level and are shown in parentheses. *, **, and *** denote 10%, 5%, and 1% statistical significance respectively.

model will give the same price predictions if the order imbalance in fact flips from buy to sell throughout the day. In fact, we do find the flip if we look at the aggressive (aggregate) retail flows.

A different source can be the "overnight risk" channel discussed in Bogousslavsky (2021). Due to illiquidity and risk of large price moves, traders may be averse to having open positions at the end of the trading day. If retail investors are more averse to this overnight risk, our results are also consistent with this explanation.

Another idea that has not been discussed in previous sections is a mood-based explanation. Birru (2018) finds that speculative assets rise on Mondays and come back down throughout the week, and this is attributed to the rise and fall of investor mood. Similarly, retail investors may have a high mood or sentiment during the beginning of the day, which subsides over time. Circling back to our findings, stocks with higher retail investor intensity will have a larger mass of investors who are prone to this behavior.

5 Conclusion

We study the persistent and large gap between overnight and intraday returns in stock markets. Studies have pointed to both attention-induced trading and extrapolative trading—behavioral patterns often related to retail investors—in the morning as possible drivers of this global phenomenon. Still, retail flow data for the US has been either partial (e.g., data from a single brokerage) or imputed (e.g., Boehmer et al. 2021). Furthermore, studies on the overnight-intraday return decomposition often fall short of establishing causal relationships due to the lack of exogenous variation in retail trading intensity.

We bring in a new dataset that resolves the data problem and an IV to address the identification challenge. Our retail flow data are accurate and exhaustive because KRX directly provides investor type-level identifiers at high frequencies. We propose to use the number of outstanding shares, or equivalently per-share price, as an instrument for retail activity. With these contributions, we confirm that more intense retail trading causes the overnight-intraday return gap to widen at the monthly and daily frequencies. Furthermore, we provide suggestive evidence that this relationship stems from the ebb and flow of retail net buy; this daily pattern appears to arise from retail demand for daytime stock market exposure.

References

Aboody, D., Even-Tov, O., Lehavy, R., and Trueman, B. (2018). Overnight returns and firm-specific investor sentiment. *Journal of Financial and Quantitative Analysis*, 53(2):485–505.

Akbas, F., Boehmer, E., Jiang, C., and Koch, P. D. (2022). Overnight returns, daytime reversals, and future stock returns. *Journal of Financial Economics*, 145(3):850–875.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56.

Bali, T. G., Cakici, N., and Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, 99(2):427–446.

Barber, B. M., Huang, X., Jorion, P., Odean, T., and Schwarz, C. (2022). A (sub)penny for your thoughts: Tracking retail investor activity in TAQ.

Barber, B. M. and Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, 55(2):773–806.

Barber, B. M., Odean, T., and Zhu, N. (2009). Do retail trades move markets? *The Review of Financial Studies*, 22(1):151–186.

Berkman, H., Koch, P., Tuttle, L., and Zhang, Y. (2012). Paying attention: Overnight returns and the hidden cost of buying at the open. *Journal of Financial and Quantitative Analysis*, 47(4):715–741.

Birru, J. (2018). Day of the week and the cross-section of returns. *Journal of Financial Economics*, 130(1):182–214.

Boehmer, E., Jones, C. M., Zhang, X., and Zhang, X. (2021). Tracking retail investor activity. *The Journal of Finance*, 76(5):2249–2305.

Bogousslavsky, V. (2021). The cross-section of intraday and overnight returns. *Journal of Financial Economics*, 141(1):172–194.

Boyarchenko, N., Larsen, L. C., and Whelan, P. (2023). The overnight drift. *The Review of Financial Studies*, page hhad020.

Branch, B. S. and Ma, A. J. (2006). The overnight return, one more anomaly.

Branch, B. S. and Ma, A. J. (2015). Overnight return, the invisible hand behind intraday returns?

Easley, D., O'Hara, M., and Saar, G. (2001). How Stock Splits Affect Trading: A Microstructure Approach. *Journal of Financial and Quantitative Analysis*, 36(1):25–51.

Fama, E. F. and French, K. R. (2004). The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives*, 18(3):25–46.

Green, T. C. and Hwang, B.-H. (2009). Price-based return comovement. *Journal of Financial Economics*, 93(1):37–50.

Grinblatt, M. and Keloharju, M. (2009). Sensation Seeking, Overconfidence, and Trading Activity. *The Journal of Finance*, 64(2):549–578.

Heimer, R. Z. and Imas, A. (2022). Biased by Choice: How Financial Constraints Can Reduce Financial Mistakes. *The Review of Financial Studies*, 35(4):1643–1681.

Hendershott, T., Livdan, D., and Rösch, D. (2020). Asset pricing: A tale of night and day. *Journal of Financial Economics*, 138(3):635–662.

Heston, S. L., Korajczyk, R. A., and Sadka, R. (2010). Intraday patterns in the cross-section of stock returns. *The Journal of Finance*, 65(4):1369–1407.

Hong, H. and Wang, J. (2000). Trading and returns under periodic market closures. *The Journal of Finance*, 55(1):297–354.

Jones, C. S., Pyun, S., and Wang, T. (2022). Return extrapolation and day/night effects.

Kelly, M. A. and Clark, S. P. (2011). Returns in trading versus non-trading hours: The difference is day and night. *Journal of Asset Management*, 12(2):132–145.

Knuteson, B. (2020). Strikingly suspicious overnight and intraday returns.

Kumar, A. and Lee, C. M. (2006). Retail investor sentiment and return comovements. *The Journal of Finance*, 61(5):2451–2486.

Laarits, T. and Sammon, M. (2022). The retail habitat.

Lou, D., Polk, C., and Skouras, S. (2019). A tug of war: Overnight versus intraday expected returns. *Journal of Financial Economics*, 134(1):192–213.

Lu, Z., Malliaris, S., and Qin, Z. (2023). Heterogeneous liquidity providers and night-minus-day return predictability. *Journal of Financial Economics*, 148(3):175–200.

Ozik, G., Sadka, R., and Shen, S. (2021). Flattening the illiquidity curve: Retail trading during the covid-19 lockdown. *Journal of Financial and Quantitative Analysis*, 56(7):2356–2388.

Savor, P. and Wilson, M. (2014). Asset pricing: A tale of two days. *Journal of Financial Economics*, 113(2):171–201.

Appendix A Extra Figures

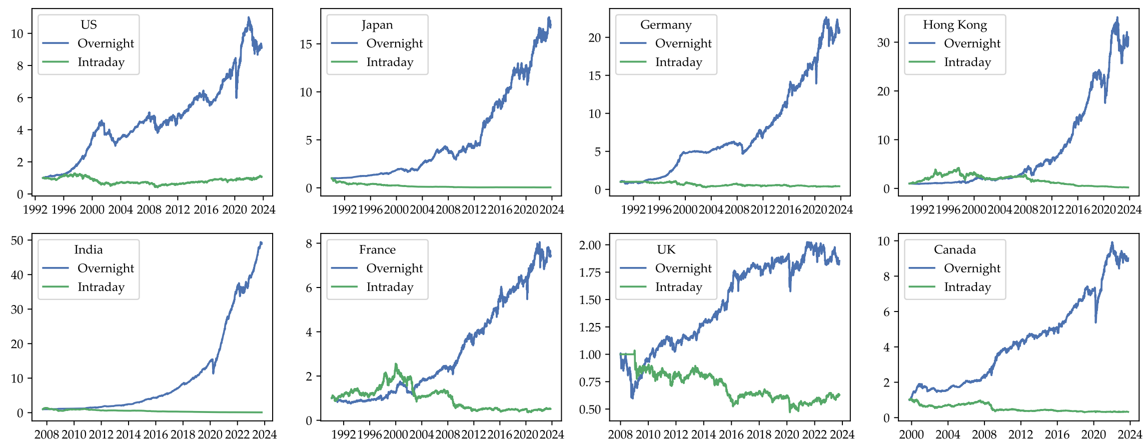
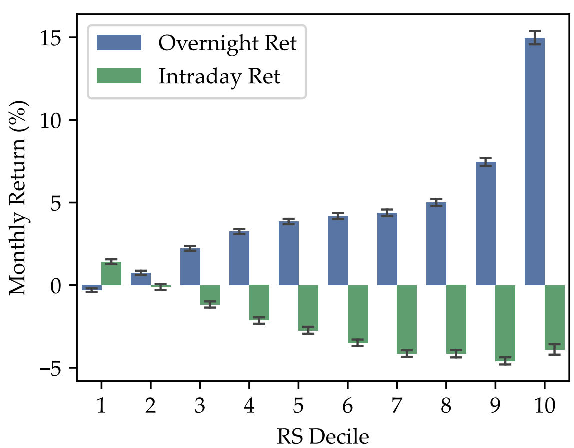
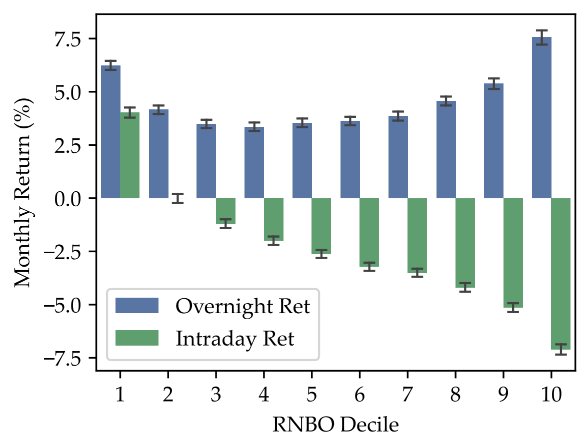


Figure A1. Cumulative Overnight and Intraday Returns

This figure shows cumulative returns for overnight and intraday returns of various stock market indexes around the world. We use the opening and closing prices of the respective stock market index ETFs. A similar figure is provided in Knuteson (2020).



(a) Sort on RS



(b) Sort on RNB

Figure A2. Monthly Overnight and Intraday Portfolio Returns

This figure repeats the same exercise described in Figure 6 with retail volume share and retail net buy during the entire day instead.

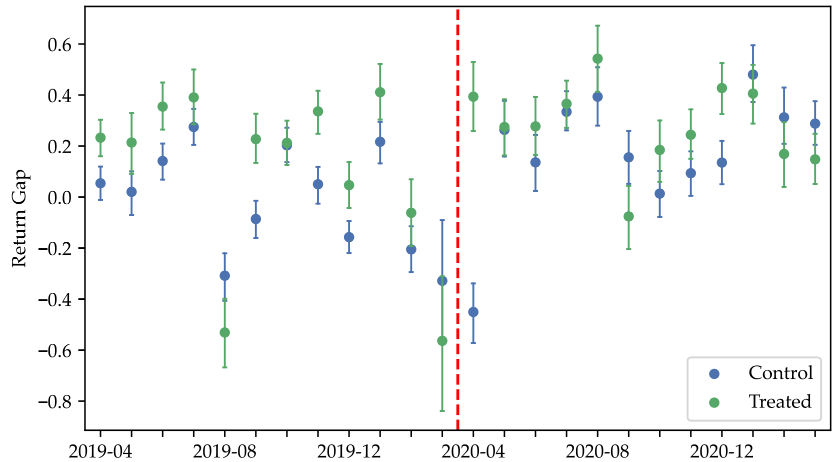
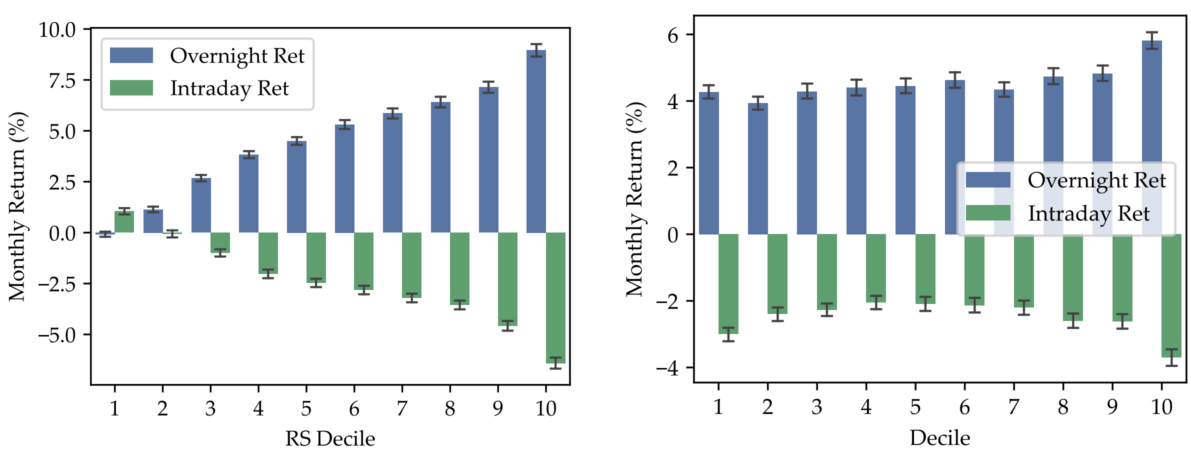


Figure A4. Monthly Return Gaps of the Treated and Control Groups

This figure shows the average monthly return gaps of the stocks in the treated and control groups. Treated group includes stocks in the first quintile of initial retail volume shares and control group includes stocks in the fifth quintile of the initial retail volume shares. 95% confidence bands are constructed using bootstrapping with 1,000 bootstrap samples.



(a) Sort on Lagged RS

(b) Sort on Lagged RNB

Figure A3. Monthly Overnight and Intraday Portfolio Returns

This figure repeats the same exercise described in Figure 7 with retail volume share and retail net buy during the entire day instead.

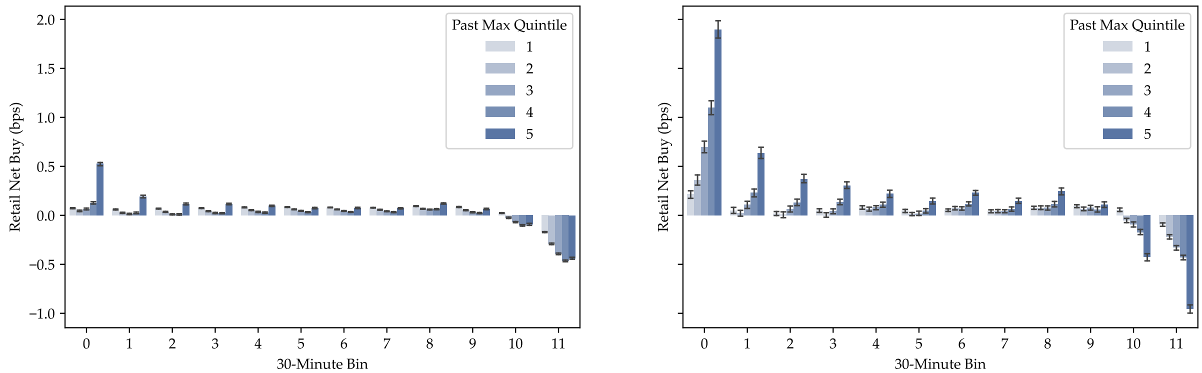


Figure A5. Intraday Retail Net Buy by Past Max Return Quintiles

This figure shows results from repeating the exercise in 13 after sorting on past max return instead. Retail net buy is normalized by shares outstanding and is expressed in basis points. 95% confidence bands are constructed using bootstrapping with 1,000 bootstrap samples.

Appendix B Extra Tables

Table A1. Variable Definitions

Variable Name	Definition
Retail Volume Share	Retail shares during any interval $t \to t+1$ is retail volume divided by total volume where the all volumes fall between $t \to t+1$.
Log Market Cap	Average closing market capitalization in KRW in the past 60 trading days, log transformed
Amihud Measure	We follow Amihud (2002) while (1) using daily volume in billions of KRW terms and (2) averaging across past 60 trading days
Market Beta	120-day rolling beta using the KOSPI return as the market return
Book-to-Market	Average book-to-market ratio in the past 60 trading days
Foreign Ownership	Latest trading day's value of shares held by foreign investors divided by shares outstanding, expressed in %
Past Volatility	Standard deviation of daily returns during the past 20 trading days
Value-Weighted Average Price	Total trading volume in KRW terms divided by total trading volume in # of shares
RS120	Retail shares between day -141 and day -21
RS20	Retail shares between day -21 and day -1
RS1	Retail shares on day -1
Retail (Inst) Turnover 120 Days	Retail (Inst) volume between day -141 and -21 divided by shares outstanding
RT20 (IT20)	Retail (Inst) turnover between day -21 and day -1 divided by RT120 (IT120)
RT1 (IT1)	Retail (Inst) turnover on day -1 divided by RT120 (IT120)

Table A2

	Ret Gap		Overnight Ret		Intraday Ret	
	(1)	(2)	(3)	(4)	(5)	(6)
Log Mkt Cap	-0.014***	0.112***	-0.017***	0.043***	-0.003	-0.069***
	(0.004)	(0.005)	(0.002)	(0.003)	(0.004)	(0.004)
Mkt Beta	0.298***	0.106***	0.157***	0.063***	-0.140***	-0.042***
	(0.019)	(0.019)	(0.012)	(0.012)	(0.016)	(0.016)
B/M	-0.066***	-0.033***	-0.027***	-0.012**	0.039***	0.022***
	(0.009)	(0.008)	(0.005)	(0.005)	(0.008)	(0.008)
Amihud	-0.083	0.006	0.037	0.099***	0.121**	0.093*
	(0.069)	(0.069)	(0.037)	(0.037)	(0.050)	(0.050)
Foreign Holdings	-0.245***	0.274***	-0.104***	0.146***	0.141***	-0.128***
	(0.030)	(0.027)	(0.018)	(0.016)	(0.025)	(0.023)
Log RT20		0.090***		0.047***		-0.043***
		(0.013)		(0.009)		(0.010)
Log IT20		-0.187***		-0.132***		0.055***
		(0.010)		(0.006)		(0.008)
Log RT1		0.266***		0.136***		-0.130***
		(0.011)		(0.007)		(0.008)
Log IT1		-0.098***		-0.002		0.096***
		(0.007)		(0.005)		(0.006)
Day FE	✓	✓	✓	✓	✓	✓
Observations	990116	990116	990116	990116	990116	990116
Overall R^2	0.003	0.011	0.002	0.010	0.001	0.004
Within R^2	-0.000	0.004	-0.000	0.005	0.000	0.001

This table reports the panel regression results using specification (17) with daily observations. Standard errors are shown in parentheses. *, **, and *** denote 10%, 5%, and 1% statistical significance respectively.

Table A3

	Ret Gap			
	(1)	(2)	(3)	(4)
RS120	-0.044	0.018	0.034	
	(0.039)	(0.059)	(0.059)	
RS20	0.561***	0.585***	0.601***	
	(0.044)	(0.063)	(0.063)	
RS1	0.862***	0.888***	0.870***	
	(0.031)	(0.039)	(0.039)	
Log RT20				0.093***
				(0.013)
Log IT20				-0.192***
				(0.010)
Log RT1				0.265***
				(0.011)
Log IT1				-0.096***
				(0.007)
RNB120			-0.354***	-0.398***
			(0.060)	(0.060)
RNB20			0.160***	0.116***
			(0.030)	(0.030)
RNB1			0.012	-0.001
			(0.016)	(0.015)
Stock Controls	✓	✓	✓	✓
Day FE	–	✓	✓	✓
Observations	990116	990116	990116	990116
Overall R^2	0.007	0.008	0.009	0.011
Within R^2	0.002	0.002	0.002	0.004

This table reports the panel regression results using specification (17) with daily observations. Standard errors are shown in parentheses. *, **, and *** denote 10%, 5%, and 1% statistical significance respectively.

Table A4

	Overnight Ret		Intraday Ret		Return Gap	
	(1)	(2)	(3)	(4)	(5)	(6)
Rtl Turnover	0.08***		-0.10***		0.17***	
	(0.01)		(0.01)		(0.01)	
Inst Turnover	-0.11***		0.16***		-0.27***	
	(0.02)		(0.02)		(0.03)	
RNB Near Open	-0.06		0.72***		-0.79**	
	(0.26)		(0.25)		(0.38)	
Retail Vlm Share		0.13***		-0.14***		0.27***
		(0.01)		(0.01)		(0.01)
Stock Controls	✓	✓	✓	✓	✓	✓
Observations	62932	62932	62932	62932	62932	62932

This table reports the Fama-MacBeth regression results using specification (17). Standard errors are shown in parentheses. *, **, and *** denote 10%, 5%, and 1% statistical significance respectively.

Table A5

	1st Stage	OLS	2SLS
	RS20	Ret Gap	Ret Gap
RS20		0.893***	1.703***
		(0.013)	(0.056)
Log Shares Oustanding	0.052***		
	(0.000)		
Stock Controls	✓	✓	✓
Observations	1034702	1034702	1034702
R^2	0.640	0.004	

This table reports the 2SLS regression results using specification (18) after instrumenting retail volume shares with log outstanding shares and using daily observations. Standard errors are shown in parentheses. *, **, and *** denote 10%, 5%, and 1% statistical significance respectively.