Price Discovery in High-Frequency Equity Markets: Evidence from Retail and Institutional Trades*

Jan Harren†

This version: December 30, 2024

Abstract

Using high-frequency trades and quotes (TAQ) data, I quantify the information content of retail and institutional trades in equity markets. I find evidence of a heterogenous price impact among retailers and institutionals. Consistent with theory, I show that information frictions, illiquidity, and information drive differences in the price impact of retail and institutional investors. A size-neutral trading strategy on institutional investors' price impact yields sizeable returns, beats the market, and is not explained by established risk factors. Furthermore, I find that episodes of coordinated trading by Robinhood investors reduce the price impact of institutional investors. This is consistent with indirect liquidity provision from retailers to institutionals via wholesalers due to internalization of retail trades.

JEL classification: G10, G12, G14

Keywords: Illiquidity, Trading costs, Payment for order flow, Asymmetric information, High-frequency markets, Cross-sectional asset pricing, Price impact, Retail trading, Robinhood trading, Market microstructure

*I gratefully acknowledge very helpful comments and suggestions from Yun Joo An, Heiner Beckmeyer, Nicole Branger, Robert Faff, Thomas Grünthaler, Larry Harris, Mete Kilic, Seyed Moussavi, Roberto Ricco, Timo Wiedemann, Tiange Ye, and Aleksandr Zotov, as well as participants at the British Accounting and Finance Association Conference 2023 (Cambridge, MA), the Southwestern Finance Association Conference 2024 (Las Vegas), the FMA European Conference 2024 (Turin), the EFMA Conference 2024 (Lisbon), the World Finance Conference 2024 (Nicosia), the German Finance Association Annual Meeting Conference 2024 (Aachen), the DGF Doctoral Seminar 2024 (Aachen), and the FMA Annual Meeting 2024 (Grapevine).

†University of Münster, Germany, jan.harren@wiwi.uni-muenster.de

1. Introduction

The recent rise in trading activity by retail investors underlines their importance alongside large institutional investors and has changed the way researchers and practitioners in financial markets think about market movements. The Gamestop (GME) example in January 2021 showed that retailers can act as an "angry mob" and move markets in unfavorable directions for institutional traders.[1] Information acquisition by retailers (e.g. on subreddit chatrooms such as r/wallstreetbets) differs from institutionals with large research departments. Furthermore, wholesalers' (high-frequency market makers) internalization of retail trades prevents direct interactions between retailers and institutionals (Barardehi, Bernhardt, Da, and Warachka, 2024; Baldauf, Mollner, and Yueshen, 2024). It enables wholesalers to provide liquidity from one group to the other, especially when liquidity is limited. Hence, information acquisition and liquidity needs are dispersed among both market participants, and they might contribute differently to price discovery. Therefore, it is crucial to measure retailers and institutionals price impact to determine how information and liquidity compensation affect prices through trades for both market participants.

Consequently, the following questions arise: How do retailers and institutionals contribute to price discovery and how do they interact with one another? How sizeable is their price impact? Is a higher price impact related to higher liquidity and information risks in equity markets? Is the price impact time varying and heterogeneous among market participants, such as retailers and institutions? Is it possible to trade on the price impact? What is the economic intuition behind the price impact of both market participants? Do retailers provide liquidity to institutional investors through wholesalers because of order flow segmentation?

This paper aims to answer these questions by measuring the price impact from high-frequency trade and quote (TAQ) data for the whole cross-section of stock returns for retailers and institutionals. I provide strong empirical evidence that the price impact is

[1] See Welch (2020) for a description on how retail investors coordinated into a collective short squeeze in January 2021.

systematic, time-varying, and heterogeneous among retailers and institutions. Furthermore, I build a size-neutral long-short strategy to trade on the price impact and show that dealers demand compensation, in terms of higher future returns, for being exposed to a high price impact from institutionals. I provide an econmoic intuition for the price impact and show that it relates to other information measures, trading costs, and informational frictions (Bali, Beckmeyer, Moerke, and Weigert, 2023). Furthermore, I show that more concentrated trading by Robinhood traders is associated with a smaller price impact for institutionals. This finding is consistent with retailers aligning the price impact of institutionals (Neuhann and Sockin, 2023), inelastic demand of institutionals (Koijen and Yogo, 2019), and liquidity provision of retailers for institutional traders (Barardehi et al., 2024). Hence, I show that adverse selection risk and liquidity provision by retailers appear to be crucial determinants of asset returns.

According to Glosten and Milgrom (1985) and Kyle (1985) trades convey information to the market if market participants have private information about the fundamental value of an asset. This leads to two testable predictions in empirical asset pricing. First, the price impact of a trade should be positively related to asymmetric information, as information about the fundamental value is transferred to trades, which then affect quotes. Second, the price impact should be permanent as order flow that contains information should be persistent to have a price impact (whereas inventory or liquidity effects should be transitory). Glosten and Harris (1988) argue that information should be reflected in spreads and asset prices. The idea is that an uninformed market maker who collects a buy (sell) order, knowing that the order might be informed, revises upward (downward) its belief about the fundamental value of the stock. Since this informed trading is ex-ante anticipated by the uninformed market maker, she adjusts the prices and the spreads to take this risk into account. Thus, under asymmetric information, all agents face the risk of being adversely selected (Easley, Hvidkjaer, and O'hara, 2002) and demand a risk premium for the risk to trade against better-informed investors (Wang, 1993, 1994). Furthermore, asymmetric information also increases the required return through allocation costs rather than bid-ask spreads (Gârleanu and Pedersen, 2004).

However, studies show that traditional information measures may not capture informed trading due to limit orders (Collin-Dufresne and Fos, 2015). Hence, the price impact might be further driven by other risks in equity markets, such as illiquidity premia and trading costs, which originate from order flow segmentation via wholesalers' (Barardehi et al., 2024; Eaton, Irvine, and Liu, 2021). However, the price impact is a good proxy for extreme positive or negative order imbalance (buy/selling pressure), as extreme buying leads on average to upward price movements and extreme selling to price drops. Hence, the price impact of retailers (institutionals) is high when there is coordinated trading (heavy buying or selling activity) by one market participant. This one-sided trading might demand liquidity from the market and expose traders in high-frequency markets to higher trading costs. Hence, institutions might avoid holding stocks, which requires them to turn to wholesalers as liquidity providers in times of high illiquidity (Barardehi et al., 2024).

This paper proceeds in three parts. First, I test the empirical predictions that order flow contains information among time and market participants. To achieve that, I measure trades' contemporaneous and persistent price effects on quotes that reveal illiquidity and information risks. Second, I relate this measure to asset prices and show that this risk is priced in the cross-section and profitable to trade on. Lastly, I show that retail trading reduces the price impact of institutionals because retailers provide liquidity to institutionals through order flow segmentation via wholesalers'.

I use intraday TAQ data from 2006 to 2020, which contains all price changes and trades that are reported in the data. I aggregate this dataset on five-minute intervals to reduce microstructure noise (Wiedemann, 2022). To distinguish between retail and institutional trades I rely on the Boehmer, Jones, Zhang, and Zhang (2021) algorithm which identifies a subset of primarily retail trades.[2] My empirical analysis builds on a vector autoregres-

[2] I am aware that the algorithm is subject to both Type I (incorrectly identifying institutional trades as retail trades) and Type II (identifying only a subset of actual retail trades) errors (Battalio and Jennings, 2023; Barber, Huang, Jorion, Odean, and Schwarz, 2023). The researchers propose identifying retail trades exactly like Boehmer et al. (2021) but modifying the buy/sell flag by using the quote rule. Since my main results are mainly based on the effects of institutionals price impact, the key findings of this paper are not prone to potential retail signing errors.

sion (VAR) that decomposes the order flow into transitory and permanent price impact components. I follow Ranaldo and Somogyi (2021) and extend the VAR in Hasbrouck (1991a) by allowing for different agents. However, in contrast to them, I distinguish between retail and institutional investors and consider equity prices instead of FX markets. I find compelling evidence that order flow systematically impacts equity markets heterogeneously across time and agents. Across agents, I observe an economically and statistically higher price impact (contemporaneous and permanent) for institutions compared to retailers. One possible explanation is that institutional investors move much larger volumes and can therefore leverage their private information in the equity markets. Furthermore, institutionals are exposed to higher implicit and explicit trading costs and are willing to pay a premium for accessing formal and informal research (Di Maggio, Egan, and Franzoni, 2022). In contrast, it is more difficult for retail investors to achieve such a price impact with a smaller order volume. Furthermore, some retailers act as noise traders (Friedman and Zeng, 2022). Over time, the price impact varies and responds to current market conditions, indicating a time variation in information and illiquidity risk when overall risk aversion is higher.

In the second part of the paper, I relate the price impact to asset prices to determine whether higher risk is associated with higher expected returns. First, I determine which agents' price impact is associated with higher subsequent returns. Independent double sorts reveal that institutions consistently get compensated with higher future returns for a higher price impact. For retailers, however, I observe this pattern only conditional on the high price impact for institutions. The high-minus-low diff-in-diff spread is consistently driven by information and illiquidity risk of institutions and amounts to 1.28% per month, being statistically significant. Thus, I conclude that only institutions' price impact poses a higher risk for a potential high-frequency market maker (wholesaler) in equity markets.

Next, I determine the forces underlying this return predictability. Competitive arbitrageurs might exploit this return predictability and drive prices to their fundamental values. However, arbitrage might be costly (Shleifer and Vishny, 1997; Pontiff, 2006) and, thus, might prevent arbitrageurs from exploiting mispricing. I hypothesize that the return predictability stems from private information from stock-based characteristics (size, liquidity, idiosyncratic volatility, age, institutional ownership, and analyst coverage) that are not directly incorporated into prices (informational frictions). I create an informational frictions index based on stock-level information (Bali et al., 2023) to test this. In line with this prediction, I show that the price impact increases with higher informational frictions for retailers and institutions. Thus, return predictability might stem from stocks with high informational frictions. Furthermore, the price impact is related to measures of informed trading and illiquidity, consistent with wholesalers' liquidity premia and costly liquidity provision (Barardehi et al., 2024).

Above return predictability enables me to inspect whether the predictability of future returns can be exploited by trading on institutions price impact. I construct a size-neutral zero-investment high-minus-low portfolio that loads on the price impact of institutions. The trading strategy yields a yearly Sharpe Ratio of 1.66 (1.30) before (after) transaction costs. Furthermore, the strategy yields sizeable and significant alphas in spanning regressions and is not explained by established risk factors in cross-sectional asset pricing (Carhart, 1997; Pástor and Stambaugh, 2003; Fama and French, 2015).

Finally, I examine the impact of retail investors' recent increase in trading activity via trading platforms such as Robinhood and how this affects institutions price impact. I use data from Robintrack and construct the retailers' crowd-wisdom portfolio (Welch, 2020) to identify stocks in which retailers had much trading activity. I compare these stocks with less heavily retail-traded stocks. I find that stocks heavily traded by retailers exhibit a smaller permanent price impact for institutions. Thus, higher retail activity reduces information and illiquidity risks from trading against institutions. This result is stable when the control and treatment groups are matched using the last quarter's fundamental data. From an economic perspective, there could be two explanations for this. Either retailers stabilize markets by providing liquidity and hence reduce the price impact of institutionals. When liquidity is scarce, wholesalers might use accumulated internalized order flows of retailers to provide liquidity to institutions, which reduces overall adverse selection and illiquidity. Thus, this explanation is consistent with retail order flow reducing institutionals price impact. On the other hand, if retail trading is pure noise, it will still disrupt institutions in price discovery, but their price impact will also reduce. Hence, all explanations are consistent with well-established theoretical models.

1.1. Literature

I contribute to the microstructure and asset pricing literature in several ways.

Market Microstructure. First, my analysis of heterogeneous asymmetric information risk across market participants and over time measures the permanent price impact in the stock market for the entire cross-section based on intraday data. I build on Ranaldo and Somogyi (2021), who measure the permanent price impact component in the foreign exchange (FX) market across 30 currency pairs and among four market participants: corporates, funds, non-bank financial firms, and banks. A trading strategy shows that there is asymmetric information risk in over-the-counter dealership FX markets. Furthermore, Barardehi et al. (2024) proposes a new illiquidity measure by using absolute number of retailers imbalance, which can be interpreted as the intensity with which wholesalers provide liquidity in less liquid conditions.

My study builds on their analysis as I extend their methodological framework using retailer and institutional trades (Hasbrouck, 1988, 1991a,b). By building a size-neutral trading strategy, I reveal that trading information and illiquidity risks in equity markets is profitable. Furthermore, I identify the drivers of the permanent price impact component in equity markets for both market participants and show that informational frictions, adverse selection, and trading costs are key drivers of the price impact in the cross-section for both, retailers and institutions. Furthermore, I extend the literature of Barardehi et al. (2024) and show that the price impact of institutionals is reduced when Robinhood investors trade coordinated in one direction, which is consistent with indirect liquidity provision from institutionals via wholesalers'. Eaton et al. (2021) interprets the price impact as a proxy for institutional trading costs. I show that there is a premium for trading on the price impact, which reflects institutional trading costs.

Asset Pricing. Second, this paper contributes to the asset pricing literature by developing a new size-neutral long-short strategy that loads on the price impact of institutions. The newly developed factor yields high long-short returns not spanned by established asset pricing factors. Several studies use high-frequency price information to construct risk factors that load on information risk and relate it to future returns in the cross-section. Easley, Kiefer, O'hara, and Paperman (1996) derive the probability of informed trading (PIN) and relate it to spreads and volume. They show that block trades are associated with lower PIN. Easley et al. (2002) use a microstructure model to derive a measure of PIN and show that information does affect returns. Easley, Hvidkjaer, and O'hara (2010) establish a long-short factor of information-based trading based on PIN. This factor is able to explain returns, especially for small stocks, and is not subsumed by established factors in equity markets. Brennan, Huh, and Subrahmanyam (2016) decompose PIN into good and bad news and show that both predict positive and negative returns around earnings announcements and that bad news drives the equity cost of capital.[3]

I contribute to this literature by establishing a factor that trades on the price impact of institutions for the cross-section of equity returns and distinguishing it for retailers and institutions. Furthermore, I build a size-neutral factor on institutional price impact and show that it yields significant returns over established risk factors after transaction costs. I uncover the channels through which this factor affects asset prices and relate it to informational frictions, illiquidity, and information. This risk interpretation challenges the well-established risk interpretation of informed trading. I show that the price impact proxies for more risk than informed trading in the stock market.

Retail trading. Third, this paper contributes to the literature on retail trading in stock markets. Welch (2020) quantifies the holdings of retail investors on the Robinhood platforms and reveals their preference for stocks with high past volume. Eaton, Green, Roseman, and Wu (2022) exploit Robinhood outages and find that even short-lived shocks to retail trading can impact large institutional investors by affecting their inventory.

[3] Their study build on Easley and O'hara (2004) who also investigate the role of information on a firm's cost of capital. Other related papers include Borochin and Rush (2016), Babenko, Boguth, and Tserlukevich (2016), and Yang, Zhang, and Zhang (2020).

I contribute to this literature and show that coordinated Robinhood retail trading (either extreme buying or selling) on the Robinhood platform significantly reduces the price impact of institutions. Retail trading thus impedes the price impact of institutions. This finding is consistent with indirect liquidity provision through order flow segmentation via wholesalers'.

The remainder of this paper is organized as follows: Section 2 explains the data sources, the classification of retail and institutional trades, the methodology of the VAR model, and provides descriptive statistics. Section 3 presents my results, while the last section 4 concludes.

2. Data & Methodology

I first describe the data sources. Second, I discuss the classification of retail and institutional trades. Third, I describe the VAR approach to estimate the permanent price impact as my measure for information and illiquidity risk. Finally, I present results and descriptive statistics on measuring the price impact for retailers and institutions.

2.1. Data sources

For the construction of the database, I merge several data sources.

TAQ Database. I use intraday data from the TAQ (Trade and Quote) database. The database contains level 1 tick data for the entire US cross-section of equities for NYSE, Nasdaq, and other regional US exchanges. The data provides the transaction price, size, exchange code, and other information for trades. The best-bid and best-ask quotes and their respective volume and exchange code are available for quotes. I merge trades and quotes and aggregate them at 5-minute intervals. The aggregation requires forwarding fill prices from the last price available within each 5-minute interval. The aggregation is made to reduce microstructure noise. In addition, bulk classification is used to classify buys and sells (Easley, López de Prado, and O'Hara, 2012b). This classification scheme requires the aggregation of trades in time intervals.

CRSP. I use monthly stock data from 2006 to 2020 and apply standard filters in cross-sectional asset pricing. I only include common shares with share codes 10 and 11 and stocks that trade on the NYSE, AMEX, and NASDAQ (exchange codes 1, 2, 3, 31, 32, 33). I include stocks with a share price larger than 5$. To ensure that my results are not driven by microcaps, I exclude the smallest quintile of the cross-sectional distribution each day after applying the filters above (Gonçalves, 2021).[4]

Robinhood Database. The main dataset for retail investor trading stems from the 2013 founded online broker Robinhood. The company pioneered zero-commission trading in equities and ETFs in the US. The app was introduced in 2015 and attracted, especially young investors.[5] The Robintrack website scraped hourly user holdings for all equities on Robinhood.[6] The API was active from May 5, 2018, to August 13, 2020, which defines my sample period in the further analysis in Section 3.4.

2.2. *Classifying retail trades*

I follow the method of Boehmer et al. (2021) to distinguish retail trades from institutional trades. Retail trades occur mostly off-exchange, either sold by the broker to a wholesaler or filled from the broker's inventory (internalization) (Battalio, Corwin, and Jennings, 2016). TAQ classifies these transactions with exchange code "D." retail trades also receive small price improvements in fractions of a cent over the National Best Bid or Offer (NBBO) for market orders, while institutional trades are in increments of whole or half cents. Therefore, I classify trades priced just above or below a round cent as retail trades. Respectively, Boehmer et al. (2021) classify a trade as a retail buy (sell) if the fractional

[4] If I exclude the smallest quintile of the NYSE distribution, I have fewer stock-month observations, hence I apply the filters above. All results are not qualitatively or statistically changed when I use stocks with a market capitalization larger than the first NYSE quintile.

[5] Robinhood traders used social media to organize the short squeeze in Gamestop stock, triggering heavy losses for short-selling hedge funds: CNBC Article.

[6] The link to the Robintrack data can be found here: Robintrack website.

component of the trade price is between 0.6 and 1 (0 and 0.4) cents, that is

$$Z_{i,t} = 100 \cdot \text{mod}\left(P_{i,t}, 0.01\right), \quad \text{where } Z_{i,t} \in [0, 1) \tag{1}$$

$$\text{Trade}_{i,t} = \begin{cases} \text{Retail sell} & \text{if } Z_{i,t} \in (0, 0.4) \\ \text{No retail trade} & \text{if } (0.4 \le Z_{i,t} \le 0.6) \text{ or } (Z_{i,t} = 0) \\ \text{Retail buy} & \text{if } Z_{i,t} \in (0.6, 1), \end{cases} \tag{2}$$

where $Z_{i,t}$ is a fraction of a penny cent.

I assume that all trades that are not retail trades according to equation (2) belong to institutional trades.

2.3. *Classifying institutional trades*

I follow Easley et al. (2012b) and classify non-retail trades using bulk classification. As I do not work with level 1 tick data, but aggregate trades on 5min intervals, the bulk classification rule proposed by Easley, Lopez de Prado, and O'Hara (2012a) seems most reasonable. The classification scheme relies on the idea that trade time is more informative in high-frequency markets than clock time.

Classification of trades has always been problematic. Reporting conventions treat orders differently depending on the buy/sell indicator. The New York Stock Exchange reports only one trade if a large sell trade was completed by multiple buys, but multiple trades if a large buy block was crossed with multiple sell orders. These conventions constrain tick-based reporting algorithms (Lee and Ready, 1991; Ellis, Michaely, and O'Hara, 2000; Chakrabarty, Li, Nguyen, and Van Ness, 2007). In my analysis, I aggregate trades into five-minute intervals to circumvent this problem and eliminate microstructure noise.[7] To determine the percentage of buying and selling volume, I use the standardized price change between the beginning and the end of the interval. I define the daily volume

[7] Results are qualitatively the same when using 30 minute time bars.

as V and calculate daily buy and sell volume as

$$V_\tau^B = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i \cdot Z\left(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}\right) \tag{3}$$

$$V_\tau^S = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i \cdot \left[1 - Z\left(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}\right)\right] = V - V_\tau^B, \tag{4}$$

where I define $\tau = t(\tau) - t(\tau - 1)$ as one trading day and Z is the cumulative distribution function of a standard normal distribution. The price change between five-minute intervals is standardized with the daily standard deviation of price changes. This procedure allows to create a buy/sell indicator for each five-minute volume interval using bulk classification. When $\Delta P > 0$, the volume is more weighted towards buys. The weighting depends on how large the price change is in relation to the daily distribution of price changes. This procedure is more appropriate for my application, as I aggregate trades on five minute intervals.

Table A21 compares different classification schemes. I use the Boehmer et al. (2021) buy/sell classification scheme for retail trades and observe how many retail trades are classified in the same way by the different algorithms as by Boehmer et al. (2021). Lee and Ready (1991) and Chakrabarty et al. (2007) outperform Ellis et al. (2000). Panel B reports the percentage of correctly classified volume. The bulk classification is close to Lee and Ready (1991) with 72.1 (71.9) of truly classified retail volume for buys (sell) vs. 79.5 (79.9) using LR. Since my application aggregates trades in five-minute intervals, I classify non-retail trades using the bulk classification (Easley et al., 2012a).

2.4. Methodology

VAR approach. This section describes the methodology used to examine whether retail and institutional traders have divergent price impacts. The VAR approach follows Hasbrouck (1988, 1991a,b). Ranaldo and Somogyi (2021) use a similar VAR approach. They provide a model that is able to separate the transitory price impact (inventory effect) from the permanent price impact (information effect). The information is inferred from observed trades and quotes. An informed trade should have a permanent effect on the price, quoted by the market maker. The permanent component is an estimate of the incorporation of better fundamental information into prices, while the transitory effect could also be due to temporary liquidity effects. The setting is model-free and accounts for serial dependence of trades and returns from mid-quotes, delays in the price impact of a trade on the quoted price, short-term mean reversion in returns from mid-quotes, nonlinearities between order size and quote revision, and half-hour seasonalities.

Equation (7) describes the evolution of r_t, the midpoint-return from quotes. Equation (8) shows the positive persistent effect of trades. I aggregate trades on 5 minute intervals and calculate signed net volume z_t, which is buy orders minus sell orders.[8] T_t is a buy/sell indicator variable, which is one if $z_t > 0$, minus one if $z_t < 0$ and zero otherwise. To account for nonlinearities between order size and quote revisions, I follow Hasbrouck (1988) and calculate logarithms of z_t

$$v_t = \begin{cases} +log(z_t) & \text{if } z_t > 0 \\ 0 & \text{if } z_t = 0 \\ -log(-z_t) & \text{if } z_t < 0. \end{cases} \tag{5}$$

For interpretability of the regression coefficients, I perform the following regression

$$v_t = c + \sum_{i=0}^{10} \theta_i T_{t-i} + \tilde{S}_t, \tag{6}$$

where $\tilde{S}_t$ is the error term which is orthogonal to T_t. I perform above steps for institutional investors (IN) and retail investors (RE) and define the agents with $j \in C$, where $C = \{IN, RE\}$. I include lagged returns and order flow in Equations (7) and (8) to account for possible inventory effects, lagged timely arrival of information, adjustment of information, and order splitting. I choose a lag length of ten, based on the arguments in Hasbrouck

[8] For institutional trades, I classify trades with bulk classification (Easley et al., 2012b).

(1991a,b)

$$r_t = \sum_{i=1}^{10} \rho_i r_{t-i} + \sum_{j \in C} \left(\sum_{i=0}^{10} \beta_i^j T_{t-i}^j + \sum_{i=0}^{10} \phi_i^j \tilde{S}_{t-i}^j \right) + \zeta_{1,l} D_{l,t} + \epsilon_{r,t} \tag{7}$$

$$T_t = \sum_{i=1}^{10} \gamma_i r_{t-i} + \sum_{j \in C} \left(\sum_{i=1}^{10} \delta_i^j T_{t-i}^j + \sum_{i=1}^{10} \omega_i^j \tilde{S}_{t-i}^j \right) + \zeta_{2,l} D_{l,t} + \epsilon_{T,t}, \tag{8}$$

where $D_{l,t}$ is a dummy variable for the time of the day from 9:30h to 16:00h in 5 minute intervals (14 dummies per day), and $\epsilon_{r,t}$ and $\epsilon_{T,t}$ denote error terms for the return and order flow equations. I estimate Equations (7) and (8) for each stock k.[9] The contemporaneous T_t in Equation (7) ensures that the system of equations is exactly identified.

Permanent price impact. The permanent price impact for each stock k is calculated as the sum of the beta coefficients in Equation (7). The permanent price impact for IN and RE can be calculated for each stock k as

$$\alpha_m^{j,k} = \sum_{t=0}^{m} \beta_t^{j,k}, \tag{9}$$

where m indicates the number of lags (ten in my case).[10] As the error terms in Equations (7) and (8) can be interpreted as the unexpected public and private information components, i.e., the persistent price impact of the trade innovation, the permanent price impact in Equation (9) can be interpreted as the (expected) asymmetric/private information (Hasbrouck, 1991b). Furthermore, I can calculate the price impact within agents j, capturing superior information within stock k as

$$\bar{\alpha}_m^k = \frac{1}{|C|} \sum_{j \in C} \sum_{t=0}^{m} \beta_t^{j,k} = \frac{1}{|C|} \sum_{j \in C} \alpha_m^{j,k}. \tag{10}$$

The permanent price impact estimates the effect of trades on quote corrections net of transitory effects on global equity markets. In addition, the measure considers the persistence of order flow and possible feedback effects.

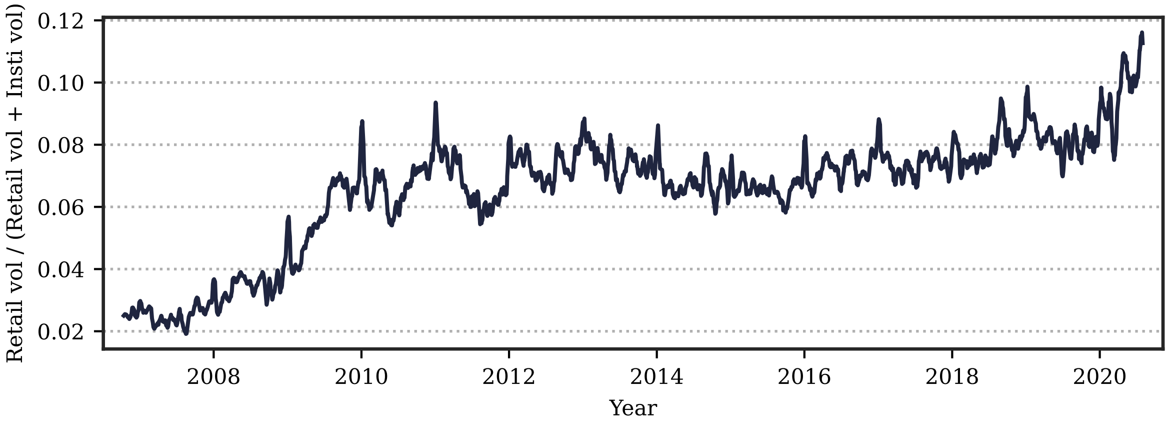
[9] For the sake of clarity, I suppress k in Equations (7) and (8).

[10] Lower lags ($m < 10$) would overestimate the price impact, as it would capture the positive initial price impact of trade on the quote. Still, they would miss a potential subsequent reversal (Ranaldo and Somogyi, 2021).

2.5. *Descriptive statistics*

This section analyzes if the permanent price impact in the US equity market systematically varies across stocks, market participants, and time. Figure 1 shows the retail share of all trades. The retail volume increased from 2006 to 2010, remained comparatively

Fig. 1. Share of retail trading volume relative to total volume



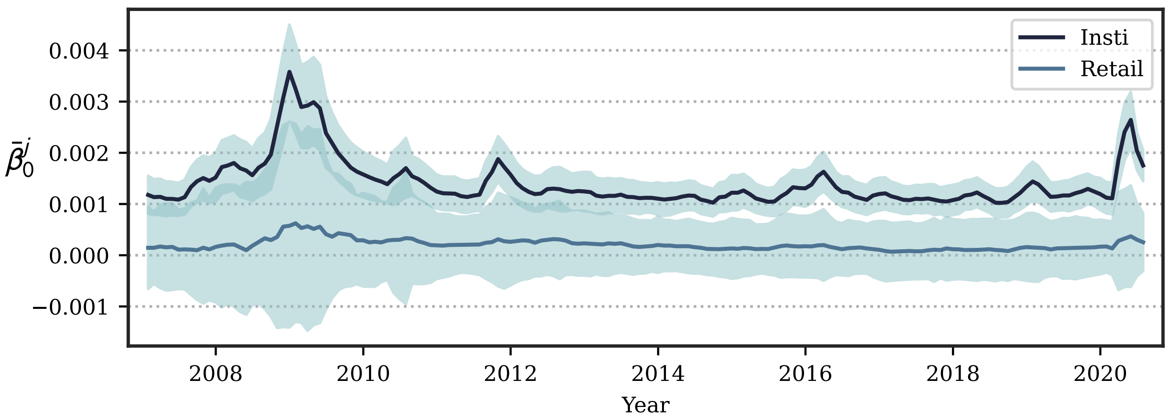
Note. The figure shows the retail trading volume over the whole trading volume for the time period January 2007 until July 2020.

stable until 2018, and increased again in the subsequent years. When interpreting my results, it is important to remember that retailers move much less volume and, therefore, cannot leverage high-volume private information compared to institutions. I split these groups because I believe that asymmetric information risk is more prevalent among experienced traders. However, it also exists for retail traders, although the impact is smaller due to the lower volume of trades.

2.5.1. *Contemporaneous price impact*

Figure 2 shows contemporaneous price impact as the cross-sectional mean over all stocks at each point in time $\bar{\beta_0}^j = (1/K)\sum_{k=1}^{K} \beta_0^{j,k}$. The average confidence intervals are also displayed. On average, the coefficients for institutional and retailers show the expected positive sign. In line with market microstructure theory, prices move in the direction of trades, and prices show recent changes in the direction of trades (Kyle, 1985; Glosten

Fig. 2. Contemporaneous price impact

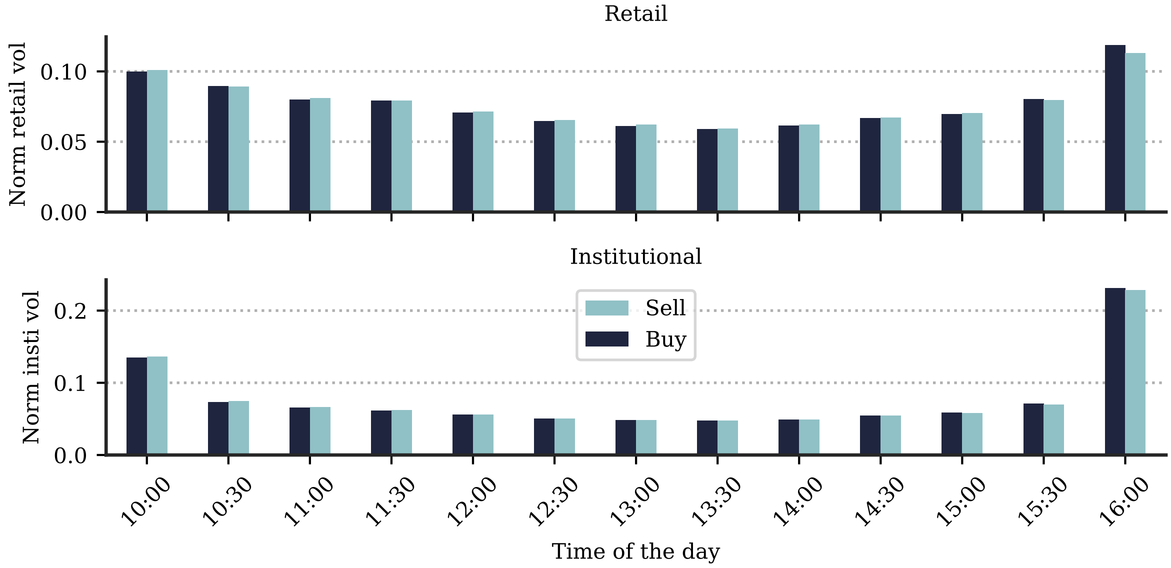


Note. The figure shows cross-sectional mean estimates of the contemporaneous price impact for institutionals and retailers. The estimate is the coefficient $\beta_0^{j,k}$ from Equation (7) for $i = 0$. The cross-sectional mean is caluclated as $\bar{\beta_0}^j = (1/K)\sum_{k=1}^{K}\beta_0^{j,k}$. 5% mean confidence intervals of the regression estimates are shown in the shaded area around the mean. The time period is January 2007 until July 2020.

and Milgrom, 1985). A positive price impact means that there is a positive imbalance in trades and stock prices go up or that there is a negative imbalance in trades and stock prices react by going down. The level of institutional contemporaneous price impact is always above that of retailers and statistically significant, whereas retailers do not exhibit a significant contemporaneous impact over time. Given their smaller trades, this is not surprising. The average confidence interval shows that retailers' contemporaneous price impact appears to be negative for some stocks. The rationale for this finding is retailers might trade against dealers for liquidity reasons and demand immediacy (Grossman and Miller, 1988). The negative β_0^{RE} also aligns with retailers trading against better-informed investors, such as institutions. A risk-averse dealer would offset the uninformed order flow (e.g., retailer) with that of the informed institution to reduce its own asymmetric information risk (Liu and Wang, 2016).

The contemporary price impact can further be rationalized when looking at the intraday mean trading volume. Figure 3 shows the mean volume over 30min intervals. Retailers trade small volumes and exhibit a U-shaped trading pattern over the day. Institutions trade large volumes over the whole trading day, with peaking activities at the beginning and the end of the trading day, accounting for over 20% of the trading volume.

Fig. 3. Intraday trading volume for 30 minute intervals



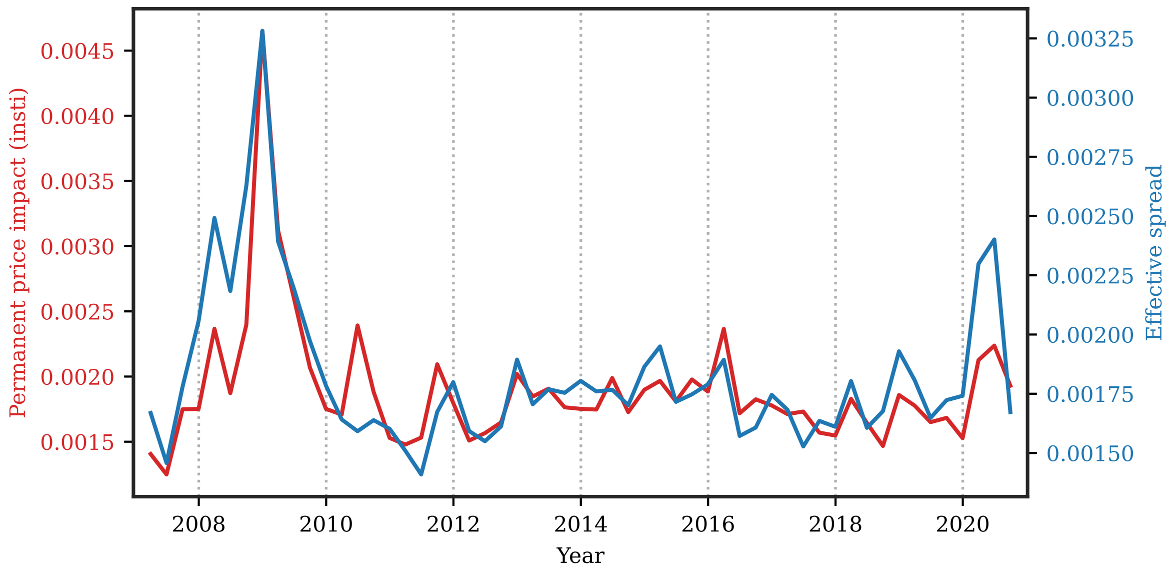
Note. The figure shows the mean fraction of retail and institutional trading volume during the day for buys and sells for half hour intervals.

This might be due to hedging needs and inventory features (Stoll, 1978). Furthermore, delta hedging of market makers in options markets and ETF rebalancing might lead to extreme flows in the last half hour of the trading day (Barbon, Beckmeyer, Buraschi, and Moerke, 2021). Thus, Institutions trade large volumes over the whole trading day, with peaking activities at the beginning and the end of the day. Retailers trade small volumes and exhibit a U-shaped trading pattern over the day.

2.5.2. Permanent price impact

According to Hasbrouck (1991a), the permanent component, α_m^j, can be interpreted as a measure of asymmetric/private information because trade motives are driven more by private (superior) information and liquidity needs rather than public information (Kyle, 1985). A persistent impact of a trade on prices arises from asymmetric information stemming from that trade. The error term in equation (7), $\epsilon_{r,t}$ reflects all public information associated with the quote revision and the error term in equation (8), $\epsilon_{T,t}$, captures all private information in the trade innovation. The system of equation ensures that $\epsilon_{T,t}$ reflects no public information and hence α_m^j can be interpreted as a measure of asymmetric information. Another interpretation is, that a high price impact measures illiquidity premia as times of large trading and price movements are associated with higher spreads and volatility (Glosten and Milgrom, 1985). Figure 4 shows that the cross-sectional average of the institutional price impact varies with the cross-sectional average of the effective spread and is especially high in times of market distress. The correlation when using institutionals (retailers) cross-sectional average price impact and the cross-sectional average of effective spreads is 96.30% (94.65%) when using all data and 84.16% (57.30%) when applying the size filters mentioned in Section 2.1. Hence, the price impact is higher when aggregate liquidity is low. Figure A8 provides the same plot for the retailers price impact. For retailers, especially the latter years and the period of COVID19 show a high comovement with spreads and an exceptional high increase in price impact for retailers.

Fig. 4. Price impact and spreads

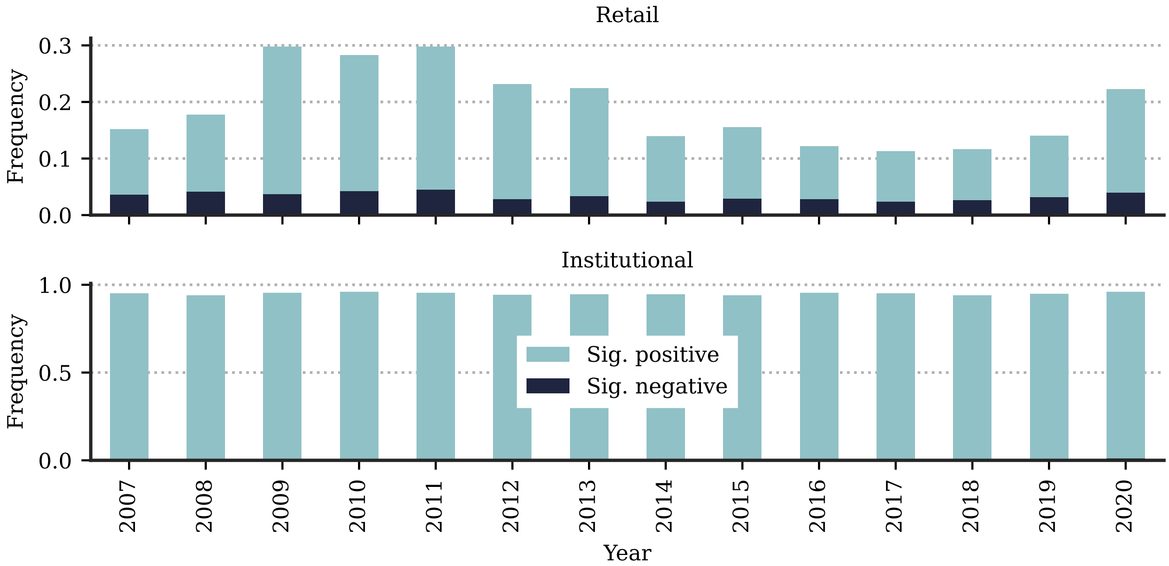


Note. The figure shows the cross-sectional average of the institutional price impact and the cross-sectional average of the effective spread. The time series correlation of the two aggregated series using the institutional (retail) price impact is 84.16 % (57.30 %). The figure excludes microcaps (share price smaller than 5%) and the smallest quintile of the cross-sectional distribution at each point in time.

I aggregate high frequency data over five minute intervals and estimate equation (7) in a rolling window fashion for each stock on each day. For each regression, I use 63 (= 252/4) days, thus I estimate the model over the course of a quarter. I measure statistical significance with a heteroskedasticity-consistent joint F-test in which the parameters in equation (9) are jointly different from zero.

Figure 5 presents the fraction of positive (negative) significant coefficients per year. Consistent with theories of asymmetric information (Glosten and Milgrom, 1985), I observe heterogeneously informed traders in equity markets. Depending on the market participant, I observe different impacts of order flow on prices. On average, institutions (retailers) have a positive significant price impact with an average fraction of 94.24% (15.73%). Over time, this fraction varies between 93.23% and 95.32% for institutions and between 8.87% and 26.03% for retailers. A positive coefficient means, that order flow and prices move in the same direction, which is, what I would expect for better informed investors. Institutions appear to have superior information across almost all stocks in the cross-section. This might be due to their preferred access to information and their central role in equity markets compared to retail investors.

Fig. 5. Fraction of significant permanent price impact coefficients per year

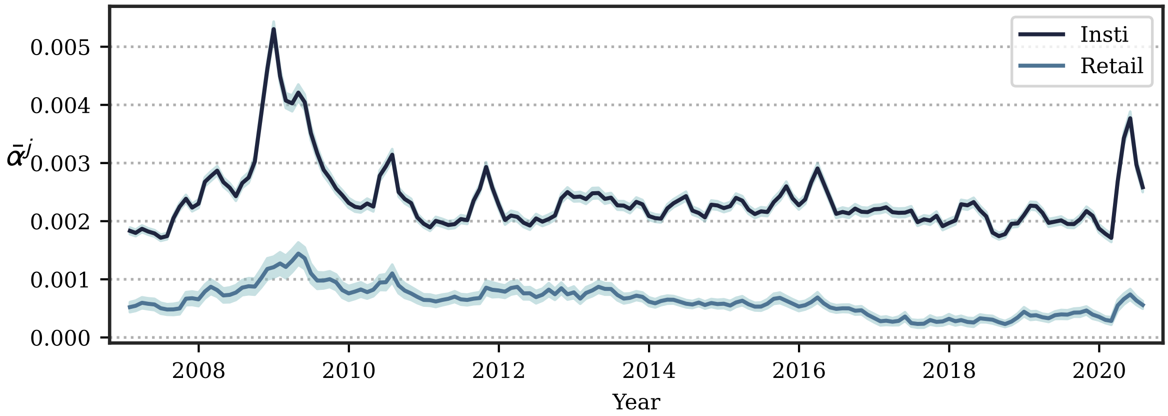


Note. The figure shows the fraction of significant coefficients per year for retailers' and institutionals' permanent price impact (α_t^j). A value of one indicates, that all α_t^j coefficients in this year were significant for the given year for the whole cross-section.

To test whether the permanent price effects for retailers differ from those observed for institutions, I examine whether all coefficients in equation (9) for $j = RE$ differ from those obtained for $j = IN$. For 81.72% of all observations, I find economically larger price effects for institutions where the difference is statistically significant.

Figure 6 plots the cross-sectional mean permanent price impact. Institutions exhibit a

Fig. 6. Permanent price impact



Note. The figure shows cross-sectional mean estimates of the permanent price impact for institutionals and retailers. The estimate is the coefficient $\alpha_m^{j,k} = \sum_{t=0}^{10} \beta_t^{j,k}$ from Equation (7). I calculate the cross-sectional mean as $\bar{\alpha}_t^j = \sum_{k=1}^{N} \alpha_t^{j,k}$. 5% confidence intervals of the cross-sectional distribution of $\alpha_t^{j,k}$ are shown in the shaded area around the mean. The time period is January 2007 until July 2020.

higher economic price impact compared to retailers at each point in time. However, the confidence intervals show that retailers price impact is more cross-sectionally dispersed compared to institutions. This is in line with the assumption that more experienced market participants have better access to the stock markets, which allows them to split orders and smooth their price impact (Van Kervel and Menkveld, 2019). Over time, the price impact varies and responds to current market conditions, indicating a time variation in asymmetric information risk when overall risk aversion is higher. The great financial crises and the recent COVID19 pandemic have sharply raised the level of asymmetric information risk for both retailers and institutions. Furthermore, the price impact of institutional investors is relatively stable, while it decreases for retailers over time, which could be due to increasing market efficiency. Markets with higher information efficiency imply that it is more difficult for investors to incorporate private information into prices. This might be especially hard for less sophisticated retail traders.

Thus, a significant price impact is present in stock markets, strongly varies among market participants and fluctuates over time and with the business cycle.

3. Results

In the last section, I derive the price impact for two groups of traders, retailers and institutions. In this section, I relate these measures to future returns to get a sense of whether information and illiquidity risks are reflected in subsequent returns. Furthermore, I conduct a trading strategy on the institutional price impact. I also analyze the economic drivers of the price impact in equity markets. Finally, I show how increased retail activity affects the price impact of institutions and relate it to indirect liquidity provision of retailers to institutionals via wholesalers'.

3.1. Illiquidity and information risks and future returns

From a theoretical asset pricing perspective, higher adverse selection risk should be rewarded with higher subsequent returns, as the investor demands compensation for trading against better informed investors (Kyle, 1985; Glosten and Milgrom, 1985; Wang, 1993, 1994). Easley et al. (2002) show theoretically and empirically that private information positively affects asset prices. Specifically, the costs of adverse selection are associated with biases in trading decisions, resulting in higher allocation costs and hence higher returns (Gârleanu and Pedersen, 2004). Therefore, the bid-ask spread alone should not fully capture traders' adverse selection risk, but expected returns should also compensate for higher asymmetric information risk. Furthermore, the price impact can be understood as a proxy for expected trading costs in the sense of Barardehi et al. (2024). When retailers trade in one direction, this order flow is internalized to a wholesaler and might be used as liquidity provision to institutionals order flow when liquidity is scarce. I distinguish between institutions and retailers and hypothesise, that these risks' should be more pronounced for institutions as they face larger volume and hence higher allocation costs.

To test this hypothesis, I perform independent double sorts for each of the agents to determine for which agent a higher price impact is associated with higher future returns.

For each month, I independently sort stocks into quintile portfolios based on α_t^{RE} and α_t^{IN} and calculate the next month's value-weighted return above the risk-free rate. The average of the monthly return time series is the reported portfolio return. Furthermore, I calculate high-minus-low portfolio spreads and the corresponding t-statistics with Newey and West (1986) robust standard errors for a lag length of ten.

Table 1 reports the results for the 25 (5x5) portfolios sorted independently by α^{IN} and α^{RE}. Future quintile portfolio returns increase when sorting on α_t^{IN}. The HmL α_t^{IN}

Table 1: Independent double sort

Independent double sort							
r_{t+1} (in %)	Low α_t^{IN}	2	3	4	High α_t^{IN}	HmL α_t^{IN}	HmL t-Stat.
Low α_t^{RE}	1.02	0.52	0.88	0.95	1.14	0.13	0.25
2	0.79	0.77	1.13	0.87	1.96	1.17	2.07
3	0.72	0.80	1.07	1.08	1.55	0.83	2.34
4	0.37	0.63	0.73	1.15	1.61	1.24	2.77
High α_t^{RE}	-0.06	0.78	0.58	0.88	1.35	1.41	4.14
HmL α_t^{RE}	-1.08	0.25	-0.31	-0.07	0.21	1.28	
HmL t-Stat	-3.00	1.11	-1.41	-0.29	0.92	2.87	

Note. The table shows monthly value-weighted returns to independently double sorted portfolios on α_t^{IN} and α_t^{RE} and the corresponding high-minus-low (HmL) portfolios and the Diff-in-diff HmL-portfolio with corresponding t-Statistics.

spread is positive and significant for all possible HmL α_t^{IN} combinations except for low α_t^{RE}. For the lowest level of α_t^{RE}, I observe an HmL spread of 0.13% per month and for the highest level of α_t^{RE}, a spread of 1.41% per month, resulting in an economically and statistically significant diff-in-diff return spread of 1.28% per month. However for HmL α_t^{RE}, there is no clear relationship with future returns, except for the HmL return spread for low α_t^{IN}, which is negative. The low α_t^{IN}/ low α_t^{RE} portfolio shows a return of 1.02%. This is the portfolio where both, retailers and institutionals have a negative price impact and thus trading against price movements. Hence, this portfolio can be interpreted as compensation for liquidity provision to other market participants. However, this premium is only reflected for retailers in the HmL α_t^{RE} for low α_t^{IN} with -1.08% being statistically significant with a t-statistic of -3. Thus, liquidity provision costs of wholesalers' also drive the compensation for price impact.

I conclude that institutions are compensated for information and trading costs, while retailers are not. Institutions could be more concerned about the risk of trading against better-informed traders, as they move much larger volumes in equity markets. Furthermore, institutions exhibit higher capital allocation costs which originate from higher adverse selection risk (Gârleanu and Pedersen, 2004). The smallest α_t^{RE} and α_t^{IN} portfolio shows that wholesalers' liquidity provision is also reflected in the price impact of institutions.[11]. Hence, in the following analyses I mainly use α_t^{IN} as a proxy for information and illiquidity risks' in the cross-section of stock returns. In the next subsection, I aim to determine the drivers of the price impact for the cross-section of stock returns.

3.2. Drivers of the price impact

In the last section, I showed that α_t^{IN} is able to forecast future returns for the next month. In this section, I want to determine the underlying forces of this return predictability. Competitive arbitrageurs might identify this return anomaly and drive prices to their fundamental values. However, arbitrage might be costly and not free to conduct because it might be risky and requires costly capital (Shleifer and Vishny, 1997; Pontiff, 2006). Thus, limits of arbitrage might prevent competitive arbitrageurs to exploit mispricing in stock markets. Due to limited investor attention and informational constraints, new informative signals are partially incorporated into asset prices because some investors who are subject to informational constraints do not adjust their demand by retrieving informative signals from observed prices. Hence, asset prices exhibit predictability. I test whether informational frictions provide an explanation for the trading signal α_t^{IN}. I want to determine the drivers of the price impact and hypothesize that information, and its associated return predictability, stems from informational frictions. That is, private information from stock-based characteristics that is not directly incorporated into stock prices (Bali et al., 2023).

To measure informational frictions, I construct the arbitrage index of Atilgan, Bali,

[11] This is consistent with the finding in Section 3.4 and Figure 9.

Demirtas, and Gunaydin (2020) which does not rely on one proxy for informational frictions but instead uses several variables that capture limits-to-arbitrage. For the informational frictions index, I include firm age, analyst coverage, size, institutional ownership, idiosyncratic volatility, and the Amihud (2002) measure. To construct the index I sort stocks in quintile portfolios in increasing order based on idiosyncratic volatility and illiquidity. Similar, I sort stocks in decreasing order based on their level of firm age, analyst coverage, size, and institutional ownership, as lower values indicate higher costs of arbitrage. The arbitrage cost index is the sum of the six scores, ranging from 6 to 30. The higher the arbitrage costs, the tighter limits-to-arbitrage. The rational behind the index is that higher illiqudity reflcets higher transaction costs (Amihud, 2002), a lower institutional ownership is associated with higher short sale constraints (Nagel, 2005), low analyst coverage and low firm age reflect higher information uncertainty (Zhang, 2006), and higher idiosyncratic volatility and smaller firms exhibit higher arbitrage costs.

To test whether the trading signal α_t^{IN} is driven by informational frictions, I perform a panel regression of the trading signal α_t^{IN} on the constituents of the index. I control for the permanent price impact of retailers α_t^{RE} and for VPIN, which is the probability of informed trading from high-frequency markets (Easley et al., 2012b). VPIN takes the daily volume from equation (3) and (4) and calculates absolute standardized order imbalance

$$VPIN_t = \frac{\alpha\mu}{\alpha\mu + 2\xi} = \frac{\mathbb{E}\left[|V_\tau^S - V_\tau^B|\right]}{\mathbb{E}(V_\tau^S + V_\tau^B)} \approx \frac{\sum_{\tau=1}^{n} |V_\tau^S - V_\tau^B|}{nV}, \tag{11}$$

where V is the number of daily volume buckets (here 50), n the number of trades in each bucket and $(\alpha\mu)$ the arrival rate of informed trades, and $(\alpha\mu + 2\cdot\xi)$ the arrival rate of all trades. Furthermore, I use effective spreads (obtained from 5 minute trades) and quoted spreads (obtained from end-of-day CRSP data) in my analysis.

Table 2 reports the results. All variables are cross-sectional standardized. I use the last variable of each quarter to avoid overlapping observations when using the price impact measures in the regression. Higher informational frictions directly translate to higher levels of price impact for institutions (columns 2), confirming my hypothesis. A one

Table 2: What explains the price impact?

	α_{10}^{IN}					α_{10}^{RE}				
(in bps)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
α_{10}^{IN}						1.20	1.12	1.28	0.96	0.76
						[7.97]	[8.37]	[9.12]	[7.24]	[3.98]
α_{10}^{RE}	1.58	1.65	1.30	1.38	0.72					
	[8.17]	[8.29]	[6.83]	[7.49]	[4.85]					
Age	−1.02					−1.57				
	[−0.65]					[−4.09]				
Illiq	7.79					43.09				
	[1.12]					[3.83]				
IVOL	−0.04					0.14				
	[−0.21]					[0.97]				
Size	−22.59					−0.70				
	[−24.02]					[−2.72]				
Analyst Cov.	−2.04					−0.05				
	[−12.65]					[−1.09]				
Insti Own.	−2.58					−0.57				
	[−10.13]					[−7.12]				
Inf. Friction		10.80		9.75			1.17		0.96	
		[25.19]		[21.46]			[9.71]		[8.16]	
$\ln(VPIN_t)$			4.80	3.24				1.16	0.98	
			[16.01]	[13.17]				[7.93]	[9.82]	
ES_t					10.22					1.39
					[12.77]					[1.50]
QS_t					−1.23					1.24
					[−2.46]					[1.78]
Entity effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs.	121015	121015	126427	121002	126426	121015	121015	126427	121002	126426
within R^2 [%]	18.8	20.8	7.4	22.7	20.6	-1.6	3.5	2.5	4.0	3.9

Note. The table shows standardized regression coefficients in basis points from panel regressions that regress the permanent price impact (α_t^j) on the constituents of the index of information frictions (columns (1) and (6)), on the index of information frictions itself (columns (2) and (7)), on the logarithm of $VPIN_t$ (columns (3) and (8)), on the logarithm of $VPIN_t$ and the index of information frictions (columns (4) and (9)), and on effective spreads obtained from 5 minute trade data and quoted spreads obtained from end-of-day CRSP data (columns (5) and (10)).

standard deviation increase in the arbitrage costs index translates to a 10.80bp increase in price impact for institutions. The R^2 is high with a value of 20.8%. When looking at the index constituents in column (1), the price impact measure α_t^{IN} heavily loads on size with a coefficient of -22.59bps, analyst coverage, and institutional ownership with the expected sign. Thus, small stocks, stocks with lower analyst coverage (higher information costs), and stocks with lower institutional ownership (higher short-sale costs) exhibit higher asymmetric information risk for institutions. VPIN in column (3) is positively correlated with the price impact, confirming that the price impact loads on information and adverse selection. Furthermore, the price impact positively correlates with volume weighted effective spreads calculated from level 1 trade data (TAQ). End-of-day quotes are not able to capture the price impact and are even negatively correlated, showing the importance of using illiquidity measures from high-frequency data. The results for α_t^{RE} show the expected sign for the constituents of the arbitrage cost index in column (6). Furthermore, age and illiquidity are highly significant and show the expected sign discussed above. Generally, the results for α_t^{RE} are economically weaker in magnitude and exhibit lower R^2, confirming my hypothesis, that limits-to-arbitrage, trading costs and asymmetric information are more of an issue for institutions than for retailers.

Overall, these results suggest, that the price impact is largely driven by limits-to-arbitrage, trading costs, and information. However, risk-based explanations are not ruled out by the arbitrage cost index. For example, illiquid stocks tend to have high betas, high idiosyncratic risks, and skewed fat-tailed distributions with volatility and jump risk premia (Bali et al., 2023). Thus, higher values of the arbitrage cost index indicate higher limits-to-arbitrage and a higher level of riskiness. The channels through which the price impact explains stock returns are thus consistent with both a risk-based explanation and the limits-to-arbitrage argument. The results for effective spreads (trading costs) are in line with the arguments of Barardehi et al. (2024) and might reflect wholesalers' compensation for liquidity provision when institutionals indirectly trade against retailers. The correlation with the VPIN indicates that the price impact also is related to informed trading in the stock market. Hence, the price impact is related to many risks in modern equity markets.

3.3. *Trading strategy on institutional price impact*

Next, I investigate whether excess returns generated by a size-neutral zero-cost long-short strategy, based on market equity and α_t^{IN}, are profitable and how it relates to traditional risk factors in the cross-section of stock returns. Table 1 shows that only α_t^{IN} has predictive power for future returns, while α_t^{RE} is neglectable in terms of cross-sectional predictability.[12] Hence, I perform the trading strategy only on institutionals price impact. First, I conduct a size-neutral long-short strategy based on α_t^{IN}. Specifically, on the last day of each month, I sort the stock universe in decile portfolios based on market capitalization. Within each size decile, ten portfolios are formed based on α_t^{IN}. I calculate value-weighted decile portfolio returns. Subsequently, I build a long-short strategy, by equally going long (short) the ten size portfolios in the highest (lowest) α_t^{IN} decile for each month. I calculate the HmL spread for each month. I hold the stocks for one month and close the position at the end of next months, such that the strategy is active for one month. To account for transaction costs, I substract the quoted half spread when entering the strategy and substract another quoted half spread when closing the position in the next month.[13]

Table 3 shows monthly value-weighted stock portfolio returns to dependent double sorted portfolios, first sorted on market capitalization (size) and then within each size portfolio on α_t^{IN}. The resulting HmL-spread conditional on size is positive and significant for all size quintiles except for the highest size quintile, confirming the results of (Easley et al., 2010). They are 1.72, 1.75, 1.58, 1.07 with t-values of 7.91, 4.58, 5.13, and 2.95 for the four smallest size quintiles, respectively. The highest size decile shows the expected

[12] Figure A3 shows that investing in the same trading strategy for retailers' price impact only yields small positive returns. When accounting for transaction costs, these returns are neglectably small.

[13] Furthermore, I implement transaction costs proposed by Frazzini, Israel, and Moskowitz (2018). The authors estimate that the approximate trading cost for value-weighted U.S. equities is about 12bps. When being conservative and changing 100% of the positions, rebalancing costs amount to $2 \cdot 12 = 24$bps per month (2.88% annual). When this adjustment is made, my trading strategy works better than using quoted half spreads as my HmL strategy exhibits effective costs of $20.29\% - 15.89\% = 4.39\%$ per year. Therefore, using quoted half spreads places a lower bound for my results.

Table 3: Dependent double sort on size and α_t^{IN}

r_{t+1} (in %)	Low Size_t	2	3	4	High Size_t	HmL Size_t	HmL t-Stat.
Low α_t^{IN}	-0.21	-0.11	-0.08	0.26	0.78	0.99	3.49
2	0.22	0.33	0.40	0.57	0.71	0.49	2.14
3	0.68	0.69	0.68	0.78	0.85	0.16	0.65
4	0.78	1.07	1.06	1.03	0.87	0.09	0.37
High α_t^{IN}	1.51	1.64	1.50	1.33	1.01	-0.50	-1.96
HmL α_t^{IN}	1.72	1.75	1.58	1.07	0.23	-1.49	
HmL t-Stat	7.91	4.58	5.13	2.95	0.59	-3.98	

Note. The table shows monthly value-weighted stock portfolio returns to dependent double sorted portfolios, first sorted on market capitalization (size) and then within each size portfolio on α_t^{IN}. The relevant HmL-portfolio is shown at in the last row, where I see the HmL return of investing in high-minus-low stocks of α_t^{IN} conditional on size.

sign but is statistically insignificant. This indicates that the expected returns of very large stocks are not affected. Table A1 shows the same results for retailers. The results are smaller in magnitude for retailers and only significant on a 10 percent level.[14]

Table 4 reports descriptive statistics on the HmL-strategy before (Panel A) and after (Panel B) transaction costs. $\text{HmL}_{\alpha_{10}^{IN}}$ yields an average excess return of 20.29% (15.89%)

Table 4: Size-neutral trading strategy before (Gross) and after (Net) transaction cost

	Avg. ret	t-Stat	Std	SR	Skew	Kurt
Panel A: Gross						
$\text{HmL}_{\alpha_{10}^{IN}}$	20.29	5.01	12.20	1.66	1.17	8.57
$\text{Long}_{\alpha_{10}^{IN}}$	18.31	2.85	24.30	0.75	-0.23	1.82
$\text{Short}_{\alpha_{10}^{IN}}$	1.99	0.35	18.11	0.11	0.68	2.43
$(r_m - r_f)$	8.93	1.97	15.87	0.56	-0.68	1.53
Panel B: Net						
$\text{HmL}_{\alpha_{10}^{IN}}$	15.89	4.06	12.18	1.30	1.07	8.33
$\text{Long}_{\alpha_{10}^{IN}}$	15.59	2.37	24.39	0.64	-0.27	1.84
$\text{Short}_{\alpha_{10}^{IN}}$	0.30	0.05	18.04	0.02	0.65	2.37
$(r_m - r_f)$	8.24	1.77	15.96	0.52	-0.72	1.66

Note. The table shows the economic value of buying (selling) size-neutral stock portfolios with high (low) asymmetric information. Specifically, I sort stocks in deciles by market capitalization, and within each decile, I sort deciles on α_t^{IN}. I calculate value-weighted returns. The strategy goes long in the ten highest α_t^{IN} portfolios and short the ten lowest α_t^{IN} portfolios each month. All values are annualized.

[14] Table A3 shows a similar plot on BE/ME to show that α_t^{IN} is not subsumed by value. Table A4 shows the same results for retailers.

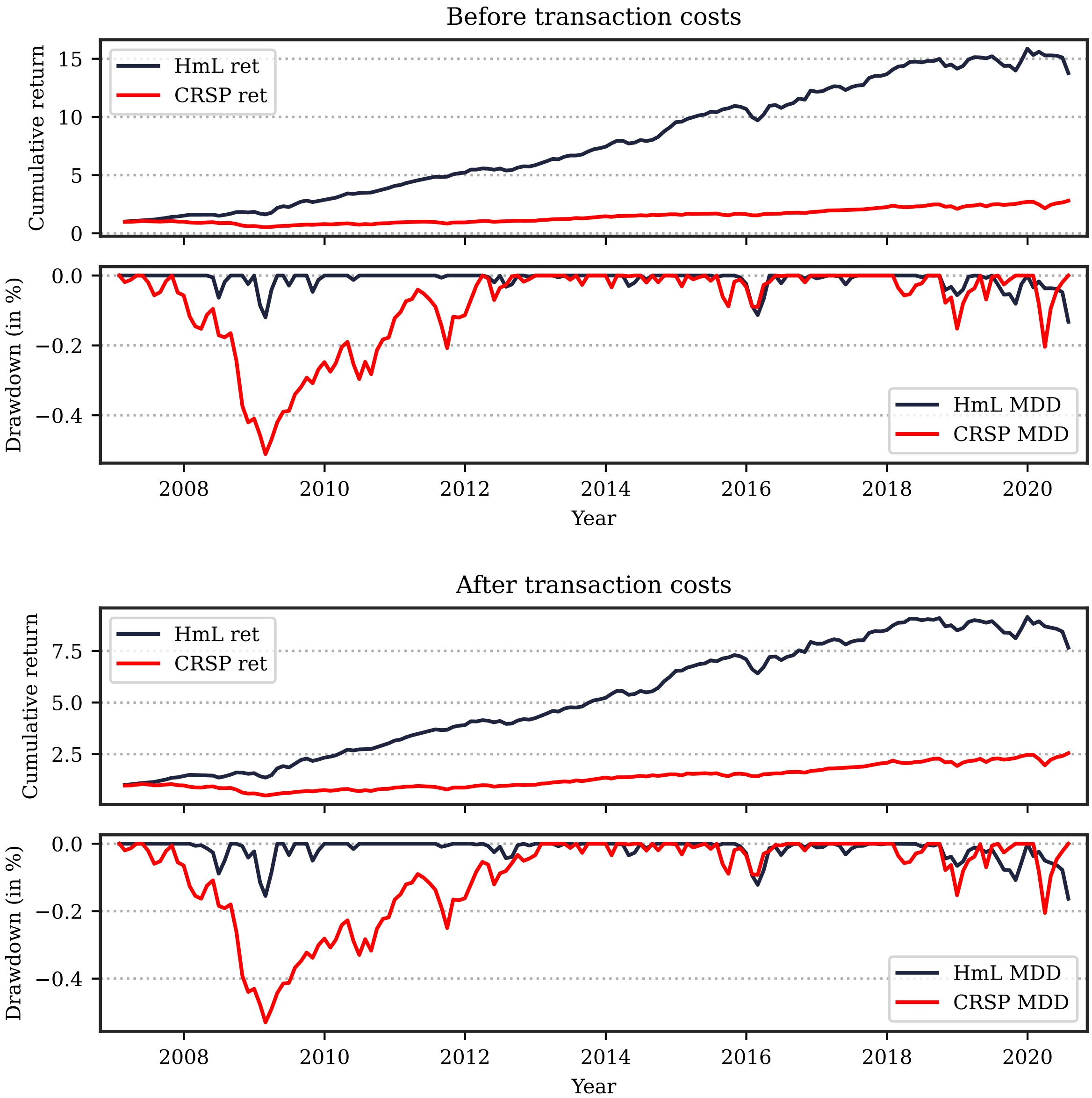
per year (after transaction costs), which is significant at the 1% level. The strategy outperforms a value-weighted investment in all stocks in my universe (20.29% vs. 8.93% and 15.89% vs. 8.24% after transaction costs) and yields a higher Sharpe Ratio (1.66 vs. 0.56 and 1.30 vs. 0.52 after transaction costs). The returns of the HmL strategy are positively skewed (compared to a negatively skewed market investment) and exhibit higher kurtosis compared to a normal distribution as I consider the deciles and hence the returns at the tails of the distribution. The strategy's performance is driven by the long portfolio, as the short portfolio is not significantly profitable. This is consistent with the reasoning that a high positive α_{10}^{IN} indicates high buying (selling) pressure with quotes rising (falling), while a low (and negative) α_{10}^{IN} signals that order flow and quotes appear to follow different directions. The size-neutral strategy for trading on retailers price impact is shown in Table A7. The strategy yields a sizeable SR of 0.85 (-0.18) before (after) transaction costs, showing that retailers' strategy is only profitable and significant before transaction costs.[15]

Figure 7 depicts the performance and drawdown curve of the size-neutral trading strategy and the value-weighted market excess return between 2007 and 2020. Over my 14-year sample period, the investment of 1$ would have grown to 13.77$ vs. a CRSP value-weighted return of 2.80$ (7.66$ vs. 2.55$ after transaction costs). Cumulative returns are smooth over time and do not exhibit significant structural breaks. The maximum drawdown shows that my trading strategy performs exceptionally well during crises and indicates fewer negative returns than the market. The strategy reveals a maximum drawdown of 13.25% vs. 51.13% of the market (16.32% vs. 52.98% after transaction costs). Trading on the price impact appears more profitable than investing in the value-weighted market portfolio.[16] Figure A3 shows the strategies performance for retailers with deciles and quintiles and Figure A4 with quintiles. The strategy cannot outperform a value-weighted investment in all stocks I consider in my analysis before transaction costs but still yields positive returns. The strategy becomes unprofitable after transaction costs.

[15] Table A8 shows the strategy's performance for retailers with quintiles instead of decile sorts.

[16] The strategy is robust to the number of portfolios. Figure A1 and Table A6 shows the strategy by first sorting quintiles on size and then quintiles on α_t^{IN}. Furthermore, Figure A2 shows the number of stocks considered in the long and short portfolios over time.

Fig. 7. Performance and drawdown curve



Note. The figure shows the cumulative performance of a one dollar investment into the size-neutral trading strategy as well as the drawdonw curve for the respective trading strategies. The black line (HmL ret) is the economic value of buying (selling) size-neutral stock portfolios with high (low) asymmetric information. Specifically I sort stocks in deciles by market capitalization and within each decile I sort stocks into deciles on α_t^{IN}. I calculate value-weighted returns. The strategy goes long in the ten highest α_t^{IN} portfolios and short the ten lowest α_t^{IN} portfolios each month. The red line (CRSP ret) depicts a value-weighted investment in the market. The sample period is January 2007 − July 2020.

Single sorts are shown in Table A12 and Figure A5 and A6 and reveal that the strategy is profitable for institutions and beats the market up until 2018-06-01 with a Sharpe Ratio of 0.77 (0.55) vs. the market 0.57 (0.52) before (after) transaction costs.

How many of these factors are explained by established risk factors in the equity market? Table 5 reveals monthly spanning regressions of HmL returns on the excess market return, the three-factor model (Fama and French, 1993), the three-factor model with momentum (Jegadeesh and Titman, 1993), the five-factor model (Fama and French, 2015), the liquidity factor (Pástor and Stambaugh, 2003), and a combination of all mentioned factors.[17] The intercept is always positive and highly significant, with an average monthly return of 1.69 (1.32 after transaction costs).[18] The trading strategy comoves positively with the market and loads on small, value, and illiquid stocks. Furthermore, it negatively comoves with momentum. The R^2 sharply increases when SMB is included in the regression (columns (2) and (9)). Furthermore, the R^2 is further increased when the liquidity factor is considered (columns (6) and (12)). With a maximum R^2 of 50.0%, common risk factors in the cross-section of equity returns cannot fully explain the returns of trading on the price impact, as the intercept remains positive and statistically significant in all regressions. Hence, there appears to be a premium for trading on the price impact in equity markets. The results for retailers are shown in Table A5. The strategy results significant alphas before transaction costs which vanish when accounting for transaction costs.

3.4. Retail trading reduces institutional risks

Retail trading increased tremendously in recent years. Brokerage platforms such as Robinhood (RH) allow small investors to participate in the stock market cheaper and

[17] Table A11 shows robustness tests for the spanning regressions in each size decile. The portfolio returns show large alphas. Table A10 shows that the HmL-betas do not differ significantly among different size portfolios, meaning that systematic risks does not differ too much.

[18] Table A11 shows spanning regressions within each size decile. The strategy yields positive and significant alphas for all size decile, except for the largest size decile. Table A9 further shows Fama and MacBeth (1973) regressions for the factor and the price impact itself, showing that the price impact is priced in equity markets.

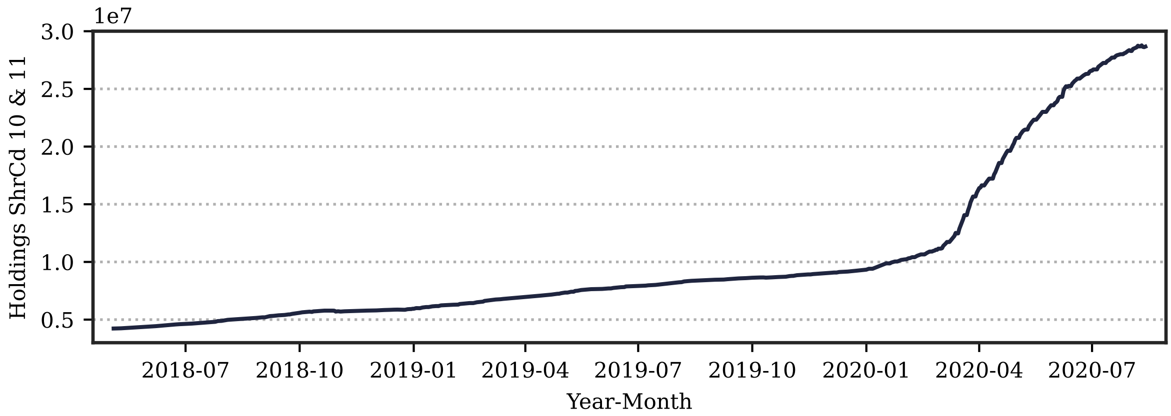
Table 5: Risk-adjusted returns

	$\text{HML}^{Gross}_{\alpha^{IN}_{10}}$						$\text{HML}^{Net}_{\alpha^{IN}_{10}}$					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
a (in %)	1.69	1.47	1.77	1.78	1.90	1.86	1.32	1.09	1.40	1.41	1.53	1.49
	[5.01]	[4.64]	[5.30]	[5.96]	[5.70]	[6.17]	[4.06]	[3.71]	[4.54]	[5.10]	[5.00]	[5.30]
$(r_m - r_f)$		0.30	0.13	0.07	0.09	0.07		0.32	0.15	0.08	0.10	0.08
		[2.79]	[1.64]	[0.86]	[1.13]	[1.04]		[2.88]	[1.78]	[1.07]	[1.28]	[1.24]
SMB			0.43	0.40	0.38	0.40			0.41	0.38	0.36	0.38
			[4.70]	[4.75]	[4.00]	[4.67]			[4.64]	[4.59]	[3.92]	[4.58]
HML			0.35	0.17	0.42	0.18			0.36	0.19	0.43	0.20
			[2.08]	[1.22]	[2.39]	[1.57]			[2.15]	[1.31]	[2.48]	[1.70]
RMW					−0.30	−0.30					−0.33	−0.32
					[−1.46]	[−1.70]					[−1.56]	[−1.86]
CMA					−0.23	−0.20					−0.23	−0.19
					[−1.66]	[−1.35]					[−1.70]	[−1.39]
UMD				−0.28		−0.25				−0.28		−0.25
				[−3.32]		[−2.85]				[−3.37]		[−2.86]
Liq						−0.13						−0.12
						[−2.10]						[−2.22]
Obs.	162	162	162	162	162	162	162	162	162	162	162	162
R^2 [%]	-0.0	15.7	35.1	45.6	37.4	49.1	0.0	17.1	36.4	46.3	39.0	50.0

Note. The table shows the estimation results from regressing HmL-returns on established risk-factors in the equity market, which are equity market excess return $(r_m - r_f)$, size (SMB), book-to-market (HML), profitability (RMW), investment (CMA), momentum (UMD), and Pástor and Stambaugh (2003) liquidity factor (Liq). Results are depicted before (gross) and after (net) transaction costs. The HmL-strategy sorts stocks in deciles by market capitalization and within each decile I sort stocks into deciles on α_t^{IN}. I calculate value-weighted returns. The strategy goes long in the ten highest α_t^{IN} portfolios and short the ten lowest α_t^{IN} portfolios each month. Newey and West (1986) robust t-statistics are in parentheses.

more accessible. These platforms attracted many investors in recent years, especially during the outbreak of COVID-19, when markets fell and recovered afterward (Welch, 2020). Hence, retailers have shown more extensive trading activities in recent years, as shown in Figure 1. Robinhood offered an API that made it possible to query the number of RH investors who held a particular stock at a specific time. In addition, the API provided hourly holdings for the entire cross-section of equities. I aggregate these holdings in Figure 8.[19] The sample period is May 2018 to August 2020, which was the period when the API was active. Retailers increased their holdings steadily, especially from March

Fig. 8. Daily Robinhood holdings



Note. The figure shows the sum of the total trading volume on Robinhood platform between May 2018 and August 2020 for all stocks with Sharecode 10 and 11.

2020 onwards (outbreak of COVID-19). Recently, Welch (2020) show that RH investors buy attention-grabbing stocks and stocks with high past share (dollar) volume. Boehmer et al. (2021) provides suggestive evidence that retail trades contain information not yet incorporated into prices. Barrot, Kaniel, and Sraer (2016) shows that retailers act as liquidity providers. Barardehi et al. (2024) shows that liquidity provision by retailers is conducted indirectly via wholesalers' who internalize retailer order flow and use this order flow to provide liquidity to institutionals.

This section aims to link the price impact for institutions to the trading activity of retailers. Does the trading activity of retailers suggest infored trading or is their coor-

[19] Data is available at the Robintrack Website.

dinated trading activity more in line with liquidity provision arguments? Welch (2020) shows that RH investors can move markets as a crowd and that this crowd-wisdom portfolio performs well in terms of alpha and timing. In this section, I select stocks that are widely bought or sold by retailers and examine whether these "crowd trades" affect the price impact of institutions. I find that the price impact of institutions is reduced heavily when retailers buy or sell a lot. Furthermore, I show that if retailers buy a lot, this reduces the price impact of retailers, which is consistent with liquiditiy provision of retailers via wholesalers' (Barardehi et al., 2024).

Crowd-wisdom portfolio. I follow Welch (2020) and define the alternative crowd (ARH) portfolio weight as

$$w_{i,t}^{ARH} = \frac{n_{i,t} \cdot P_{i,t}}{\sum_i n_{i,t} \cdot P_{i,t}}, \tag{12}$$

where $n_{i,t}$ are the number of RH investors investing in stock i at time t and $P_{i,t}$ the price of stock i at time t. I weight the RH holdings by price to obtain the dollar volume of the investment.[20] I calculate the absolute difference in weight changes as

$$|\Delta w_{i,t}^{ARH}| = |w_{i,t}^{ARH} - w_{i,t-1}^{ARH}|, \tag{13}$$

for each stock i. I can interpret $|\Delta w_{i,t}^{ARH}|$ as a specific stock's trading intensity measure. This measure is large when retailers buy or sell a lot. In the latter analysis, I aggregate my data quarterly to control for overlapping observations. Hence, I calculate the mean of $|\Delta w_{i,t}^{ARH}|$ over the last quarter.

On the last day of each quarter, I sort stocks into 100 portfolios based on $|\Delta w_{i,t}^{ARH}|$, such that each stock is assigned a number $pf_{i,t}$ between 1 and 100. I calculate the change

[20] One drawback is that each variable correlated with price (market capitalization, dollar trading volume,...) would be mechanically correlated with this portfolio investment weight. However, using $w_{i,t}^{ARH} = n_{i,t}/\sum_i n_{i,t}$ would mean that an investor holding two stocks worth 1\$ and 100\$ would assign a weight of 50% to each stock, which is counterintuitive. I tackle this issue with a matching approach that controls for size, return, and return standard deviation when comparing stocks.

in $pf_{i,t}$ as

$$\Delta pf_{i,t} = pf_{i,t} - pf_{i,t-1}. \tag{14}$$

If $\Delta pf_{i,t}$ is larger than 20, i.e., if a stock's $\Delta w_{i,t}^{ARH}$ moves up by more than 20% in the cross-sectional distribution, the stock is assigned to the treatment group (dummy of one). Afterwards, I calculate the change in α_{10}^{IN} and run the following cross-sectional regression

$$\Delta\alpha_{i,t}^{IN,10} = \hat{\beta} \cdot \mathbb{1}\left(\Delta pf_{i,t} > 20\right)_{i,t} + \hat{e}_{i,t}. \tag{15}$$

The results are depicted in Table 6 column (1). Stocks that retail investors heavily traded

Table 6: EOQ: DiD with 1-3 matching bulk fundamental

	$\lvert\Delta \mathrm{w}_t\rvert$							
	α_{10}^{IN}							
in (%)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{1}_{\Delta 20}$	−0.06	−0.06	−0.05	−0.06	−0.07	−0.05	−0.06	−0.06
	[−5.26]	[−3.66]	[−3.66]	[−5.78]	[−4.78]	[−4.77]	[−4.10]	[−4.17]
Obs.	17731	17731	17730	17730	17730	17731	17731	17730
R^2 [%]	1.3	0.4	0.4	0.5	0.6	0.4	0.5	0.5
ln(Size)		X		X	X		X	
Ret.			X	X	X			X
Std.					X	X	X	X
Entity effects	Y	Y	Y	Y	Y	Y	Y	Y
Time effects	Y	Y	Y	Y	Y	Y	Y	Y

Note. The table shows coefficients from panel regressions of regressing the change in $\alpha_{i,t}^{IN}$ on a dummy which is one when the absolute weight of the crowd-wisdom portfolio ($\lvert\Delta \mathrm{w}_{t,i}\rvert$) of stock i moves up by more than 20% in the cross-sectional distribution. Columns (2) to (8) show similar results by comparing stocks in a 1-3 matching approach on firm fundamentals, such as ln(size), return (Ret.), and return standard deviation (Std.).

via Robinhood (treatment group) have a lower price impact for institutions by −6bps. If the price impact of institutions is interpreted as a measure of informed trading and institutional trading costs, this interpretation is consistent with greater trading intensity by retailers leading to lower trading costs for institutions as retailers provide liquidity to institutional buying or selling pressure via wholesalers' which internalize retail trades.

Matching approach. I follow Cao, Goyal, Ke, and Zhan (2022) and conduct a matching approach. I match stocks according to characteristics in $t-1$ and compare their price impact in t. Specifically, I match according to market capitalization, return, and return volatility. Market capitalization captures the size and price of the firm and is thus correlated with firm fundamentals, such as institutional ownership and information transparency. Hence, matching according to market capitalization allows controlling for firm heterogeneity. I measure the Euclidean distance (norm) between characteristics at each point in time and choose the stock pairs such that it minimizes the norm

$$\min_{p,q} d(p,q) = \min_{p,q} ||p-q||_2 = \min_{p,q} \sum_{j=1}^{N} (p_j - q_j)^2 , \tag{16}$$

where j are different characteristics and p and q are potential stock pairs. The stocks with the smallest distance are compared against each other. I match stocks that moved more than 20% in $|\Delta w_{i,t}^{ARH}|$ (treatment) with stocks that exhibit smaller moves. The matching procedure is 1-3,[21] i.e., one treatment α_{10}^{IN} is matched with three control α_{10}^{IN}. I calculate the average price impact of all three α_{10}^{IN} in the control group in Equation (17) and take the time difference in Equation (18) as

$$\Delta\alpha_{i,t}^{IN,10} = \alpha_{i,t}^{IN,10,treat} - 1/C \sum_{c}^{C} \alpha_{i,t}^{IN,10,c} \tag{17}$$

$$\Delta\Delta\alpha_{i,t}^{IN,10} = \Delta\alpha_{i,t}^{IN,10} - \Delta\alpha_{i,t-1}^{IN,10}, \tag{18}$$

where C is the maximum number of firms in the control group. With 1-3 matching, this results $C = 3$. Afterward, I run the following regression

$$\Delta\Delta\alpha_{i,t}^{IN,10} = \hat{\beta} \cdot \mathbb{1}\left(\Delta pf_{i,t} > 20\right)_{i,t} + \hat{e}_{i,t}. \tag{19}$$

All regressions include day and firm fixed effects.

Table 6 shows the results in columns (2) – (8). The coefficient for $\mathbb{1}_{\Delta 20}$ is negative and

[21] All results are qualitatively the same if I do 1-1 matching.

highly statistically significant, even if I match according to market capitalization, stock return, and return volatility. Thus, a stock heavily traded by retailers from one quarter to the other exhibits 5bps to 7bps smaller price impact for institutions on average. Hence, I conclude that higher retail activity in stocks reduces trading costs and information risks for institutions. This finding is consistent with retailers indirectly providing liquidity to equity markets because retail trades are internalized by wholesalers. The wholesaler accumulates inventory and is able to offset this inventory with institutional buy/sell pressure (Barardehi et al., 2024). The finding is also consistent with retailers acting as a stabilizing force for equity markets and aligning the price impact of all market participants (Neuhann and Sockin, 2023). When interpreting the price impact as a measure of asymmetric information, this finding is in line with retailers impeding the price discovery process of institutionals and hence, decrease their share of informed trading. The finding is robust for using different measures of trading intensity by retailers and different control variables for the matching as can be seen from Table A13 to A17. It is also robust to using Lee and Ready (1991) instead of bulk classification (Table A18 and Table A19).

Signed retail trading. If institutionals are disturbed by retail trading, the sign of their trading activity (buying vs. selling) should also be important. My hypothesis is that more buying (selling) pressure from retailers should lead institutionals to trade in opposite directions of retailers, hence reducing institutionals price impact. Consider the GME example in January 2021. Retailers were heavily buying GME, while institutionals were mainly short in this stock. Hence, when retailers buy a lot, institutionals might respond with less buying pressure (and even shorting stocks which retailers buy). This contrarian trading is also in line with liquidity provision by retailers for institutional buying/selling pressure (Barrot et al., 2016). Hence, institutionals price impact should react differently when using Δw_t, instead of $|\Delta w_t|$. Table 7 shows the results. Higher directional trading on the Robinhood platform significantly reduces the price impact of institutionals by 2bps per quarter. This is in line with retailers aligning the price impact of all market participants (Neuhann and Sockin, 2023) and retailers providing liquidity

Table 7: EOQ: DiD with 1-3 matching bulk fundamental

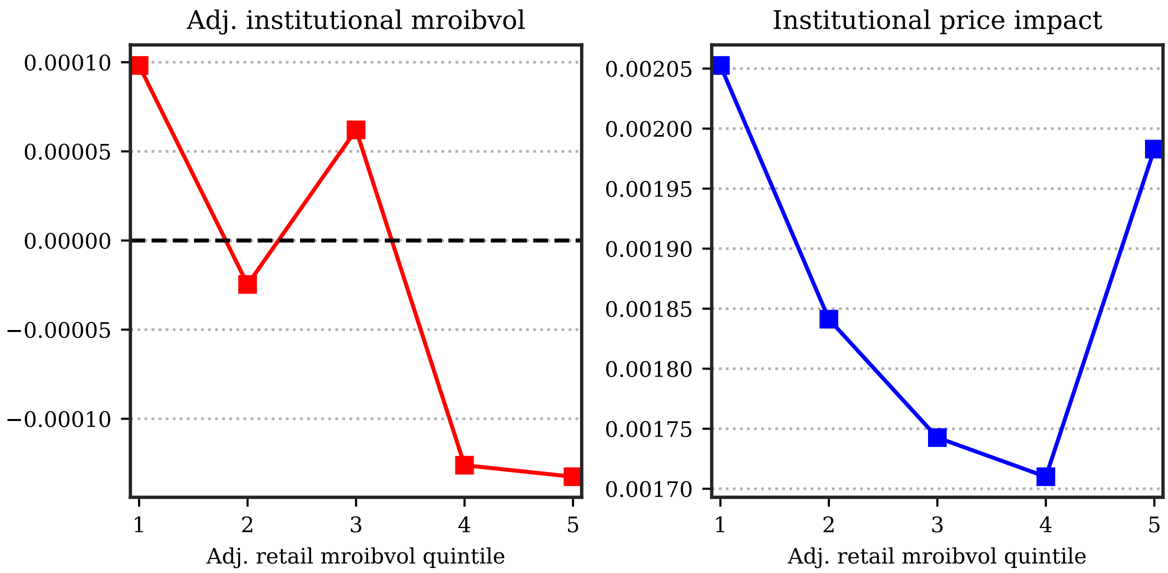
	Δw_t							
	α_{10}^{IN}							
in (%)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{1}_{\Delta 30}$	-0.02	-0.01	-0.02	-0.01	-0.01	-0.02	-0.02	-0.02
	$[-4.56]$	$[-2.08]$	$[-4.45]$	$[-2.50]$	$[-2.35]$	$[-4.08]$	$[-2.96]$	$[-3.58]$
Obs.	17731	17731	17730	17730	17730	17731	17731	17730
R^2 [%]	0.5	0.1	0.2	0.1	0.1	0.2	0.2	0.2
ln(Size)		X		X	X		X	
Ret.			X	X	X			X
Std.					X	X	X	X
Entity effects	Y	Y	Y	Y	Y	Y	Y	Y
Time effects	Y	Y	Y	Y	Y	Y	Y	Y

Note. The table shows coefficients from panel regressions of regressing the change in $\alpha_{i,t}^{IN}$ on a dummy which is one when the weight of the crowd-wisdom portfolio ($\Delta w_{t,i}$) of stock i moves up by more than 30% in the cross-sectional distribution. Columns (2) to (8) show similar results by comparing stocks in a 1-3 matching approach on firm fundamentals, such as ln(size), return (Ret.), and return standard deviation (Std.).

to institutionals via wholesalers (Barardehi et al., 2024). Furthermore, it is in line with Koijen and Yogo (2019) and shows that the demand of institutionals reacts more inelastic compared to the demand of retail investors and goes in opposite direction. The finding is more robust when using a 30% jump (as in Table 7) in the cross-sectional distribution compared to a 20% jump in the distribution (Table A20), showing that the price impact of institutionals is smaller (larger) when retailers buy (sell) heavily.

Evidence of liquidity provision. Most retail trades are executed off-exchange, either sold by the broker to a wholesaler or filled from the broker's inventory. While obtaining data on wholesaler trades is hard, Boehmer et al. (2021) provides an algorithm that identifies a subset of retail trades executed by wholesalers in TAQ. Barber et al. (2023) find that these trades' standardized imbalances ($Mroib$) are measured with error. However, Barardehi et al. (2024) show that $|Mroib|$ is a good proxy for the intensity with which wholesalers provide liquidity in illiquid market conditions. They show that $Mroibvol$ from TAQ and institutional trades in the proprietary ANcerno data (which measures institutional trades) are negatively correlated and that a measure of institutional price impact (calculated from ANcerno) and *Mroibvol* have a U-shaped pattern. I replicate their finding for the volume of retail trades (*Mroibvol*), the institutional trade imbalances, and the price impact I measure with TAQ data. Figure 9 shows the results. The

Fig. 9. Liqudity provision of retailers



Note. The figure plots institutional standardized net volume (Institutional *Mroibvol*) and institutional price impact from TAQ against retail standardized net volume (Retail *Mroibvol*). Each day, I subtract the cross-sectional average of *Mroibvol* (market adjustment). I aggregate the volume to the quarterly frequency using the mean volume. Each quarter I sort stocks on retail *Mroibvol* and calculate the mean institutional market adjusted institutional *Mroibvol* (left) and institutional price impact (right) for each quarter and each quintile. The time series averages are plotted in above figure. The time period is January 2007 until July 2020.

highest (lowest) quintile for *Mroibvol* (very positive retail TAQ imbalance) is associated with negative (positive) institutional imbalances from TAQ. Furthermore, retailers' most positive and negative order imbalance is associated with TAQ's highest institutional price impact. Figure A7 shows that this effect is even more pronounced when looking at the Robinhood sample from May 2018 to July 2020. Table A22 shows that the difference between standardized institutional trading volume in the high retail trading volume minus low retail trading volume quintile is negative and statistically significant.

This finding shows that a high price impacts proxies for trading costs of institutions when liquidity is scarce and that institutions trade in opposing directions in the most extreme quintiles. Hence, a very high price impact can also be interpreted as a proxy for liquidity provision by retailers for institutions via wholesalers' when liquidity is scarce in equity markets (Barrot et al., 2016; Barardehi et al., 2024).

4. Conclusion

I measure the price impact in high-frequency equity markets for retail investors and institutionals using high-frequency trade and quote data from TAQ. I find a heterogeneous price impact among retailers and institutionals. I find that the price impact is driven by informational frictions, illiquidity (measured from high-frequency data), and information. Furthermore, the price impact varies with the business cycle and is high when overall economic risk aversion is high. Within the day, the price impact is strongest in the first 30 minutes and the last 30 minutes of the day.

A significant implication of my research is the profitability of a size-neutral trading strategy based on the price impact of institutionals, both before and after transaction costs. In contrast, a similar strategy based on the price impact of retailers yields a substantial Sharpe Ratio but becomes unprofitable after accounting for transaction costs. This suggests that while the price impact is priced in the cross-section for both investor groups, only the price impact of institutionals is substantial enough to be profitable for trading. This finding aligns with the concept of risk compensation for trading costs and information, offering valuable guidance for finance professionals in their trading decisions.

I find that extensive trading of Robinhood investors (which either shows coordinated buying or selling pressure) decreases the price impact of institutionals. As wholesalers internalize retail trades, this finding is consistent with liquidity provision from retail investors to institutional investors via wholesalers' when liquidity is scarce. Furthermore, coordinated buying of retail investors on Robinhood also leads to a decrease in the price impact of institutionals, consistent with the inelastic demand of institutional investors, showing that retailers prefer different stocks than institutional investors.

References

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of financial markets 5, 31–56.

Atilgan, Y., Bali, T. G., Demirtas, K. O., Gunaydin, A. D., 2020. Left-tail momentum: Underreaction to bad news, costly arbitrage and equity returns. Journal of Financial Economics 135, 725–753.

Babenko, I., Boguth, O., Tserlukevich, Y., 2016. Idiosyncratic cash flows and systematic risk. The Journal of Finance 71, 425–456.

Baldauf, M., Mollner, J., Yueshen, B. Z., 2024. Siphoned apart: A portfolio perspective on order flow segmentation. Journal of Financial Economics 154, 103807.

Bali, T. G., Beckmeyer, H., Moerke, M., Weigert, F., 2023. Option return predictability with machine learning and big data. The Review of Financial Studies .

Barardehi, Y. H., Bernhardt, D., Da, Z., Warachka, M., 2024. Uncovering the liquidity premium in stock returns using sub-penny trade executions. Working Paper .

Barber, B. M., Huang, X., Jorion, P., Odean, T., Schwarz, C., 2023. A (sub) penny for your thoughts: Tracking retail investor activity in taq. Available at SSRN 4202874 .

Barbon, A., Beckmeyer, H., Buraschi, A., Moerke, M., 2021. The role of leveraged etfs and option market imbalances on end-of-day price dynamics. Working Paper .

Barrot, J.-N., Kaniel, R., Sraer, D., 2016. Are retail traders compensated for providing liquidity? Journal of Financial Economics 120, 146–168.

Battalio, R., Corwin, S. A., Jennings, R., 2016. Can brokers have it all? on the relation between make-take fees and limit order execution quality. Journal of Finance 71, 2193–2238.

Battalio, R. H., Jennings, R. H., 2023. On the potential cost of mandating qualified auctions for marketable retail orders. Available at SSRN .

Boehmer, E., Jones, C. M., Zhang, X., Zhang, X., 2021. Tracking retail investor activity. Journal of Finance 76, 2249–2305.

Borochin, P., Rush, S., 2016. Identifying and pricing adverse selection risk with vpin. Available at SSRN 2599871 .

Brennan, M. J., Huh, S.-W., Subrahmanyam, A., 2016. Asymmetric effects of informed trading on the cost of equity capital. Management Science 62, 2460–2480.

Cao, J., Goyal, A., Ke, S., Zhan, X., 2022. Options trading and stock price informativeness. Swiss Finance Institute Research Paper .

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

Chakrabarty, B., Li, B., Nguyen, V., Van Ness, R. A., 2007. Trade classification algorithms for electronic communications network trades. Journal of Banking & Finance 31, 3806–3821.

Collin-Dufresne, P., Fos, V., 2015. Do prices reveal the presence of informed trading? The Journal of Finance 70, 1555–1582.

Di Maggio, M., Egan, M., Franzoni, F., 2022. The value of intermediation in the stock market. Journal of Financial Economics 145, 208–233.

Easley, D., Hvidkjaer, S., O'hara, M., 2002. Is information risk a determinant of asset returns? The journal of finance 57, 2185–2221.

Easley, D., Hvidkjaer, S., O'hara, M., 2010. Factoring information into returns. Journal of Financial and Quantitative Analysis 45, 293–309.

Easley, D., Kiefer, N. M., O'hara, M., Paperman, J. B., 1996. Liquidity, information, and infrequently traded stocks. The Journal of Finance 51, 1405–1436.

Easley, D., Lopez de Prado, M., O'Hara, M., 2012a. Bulk classification of trading activity. Johnson School Research Paper Series 8, 14.

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

Easley, D., O'hara, M., 2004. Information and the cost of capital. The journal of finance 59, 1553–1583.

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, 502–528.

Eaton, G. W., Irvine, P. J., Liu, T., 2021. Measuring institutional trading costs and the implications for finance research: The case of tick size reductions. Journal of Financial Economics 139, 832–851.

Ellis, K., Michaely, R., O'Hara, M., 2000. The accuracy of trade classification rules: Evidence from nasdaq. Journal of Financial and Quantitative Analysis 35, 529–551.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of financial economics 33, 3–56.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of financial economics 116, 1–22.

Fama, E. F., MacBeth, J. D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of political economy 81, 607–636.

Frazzini, A., Israel, R., Moskowitz, T. J., 2018. Trading costs. Working Paper .

Friedman, H. L., Zeng, Z., 2022. Retail investor trading and market reactions to earnings announcements. Available at SSRN 3817979 .

Gârleanu, N., Pedersen, L. H., 2004. Adverse selection and the required return. The Review of Financial Studies 17, 643–665.

Glosten, L. R., Harris, L. E., 1988. Estimating the components of the bid/ask spread. Journal of financial Economics 21, 123–142.

Glosten, L. R., Milgrom, P. R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of financial economics 14, 71–100.

Gonçalves, A., 2021. The short duration premium. Journal of Financial Economics 141, 919–945.

Grossman, S. J., Miller, M. H., 1988. Liquidity and market structure. the Journal of Finance 43, 617–633.

Hasbrouck, J., 1988. Trades, quotes, inventories, and information. Journal of financial economics 22, 229–252.

Hasbrouck, J., 1991a. Measuring the information content of stock trades. Journal of Finance 46, 179–207.

Hasbrouck, J., 1991b. The summary informativeness of stock trades: An econometric analysis. Review of Financial Studies 4, 571–595.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. The Journal of finance 48, 65–91.

Koijen, R. S., Yogo, M., 2019. A demand system approach to asset pricing. Journal of Political Economy 127, 1475–1515.

Kyle, A. S., 1985. Continuous auctions and insider trading. Econometrica: Journal of the Econometric Society pp. 1315–1335.

Lee, C. M., Ready, M. J., 1991. Inferring trade direction from intraday data. Journal of Finance 46, 733–746.

Liu, H., Wang, Y., 2016. Market making with asymmetric information and inventory risk. Journal of Economic Theory 163, 73–109.

Nagel, S., 2005. Short sales, institutional investors and the cross-section of stock returns. Journal of Financial Economics 78, 277–309.

Neuhann, D., Sockin, M., 2023. Public asset purchases and private risk sharing. Working Paper .

Newey, W. K., West, K. D., 1986. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Pástor, L., Stambaugh, R. F., 2003. Liquidity risk and expected stock returns. Journal of Political economy 111, 642–685.

Pontiff, J., 2006. Costly arbitrage and the myth of idiosyncratic risk. Journal of Accounting and Economics 42, 35–52.

Ranaldo, A., Somogyi, F., 2021. Asymmetric information risk in FX markets. Journal of Financial Economics 140, 391–411.

Shleifer, A., Vishny, R. W., 1997. The limits of arbitrage. The Journal of finance 52, 35–55.

Stoll, H. R., 1978. The supply of dealer services in securities markets. The Journal of Finance 33, 1133–1151.

Van Kervel, V., Menkveld, A. J., 2019. High-frequency trading around large institutional orders. The Journal of Finance 74, 1091–1137.

Wang, J., 1993. A model of intertemporal asset prices under asymmetric information. The Review of Economic Studies 60, 249–282.

Wang, J., 1994. A model of competitive stock trading volume. Journal of political Economy 102, 127–168.

Welch, I., 2020. The wisdom of the robinhood crowd. Tech. rep., National Bureau of Economic Research.

Wiedemann, T., 2022. Quantifying the beta estimation bias and its implications for empirical asset pricing. Working Paper .

Yang, Y. C., Zhang, B., Zhang, C., 2020. Is information risk priced? evidence from abnormal idiosyncratic volatility. Journal of Financial Economics 135, 528–554.

Zhang, X. F., 2006. Information uncertainty and stock returns. Journal of Finance 61, 105–137.

Appendix A. Appendix

Table A1: EOM: Dependent double sort

Dependent double sort							
r_{t+1} (in %)	Low Size	2	3	4	High Size	HmL Size	HmL t-Stat.
Low α_t^{RE}	0.40	0.43	0.64	0.63	0.86	0.46	1.61
2	0.46	0.64	0.60	0.79	0.78	0.32	1.14
3	0.60	0.68	0.57	0.81	0.79	0.19	0.71
4	0.66	0.90	0.77	0.80	0.78	0.12	0.59
High α_t^{RE}	0.76	0.85	0.90	0.91	0.74	-0.01	-0.05
HmL α_t^{RE}	0.36	0.42	0.26	0.28	-0.11	-0.47	
HmL t-Stat	1.87	2.88	1.86	1.88	-0.58	-1.52	

Note. The table shows the returns to value-weighted stock portfolios first sorted by market equity, and then, within each size quintile, by α_t^{RE}.

Table A2: EOM: Dependent double sort

Dependent double sort							
r_{t+1} (in %)	Low α_t^{IN}	2	3	4	High α_t^{IN}	HmL α_t^{IN}	HmL t-Stat.
Low Size_t	-0.41	0.12	0.08	0.30	0.92	1.33	5.65
2	0.07	0.55	0.44	0.46	1.04	0.97	3.94
3	0.36	0.64	0.72	0.84	1.15	0.79	2.76
4	0.57	0.83	0.87	1.27	1.68	1.11	2.76
High Size_t	0.82	0.74	1.11	1.07	1.56	0.74	1.66
HmL Size_t	1.23	0.62	1.03	0.77	0.64	-0.59	
HmL t-Stat	4.89	4.23	4.01	3.33	1.62	-1.22	

Note. The table shows the returns to value-weighted stock portfolios first sorted by α_t^{IN}, and then by market equity.

Table A3: EOM: Dependent double sort BE/ME and α_t^{IN}

Dependent double sort							
r_{t+1} (in %)	Low BE/ME	2	3	4	High BE/ME	HmL BE/ME	HmL t-Stat.
Low α_t^{RE}	0.95	0.62	0.47	0.25	0.31	-0.64	-1.46
2	1.26	0.81	0.81	0.39	0.09	-1.17	-2.79
3	1.46	1.02	0.79	0.59	0.59	-0.87	-2.10
4	1.65	1.24	1.04	0.78	0.82	-0.83	-1.73
High α_t^{RE}	1.80	1.56	1.28	0.97	1.14	-0.66	-1.12
HmL α_t^{RE}	0.85	0.94	0.81	0.72	0.83	-0.02	
HmL t-Stat	2.20	2.99	2.28	2.62	2.04	-0.03	

Note. The table shows the returns to value-weighted stock portfolios first sorted by book-to-market-equity, and then, within each BE/ME quintile, by α_t^{IN}.

Table A4: EOM: Dependent double sort BE/ME and α_t^{RE}

Dependent double sort							
r_{t+1} (in %)	Low BE/ME	2	3	4	High BE/ME	HmL BE/ME	HmL t-Stat.
Low α_t^{RE}	1.25	0.75	0.74	0.31	0.15	-1.10	-2.37
2	1.24	0.68	0.62	0.44	0.52	-0.72	-1.71
3	1.12	0.88	0.61	0.19	0.22	-0.90	-2.45
4	1.21	0.82	0.92	0.56	0.44	-0.77	-1.35
High α_t^{RE}	0.97	0.99	0.60	0.76	0.83	-0.14	-0.22
HmL α_t^{RE}	-0.28	0.25	-0.14	0.46	0.69	0.96	
HmL t-Stat	-0.69	1.22	-0.70	2.57	2.07	1.76	

Note. The table shows the returns to value-weighted stock portfolios first sorted by book-to-market-equity, and then, within each BE/ME quintile, by α_t^{RE}.

Table A5: EOM: Spanning regressions retail

	$HML^{Gross}_{\alpha^{RE}_{10}}$						$HML^{Net}_{\alpha^{RE}_{10}}$					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
a (in %)	0.34	0.30	0.29	0.29	0.30	0.31	−0.07	−0.13	−0.14	−0.14	−0.12	−0.11
	[3.84]	[3.57]	[3.77]	[3.76]	[3.47]	[3.51]	[−0.68]	[−1.25]	[−1.46]	[−1.46]	[−1.12]	[−1.11]
$(r_m - r_f)$		0.06	0.05	0.04	0.04	0.04		0.07	0.06	0.06	0.06	0.05
		[1.87]	[1.67]	[1.88]	[1.21]	[1.31]		[2.26]	[2.08]	[2.22]	[1.55]	[1.73]
SMB			0.09	0.09	0.08	0.07			0.07	0.07	0.07	0.06
			[1.59]	[1.56]	[1.72]	[1.48]			[1.31]	[1.31]	[1.32]	[1.10]
HML			−0.05	−0.05	−0.05	−0.05			−0.04	−0.04	−0.04	−0.03
			[−0.69]	[−0.76]	[−0.76]	[−0.67]			[−0.61]	[−0.59]	[−0.67]	[−0.46]
RMW					−0.04	−0.04					−0.06	−0.06
					[−0.38]	[−0.39]					[−0.63]	[−0.64]
CMA					−0.00	0.01					−0.00	0.01
					[−0.03]	[0.09]					[−0.05]	[0.07]
UMD				−0.01		−0.01				0.00		−0.00
				[−0.21]		[−0.30]				[0.05]		[−0.10]
Liq						0.02						0.02
						[0.70]						[0.81]
Obs.	162	162	162	162	162	162	162	162	162	162	162	162
R^2 [%]	0.0	3.4	5.4	5.5	5.6	5.8	0.0	5.3	6.8	6.8	7.2	7.4

Note. The table shows the economic value of buying (selling) size-neutral stock portfolios with high (low) price impact. Specifically I sort stocks in deciles by market capitalization and within each decile I sort deciles on α^{RE}_t. I calculate value-weighted returns. The strategy goes long in the ten highest α^{RE}_t portfolios and short the ten lowest α^{RE}_t portfolios each month. All values are annualized.

Table A6: EOM:Size-neutral trading strategy insti (5,5)

	Avg. ret	t-Stat	Std	SR	Skew	Kurt
Panel A: Gross						
$HmL_{\alpha^{IN}_{10}}$	15.12	4.75	9.64	1.57	1.17	7.90
$Long_{\alpha^{IN}_{10}}$	16.39	2.74	23.22	0.71	-0.30	1.82
$Short_{\alpha^{IN}_{10}}$	-1.27	-0.24	17.78	-0.07	0.76	2.46
$(r_m - r_f)$	8.93	1.97	15.87	0.56	-0.68	1.53
Panel B: Net						
$HmL_{\alpha^{IN}_{10}}$	11.31	3.71	9.62	1.18	1.09	7.62
$Long_{\alpha^{IN}_{10}}$	14.03	2.29	23.30	0.60	-0.34	1.86
$Short_{\alpha^{IN}_{10}}$	-2.73	-0.54	17.69	-0.15	0.73	2.37
$(r_m - r_f)$	8.24	1.77	15.96	0.52	-0.72	1.66

Note. The table shows the economic value of buying (selling) size-neutral stock portfolios with high (low) price impact. Specifically, I sort stocks in quintiles by market capitalization, and within each quintile, I sort quintiles on α^{IN}_t. I calculate value-weighted returns. The strategy goes long in the five highest α^{IN}_t portfolios and short the five lowest α^{IN}_t portfolios each month. All values are annualized.

Table A7: EOM:Size-neutral trading strategy retail

	Avg. ret	t-Stat	Std	SR	Skew	Kurt
Panel A: Gross						
$\text{HmL}_{\alpha_{10}^{IN}}$	4.13	3.84	4.87	0.85	0.87	3.83
$\text{Long}_{\alpha_{10}^{IN}}$	10.58	1.93	20.47	0.52	-0.38	1.24
$\text{Short}_{\alpha_{10}^{IN}}$	-6.45	-1.24	19.31	-0.33	0.55	1.19
$(r_m - r_f)$	8.93	1.97	15.87	0.56	-0.68	1.53
Panel B: Net						
$\text{HmL}_{\alpha_{10}^{IN}}$	-0.88	-0.68	4.97	-0.18	0.70	3.57
$\text{Long}_{\alpha_{10}^{IN}}$	7.58	1.33	20.59	0.37	-0.43	1.31
$\text{Short}_{\alpha_{10}^{IN}}$	-8.46	-1.69	19.21	-0.44	0.52	1.11
$(r_m - r_f)$	8.24	1.77	15.96	0.52	-0.72	1.66

Note. The table shows the economic value of buying (selling) size-neutral stock portfolios with high (low) price impact. Specifically, I sort stocks in deciles by market capitalization, and within each decile, I sort deciles on α_t^{RE}. I calculate value-weighted returns. The strategy goes long in the ten highest α_t^{RE} portfolios and short the ten lowest α_t^{RE} portfolios each month. All values are annualized.

Table A8: EOM:Size-neutral trading strategy retail (5,5)

	Avg. ret	t-Stat	Std	SR	Skew	Kurt
Panel A: Gross						
$\text{HmL}_{\alpha_{10}^{IN}}$	2.83	3.26	3.56	0.79	0.15	1.53
$\text{Long}_{\alpha_{10}^{IN}}$	9.55	1.76	20.19	0.47	-0.58	1.57
$\text{Short}_{\alpha_{10}^{IN}}$	-6.72	-1.30	19.74	-0.34	0.65	1.49
$(r_m - r_f)$	8.93	1.97	15.87	0.56	-0.68	1.53
Panel B: Net						
$\text{HmL}_{\alpha_{10}^{IN}}$	-1.28	-1.30	3.63	-0.35	0.03	1.44
$\text{Long}_{\alpha_{10}^{IN}}$	7.13	1.28	20.29	0.35	-0.62	1.65
$\text{Short}_{\alpha_{10}^{IN}}$	-8.41	-1.67	19.65	-0.43	0.62	1.41
$(r_m - r_f)$	8.24	1.77	15.96	0.52	-0.72	1.66

Note. The table shows the economic value of buying (selling) size-neutral stock portfolios with high (low) price impact. Specifically, I sort stocks in quintiles by market capitalization, and within each quintile, I sort quintiles on α_t^{RE}. I calculate value-weighted returns. The strategy goes long in the five highest α_t^{RE} portfolios and short the five lowest α_t^{RE} portfolios each month. All values are annualized.

Table A9: Fama and MacBeth (1973) regression and risk-premium

(in %) · 12	(Size neutral factor)$_t$ Insti	(Size neutral factor)$_t$ Retail	$\alpha^{10}_{i,t}$ Insti	$\alpha^{10}_{i,t}$ Retail
$\bar{\lambda}$	13.36	8.04	1.68	0.16
	[2.53]	[2.73]	[4.23]	[4.05]
Obs.	163	163	163	163
adj R^2 [%]	0.0	0.0	0.0	0.0

Note. The table shows Fama and MacBeth (1973) regressions for the size-neutral trading strategy factor and for using the price impact directly.

Table A10: EOM: Robustness: Betas across portfolios

	$HML^{Gross}_{\alpha^{IN}_{10}}$ Low	$HML^{Gross}_{\alpha^{IN}_{10}}$ (2)	$HML^{Gross}_{\alpha^{IN}_{10}}$ High
$(r_m - r_f)$	0.12	0.10	0.16
	[0.93]	[0.79]	[1.06]
Obs.	163	163	163
R^2 [%]	1.0	0.6	1.3

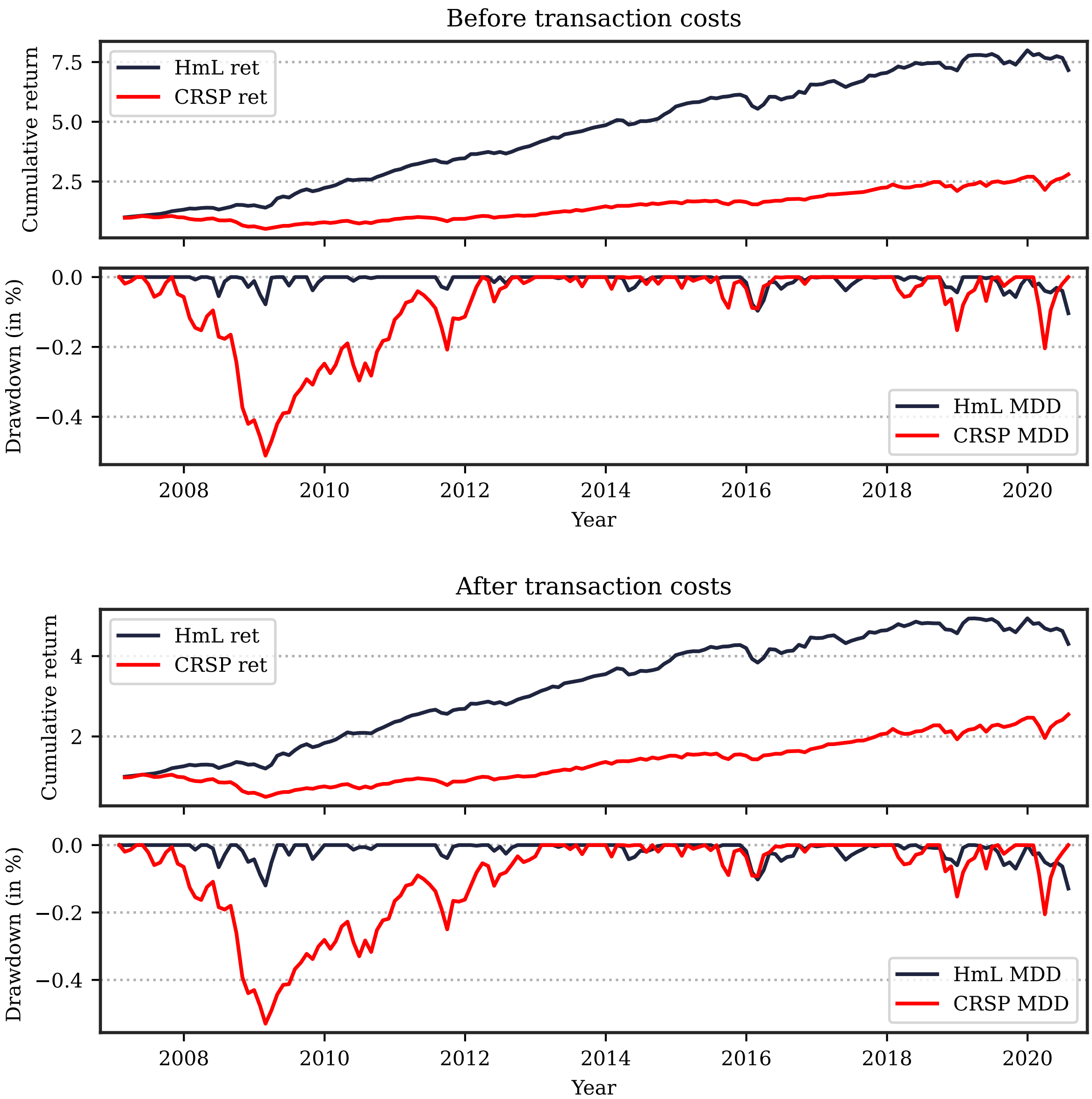
Note. The table shows beta coefficients of different portfolios. The beta w.r.t. the market do not differ too much among size portfolios.

Table A11: EOM: Alphas for each subportfolio

	$HML^{Gross}_{\alpha^{IN}_{10}}$									
	Small	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	Large
a (in %)	1.06	1.30	1.50	1.44	1.62	1.09	1.15	0.84	0.62	0.25
	[3.20]	[5.57]	[5.33]	[5.11]	[5.84]	[3.15]	[3.90]	[2.42]	[2.69]	[0.96]
$(r_m - r_f)$	0.07	0.08	0.06	−0.06	−0.04	0.06	−0.05	−0.00	0.08	0.03
	[0.54]	[1.29]	[1.12]	[−1.08]	[−0.77]	[1.25]	[−1.07]	[−0.01]	[1.46]	[0.34]
SMB	−0.00	−0.08	0.12	−0.07	−0.08	−0.06	−0.00	−0.04	−0.13	−0.15
	[−0.05]	[−0.86]	[1.29]	[−0.58]	[−0.76]	[−0.65]	[−0.04]	[−0.56]	[−1.69]	[−1.88]
HML	−0.09	0.12	−0.01	0.21	0.07	0.02	0.04	0.06	0.10	0.15
	[−1.22]	[1.03]	[−0.18]	[1.70]	[0.67]	[0.22]	[0.33]	[0.71]	[1.08]	[1.17]
RMW	0.04	−0.01	0.04	−0.23	−0.13	−0.22	−0.23	−0.07	−0.27	−0.23
	[0.23]	[−0.03]	[0.22]	[−1.71]	[−0.72]	[−1.74]	[−1.93]	[−0.63]	[−1.92]	[−2.07]
CMA	−0.32	−0.34	−0.19	−0.14	−0.07	0.02	−0.15	−0.03	−0.09	−0.25
	[−2.31]	[−1.46]	[−1.13]	[−0.79]	[−0.35]	[0.10]	[−0.97]	[−0.17]	[−0.71]	[−1.34]
Obs.	163	163	163	163	163	163	163	163	163	163
R^2 [%]	3.9	4.2	3.4	4.1	1.0	2.5	2.5	0.4	6.1	3.6

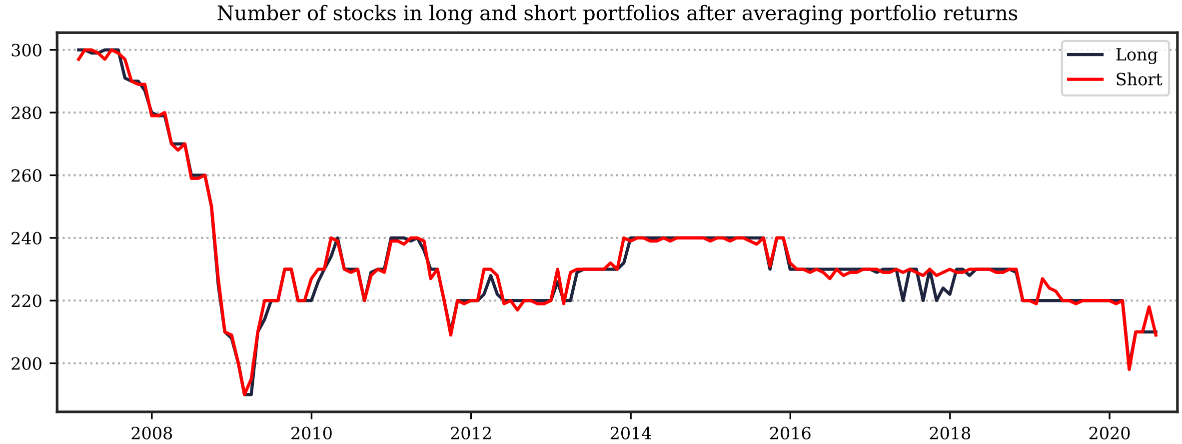
Note. The table shows spanning regressions for each size decile portfolio. Alphas are significant and positive for all decile portfolio HmL returns except for the largest decile.

Fig. A1. Performance and drawdown curve (Robustness $(5, 5)$)



Note. The figure shows the cumulative performance of a one dollar investment into the size-neutral trading strategy as well as the drawdonw curve for the respective trading strategies. The black line (HmL ret) is the economic value of buying (selling) size-neutral stock portfolios with high (low) price impact. Specifically we sort stocks in quintiles by market capitalization and within each quintile we sort stocks into deciles on α_t^{IN}. We calculate value-weighted returns. The strategy goes long in the ten highest α_t^{IN} portfolios and shorts the ten lowest α_t^{IN} portfolios each month. The red line (CRSP ret) depicts a value-weighted investment in the market. The sample period is January 2007 − July 2020.

Fig. A2. Number of stocks in long and short portfolio



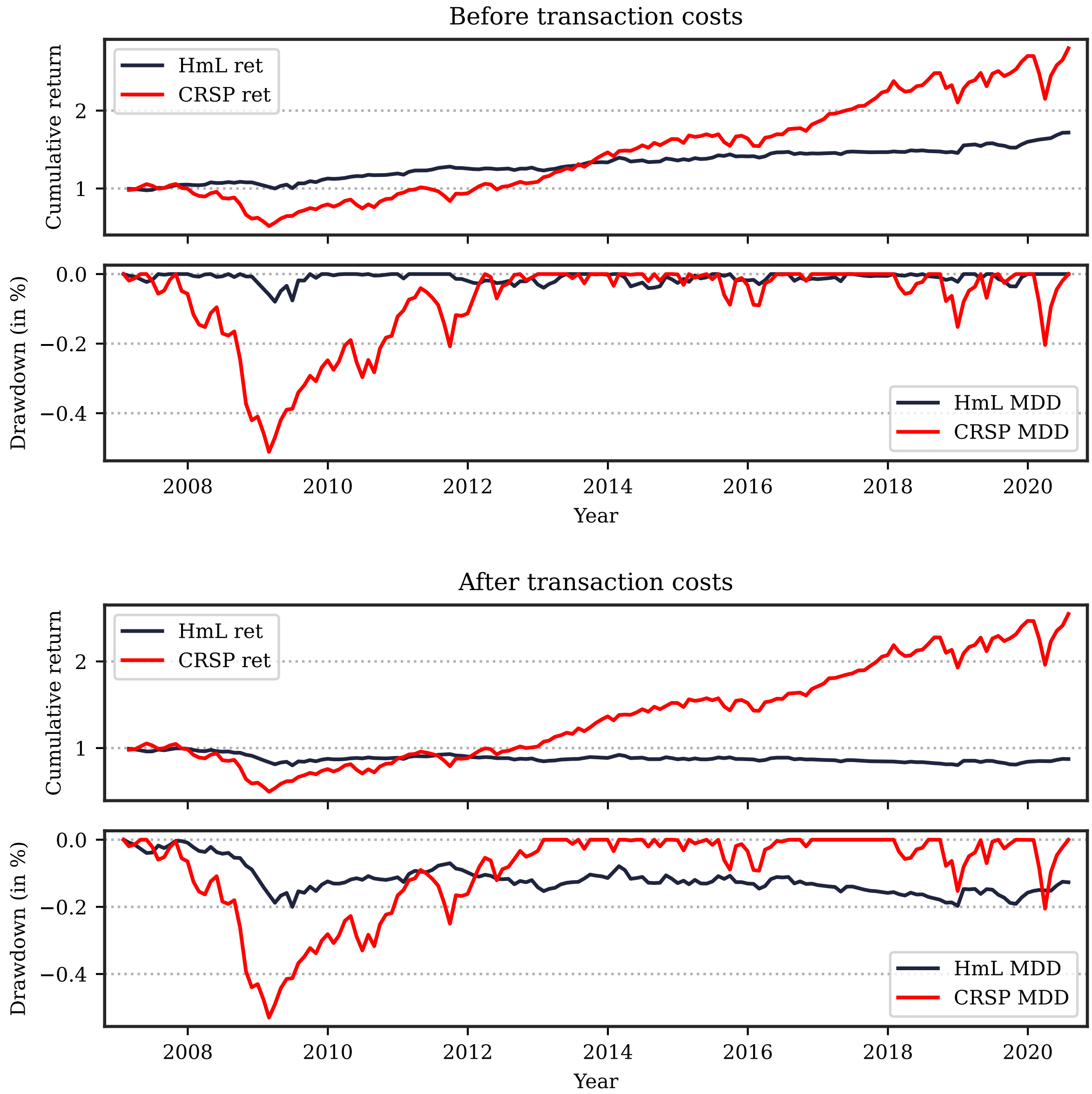
Note. The figure shows the number of stocks considered for the trading strategy in the long and short portfolio over time. The strategy goes long (short) the ten size portfolios in the highest (lowest) α_t^{IN} decile. Specifically, we sort stocks in deciles by market capitalization, and within each decile, we sort stocks into deciles on α_t^{IN}. We calculate value-weighted returns. The strategy goes long in the ten highest α_t^{IN} portfolios and shorts the ten lowest α_t^{IN} portfolios each month. The red line (CRSP ret) depicts a value-weighted investment in the market. The sample period is January 2007 – July 2020.

Table A12: EOM:Single sort

Single sort						
r_{t+1} (in %)	Low	2	3	4	High	HmL
α_t^{IN}	8.96	8.77	11.66	12.25	17.36	8.40
t-stat α_t^{IN}	2.06	1.78	2.08	2.01	2.64	1.99
α_t^{RE}	9.93	9.55	9.70	8.90	11.26	1.33
t-stat α_t^{RE}	2.05	2.18	2.11	1.54	1.90	0.48

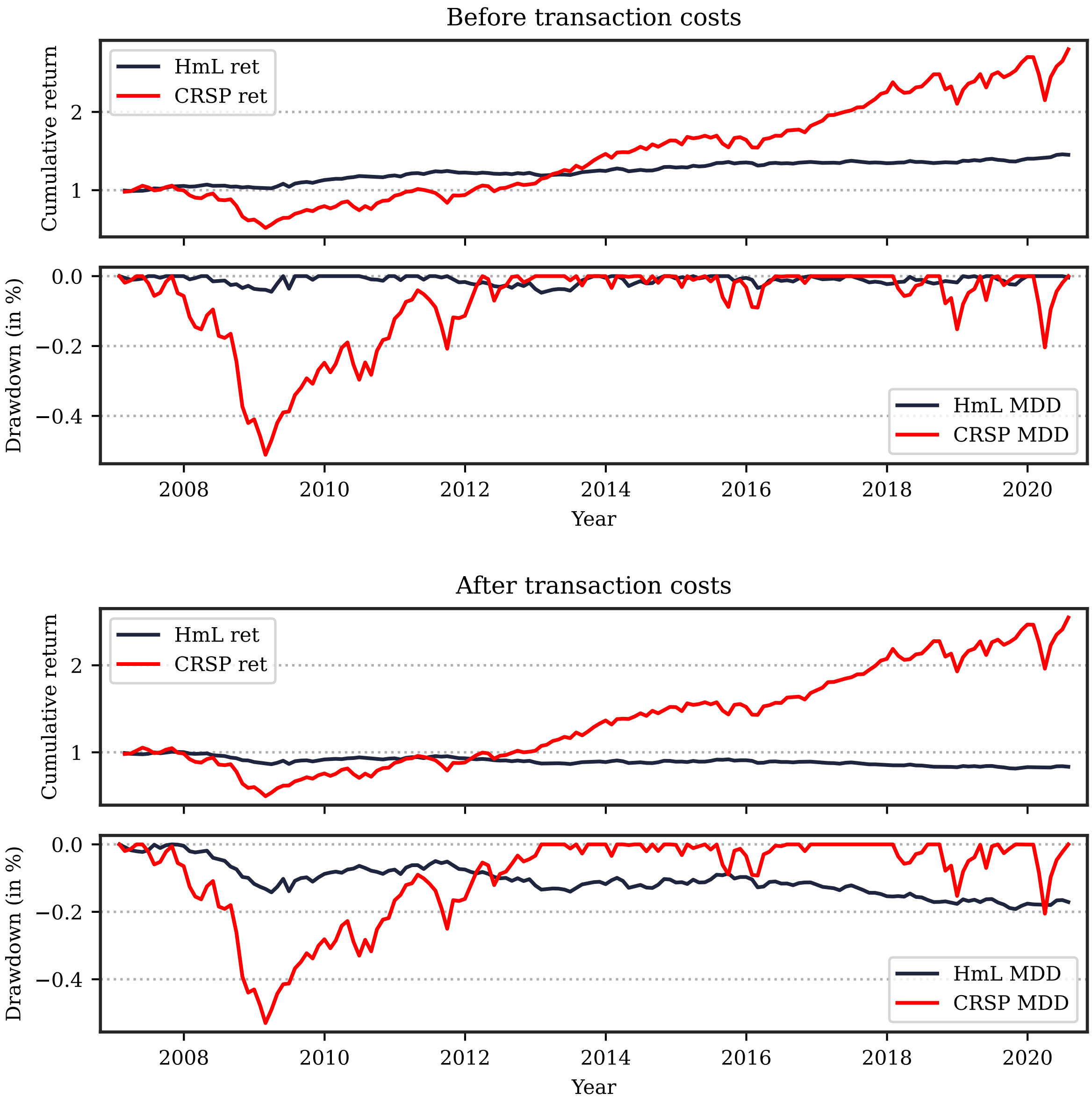
Note. The table shows single sorts on α_t^j.

Fig. A3. Performance and drawdown curve of size-neutral strategy on retailers' price impact



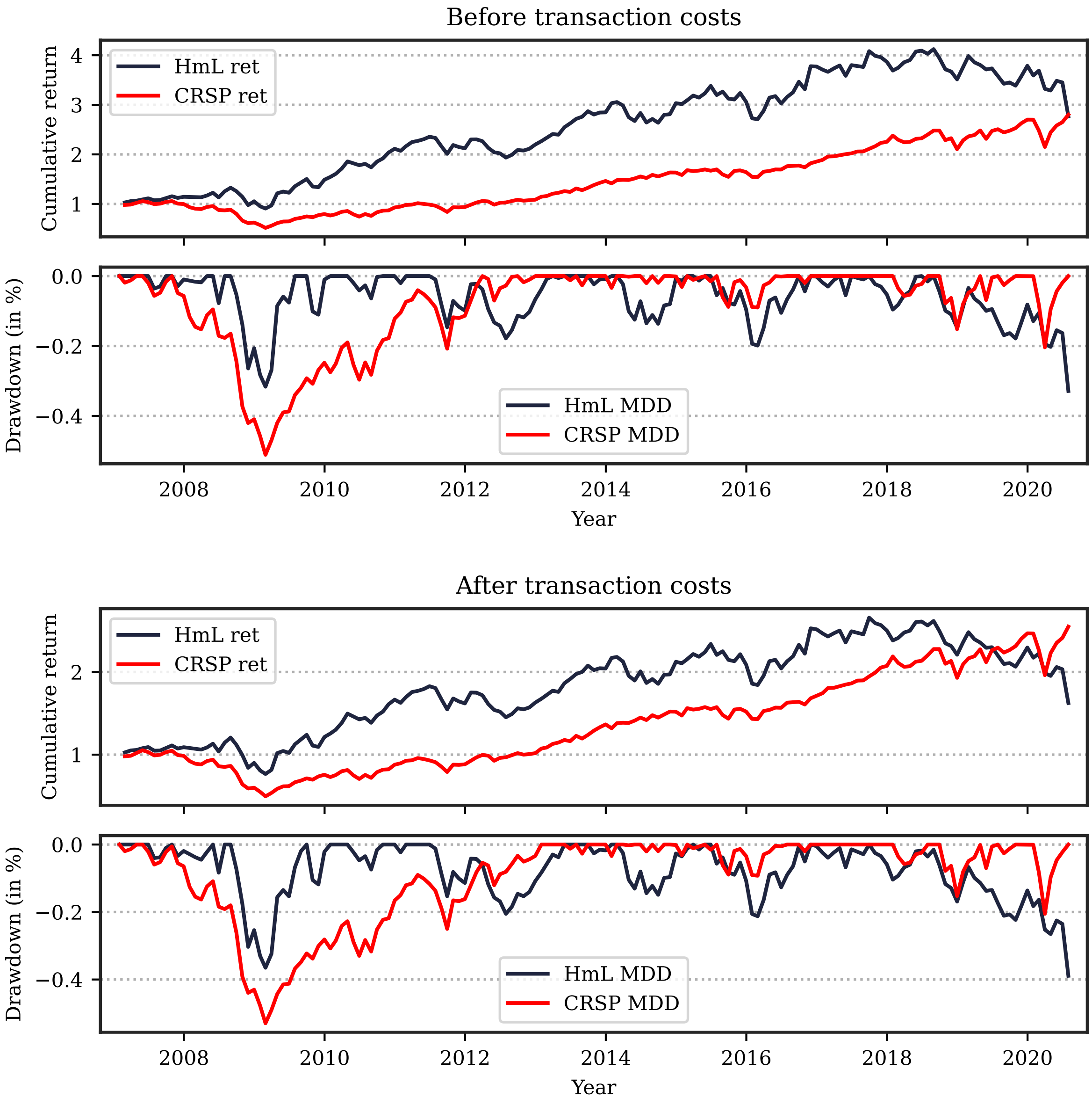
Note. The figure shows the cumulative performance of a one-dollar investment into the size-neutral trading strategy as well as the drawdown curve for the respective trading strategies. The black line (HmL ret) is the economic value of buying (selling) size-neutral stock portfolios with high (low) price impact. Specifically we sort stocks in deciles by market capitalization and within each decile we sort stocks into deciles on α_t^{RE}. We calculate value-weighted returns. The strategy goes long in the ten highest α_t^{RE} portfolios and shorts the ten lowest α_t^{RE} portfolios each month. The red line (CRSP ret) depicts a value-weighted investment in the market. The sample period is January 2007 − July 2020.

Fig. A4. Performance and drawdown curve of size-neutral strategy on retailers' price impact (5,5)



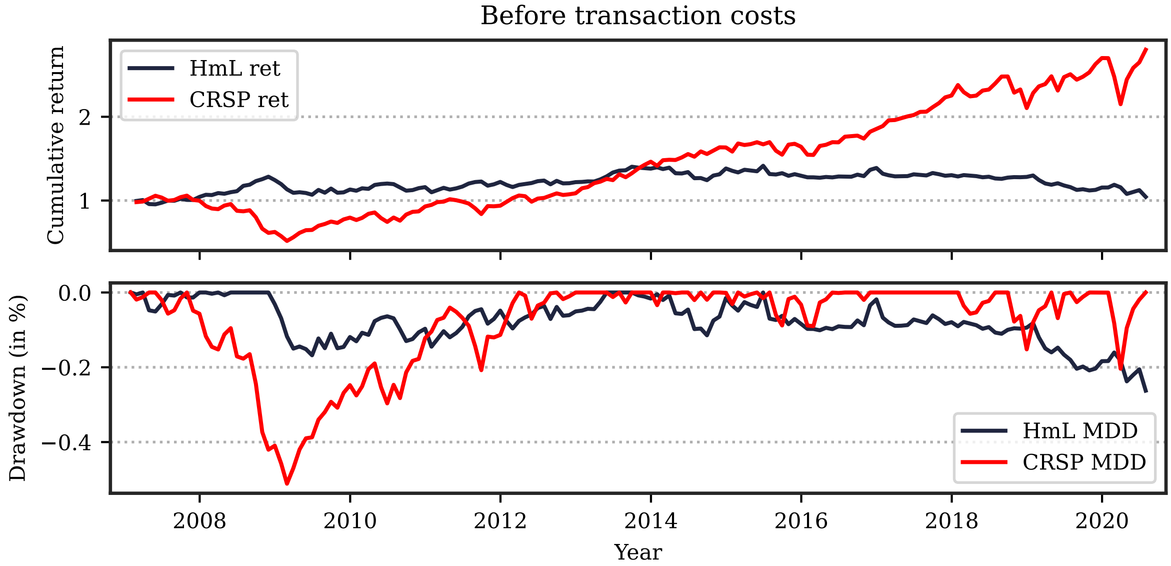
Note. The figure shows the cumulative performance of a one-dollar investment into the size-neutral trading strategy as well as the drawdown curve for the respective trading strategies. The black line (HmL ret) is the economic value of buying (selling) size-neutral stock portfolios with high (low) price impact. Specifically we sort stocks in deciles by market capitalization and within each decile we sort stocks into deciles on α_t^{RE}. We calculate value-weighted returns. The strategy goes long in the ten highest α_t^{RE} portfolios and shorts the ten lowest α_t^{RE} portfolios each month. The red line (CRSP ret) depicts a value-weighted investment in the market. The sample period is January 2007 − July 2020.

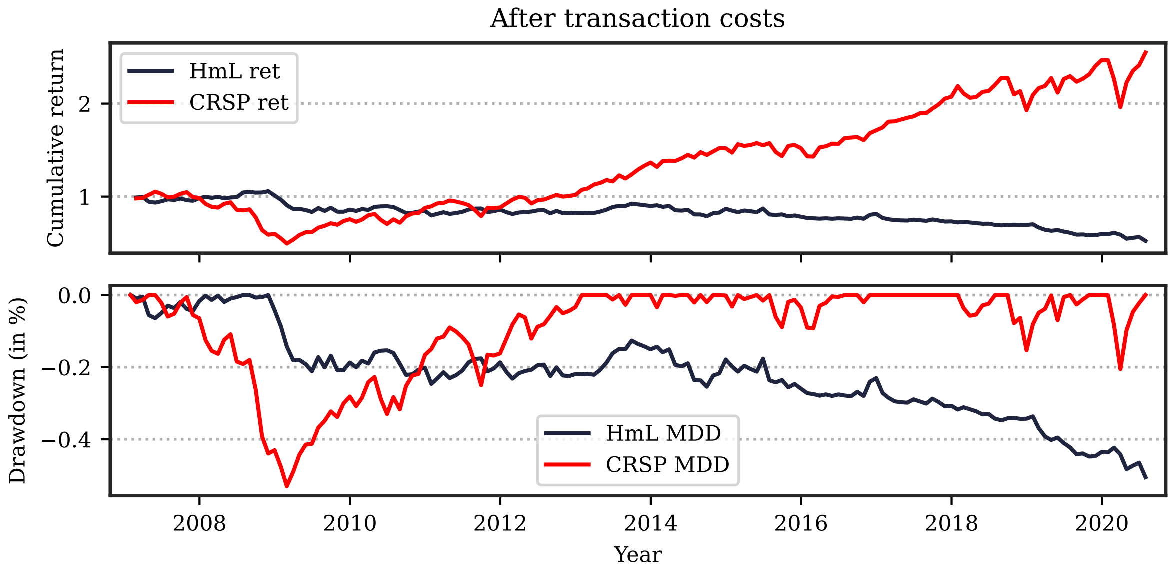
Fig. A5. Performance and drawdown curve of single sort strategy on institutionals' price impact



Note. The figure shows the cumulative performance of a one-dollar investment into the single sorted trading strategy and the drawdown curve for the respective trading strategies. The black line (HmL ret) is the economic value of buying (selling) stocks with high (low) price impact. Specifically, I sort stocks in deciles by α_t^{IN}. I calculate value-weighted returns. The red line (CRSP ret) depicts a value-weighted investment in the market. The sample period is January 2007 − July 2020. The SR up until 2018-06-01 is 0.77 (0.55) before (after) transaction costs and beats the market with a SR of 0.57 (0.52).

Fig. A6. Performance and drawdown curve of single sort strategy on retailers' price impact





Note. The figure shows the cumulative performance of a one-dollar investment into the single sorted trading strategy and the drawdown curve for the respective trading strategies. The black line (HmL ret) is the economic value of buying (selling) stocks with high (low) price impact. Specifically, I sort stocks in deciles by α_t^{RE}. I calculate value-weighted returns. The red line (CRSP ret) depicts a value-weighted investment in the market. The sample period is January 2007 − July 2020.

Table A13: EOQ: DiD with 1-1 matching bulk fundamental

	$\lvert\Delta \mathrm{w}_t\rvert$							
	α_{10}^{IN}							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{1}_{\Delta 20}$	−0.07	−0.06	−0.15	−0.08	−0.12	−0.16	−0.07	−0.17
	[−5.75]	[−4.83]	[−6.28]	[−4.79]	[−7.99]	[−5.15]	[−4.54]	[−10.88]
Obs.	21869	21869	21869	21869	21869	21869	21869	21869
R^2 [%]	0.6	0.1	0.5	0.2	0.5	0.6	0.2	0.8
ln(Size)		X		X	X		X	
Ret.			X	X	X			X
Std.					X	X	X	X
Entity effects	Y	Y	Y	Y	Y	Y	Y	Y
Time effects	Y	Y	Y	Y	Y	Y	Y	Y

Note. The table shows coefficients from panel regressions of regressing the change in $\alpha_{i,t}^{IN}$ on a dummy which is one when the absolute weight of the crowd-wisdom portfolio ($\lvert\Delta \mathrm{w}_{t,i}\rvert$) of stock i moves up by more than 20% in the cross-sectional distribution. Columns (2) to (8) show similar results by comparing stocks in a 1-1 matching approach on firm fundamentals, such as ln(size), return (Ret.), and return standard deviation (Std.).

Table A14: EOQ: DiD with 1-3 matching bulk

	$\lvert\Delta \mathrm{w}_t\rvert$							
	α_{10}^{IN}							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{1}_{\Delta 20}$	−0.07	−0.06	−0.08	−0.08	−0.08	−0.06	−0.06	−0.08
	[−5.75]	[−4.84]	[−5.89]	[−5.42]	[−5.69]	[−4.07]	[−4.27]	[−6.68]
Obs.	21869	21869	19257	19257	19256	21868	21868	19256
R^2 [%]	0.6	0.2	0.8	0.8	0.9	0.2	0.3	0.7
ln(Size)		X		X	X		X	
Inf. Fric.			X	X	X			X
ln(VPIN)					X	X	X	X
Entity effects	Y	Y	Y	Y	Y	Y	Y	Y
Time effects	Y	Y	Y	Y	Y	Y	Y	Y

Note. The table shows coefficients from panel regressions of regressing the change in $\alpha_{i,t}^{IN}$ on a dummy which is one when the absolute weight of the crowd-wisdom portfolio ($\lvert\Delta \mathrm{w}_{t,i}\rvert$) of stock i moves up by more than 20% in the cross-sectional distribution. Columns (2) to (8) show similar results by comparing stocks in a 1-3 matching approach on firm fundamentals, such as ln(size), informational frictions (Inf. Fric.), and the log of $VPIN_t$ (ln(VPIN)).

Table A15: EOQ: DiD 1-3 matching (with std. of weight changes)

	$\sigma(\Delta \mathrm{w}_t)$							
	α_{10}^{IN}							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{1}_{\Delta 20}$	-0.06	-0.05	-0.07	-0.07	-0.07	-0.05	-0.05	-0.07
	[-7.14]	[-5.48]	[-7.40]	[-6.79]	[-7.98]	[-5.66]	[-4.32]	[-7.74]
Obs.	21868	21868	19257	19257	19256	21867	21867	19256
R^2 [%]	0.8	0.3	1.1	1.1	1.2	0.3	0.3	1.0
ln(Size)		X		X	X		X	
Inf. Fric.			X	X	X			X
ln(VPIN)					X	X	X	X
Entity effects	Y	Y	Y	Y	Y	Y	Y	Y
Time effects	Y	Y	Y	Y	Y	Y	Y	Y

Note. The table shows coefficients from panel regressions of regressing the change in $\alpha_{i,t}^{IN}$ on a dummy which is one when the standard deviation of weight changes of the crowd-wisdom portfolio ($\sigma(\Delta \mathrm{w}_{t,i})$) of stock i moves up by more than 20% in the cross-sectional distribution. Columns (2) to (8) show similar results by comparing stocks in a 1-3 matching approach on firm fundamentals, such as ln(size), informational frictions (Inf. Fric.), and the log of $VPIN_t$ (ln(VPIN)).

Table A16: EOQ: DiD with 1-3 matching bulk fundamental sanity check

	$\lvert\Delta \mathrm{w}_t\rvert$							
	α_{10}^{IN}							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{1}_{\Delta -20}$	0.04	0.04	0.07	0.06	0.08	0.12	0.05	0.17
	[4.51]	[2.46]	[3.66]	[3.32]	[4.89]	[7.36]	[3.38]	[6.79]
Obs.	21869	21869	21869	21869	21869	21869	21869	21869
R^2 [%]	0.1	0.0	0.1	0.1	0.2	0.3	0.1	0.5
ln(Size)		X		X	X		X	
Ret.			X	X	X			X
Std.					X	X	X	X
Entity effects	Y	Y	Y	Y	Y	Y	Y	Y
Time effects	Y	Y	Y	Y	Y	Y	Y	Y

Note. The table shows coefficients from panel regressions of regressing the change in $\alpha_{i,t}^{IN}$ on a dummy which is one when the absolute weight changes of the crowd-wisdom portfolio ($\lvert\Delta \mathrm{w}_{t,i}\rvert$) of stock i moves down by more than 20% in the cross-sectional distribution. Columns (2) to (8) show similar results by comparing stocks in a 1-3 matching approach on firm fundamentals, such as ln(size), informational frictions (Inf. Fric.), and the log of $VPIN_t$ (ln(VPIN)).

Table A17: EOQ: DiD bulk

	$\lvert\Delta \mathrm{w}_t\rvert$			
	α_{10}^{IN}		$\ln(VPIN)$	
	(1)	(2)	(3)	(4)
$\mathbb{1}_{\Delta 20}$	−0.07	0.04	−1.34	−2.25
	[−5.75]	[1.89]	[−3.49]	[−2.57]
Obs.	21869	2471	21868	2471
R^2 [%]	0.6	0.2	0.1	0.4
Entity effects	Y	N	Y	N
Time effects	Y	N	Y	N
COVID sample	N	Y	N	Y

Note. The table shows coefficients from panel regressions of regressing the change in $\alpha_{i,t}^{IN}$ on a dummy which is one when the absolute weight changes of the crowd-wisdom portfolio ($\lvert\Delta \mathrm{w}_{t,i}\rvert$) of stock i moves up by more than 20% in the cross-sectional distribution. Columns (2) and (4) show results for the COVID-19 sample where investors received COVID-19 payments.

Table A18: EOQ: DiD with 1-3 matching bulk fundamental, no bulk

The table shows

	$\lvert\Delta \mathrm{w}_t\rvert$							
	α_{10}^{IN}							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{1}_{\Delta 20}$	−0.02	−0.03	−0.02	−0.01	−0.02	−0.02	−0.03	−0.02
	[−5.62]	[−5.90]	[−2.30]	[−1.59]	[−3.11]	[−3.07]	[−4.05]	[−2.93]
Obs.	21746	21746	21746	21746	21746	21746	21746	21746
R^2 [%]	0.1	0.1	0.0	0.0	0.1	0.1	0.1	0.0
ln(Size)		X		X	X		X	
Ret.			X	X	X			X
Std.					X	X	X	X
Entity effects	Y	Y	Y	Y	Y	Y	Y	Y
Time effects	Y	Y	Y	Y	Y	Y	Y	Y

Note. The table shows coefficients from panel regressions of regressing the change in $\alpha_{i,t}^{IN}$ on a dummy which is one when the absolute weight changes of the crowd-wisdom portfolio ($\lvert\Delta \mathrm{w}_{t,i}\rvert$) of stock i moves down by more than 20% in the cross-sectional distribution. Columns (2) to (8) show similar results by comparing stocks in a 1-3 matching approach on firm fundamentals, such as ln(size), return (Ret.), and return standard deviation (Std.). The permanent price impact is measured with Lee and Ready (1991) instead of bulk classification.

Table A19: EOQ: DiD with 1-3 matching bulk fundamental, no bulk

	$\lvert\Delta \mathrm{w}_t\rvert$							
	α_{10}^{IN}							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{1}_{\Delta 20}$	-0.02	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03
	$[-5.62]$	$[-5.90]$	$[-5.27]$	$[-4.56]$	$[-4.85]$	$[-3.87]$	$[-3.29]$	$[-4.23]$
Obs.	21746	21746	19135	19135	19134	21745	21745	19134
R^2 [%]	0.1	0.1	0.3	0.3	0.3	0.1	0.1	0.2
ln(Size)		X		X	X		X	
Inf. Fric.			X	X	X			X
ln(VPIN)					X	X	X	X
Entity effects	Y	Y	Y	Y	Y	Y	Y	Y
Time effects	Y	Y	Y	Y	Y	Y	Y	Y

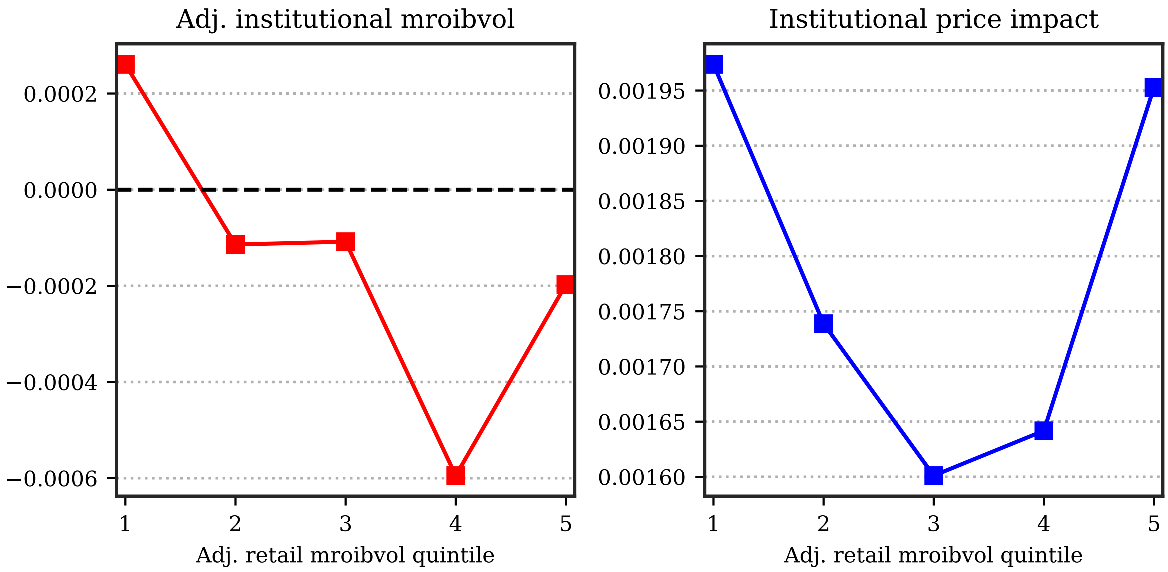
Note. The table shows coefficients from panel regressions of regressing the change in $\alpha_{i,t}^{IN}$ on a dummy which is one when the absolute weight changes of the crowd-wisdom portfolio ($\lvert\Delta \mathrm{w}_{t,i}\rvert$) of stock i moves down by more than 20% in the cross-sectional distribution. Columns (2) to (8) show similar results by comparing stocks in a 1-3 matching approach on firm fundamentals, such as ln(size), informational frictions (Inf. Fric.), and the log of $VPIN_t$ (ln(VPIN)). The permanent price impact is measured with Lee and Ready (1991) instead of bulk classification.

Table A20: EOQ: DiD with 1-3 matching bulk fundamental

	$\Delta \mathrm{w}_t$							
	α_{10}^{IN}							
in (%)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{1}_{\Delta 20}$	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	$[-2.83]$	$[-2.24]$	$[-2.62]$	$[-1.83]$	$[-1.51]$	$[-2.79]$	$[-2.00]$	$[-2.18]$
Obs.	17731	17731	17730	17730	17730	17731	17731	17730
R^2 [%]	0.4	0.1	0.2	0.1	0.1	0.2	0.1	0.1
ln(Size)		X		X	X		X	
Ret.			X	X	X			X
Std.					X	X	X	X
Entity effects	Y	Y	Y	Y	Y	Y	Y	Y
Time effects	Y	Y	Y	Y	Y	Y	Y	Y

Note. The table shows coefficients from panel regressions of regressing the change in $\alpha_{i,t}^{IN}$ on a dummy which is one when the absolute weight of the crowd-wisdom portfolio ($\Delta \mathrm{w}_{t,i}$) of stock i moves up by more than 20% in the cross-sectional distribution. Columns (2) to (8) show similar results by comparing stocks in a 1-3 matching approach on firm fundamentals, such as ln(size), return (Ret.), and return standard deviation (Std.).

Fig. A7. Liqudity provision of retailers



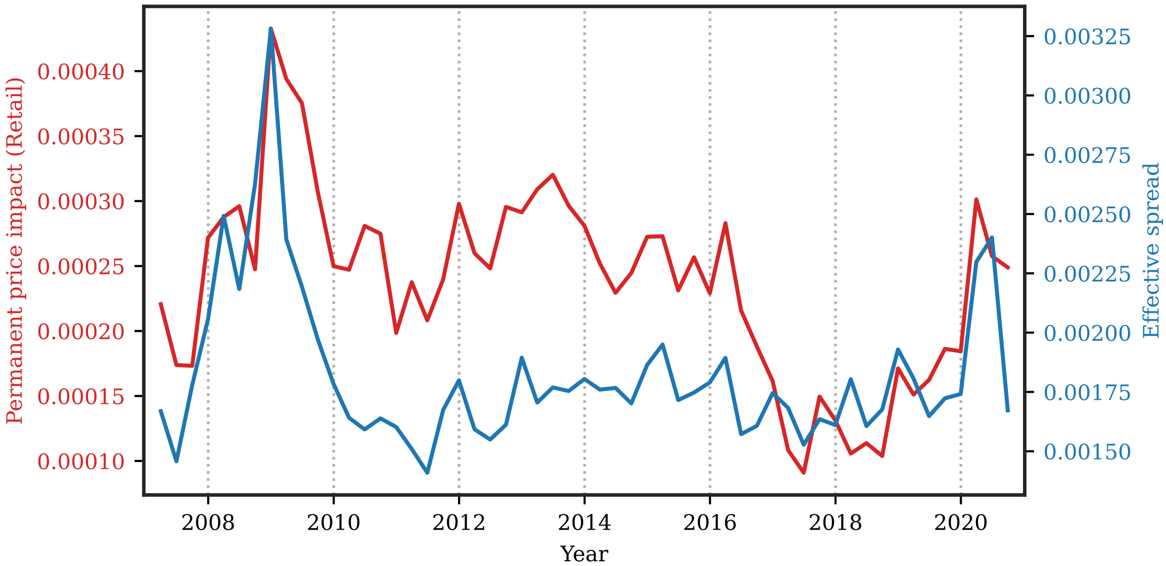
Note. The figure plots institutional standardized net volume (Institutional *Mroibvol*) and institutional price impact from TAQ against retail standardized net volume (Retail *Mroibvol*). Each day, I subtract the cross-sectional average of *Mroibvol* (market adjustment). I aggregate the volume to the quarterly frequency using the mean volume. Each quarter I sort stocks on retail *Mroibvol* and calculate the mean institutional market adjusted institutional *Mroibvol* (left) and institutional price impact (right) for each quarter and each quintile. The time series averages are plotted in above figure. The time period is May 2018 until July 2020.

Table A21: Trade classification full

	LR			EMO			CLNV			BULK	
	C	NC	FC	C	NC	FC	C	NC	FC	C	FC
Panel A: % Number of trades											
Retail buy	0.794	0.000	0.205	0.489	0.091	0.420	0.780	0.021	0.199	0.613	0.387
Retail sell	0.792	0.001	0.207	0.485	0.030	0.485	0.778	0.022	0.200	0.613	0.387
Panel B: % Volume											
Retail buy	0.795	0.000	0.205	0.503	0.076	0.421	0.783	0.015	0.202	0.721	0.279
Retail sell	0.799	0.000	0.200	0.501	−0.001	0.501	0.785	0.016	0.199	0.719	0.281

Note. The table shows percentage of trades classified (C), not classified (NC) and falsely classified (FC) by the Lee and Ready (1991) (LR), Ellis et al. (2000) (EMO), and Chakrabarty et al. (2007) (CLNV), using the Boehmer et al. (2021) algorithm (Retail buy, Retail sell) as benchmark. In Easley et al. (2012b) (BULK) classification, the aggregation level is 30min.

Fig. A8. Price impact and spreads



Note. The figure shows the cross-sectional average of the retail price impact and the cross-sectional average of the effective spread. The time series correlation of the two aggregated series using the institutional (retail) price impact is 84.16 % (57.30 %). The figure excludes microcaps (share price smaller than 5%) and the smallest quintile of the cross-sectional distribution at each point in time.

Table A22: Trading of institutionals vs. retailers

	$\text{HML}_{pf5-pf1}$	
	Full sample	Robinhood sample
a (in bps)	-2.31	-4.58
	$[-1.97]$	$[-1.55]$
Obs.	55	10
R^2 [%]	0.0	0.0

Note. The table shows differences between trading volume of retailers vs. institutionals and especially the differences between portfolio 5 and portfolio 1 of Figure 9 and A7. Standard errors are HAC robust with optimal lag length.