Asymmetric information risk in high-frequency equity markets*

Jan Harren†

This version: January 15, 2024

Abstract

We quantify the information content of trades in equity markets for retail investors and institutionals using high-frequency data for the cross-section of stock returns. We find evidence of heterogenous asymmetric information across market participants, time, and stocks, consistent with information models that include asymmetric information risk. Informational frictions drive asymmetric information risk in equity markets. A size-neutral trading strategy on institutionals' asymmetric information risk yields sizeable returns, beats the market, and is not explained by established risk factors. As retailers trade significant volumes in one direction on the Robinhood brokerage platform, the institutions' asymmetric information risk is reduced, consistent with models where retail trading aligns the price impact of all market participants in markets where institutions have particular market power. We argue that information remains an important determinant of equity returns.

JEL classification: G10, G12, G14
EFM classification: 310, 330, 350, 360, 380

Keywords: Asymmetric information risk, Cross-sectional asset pricing, Market microstructure, Retail trading, Robinhood trading

*I gratefully acknowledge very helpful comments and suggestions from Heiner Beckmeyer, Nicole Branger, and Timo Wiedemann, as well as participants at the British Accounting and Finance Association Conference 2023 (Harvard University), and the Southwestern Finance Association Conference 2024 (Las Vegas, scheduled).

†University of Münster, Finance Center Münster, Universitätsstraße 14-16, 48143 Münster, Germany, jan.harren@wiwi.uni-muenster.de

1. Introduction

One of the key questions in asset pricing is how security prices are determined and what moves prices and why. Fama (1970) introduced the efficient market hypothesis and states that all information should be incorporated into asset prices. However, capital allocation, trading frictions of dealers, and dealer's risk bearing capacity affect prices in certain ways and might lead to deviations from perfectly efficient prices. For example, inventory models (Garman, 1976; Ho and Stoll, 1981; O'hara and Oldfield, 1986) show that there is a transitory price impact if dealers deviate from their optimal inventory. Furthermore, asymmetric information models reveal that the impact of information on prices is persistent and that the information content of trades affects quotes of dealers as they want compensation for being exposed to adverse selection risk (Bagehot, 1971; Kyle, 1985; Glosten and Milgrom, 1985). Next to these microstructure models, asset pricing research focuses on the role of the market and other aggregated risks. We want to bridge the fields of asset pricing and market microstructure by showing that asymmetric information risk in trades affect asset prices and the way dealers determine prices in equity markets.

Consequently, the following questions arise: Is asymmetric information risk present in equity markets? Is it time varying and heterogeneous among market participants, such as retailers and institutions? Is it possible to trade on asymmetric information risk? Is asymmetric information risk priced in the cross-section of stock-returns? We aim to answer these questions by developing a novel way of measuring asymmetric information risk from high-frequency trade and quote (TAQ) data for the whole cross-section of stock returns for different market participants. We provide strong empirical evidence, that asymmetric information risk is systematic, time varying and heterogeneous among retailers and institutions. Furthermore, we build a long-short strategy to trade on asymmetric information risk and show, that dealers demand compensation, in terms of higher future returns, for being exposed to adverse selection risk. Hence, adverse selection appears to be a crucial determinant of asset prices and the strong-form market efficiency hypothesis of Fama (1970) might be called into question.

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 and asymmetric information risk should be positively related, as information about fundamental value is transferred to trades, which then affect quotes. Second, this 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 this effect should be reflected in spreads as well as in 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. As this revision in expectations, conditional on the type of order, can be anticipated, the market maker incorporates it into prices and spreads. 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). This paper proceeds in two parts. First, we test above empirical predictions, whether order flows convey superior information among time and market participants. Therefore, we measure the contemporaneous and persistent price effects of trades on quotes that reveal asymmetric information risk. Second, we relate this measure to asset prices and show that this risk is priced and profitable to trade on.

We use intraday TAQ data from 2006 to 2020 which contains level 1 tick data. We aggregate this dataset on five-minute intervals to reduce microstructure noise. To distinguish between retail and institutional trades we rely on the Boehmer, Jones, Zhang, and Zhang (2021) algorithm.[1] Our empirical analysis builds on a vector autoregression (VAR)

[1] We are 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). In our analysis, however, we

that decomposes the order flow into transitory and permanent price impact components. We follow Ranaldo and Somogyi (2021) and extend the VAR in Hasbrouck (1991a) by allowing for different agents. However, in contrast to them, we distinguish between private and institutional investors and consider equity prices. We find compelling evidence that order flow systematically impacts equity markets heterogeneously across time, and agents. Across agents we observe an econcomically 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 are therefore able to leverage their private information in the equity markets, whereas it is more difficult for retail investors to achieve such price impact with a smaller order volume. 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.

In the second part of the paper, we relate our measure of asymmetric information risk to asset prices to determine whether higher risk is associated with higher subsequent returns. First, we want to determine which agents' asymmetric information risk is associated with higher subsequent returns. Independent double sorts reveal that institutions consistently get compensated with higher future returns for higher asymmetric information risk. For retailers however, we observe this pattern only conditional on asymmetric information risk being high for institutions. The high-minus-low diff-in-diff spread is consistently driven by asymmetric information risk of institutions and amounts to 1.28% per month being statistically significant. Thus, we conclude that only institutions get compensated for carrying higher asymmetric information risk in equity markets.

Next, we determine the forces underlying this return predictability through institutions' asymmetric information risk. 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 to exploit mispricing. We hypothesize that the return predictability stems from private

focus on the institutional measure. Therefore, the algorithm helps to filter out potential retail trades that could disturb the institutional asymmetric information risk measure.

information from stock-based characteristics (size, liquidity, institutional ownership, ...) that is not directly incorportated into prices (informational frictions). To test this, we create an informational frictions index that is based on stock-level information (Bali, Beckmeyer, Moerke, and Weigert, 2023). In line with this prediction, we show that our measure of asymmetric information increases with higher informational frictions, for retailers and institutions. Thus, above return predictability might stem from stocks with high informational frictions.

Above return predictability enables us to inspect whether the predictability of future returns can be exploited by trading on asymmetric information risk of institutions. We construct a size-neutral zero-investment portfolio that loads on asymmetric information risk 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 fully explained by established risk-factors of cross-sectional asset pricing (Carhart, 1997; Pástor and Stambaugh, 2003; Fama and French, 2015).

Finally, we examine the impact of the recent increase in trading activity by retail investors via trading platforms such as Robinhood and how this affects our measure of institutions' asymmetric information risk. We use data from Robintrack and construct the retailers' crowd-wisdom portfolio (Welch, 2020) to identify stocks in which retailers had much trading activity. We compare these stocks with less-heavily traded stocks by retailers using a diff-in-diff approach. We find that stocks that are heavily traded by retailers exhibit a smaller permanent price impact for institutions. Thus, higher retail activity reduces asymmetric information risk of 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 result. Either retailers stabilize markets and their activity reduces adverse selection in stock markets, or retailers are not perceived to be informed (noise traders) by institutions, so institutions do not care about retailers' trading activities at all. The first explanation is consistent with models in which institutions have particular market power that translates into higher price impact (Neuhann and Sockin, 2023). In this model, the price impact of institutions is reduced by the government, which encourages players in inelastic markets to trade more, thereby equalizing the price impact of market participants. Substituting the government with retail investors, this framework is consistent with a stabilizing force of retail traders that disrupts institutional market participants in price discovery and leads to more risk sharing among market. The trading channel on noise can be ruled out because the price impact of institutions is stronger and more significant, while the price impact of retailers is not.

1.1. Literature

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

Market Microstructure. First, our 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. We build on Ranaldo and Somogyi (2021), which 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 FX markets.

Our study builds on their analysis as we extend their methodological framework (Hasbrouck, 1988, 1991a,b) and quantify asymmetric information risk for retailers and institutions and show that it also exists in equity markets. In a trading strategy, we reveal that it is profitable to trade asymmetric information risk in equity markets. Furthermore, we identify the drivers of the permanent price impact component in equity markets for both market participants and show that informational frictions are key drivers of asymmetric information risk in the cross-section for both, retailers and institutions.

Asset Pricing. Second, our paper contributes to the asset pricing literature by developing a new size-neutral long-short strategy that loads on asymmetric information risk. The newly developed factor yields high long-short returns which are not spanned by established asset pricing factors. There are several studies that relate asymmetric information risk 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) build 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.[2]

We contribute to this literature by establishing a measure of informed trading for the cross-section of equity returns and distinguishing it for retailers and institutions. Furthermore, we provide a trading strategy that generates significant returns for trading on high asymmetric information risk of institutions. We uncover the channels through which asymmetric information risk affects asset prices and relate it to informational frictions.

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 our results, while the last section 4 concludes.

2. Data & Methodology

We first describe the data sources. Second, we discuss the classification of retail and institutional trades. Third, we describe the VAR approach to estimate the permanent price impact as our measure for asymmetric information risk. Finally, we present results

[2] Their study build on Easley and O'hara (2004) who also investigate the role of information on a firm's cost of capital.

and descriptive statistics on measuring asymmetric information risk for retailers and institutions.

2.1. Data sources

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

TAQ Database. We 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. We 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. We use monthly stock data from 2006 to 2020 and apply standard filters in cross-sectional asset pricing. We 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). We only include stocks with a share price larger than 5$. To ensure that our results are not driven by microcaps, we exclude the smallest quintile of the cross-sectional distribution each day after applying the filters above (Gonçalves, 2021).[3]

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.[4] The Robintrack website scraped hourly user holdings for all equities on

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

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

Robinhood.[5] The API was active from May 5, 2018, to August 13, 2020, which defines our sample period in the further Diff-in-diff analysis in Section 3.4.

2.2. *Classifying retail trades*

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

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

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

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

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

2.3. *Classifying institutional trades*

We follow Easley et al. (2012b) and classify non-retail trades using bulk classification. As we do not work with level 1 tick data, but aggregate our trades on 5min intervals, the

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

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 our analysis, we aggregate trades into five-minute intervals to circumvent this problem and eliminate microstructure noise.[6] To determine the percentage of buying and selling volume, we use the standardized price change between the beginning and the end of the interval. We define the daily volume 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 we define $\tau = t(\tau) - t(\tau - 1)$ as one trading day. The price change between five-minute intervals is standardized with the daily standard deviation of price changes. This procedure allows us 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 our application, as we aggregate our trades on five minute intervals.

Table 1 compares different classification schemes. We 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

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

Table 1: 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.

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 our application aggregates trades in five-minute intervals, we classify our 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. We aggregate trades on 5 minute intervals and calculate signed net volume z_t, which is buy orders minus sell orders.[7] 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, we 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, we 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. We perform above steps for institutional investors (IN) and retail investors (RE) and define the agents with $j \in C$, where $C = \{IN, RE\}$. We 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. We choose a lag length of ten, based on the arguments in Hasbrouck (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

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

order flow equations. We estimate Equations (7) and (8) for each stock k.[8] 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 asymmetric information 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 our case).[9] 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, we 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.

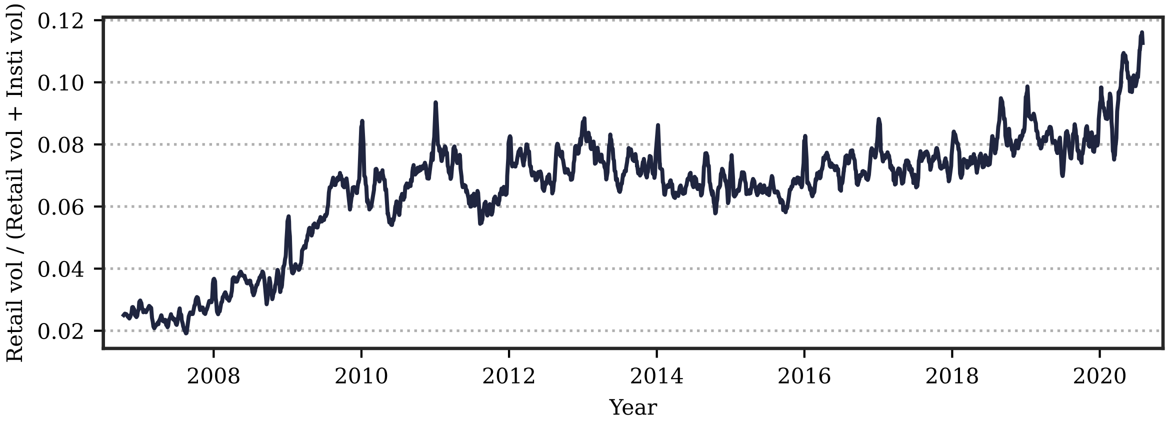
2.5. Descriptive statistics

This section analyzes if asymmetric information 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 stable, and increased again in 2019. When interpreting our results, it is important to remember

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

[9] 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).

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



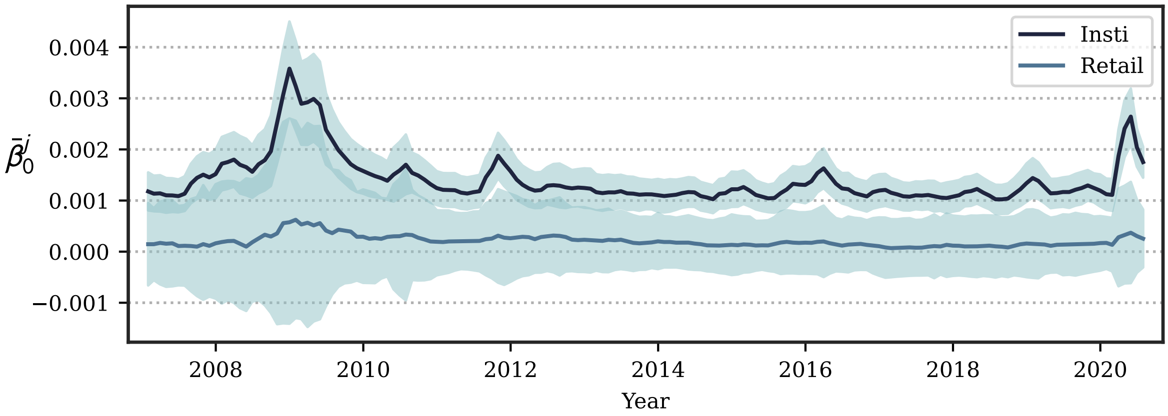
Note. The figure shows the retail trading volume over the whole trading volume for the time period January 2007 until July 2020.

that retailers move much less volume and, therefore, cannot leverage high-volume private information compared to institutions. We split these groups because we 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 and Milgrom, 1985). 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

Fig. 2. Contemporaneous price impact

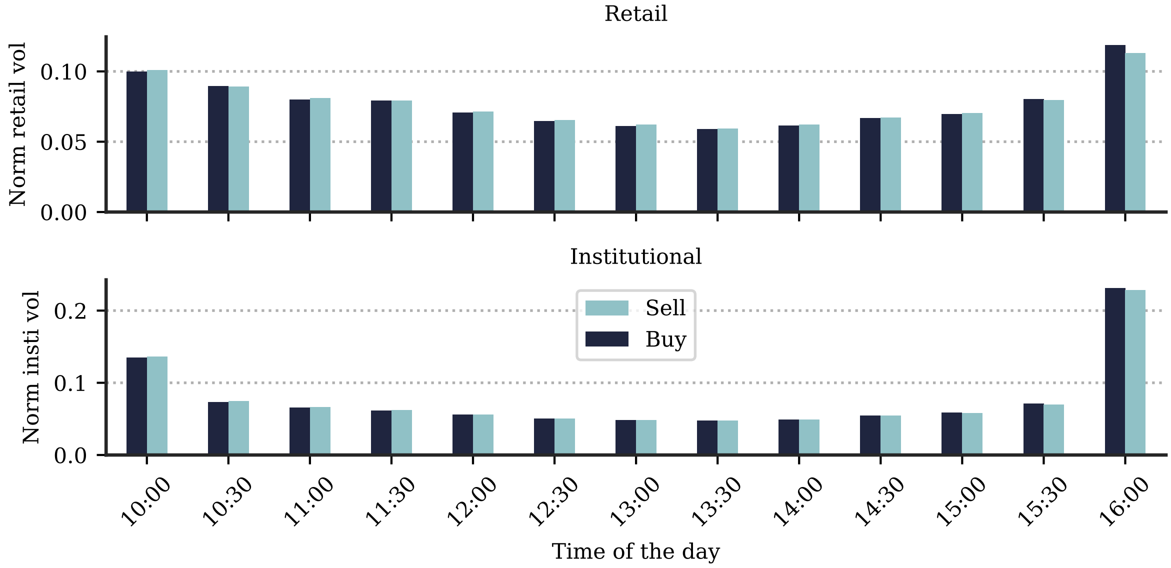


Note. The figure shows cross-sectional mean estimates of the contemporaneous price impact for institutionals and retailers. The estimate is the coefficient $\beta_0^{j,k}$ from Equation (7) for $i = 0$. The cross-sectional mean is caluclated as $\bar{\beta_0}^j = (1/K) \sum_{k=1}^{K} \beta_0^{j,k}$. 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.

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. 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.

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.

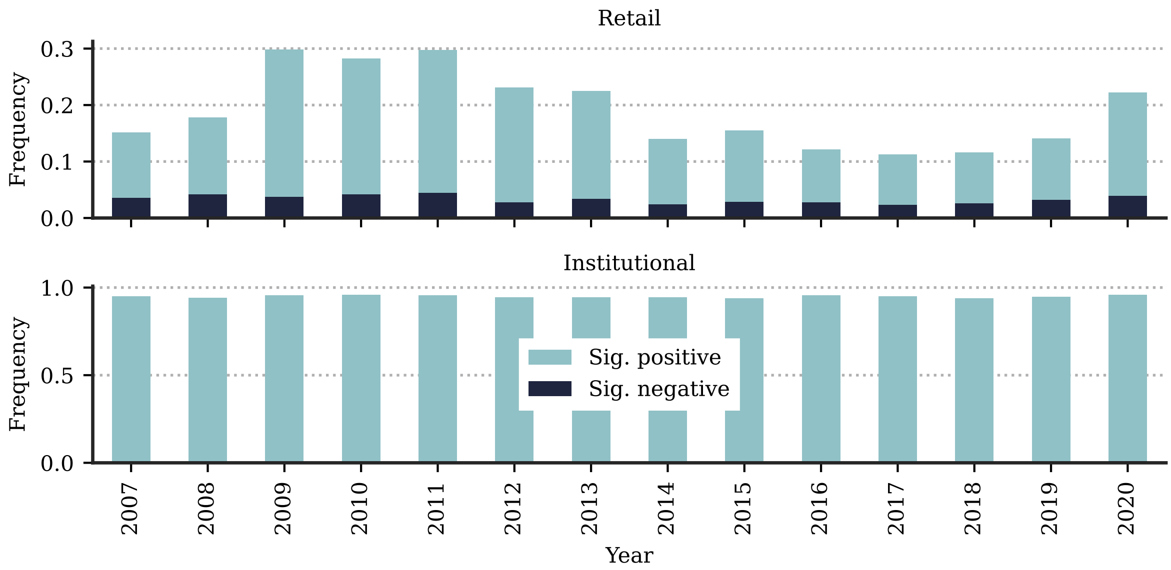
2.5.2. Permanent price impact

The permanent component, α_m^j, in Hasbrouck (1991a) 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.

We aggregate our 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, we use 63 (= 252/4) days, thus we estimate our model over the course of a quarter. We measure statistical significance with a heteroskedasticity-consistent joint F-test in which the parameters in equation (9) are jointly different from zero.

Figure 4 presents the fraction of positive (negative) significant coefficients per year. Consistent with theories of asymmetric information (Glosten and Milgrom, 1985), we

Fig. 4. Fraction of significant permanent price impact coefficients per year



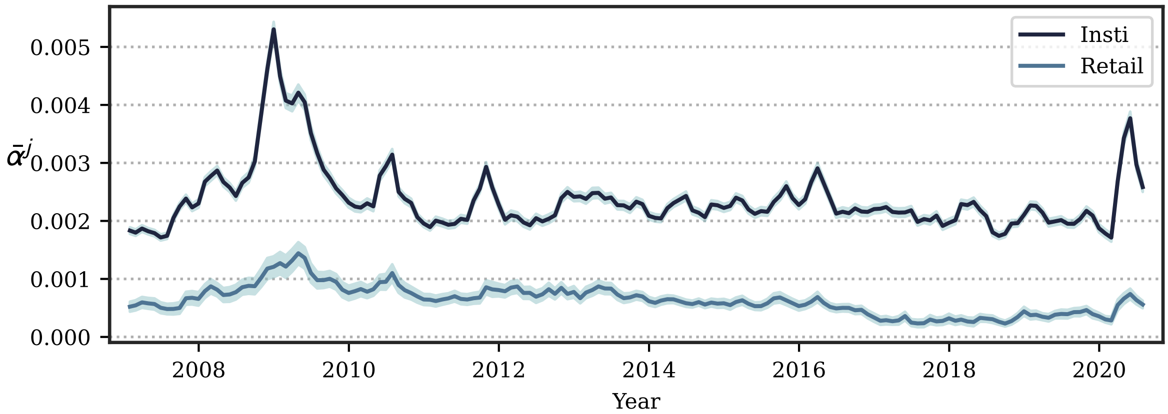
Note. The figure shows the fraction of significant coefficients per year for retailers' and institutionals' permanent price impact (α_t^j). A value of one indicates, that all α_t^j coefficients in this year were significant for the given year for the whole cross-section.

observe heterogeneously informed traders in equity markets. Depending on the market participant, we see 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 we 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.

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

Figure 5 plots the cross-sectional mean permanent price impact. Institutions exhibit a

Fig. 5. 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). We 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, asymmetric information 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, we derived our measure of asymmetric information risk for two groups of traders, retailers and institutions. In this section, we relate these measures of asymmetric information risk to future returns to get a sense of whether this risk is reflected in subsequent returns. Furthermore, we conduct a trading strategy on the institutional price impact measure. We also analyze the economic drivers of asymmetric information risk in equity markets. Finally, we use a diff-in-diff approach to show how increased retail activity affects the price impact of institutions.

3.1. Asymmetric information risk 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. We distinguish between institutions and retailers and hypothesise, that asymmetric information risk should be more pronounced for institutions as they face larger volume and hence higher allocation costs.

To test our hypothesis, we perform independent double sorts for each of our agents to determine for which agent asymmetric information is associated with higher future returns. For each month, we independently sort our 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, we calculate high-minus-low portfolio spreads and the corresponding t-statistics with Newey West robust standard errors (Newey and West, 1986) for a lag length of ten.

Table 2 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 2: 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}, we 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}.

We conclude that institutions are compensated for the risk of asymmetric information, 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). Hence, in the following analyses we mainly use α_t^{IN} as a proxy for asymmetric information risk in the cross-section of stock returns. In the next subsection, we aim to determine the drivers of asymmetric information risk for the cross-section of stock returns.

3.2. *Informational frictions*

In the last section, we showed that α_t^{IN} is able to forecast future returns for the next month. In this section, we 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. We test whether informational frictions provide an explanation for the trading signal α_t^{IN}.

We want to determine the drivers of asymmetric information risk and hypothesize that asymmetric 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, we construct the arbitrage index of Atilgan, Bali, 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, we include firm age, analyst coverage, size, institutional ownership, idiosyncratic volatility, and the Amihud (2002) measure.

To construct the index we sort stocks in quintile portfolios in increasing order based on idiosyncratic volatility and illiquidity. Similar, we 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, we perform a panel regression of our trading signal α_t^{IN} on the constituents of the index. We 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.

Table 3 reports the results. All variables are cross-sectional standardized. We use the last variable of each quarter to avoid overlapping observations when using our price impact measures in the regression. Higher informational frictions directly translate to higher levels of asymmetric information risk for institutions (columns 2), confirming our hypothesis. A one standard deviation increase in the arbitrage costs index translates to a 26.27bp increase in asymmetric information for institutions. The R^2 is high with a value of 21.9%. When looking at the index constituents in column (1), our price impact measure α_t^{IN} heavily loads on size with a coefficient of -34.58bps, analyst coverage, institutional ownership, and IVOL with the expected sign. Thus, small stocks, stocks with lower analyst coverage (higher information costs), stocks with lower institutional ownership (higher short-sale costs), and stocks with higher IVOL exhibit higher asymmetric information risk for institutions. VPIN in column (3) is positively correlated with our measure of asymmetric information risk, confirming our measure as a measure of private information/adverse selection. The results for α_t^{RE} show the expected sign for the constituents of the arbitrage cost index in column (5). Furthermore, age is highly significant

Table 3: What explains the price impact?

	α_{10}^{IN}				α_{10}^{RE}			
(in bps)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
α_{10}^{IN}					6.00	6.43	6.53	5.69
					[9.84]	[10.66]	[14.85]	[9.77]
α_{10}^{RE}	5.15	6.16	4.25	5.35				
	[9.88]	[10.70]	[14.09]	[9.86]				
Age	−3.13				−5.63			
	[−0.92]				[−3.28]			
Illiq	1.01				2.13			
	[0.91]				[1.02]			
IVOL	1.22				1.03			
	[3.77]				[5.07]			
Size	−34.58				−5.01			
	[−19.97]				[−9.55]			
Analyst Cov.	−1.79				0.55			
	[−5.89]				[4.61]			
Insti Own.	−3.65				−1.10			
	[−10.20]				[−5.67]			
Inf. Friction		26.27		22.20		4.45		2.89
		[21.06]		[18.08]		[10.72]		[7.51]
$\ln(VPIN_t)$			16.56	10.23			6.95	5.02
			[21.06]	[15.53]			[15.46]	[12.33]
Entity effects	Y	Y	Y	Y	Y	Y	Y	Y
Time effects	Y	Y	Y	Y	Y	Y	Y	Y
Obs.	166390	166390	205076	166374	166390	166390	205076	166374
R^2 [%]	28.4	21.9	8.1	24.5	5.1	4.8	3.1	5.5

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 (5)), on the index of information frictions itself (columns (2) and (6)), on the logarithm of $VPIN_t$ (columns (3) and (7)), and on the logarithm of $VPIN_t$ and the index of information frictions (columns (4) and (8)).

and shows the expected sign discussed above, whereas higher analyst coverage translates to higher asymmetric information risk for retailers. Generally, the results for α_t^{RE} are economically weaker in magnitude and exhibit lower R^2, confirming our hypothesis, that limits-to-arbitrage and asymmetric information is more of an issue for institutions than for retailers.

Overall, these results suggest, that asymmetric information in the cross-section of stock returns is largely driven by information frictions, limits-to-arbitrage, and mispricing. 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 asymmetric information risk explains stock returns are thus consistent with both a risk-based explanation and the limits-to-arbitrage argument.

3.3. Trading strategy on asymmetric information risk

Next, we 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. First, we conduct a size-neutral long-short strategy based on α_t^{IN}. Specifically, on the last day of each month, we sort our stock universe in decile portfolios based on market capitalization. Within each size decile, ten portfolios are formed based on α_t^{IN}. We calculate value-weighted decile portfolio returns. Subsequently, we 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. We calculate the HmL spread for each month. We 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, we substract the quoted half spread when entering the strategy and substract another quoted half spread when closing the position in the next month.[10]

Table 4 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 smalles size quintiles, respectively. The highest size decile shows the expected sign but is statistically insignificant. This suggests that does not affect expected returns of very large stocks.

Table 4: 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 we see the HmL return of investing in high-minus-low stocks of α_t^{IN} conditional on size.

Table 5 reports descriptive statistics on the HmL-strategy before (Panel A) and after (Panel B) transaction costs. HmL$_{\alpha_{10}^{IN}}$ yields an average excess return of 20.29% (15.89%) per year (after transaction costs), which is significant at the 1% level. The strategy outperforms an investment in the market index (20.29% vs. 16.01% and 15.89% vs. 15.32% after transaction costs) and yields a higher Sharpe Ratio (1.66 vs. 1.03 and 1.30 vs. 0.98 after transaction costs). The returns of the HmL strategy are positively

[10] Furthermore, we 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, our trading strategy works better than using quoted half spreads as our HmL strategy exhibits effective costs of $29.51\% - 19.59\% = 9.92\%$ per year. Therefore, using quoted half spreads places a lower bound for our results.

Table 5: Size-neutral trading strategy before (Gross) and after (Net) transaction cost

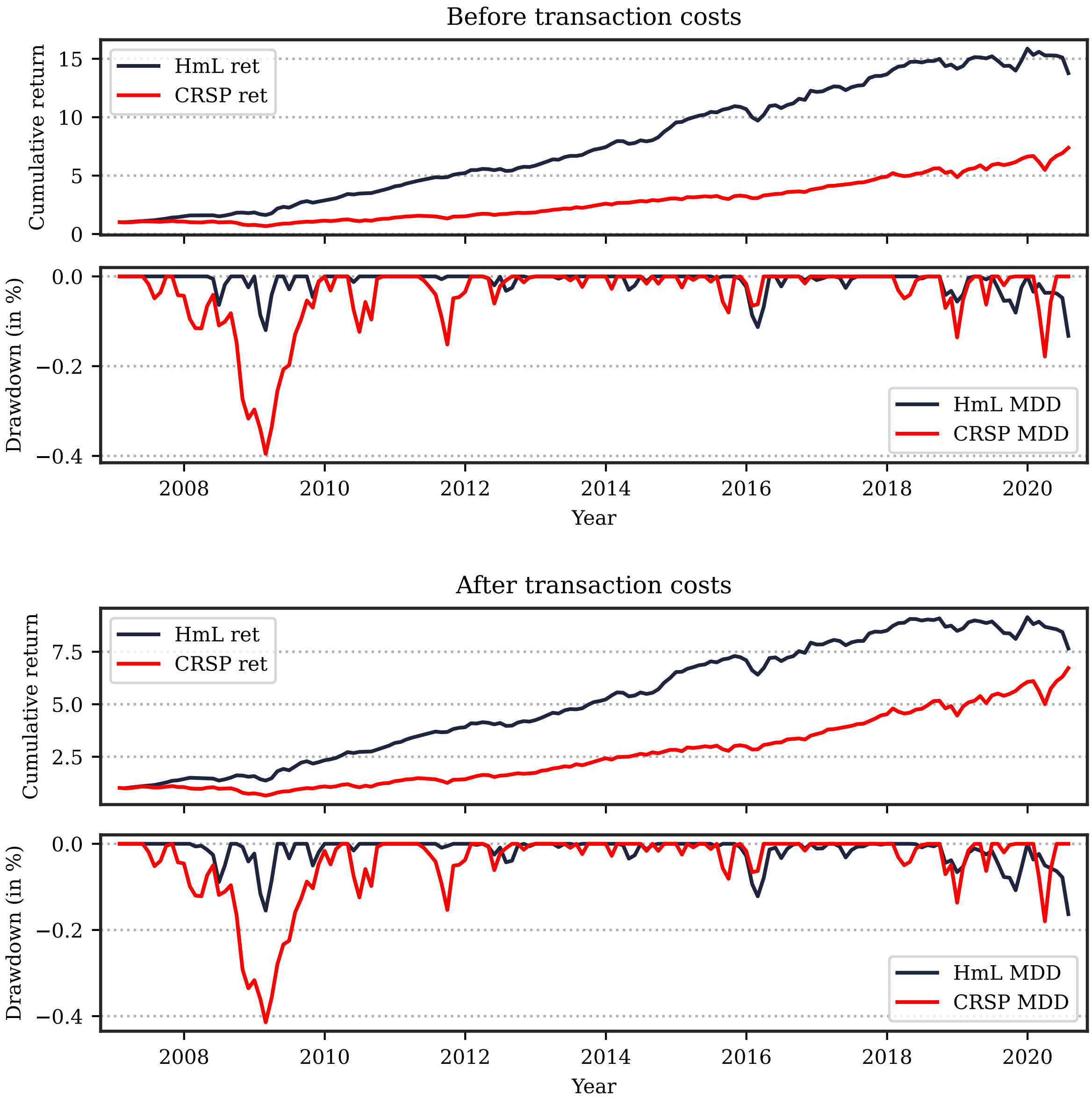
	Avg. ret	t-Stat	Std	SR	Skew	Kurt
Panel A: Gross						
$\text{HmL}_{\alpha_{10}^{IN}}$	20.29	5.01	12.20	1.66	1.17	8.57
$\text{Long}_{\alpha_{10}^{IN}}$	18.31	2.85	24.30	0.75	-0.23	1.82
$\text{Short}_{\alpha_{10}^{IN}}$	1.99	0.35	18.11	0.11	0.68	2.43
$(r_m - r_f)$	16.01	4.00	15.53	1.03	-0.40	1.19
Panel 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
$(r_m - r_f)$	15.32	3.73	15.58	0.98	-0.42	1.22

Note. The table shows the economic value of buying (selling) size-neutral stock portfolios with high (low) asymmetric information. Specifically we sort stocks in deciles by market capitalization and within each decile we sort deciles on α_t^{IN}. We 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.

skewed (compared to a negatively skewed market investment) and exhibit higher kurtosis compared to a normal distribution as we 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 private information, while a low (and negative) α_{10}^{IN} signals that order flow and quotes appear to follow different directions.

Figure 6 depicts the performance and drawdown curve of the size-neutral trading strategy and the value-weighted market excess return between 2007 and 2020. Over our 14-year sample period, the investment of 1\$ would have grown to 13.77\$ vs. a CRSP value-weighted return of 7.38\$ (7.66\$ vs. 6.72\$ after transaction costs). Cumulative returns are smooth over time and do not exhibit significant structural breaks. The maximum drawdown shows that our 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. 39.54% of the market (16.32% vs. 41.42% after transaction costs). Trading on asymmetric information risk appears more profitable than investing in the value-weighted market portfolio.

Fig. 6. 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 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 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.

How many of these factors are explained by established risk factors in the equity market? Table 6 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. The intercept is always positive and highly significant, with an average monthly

Table 6: Risk-adjusted returns

	$HML^{Gross}_{\alpha^{IN}_{10}}$						$HML^{Net}_{\alpha^{IN}_{10}}$					
	(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

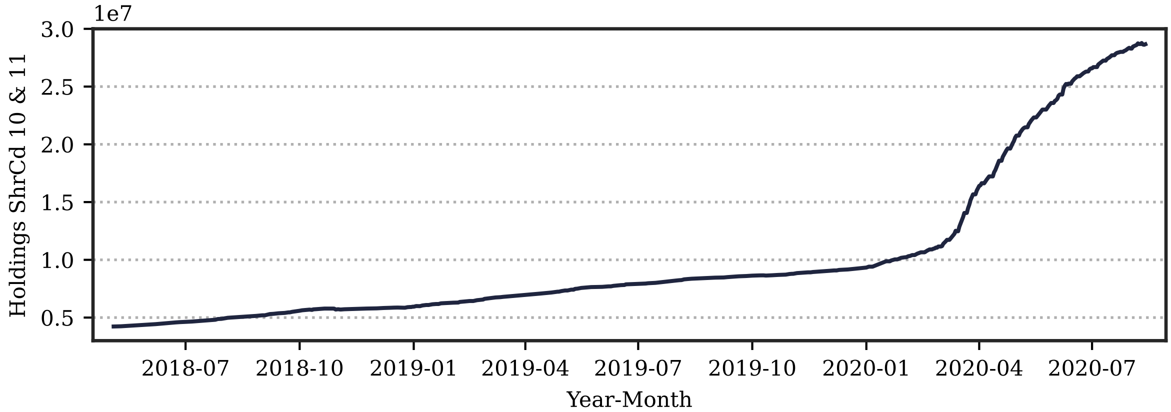
Note. The table shows the estimation results from regressing HmL-returns on established risk-factors in the equity market, which are equity market excess return $(r_m - r_f)$, size (SMB), book-to-market (HML), profitability (RMW), investment (CMA), momentum (UMD), and Pástor and Stambaugh (2003) liquidity 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 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 short the ten lowest α_t^{IN} portfolios each month. Newey and West (1986) robust t-statistics are in parentheses.

return of 1.69 (1.32 after transaction costs). 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 asymmetric information, as the intercept remains positive and statistically significant in all regressions. Hence, asymmetric information risk is priced in equity markets.

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 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 provides hourly holdings for the entire cross-section of equities. We aggregate these holdings in Figure 7.[11] The sample period is May 2018 to August 2020. Retailers increased their

Fig. 7. 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.

holdings steadily, especially from March 2020 onwards (outbreak of COVID-19). Recently, Welch (2020) show that RH investors buy attention-grabbing stocks and stocks with high

[11] Data is available at the Robintrack Website.

past share (dollar) volume. Boehmer et al. (2021) provides suggestive evidence that retail trades contain information not yet incorporated into prices.

This section aims to link our measure of asymmetric information risk for institutions to the trading activity of retailers. How does retail activity influence private information for institutions in stock markets? 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, we perform a diff-in-diff approach by looking at stocks bought or sold heavily by retailers to test if these crowd trades influence our measure of private information. The price impact of institutionals' reduces with larger trading activity of retailers, consistent with models where retail institutionals' market power via price impact is reduced by aligning the price impact of all market participants. In this analysis, the alignment stems from the trading activity of retailers, which reduces institutional market power in terms of price impact (Neuhann and Sockin, 2023).

Diff-in-diff approach. We 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. We weight the RH holdings by price to obtain the dollar volume of the investment.[12] We 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. We can interpret $|\Delta w_{i,t}^{ARH}|$ as a specific stock's attention measure. In the latter analysis, we aggregate our data quarterly to control for overlapping observations. Hence, we calculate the mean of $|\Delta w_{i,t}^{ARH}|$ over the last quarter.

[12] 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.

On the last day of each quarter, we sort our 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. We calculate the change 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, we 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 7 column (1). Stocks that retail investors heavily traded

Table 7: 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)
$\mathbb{1}_{\Delta 20}$	-0.07	-0.08	-0.08	-0.06	-0.08	-0.08	-0.08	-0.07
	$[-5.75]$	$[-5.58]$	$[-4.92]$	$[-5.15]$	$[-5.67]$	$[-4.72]$	$[-4.79]$	$[-4.90]$
Obs.	21869	21869	21869	21869	21869	21869	21869	21869
R^2 [%]	0.6	0.4	0.3	0.2	0.4	0.3	0.4	0.3
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.).

via Robinhood (treatment group) have lower asymmetric information risk for institutions by -0.07bps. Thus, high trading activity by retail investors appears to reduce asymmetric information risk for institutions. However, institutions may not consider retail investors to be informed. As a result, institutions may not believe that a higher retail holding in certain stocks is associated with a higher adverse selection risk. Examples like Gamestop in January 2021 have shown that retailers can act like an “angry mob” and that this crowd is capable of moving markets. However, for the sample period from mid-2018 to mid-2020, institutions could not assume that retailers are capable of moving markets, given the lack of past examples such as Gamestop. Therefore, we conclude that retail trading reduces adverse selection risk for institutions for the given sample period. Furthermore, the market power of institutionals is reduced because retailers' act as a stabilizing force by aligning the high price impact of institutionals, with the small price impact of retail investors (Neuhann and Sockin, 2023).

Diff-in-diff matching approach. We follow Cao, Goyal, Ke, and Zhan (2022) and conduct a diff-in-diff matching approach. We match stocks according to characteristics in $t-1$ and compare their asymmetric information component in t. Specifically, we match according to market capitalization, return volatility, and return. 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. We measure the Euclidian distance (norm) between all three 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. We 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,[13] i.e., one treatment α_{10}^{IN} is matched with three control α_{10}^{IN}. We take

[13] All results are qualitatively the same if we do 1-1 matching.

the average of all three α_{10}^{IN} in the control group (17) and take the time difference (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 gives $C = 3$. Afterward, we 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 time and firm fixed effects.

Table 7 column (2) – (8) show the results. The coefficient on $\mathbb{1}_{\Delta 20}$ is negative and highly statistically significant, even if we match according to market capitalization, stock return, and return volatility. Thus, a stock heavily traded by retailers from one quarter to the other exhibits 6bps to 8bps smaller asymmetric information risk for institutions on average. Hence, we conclude that higher retail activity in stocks reduces asymmetric information risk for institutions. This could be because retailers make markets more efficient and their trading activity reduces adverse selection in the whole market or because institutions do not perceive retailers to be informed traders who pose a higher risk for institutional traders in equity markets. In addition, the market power of institutions (i.e., the price impact) is reduced because retailers act as a stabilizing force that aligns their small price impact with that of institutions in markets where institutions have particular market power (Neuhann and Sockin, 2023).

4. Conclusion

An extensive literature tries to find risk factors in cross-sectional asset pricing (Harvey, Liu, and Zhu, 2016). A different stand of literature tries to understand the determination of prices via trading activity and information. We bridge this research and measure asymmetric information risk in equity markets using high frequency data. Our study builds on Ranaldo and Somogyi (2021) who use a similar methodology for the FX market. We address the following questions: First, does order flow convey asymmetric information across time and market participants? Second, is asymmetric information risk priced in the cross-section of stock returns?

First, we find evidence that asymmetric information risk is present in equity markets and varies among market participants and across time. Institutional investors are consistently better informed than retail investors, which provides evidence that there is asymmetric information risk in equity markets. Over time, asymmetric information risk varies with the business cycle, providing evidence that asymmetric information risk is higher when overall risk aversion is high.

Second, we find that asymmetric information risk of institutions is priced in the cross-section and predicts future returns. This predictability originates partly from informational frictions. Our results reveal that a long-short trading strategy on asymmetric information risk of institutions is highly profitable and remains statistically and economically significant even after accounting for high levels of transaction costs.

Finally, we relate the recent rise in retail activity to our measure of institutional asymmetric information risk and show that stocks that are heavily traded by retailers exhibit lower institutional asymmetric information risk because institutions may not perceive retailers as informed or because retailers increase market efficiency, thereby reducing the risk of adverse selection in equity markets overall. In addition, the price impact of institutions is reduced because retailer trading aligns the price impact of all market participants in markets where institutions have particular market power (Neuhann and Sockin, 2023).

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.

Bagehot, W., 1971. The only game in town. Financial Analysts Journal 27, 12–14.

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

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 .

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.

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.

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.

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., 1970. Efficient capital markets: A review of theory and empirical work. The journal of Finance 25, 383–417.

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.

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

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

Garman, M. B., 1976. Market microstructure. Journal of financial Economics 3, 257–275.

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.

Harvey, C. R., Liu, Y., Zhu, H., 2016. … and the cross-section of expected returns. The Review of Financial Studies 29, 5–68.

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.

Ho, T., Stoll, H. R., 1981. Optimal dealer pricing under transactions and return uncertainty. Journal of Financial economics 9, 47–73.

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

Jensen, T. I., Kelly, B., Pedersen, L. H., 2022. Is there a replication crisis in finance? The Journal of Finance .

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.

O'hara, M., Oldfield, G. S., 1986. The microeconomics of market making. Journal of Financial and Quantitative analysis 21, 361–376.

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.

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

Appendix A. Appendix

Table A1: EOM: Independent double sort ret exc lead1m

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	0.53	0.89	0.96	1.15	0.12	0.25
2	0.80	0.77	1.13	0.87	1.96	1.16	2.07
3	0.72	0.80	1.08	1.09	1.55	0.82	2.31
4	0.38	0.64	0.73	1.15	1.62	1.24	2.77
High α_t^{RE}	-0.04	0.78	0.58	0.88	1.30	1.34	4.08
HmL α_t^{RE}	-1.07	0.25	-0.31	-0.07	0.15	1.22	
HmL t-Stat	-2.98	1.12	-1.41	-0.30	0.61	2.68	

Note. The table shows independent double sorts on α_t^{IN} and α_t^{RE} with the leading return variable from Jensen, Kelly, and Pedersen (2022).

Table A2: EOM: Single sort

Single sort					
α_t^{IN} portfolio	1	2	3	4	5
$Prob(\alpha_t^{IN} > 0 \,\&\, p(\alpha_t^{IN}) < 0.05)$	95.76	99.76	99.76	99.76	99.60
$Prob(\alpha_t^{RE} > 0 \,\&\, p(\alpha_t^{RE}) < 0.05)$	13.07	14.71	15.31	16.10	17.08

Note. The table shows the probabaility that the permanent price impact is positive and significant at the 5% level for different portfolios of α_t^{IN}.

Table A3: 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 OI_t	0.81	0.76	1.20	0.98	1.71	0.90	2.07
2	0.49	0.66	0.76	0.86	1.13	0.64	1.81
3	0.48	0.68	0.59	0.63	0.78	0.30	1.16
4	0.36	0.84	0.73	0.92	1.08	0.72	2.25
High OI_t	0.84	0.71	0.98	1.17	1.79	0.95	2.30
HmL OI_t	0.03	-0.05	-0.22	0.18	0.08	0.05	
HmL t-Stat	0.32	-0.45	-1.11	0.64	0.34	0.21	

Note. The table shows the returns to value-weighted stock portfolios first sorted by α_t^{IN}, and then by order imbalance (OI_t^{IN}), which is defined as buy minus sell.

Table A4: EOM: Dependent double sort

Dependent double sort

r_{t+1} (in %)	Low OI_t	2	3	4	High OI_t	HmL OI_t	HmL t-Stat.
Low α_t^{IN}	0.72	0.52	0.39	0.60	0.74	0.02	0.12
2	0.79	0.65	0.60	0.75	0.78	-0.00	-0.03
3	0.80	0.81	0.61	0.90	0.88	0.08	0.55
4	1.04	1.18	0.57	1.04	0.84	-0.19	-1.14
High α_t^{IN}	1.25	1.29	1.45	1.61	1.33	0.09	0.37
HmL α_t^{IN}	0.52	0.77	1.05	1.01	0.60	0.07	
HmL t-Stat	1.42	2.41	4.58	2.93	1.38	0.29	

Note. The table shows the returns to value-weighted stock portfolios first sorted by OI_t (order imbalance (buy minus sell)), and then by α_t^{IN}.

Table A5: EOM: 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 OI_t	0.77	0.75	1.20	0.97	1.82	1.05	1.87
2	0.47	0.67	0.80	1.00	1.47	1.00	2.65
3	0.25	0.57	0.56	0.56	0.97	0.72	3.00
4	0.47	0.87	0.68	0.90	1.60	1.13	2.79
High OI_t	0.77	0.70	0.98	1.21	1.89	1.12	2.53
HmL OI_t	0.00	-0.05	-0.22	0.25	0.07	0.06	
HmL t-Stat	0.04	-0.51	-1.15	0.74	0.17	0.16	

Note. The table shows the returns to value-weighted stock portfolios sorted independently by α_t^{IN}, and by order imbalance (OI_t), which is buy minus sell.

Table A6: 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 α_t^{RE}	0.89	0.54	0.99	0.91	1.28	0.39	0.83
2	0.73	0.87	1.21	0.98	1.77	1.04	2.34
3	0.77	0.67	0.93	1.08	1.45	0.69	1.53
4	0.61	0.71	0.72	1.07	1.50	0.89	3.11
High α_t^{RE}	0.33	0.60	0.78	0.90	0.90	0.57	1.51
HmL α_t^{RE}	-0.56	0.07	-0.22	-0.01	-0.38	0.18	
HmL t-Stat	-2.02	0.38	-1.39	-0.07	-1.09	0.43	

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

Table A7: EOM: Dependent double sort

Dependent double sort							
r_{t+1} (in %)	Low α_t^{RE}	2	3	4	High α_t^{RE}	HmL α_t^{RE}	HmL t-Stat.
Low α_t^{IN}	0.91	0.83	0.70	0.42	0.48	-0.43	-1.30
2	0.56	0.68	0.76	0.66	1.03	0.47	2.06
3	0.85	0.85	0.91	0.85	0.95	0.09	0.45
4	0.89	1.02	1.05	1.26	1.32	0.42	1.64
High α_t^{IN}	1.16	1.06	1.24	1.47	1.58	0.41	1.51
HmL α_t^{IN}	0.25	0.22	0.53	1.05	1.09	0.84	
HmL t-Stat	0.56	0.54	1.60	2.43	4.37	2.15	

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

Table A8: 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 A9: EOM: 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 $Size_t$	-0.85	-0.56	-0.08	0.31	1.00	1.85	4.88
2	-0.39	0.20	0.31	0.73	1.60	1.99	4.07
3	-0.36	0.47	0.70	1.10	1.70	2.06	5.52
4	0.30	0.70	1.00	1.35	1.49	1.18	2.59
High $Size_t$	0.79	0.75	1.12	1.04	2.07	1.40	1.57
HmL $Size_t$	1.64	1.32	1.20	0.74	0.98	-0.49	
HmL t-Stat	4.78	5.26	3.84	2.65	1.12	-0.47	

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

Table A10: EOM: Robustness: Betas across portfolios

	$HML^{Gross}_{\alpha^{IN}_{10}}$		
	Low	(2)	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 the .

Table A11: EOM: Alphas for each subportfolio

	$\text{HML}^{Gross}_{\alpha^{IN}_{10}}$									
	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 the .

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.

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

	$\|\Delta \mathrm{w}_t\|$			
	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 ($|\Delta \mathrm{w}_{t,i}|$) 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

	$\|\Delta \mathrm{w}_t\|$							
	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 ($|\Delta \mathrm{w}_{t,i}|$) 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.