ARTICLE IN PRESS

Journal of Financial Markets xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Financial Markets

journal homepage: www.elsevier.com/locate/finmar

Retail trading and analyst coverage

Charles Martineau [a,b,1], Marius Zoican [c,b,*,1]

[a] University of Toronto Scarborough, 1095 Military Trail, Toronto, Ontario, Canada M1C 1A4
[b] Rotman School of Management, 105 St. George Street, Toronto, Ontario, Canada M5S 3E6
[c] University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, Ontario, Canada L5L 1C6

ARTICLE INFO

JEL classification:
G11
G24
G40
D83
M41

Keywords:
Retail trading
Analyst coverage
Liquidity
Price efficiency

ABSTRACT

How does retail trading impact information supply in financial markets? We build a trading model with endogenous information supply where analysts maximize trading volume by institutional investors. In equilibrium, sell-side analysts provide higher quality signals in stocks with large retail interest, as institutional investors can trade more aggressively without revealing information. We provide empirical evidence supporting the main prediction of the model: A one standard deviation increase in retail trading leads to an additional 0.6 analysts covering the stock. To establish causality, we confirm our results using stock splits as a plausibly exogenous shock to retail trading.

1. Introduction

Retail traders are increasingly important in the financial markets ecosystem. Over the past decade, the transition from brick-and-mortar asset management firms to cost-efficient digital brokerages – boasting low fees and low minimum account sizes – has greatly improved access to financial markets for individual investors. As a result, in July 2020 retail traders accounted for more than 25% of the U.S. equity market, compared to only 10.1% in 2010 (Osipovich, 2020). Further, McCrank (2021) documents that retail traders are now the second-largest market segment after high-frequency market-makers, but ahead of quantitative investors (15.9%), hedge funds (9%) or bank-affiliated traders (5.8%). Two leading retail brokerages, Fidelity and Charles Schwab, manage more than $15 trillion U.S. dollars as of January 2021.

A shift in investor composition is likely to have a far-reaching impact on markets. How does retail trading impact the supply of information in financial markets, as measured, for example, by sell-side analyst coverage? In this paper, we investigate, both theoretically and empirically, the link between retail trading and information production in financial markets.

* Corresponding author at: University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, Ontario, Canada L5L 1C6.
E-mail addresses: charles.martineau@rotman.utoronto.ca (C. Martineau), marius.zoican@rotman.utoronto.ca (M. Zoican).

[1] We have greatly benefited from discussion on this research with Daniel Andrei, Michael Brolley, Alexandre Corhay, Peter Cziraki, Craig Doidge, Jerôme Dugast, Thierry Foucault, Nicolas Inostroza, Itay Goldstein, Jim Goldman, Vincent Grégoire, Vincent Van Kervel, Tom McCurdy, Jordi Mondria, Shannon Liu, Chayawat Ornthanalai, Christine Parlour, Jorge Sabat, Ioanid Roşu, Liyan Yang, Mao Ye, and Yan Xiong. We are grateful to seminar participants at Université Paris-Dauphine, University of Sydney, Bank of Lithuania, University of Toronto Rotman School of Management and Department of Economics, 2019 Santiago Finance Workshop, 2019 Paris Finance December Meeting, 2020 Future of Financial Information, 2020 Northern Finance Association for insightful comments. We thank Xiaopeng Wu for excellent research assistance. The authors gratefully acknowledge the Financial Innovation Lab (FinHub) at Rotman School of Management for financial support and the Canadian Social Sciences and Humanities Research Council (SSHRC) for an Insight Development research grant (430-2020-00014). Marius Zoican acknowledges financial support from Quantitative Management Research Initiative (QMI) under the aegis of the Fondation du Risque, a joint initiative by Université Paris-Dauphine, l'École Nationale de la Statistique et de l'Administration ParisTech (ENSAE), and LFIS, France.

https://doi.org/10.1016/j.finmar.2023.100849
Received 29 June 2022; Received in revised form 17 May 2023; Accepted 22 May 2023
Available online 29 May 2023

Our first contribution is to build a noisy rational expectation equilibrium model (NREE) with endogenous information supply. In the model, uninformed retail investors trade alongside sophisticated institutional investors and liquidity-driven noise traders. Retail and institutional traders have access to different information: while institutional investors observe both the clearing price and an informative analyst signal, retail traders can only condition their demand on the clearing price. In this sense, retail and institutional investors correspond, respectively, to uninformed and informed traders in a standard NREE framework such as, for example, Grossman and Stiglitz (1980). The presence of noise traders implies that the analyst's public signal is only imperfectly revealed in prices, which leads to an informational advantage for institutional traders. We argue that this informational gap in the model captures a real-world friction: institutional investors extract a higher value than their retail counterparts from analyst coverage, through enhanced client service (e.g., discussing ideas with analysts) and access to company management.

We contribute to the NREE literature by explicitly modeling the analyst's information supply problem. In our setup, the analyst endogenously chooses the precision of their signal to maximize institutional trading volume for its affiliated brokerage house. The objective is consistent with the empirical evidence in Bhushan (1989), Frankel et al. (2006), or Groysberg et al. (2011) on analyst career concerns. To provide a higher precision signal, the analyst incurs a convex effort cost.

How does retail trading impact institutional trading volume in our model? First, there is a direct substitution effect, since retail and institutional traders compete to provide liquidity to noise traders. A larger mass of retail traders crowds out liquidity provision by institutional investors and leads to lower institutional trading volume.

Second, there is an indirect effect driven by the precision of analyst forecasts. More precise analyst signals reduce asset payoff uncertainty, leading to an increase in institutional trading volume. We show that this effect is greater for stocks with a larger mass of retail traders, as higher retail participation enables institutional investors to trade more aggressively on the signal without revealing information.

In equilibrium, we show that analysts produce more precise signals for stocks with a larger mass of retail traders, where trading volumes are more elastic with respect to information quality. This information-driven increase in institutional volume partially offsets the direct substitution effect arising from liquidity competition.

Our second contribution is that we empirically document a strong correlation between retail trading and analyst coverage, even after controlling for factors associated with analyst coverage, such as market capitalization, investor attention, or institutional ownership. We proxy the strength of analyst coverage by the quarterly number of analyst earnings forecasts in a given stock, which we collect from I/B/E/S for the 2014 to 2020 period. To identify retail trades from high-frequency data, we use the algorithm developed by Boehmer et al. (2021). We then construct a stock-quarter retail trading measure as the share of total retail volume in dollars divided by the total dollar volume.

In line with the model's key prediction, we find that retail trading is strongly correlated with analyst coverage: a one percentage point increase in the retail share leads to 0.068 more analyst reports. In relative terms, if the retail volume share increases by one standard deviation (9.33 percentage points), a stock is covered by 0.63 more analysts. The impact is economically significant, as the median U.S. stock only receives two analyst reports.

We use stock split events to instrument for the retail trading share to provide suggestive evidence of a causal link between retail trading and analyst coverage. Following the intuition in Brandt et al. (2010) and Cox et al. (2022), stock splits have a direct impact on retail trading interest. Since a stock split reduces the nominal share price, it makes the stock more affordable to retail traders who are typically capital-constrained. As in Brandt et al. (2010), we find that stocks that experience a forward (reverse) stock split result in an increase (decrease) in retail trading. We implement the instrumental variable estimator using two-stage least squares (2SLS) and find strong support for the relationship between analyst coverage and retail trading.

In Section 2, we discuss the related literature. In Section 3, we discuss our theoretical model of analyst coverage. In Section 4, we provide empirical evidence on the relation between retail trading and analyst coverage. Finally, Section 5 concludes.

2. Related literature

Our paper is closely linked to the literature on information disclosure and information supply in financial markets, building on the noisy rational expectation equilibrium models of Grossman and Stiglitz (1980) and Verrecchia (1982). A review paper by Goldstein and Yang (2017) provides a comprehensive survey of this extensive literature. Our novel contribution is to model the endogenous supply of public information, in particular the analysts' incentive to choose a signal precision that maximizes trading volume in the stock. We follow Admati and Pfleiderer (1988) and Mondria et al. (2022) to define trading volume as the difference in holding between the start and the end of the trading game, where the initial holdings are set to zero.

Our paper also relates to the literature on retail trading [see Barber and Odean (2013) for a comprehensive survey]. Closest to our paper, Kaniel et al. (2008) and Barrot et al. (2016) argue that retail traders act as uninformed liquidity providers who reduce the price impact of trades. We complement this view with the argument that institutional investors can maximize the value of analyst reports by aggressively trading in stocks with high retail participation, for which they reveal less information through trading. Our assumption that retail traders are uninformed is in line with the majority of empirical findings. Barber and Odean (2000, 2007) document that retail traders exhibit behavioral biases that hurt their performance. Using Taiwanese data, Barber et al. (2008) estimate a 3.8% annual performance penalty for individual portfolios. More recently, Barber et al. (2022) show that users of Robinhood's trading app (i.e., the *Top Movers* tab) are drawn to stocks with extreme returns, leading to portfolio underperformance. Eaton et al. (2022) show that exogenous shocks to participation in the Robinhood platform are uncorrelated with future returns, and conclude that retail investors behave as noise traders. However, Welch (2021) documents a good risk-return performance of retail investors, as a group, between 2018 and 2020.

Finally, we contribute to the literature on analysts' coverage choices. What determines the amount and quality of coverage a stock receives from an analyst? Bhushan (1989) presents a simple model of total expenditure by investors on analyst services and documents that firm size, institutional ownership, and stock price volatility are some of the most common characteristics correlating with analyst coverage. Since then, the literature on analyst coverage examines the determinants of coverage from the investor and the analyst perspective. From the investor perspective, Barth et al. (2001) find that investors demand more analyst coverage for firms with high intangibles (e.g., high R&D expenditure) and those that are more difficult to examine. Hameed et al. (2015) study how heterogeneity in coverage value, from the investors' perspective, drives analyst coverage decisions. We complement the literature by documenting a new factor driving analysts' coverage choices: the intensity of retail trading in a particular stock.

Our assumption that analysts maximize volume-driven commissions is grounded in the accounting literature on analyst career concerns. Groysberg et al. (2011) find that analysts' compensation and future employment prospects depend on their reputation ("all star" status recognized by the *Wall Street Journal*) and ability to generate commission revenue. Analysts are more likely to choose to follow firms with significant trading volume and institutional ownership as they represent a more lucrative source of income for their brokerage.

3. A model of analyst coverage and retail trading

In this section, we discuss our noisy rational expectations model that is based on Grossman and Stiglitz (1980) and Goldstein and Yang (2017). We extend these seminal models by introducing endogenous information supply — starting from the assumption that sell-side analysts aim to maximize their clients' trading volume. In Appendix A, we list all exogenous parameters and endogenous quantities in the model.

3.1. Model primitives

3.1.1. Assets and markets

Consider a three-date economy, where time is indexed by $t \in \{0, 1, 2\}$. A single risky asset pays off a stochastic dividend $v \sim \mathcal{N}\left(0, \tau_v^{-1}\right)$ at $t = 2$. Additionally, agents can lend and borrow freely at a risk-free rate of r, which we normalize to one. The risky asset is traded on an auction market that clears at $t = 2$, just before the stochastic dividend is revealed. For tractability, we assume the asset is in zero net supply; in Appendix B, we solve the model numerically for the case of positive net supply and obtain qualitatively identical results. All trades are executed at the market clearing price.

3.1.2. Investors

There are two types of investors in the economy: a unit mass of price-sensitive institutional investors (**I**), as well as a mass $\mu \leq 1$ of retail investors (**R**).[2] Both types have CARA utility over wealth at the terminal date $t = 2$, and a risk-aversion coefficient of γ. That is, the investors' expected utility can be written as:

$$U_{\{\mathbf{I},\mathbf{R}\}} = -\mathbb{E}\left[\exp\left(-\gamma W_2\right)\right]. \tag{1}$$

The two investor types differ in their information set. In particular, institutional investors have access to analyst coverage, whereas retail investors do not (but can partially infer its content through prices). We motivate the assumption as retail traders having impaired access to analyst services. For institutional investors, the value added by analysts goes well beyond the public report and may include client services (the ability to discuss ideas with the analysts), as well as enhanced access to the company management. In addition, retail traders may lack the financial sophistication required to extract trading signals from reports.

At $t = 1$, each investor j of type $i \in \{\mathbf{I}, \mathbf{R}\}$ chooses her demand $\mathcal{D}_{ij}(\cdot)$ as a function of the market clearing price. Noise traders participate in the asset market alongside investors. The noise trader demand u is distributed as a random normal variable with variance τ_u^{-1}: $u \sim \mathcal{N}\left(0, \tau_u^{-1}\right)$. The presence of noise traders guarantees that the market clearing price does not perfectly reveal the analyst's signal.

3.1.3. Discussion: Who are the noise traders?

In line with the literature on NREE models, noise traders in our setup lead to prices reflecting information imperfectly, thus circumventing the "no trade" result of Grossman and Stiglitz (1980). How do noise traders in the model map to real-life agents? We note that in our model both **I** and **R** traders are price-sensitive in the sense that their demand is a function of the clearing price. Therefore, we interpret noise traders in the model as price-*insensitive* investors who follow mechanical portfolio re-balancing rules: for example, they can stand in for exchange-traded funds (ETFs) that passively track an index. This interpretation is in line with recent literature: Chinco and Fos (2021) show that if many funds employ threshold-based rules to select securities, predicting how the rules interact with each other quickly becomes computationally infeasible, which translates to unpredictable noise. Since such passive funds do not (typically) trade on information, we assume the analysts' objective function does not depend on their volume.

It is possible to recast the source of noise in our model such that all traders are discretionary. Grossman (1976, p. 574) interprets noise in a similar model as "an uncertain stock of the risky asset", interpretation which is also used in Grossman and Stiglitz (1980). In this case, noise arises from a random asset supply rather than from trading demand, which leaves **I** and **R** investors as the only agents in the model.

[2] We assume that $\mu < 1$ for tractability purposes, as it allows us to prove the main result in Proposition 1 analytically. The assumption is consistent with the fact that retail traders amount to less than half of aggregate volume in equity markets, even throughout the recent retail boom of 2020–2021.

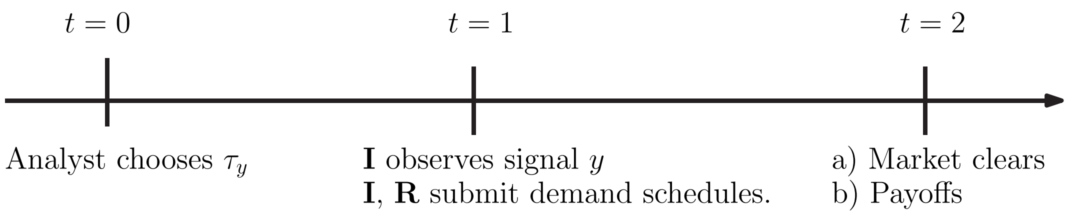

Fig. 1. Model timing.

3.1.4. Analyst coverage

There is a single analyst who provides at $t = 1$ an unbiased signal about the asset's payoff: $y = v + \epsilon_y$, where $\epsilon_y \sim \mathcal{N}\left(0, \tau_y^{-1}\right)$. Better signals require the analyst to exert more effort: The cost of producing a signal with precision τ_y is $\frac{1}{2}c\tau_y^2$ for a $c > 0$.[3]

The analyst chooses the signal precision $\tau_y^\star$ at $t = 0$ to maximize the expected trading volume from **I** traders net of the signal cost, that is:

$$\tau_y^\star = \arg\max_{\tau_y} \mathbb{E}\text{Volume}_{\mathbf{I}}\left(\tau_y\right) - \frac{c}{2}\tau_y^2. \tag{2}$$

The rationale behind the analyst's objective function is driven by the structure of the financial information industry. Sell-side analysts are affiliated with brokerage houses that process investors' orders and collect volume-based fees. Therefore, sell-side analysts have an incentive to generate high volumes for their brokerage houses.[4] We note that the volume-maximization objective is consistent with analysts providing unbiased forecasts. That is, security analysts in the model do not have an incentive to move prices in any particular direction (i.e., "pump" the stock price), but rather to stimulate trade between investors. Forecast bias might arise in a different model, where analysts are employed by investment banks whose incentives are aligned with those of firms that issue equity.

In the model, the analysts' objective function can be interpreted to emerge from career concerns (Holmström, 1999). Hong and Kubik (2003) show that analysts' careers depend less on forecast accuracy and more on their ability to promote stocks and generate investment banking business. In the same spirit, Frankel et al. (2006) document that analysts direct their efforts towards researching firms with large trading volumes and institutional ownership to maximize commission fee revenue for brokerage houses. Finally, Groysberg et al. (2011) find that analyst compensation is related to investment banking contributions, and reputation effects (e.g., "All-Star" recognition) and not necessarily to the quality of the forecasts. Such reputation and business-generating incentives are consistent with analysts competing on investor attention. In particular, in light of the empirical evidence, the analysts' objective may not coincide with the investors' problem (i.e., reduce variance across the portfolio).

3.1.5. Timing

Fig. 1 provides the sequence of events at each time $t \in \{0, 1, 2\}$.

We solve the game by backward induction. In Section 3.2, we solve for the market clearing prices at $t = 2$ and the optimal investor demand schedule at $t = 1$, taking the investor signal precision as given. Next, in Section 3.3 we solve for the analysts' optimal signal precision. We provide proofs for the formal results of our model in Appendix C.

3.2. Market clearing and optimal demand schedules

Following the literature (e.g., Grossman and Stiglitz, 1980; Goldstein and Yang, 2017), we conjecture that the market clearing price p is a linear combination between the analyst forecast y and the noise trading demand u, with weights p_y and p_u, respectively:

$$p = p_y y + p_u u. \tag{3}$$

Next, we verify the linear price conjecture and obtain closed-form expressions for p_y and p_u. Since investors have CARA utility and stochastic quantities are normally distributed, it immediately follows from Verrecchia (1982) that the optimal demand schedule from investor j is:

$$\mathcal{D}_j\left(p\right) = \frac{\mathbb{E}\left[v \mid \mathcal{F}_j\right] - p}{\gamma \text{var}\left[v \mid \mathcal{F}_j\right]}, \tag{4}$$

where $\mathcal{F}_j$ is the information set of investor j.

[3] Our results are robust to using any cost function that increases in the signal precision τ_y and does not depend on the mass of retail traders μ.

[4] The assumption is consistent with the U.S. regulatory environment, where research and trading services can be bundled together. In Europe, following MiFID II, the two services have to be offered separately, thus leading to an arguably weaker link between trading volume and incentives for information provision.

From Bayes' rule, and from the properties of the normal distribution, we obtain the posterior expectation and variance for institutional investors as follows:

$$\mathbb{E}[v \mid y] = \frac{\tau_y}{\tau_y + \tau_v} y \text{ and}$$
$$\text{var}[v \mid y] = \frac{1}{\tau_y + \tau_v}. \tag{5}$$

We note that since all institutional investors observe the analyst forecast y, and the conjectured price does not depend on any other informative signal, the posterior distribution of the asset payoff does not depend on the price. That is, $\mathbb{E}[v \mid y] = \mathbb{E}[v \mid y, p]$. Consequently, the optimal demand schedule for **I** investors obtains by replacing the quantities from Eq. (5) in Eq. (4):

$$\mathcal{D}_{\mathbf{I}} = \frac{1}{\gamma}\left(\tau_y y - \left(\tau_y + \tau_v\right) p\right). \tag{6}$$

We turn next to the retail traders' demand. Since the market clearing price is a function of the analyst signal, it becomes informative for retail traders. In particular, the price is equivalent to a signal s_p, where:

$$s_p = v + \epsilon_y + \frac{p_u}{p_y} u, \tag{7}$$

with variance $\tau_y^{-1} + h^2 \tau_u^{-1}$, where $h = \frac{p_u}{p_y}$. It follows that the posterior expectation and variance of the asset payoff for retail traders becomes:

$$\mathbb{E}[v \mid p] = \frac{1}{p_y} \frac{\phi \tau_y}{\tau_v + \phi \tau_y} p$$
$$\text{var}[v \mid p] = \left(\tau_v + \phi \tau_y\right)^{-1}, \tag{8}$$

where we define $\phi = \frac{\tau_u}{\tau_u + h^2 \tau_y}$ for simplicity of notation. Since $\phi < 1$, the retail traders benefit less than institutional investors from analyst coverage, since they only learn about the signal indirectly through prices. Therefore, the retail trader demand schedule is:

$$\mathcal{D}_{\mathbf{R}} = \frac{1}{\gamma}\left(\frac{1}{p_y} \phi \tau_y p - \left(\phi \tau_y + \tau_v\right) p\right). \tag{9}$$

The equilibrium price is pinned down by the market clearing condition:

$$\int_0^1 \mathcal{D}_{\mathbf{I}} \mathrm{d}i + \int_0^\mu \mathcal{D}_{\mathbf{R}} \mathrm{d}i + u = 0, \tag{10}$$

where all investors of a given type have identical information sets and therefore identical demands. The resulting price verifies the original linear conjecture since:

$$p = \underbrace{\frac{p_y \gamma}{p_y\left((1+\mu)\tau_v + \tau_y + \mu \tau_y \phi\right) - \mu \tau_y \phi}}_{p_u} u + \underbrace{\frac{p_y \tau_y}{p_y\left((1+\mu)\tau_v + \tau_y + \mu \tau_y \phi\right) - \mu \tau_y \phi}}_{p_y} y. \tag{11}$$

The coefficients in Eq. (11) define a two-equation system in p_y and p_u. We solve the system to find:

$$p_y = \tau_y \frac{1 + \mu\phi}{\tau_v(1+\mu) + \tau_y(1+\mu\phi)} \text{ and} \tag{12}$$
$$p_u = \gamma \frac{1 + \mu\phi}{\tau_v(1+\mu) + \tau_y(1+\mu\phi)}. \tag{13}$$

We follow Goldstein and Yang (2017) to define liquidity as the inverse of p_u, that is the price impact of noise trading u. Lemma 1 states that a larger mass of retail traders reduces the price impact on liquidity demanding trades. The result is intuitive since a larger retail trader sector is better able to absorb liquidity shocks, as documented empirically by Kaniel et al. (2008) and Barrot et al. (2016).

Lemma 1. *Conditional on the analyst signal precision, market liquidity increases in the mass of retail traders μ.*

Proof. The result is immediate since:

$$\frac{\partial p_u^{-1}}{\partial \mu} = \frac{\tau_v(1-\phi)}{\gamma(1+\mu\phi)^2} > 0. \quad \square \tag{14}$$

Further, the price loading on the analyst signal p_y also decreases in the mass of retail traders μ. That is, higher retail trading interest reduces the sensitivity of prices to (public) news and allows competitive informed investors to trade more aggressively on the analysts' signals.

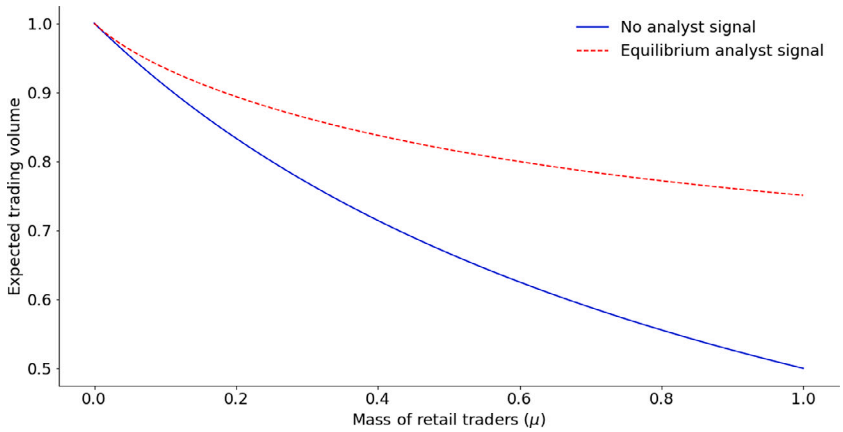

Fig. 2. Equilibrium trading volume and retail trading.
This figure illustrates the expected trading volume (i) in the absence of analyst coverage and (ii) at the equilibrium analyst signal precision $\tau_y^\star$, as a function of the mass of retail traders μ. Parameter values: $\tau_v = 1$, $\gamma = 5$, $\tau_u = 0.5$, $c = 0.2$.

3.3. Trading volume and endogenous analyst signal

In this subsection, we compute the expected trading volume from institutional investors. We then solve for the analysts' optimal signal precision at $t = -1$.

Trading volume can be computed as the absolute difference in holdings between the beginning and the end of the trading game (Admati and Pfleiderer, 1988; Wang, 1994). The final portfolio size is straightforward: each **I** investor holds $\mathcal{D}_{\mathbf{I}}(p)$ units of the asset — in a CARA-normal framework without wealth effects, the optimal portfolio is unrelated to the prior holdings. We follow, for example, Mondria et al. (2022) and assume that at $t = 0$ each institutional investor holds zero units of the asset. In this case, the expected volume is given by:

$$\mathbb{E}\text{Volume}_{\mathbf{I}} = \mathbb{E}\int_0^1 \left\|\mathcal{D}_{\mathbf{I}i}(p) - 0\right\| \mathrm{d}i = \left\|\mathcal{D}_{\mathbf{I}}(p)\right\|. \tag{15}$$

The assumption allows us to focus sharply on the impact of retail trading on analyst signals, abstracting from confounding channels such as risk sharing between institutional investors. To compute the expression in Eq. (15), we substitute the equilibrium price in Eq. (11) in the institutional investors' demand schedule in Eq. (6):

$$\begin{aligned}\mathcal{D}_{\mathbf{I}}(p) &= \frac{1}{\gamma}\left[\left(\tau_y - \left(\tau_y + \tau_v\right)p_y\right)y - \left(\tau_y + \tau_v\right)p_u u\right] \\ &\sim \mathcal{N}\left(0, \frac{1}{\gamma^2}\left[\frac{\left(\tau_y - \left(\tau_y + \tau_v\right)p_y\right)^2\left(\tau_y + \tau_v\right)}{\tau_y\tau_v} + \frac{\left(\left(\tau_y + \tau_v\right)p_u\right)^2}{\tau_u}\right]\right),\end{aligned} \tag{16}$$

where we use that $\text{var}(y) = \frac{\tau_y+\tau_v}{\tau_y\tau_v}$ and $\text{var}(u) = \tau_u^{-1}$. From the properties of the normal distribution, we further have that:

$$\mathbb{E}\text{Volume}_{\mathbf{I}}\left(\tau_y\right) = \sqrt{\frac{2}{\pi}}\frac{1}{\gamma}\sqrt{\frac{\left(\tau_y - \left(\tau_y + \tau_v\right)p_y\right)^2\left(\tau_y + \tau_v\right)}{\tau_y\tau_v} + \frac{\left(\left(\tau_y + \tau_v\right)p_u\right)^2}{\tau_u}}. \tag{17}$$

Fig. 2 illustrates that the expected volume decreases in the mass of retail traders. To understand the intuition, we focus first on the case with no analyst coverage (or equivalently $\tau_y = 0$). In this scenario, institutional and retail traders have identical information sets and compete to provide liquidity to noise traders. The expected volume in Eq. (17) becomes:

$$\mathbb{E}\text{Volume}_{\mathbf{I}}(0) = \frac{\gamma\tau_v}{\tau_v(1+\mu)\sqrt{\tau_u}}, \tag{18}$$

which decreases in μ. Intuitively, if all trading is motivated by liquidity provision, institutional and retail traders are substitutes. Therefore, a higher mass of retail traders reduces institutional volume (i.e., a direct effect). The indirect effect is driven by endogenous information supply. A higher mass of retail traders enables institutional investors to trade more aggressively on analyst signals. Therefore, the expected institutional volume increases.

We turn next to endogenous information supply. At $t = 0$, the analyst chooses its signal precision $\tau_y^\star$ to maximize:

$$\tau_y^\star \equiv \max\left\{0, \arg\max_{\tau_y}\sqrt{\frac{2}{\pi}}\frac{1}{\gamma}\sqrt{\frac{\left(\tau_y - \left(\tau_y + \tau_v\right)p_y\right)^2\left(\tau_y + \tau_v\right)}{\tau_y\tau_v} + \frac{\left(\left(\tau_y + \tau_v\right)p_u\right)^2}{\tau_u}} - \frac{c\tau_y^2}{2}\right\}, \tag{19}$$

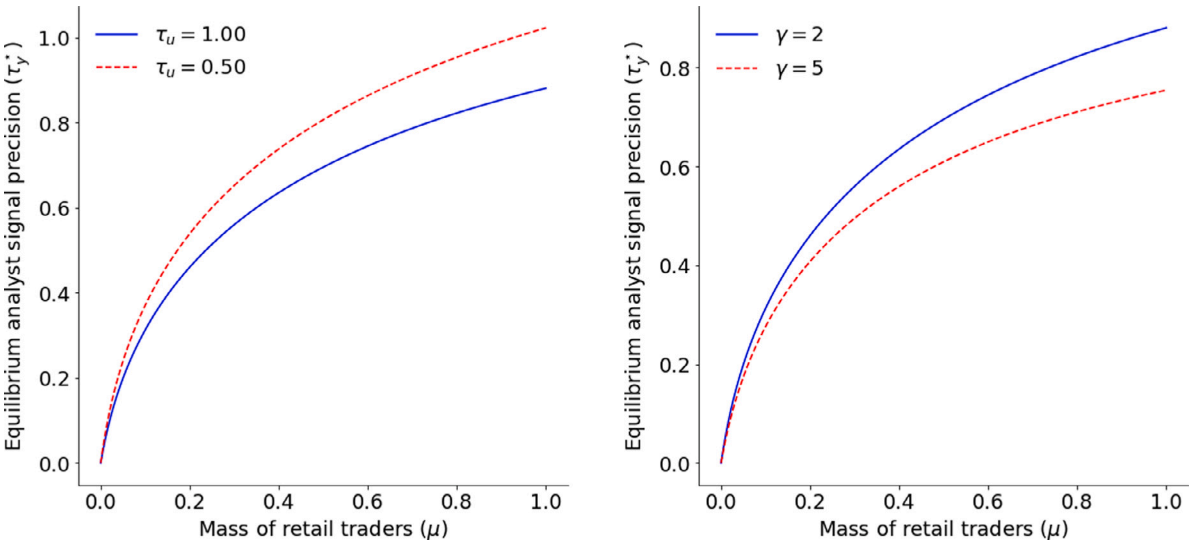

Fig. 3. Equilibrium analyst signal precision.
This figure illustrates the equilibrium analyst signal precision $\tau_y^\star$ as a function of the mass of retail traders μ. Parameter values: $\tau_v = 1$, $\gamma = 5$, $\tau_u = 0.5$, $c = 0.2$.

where p_y and p_u are given in Eq. (12).

Proposition 1 states the main result of our model, which we illustrate in Fig. 3. A higher mass of retail traders leads to a lower elasticity of price with respect to news (p_y). As a result, **I** investors are able to trade more aggressively on the information in the analyst's report. Consequently, the analyst has an incentive to increase the precision of the signal and boost trading volume.

Proposition 1. *The analyst signal precision (that is, $\tau_y^\star$) increases in the mass of retail traders μ.*

Fig. 3 highlights that larger liquidity trading (i.e., lower τ_u) boosts the analyst's signal precision. The rationale is that liquidity traders introduce noise into the price signal, thereby preventing retail traders from accurately inferring the information in analyst reports and further boosting the informational advantage of institutional traders.

Corollary 1 states that higher retail trader participation should improve market efficiency. The rationale is that an increase in the mass of retail traders stimulates information production by analysts who aim to maximize traded volume, and therefore the clearing price becomes more informative about the asset payoff.

Corollary 1. *Price efficiency, as measured by the asset payoff precision conditional on price, increases in the mass of retail traders. That is,*

$$\frac{\partial var^{-1}(v \mid p)}{\partial \mu} > 0. \tag{20}$$

Corollary 2 highlights the relationship between retail trading activity and liquidity. The result is driven by two forces. First, a higher mass of retail traders increases the risk-taking capacity of the market and allows noise traders to spread out the risk across more counterparties. Second, a jump in retail trading increases the signal precision of the analyst and reduces uncertainty for both informed and retail traders (since the clearing price is more informative), which allows them to provide more liquidity. Fig. 4 illustrates the results on liquidity and price efficiency.

Corollary 2. *Liquidity increases in the mass of retail traders. That is,*

$$\frac{\partial p_u^{-1}\left(\tau_y^\star\right)}{\partial \mu} > 0. \tag{21}$$

A subtle prediction of the model is that the separating research and trading services should weaken the link between retail trading and analyst coverage. The prediction could be tested, for example, in the context of European markets where the MiFID II regulation mandates the un-bundling of research and trading commission.

4. Evidence on retail trading and analyst coverage

In this section, we empirically examine the link between analyst coverage and retail trading. We also test the main prediction of our model in Proposition 1.

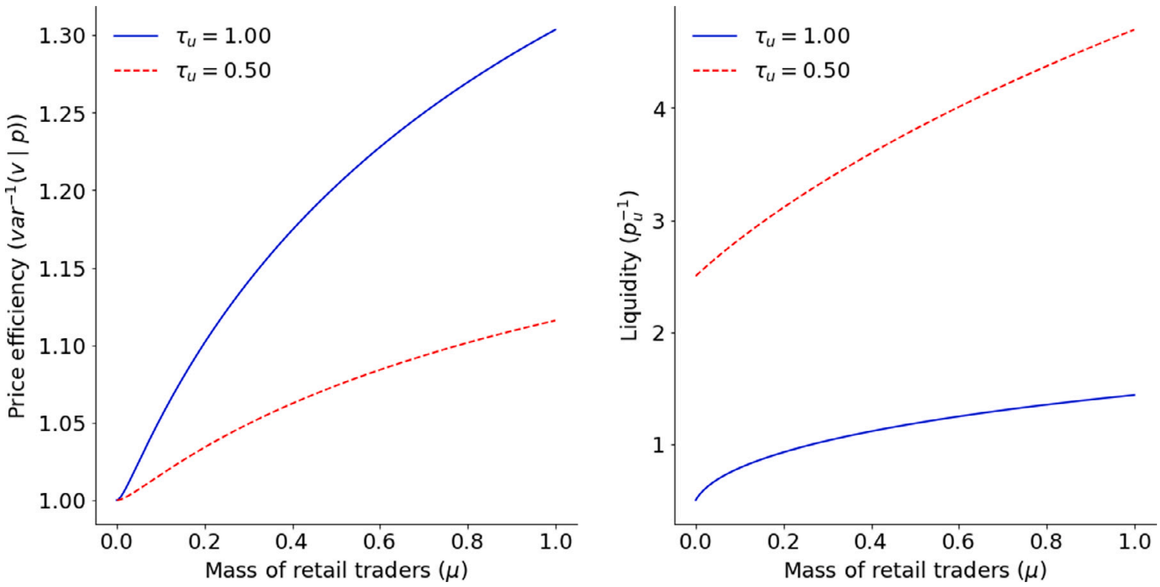

Fig. 4. Equilibrium price efficiency and market liquidity.
This figure illustrates the equilibrium price efficiency $\text{var}^{-1}(v \mid p)$ and market liquidity p_u^{-1} as a function of the mass of retail traders μ. Parameter values: $\tau_v = 1$, $\gamma = 5$, $\tau_u = 0.5$, $c = 0.2$.

4.1. Data

4.1.1. Analyst coverage

We retrieve analyst coverage data based on quarterly analyst earnings forecasts from I/B/E/S. We select only analyst forecasts issued over the 90 days before the earnings announcement date. If analysts revise their forecasts during this interval, we use only their most recent forecasts. We follow Livnat and Mendenhall (2006) and impose the following selection criteria for each firm's earnings announcement i for firm quarter q:

1. The earnings announcement date is reported in Compustat.
2. The price per share is available from Compustat as of the end of quarter q and is greater than \$1 and the stock market capitalization is greater than \$5 million.
3. The firm's shares are traded on the New York Stock Exchange (NYSE), American Stock Exchange, or NASDAQ.
4. Accounting data, specifically total assets and market capitalization, are available in Compustat at the end of December of the previous calendar year.

We measure the quality of information supply by the number of analyst forecasts for a given stock and quarter. Our measure aligns with results in the accounting and finance literature documenting that greater coverage leads to a faster and more complete price adjustment process (see, for example, the evidence in Brennan et al., 1993; Gleason and Lee, 2003). We argue that the number of analyst reports is a "cleaner" measure of information supply than, for example, standardized unexpected earnings (SUE), since the latter are a function of endogenous market prices. Further, using the number of analysts allows us to capture the decision of *not* supplying information: Indeed, there are a significant number of stock-quarters in our sample with zero coverage.

4.1.2. Retail trading

To measure retail volume, we use the TAQ database to identify retail trades using the algorithm developed by Boehmer et al. (2021). The algorithm relies on the observation that retail trades often receive price improvement in fractions of a penny and are routed to a FINRA trade reporting facility. The algorithm identifies retail trades as those reported to a FINRA trade reporting facility (exchange code "D" in TAQ) with fractional penny prices between 0.006 and 0.01 for retail buys and between 0.00 and 0.004 for retail sells.

In Fig. 5, we plot the 21-day moving average of retail volume as a fraction of total equity dollar volume. A salient feature of Fig. 5 is the sharp increase in retail trading at the beginning of 2020.

We argue that the retail surge is not driven by information supply, but rather by a combination of two exogenous events. First, three major online brokerages in the U.S. (Charles Schwab, TD Ameritrade Holding Corp., and E*Trade Financial Corp) eliminated trading commissions in October 2019. In a survey study, Lush et al. (2021) show that greater access to financial markets (e.g., brokerages opening accounts with zero or low minimum balance) was a primary driver of new account openings in 2020. Second, the COVID-19 pandemic generated a shift in work patterns and entertainment opportunities leading to a heightened trading appetite for individuals (Ozik et al., 2021).

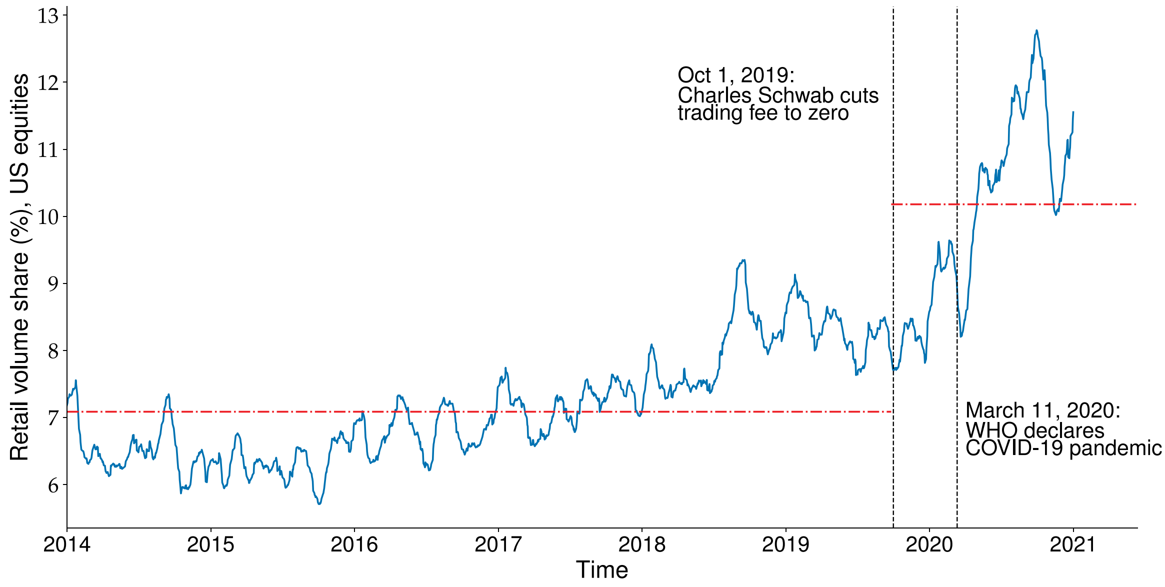

Fig. 5. The growth in retail trading volume.
This figure illustrates the 21-day moving average of retail dollar volume share on U.S. equity markets, as a fraction of total equity dollar trading volume. Retail volume is identified using the Boehmer et al. (2021) algorithm, i.e., a marketable order is identified as retail if (i) is reported to a FINRA trade reporting facility (exchange code "D" in TAQ), and (ii) it receives price improvement relative to the tick size (i.e., has a fractional penny price). We highlight two events that spurred rapid growth in retail volumes: a trading commission race-to-zero between U.S. online brokerages at the beginning of 2019Q4, and the start of the COVID-19 pandemic on March 11, 2020. Further, we use horizontal dash-dot lines to display the average retail volume share between 2014Q1 and 2019Q4, and 2020Q1 through 2020Q4, respectively. Further, for each stock and quarter, we compute their respective *retail share* as the proportion of retail dollar volume for stock i over total dollar volume for stock i.

4.1.3. Other controls

We consider additional variables known to influence analyst coverage. An increase in information demand for a particular stock (e.g., triggered by a news event) could simultaneously lead to higher retail interest and analyst coverage. To cleanly identify the effect of retail trading on analyst coverage, we further control for institutional investor attention as a proxy for information demand. We follow Ben-Rephael et al. (2017) and construct a measure of investor attention based on demand for information from the Bloomberg terminal, which is available since 2012. The Bloomberg measure of attention combines the number of times terminal users actively search, and subsequently read, news articles about a particular stock.[5] Bloomberg assigns a score of 10 when users search for news and a score of 1 when users read a news article. The scores are then aggregated into hourly counts for each stock. Bloomberg then creates an abnormal attention score. If the score is in the bottom 80% relative to the past 30 days, the abnormal attention score is 0. If the score is in the top 20% (10%, 6%, 4%), the abnormal attention score is 1 (2, 3, 4, respectively). Only the abnormal attention measures are retrievable from Bloomberg. We retrieve data on 3018 U.S. stocks with attention scores from Bloomberg Terminals. If a stock has no attention score, it is assigned a score of 0 (i.e., no attention). If investors redistribute part of their limited attention to a particular stock, that stock will appear to have abnormal attention in our data set. Therefore, if investors have limited capacity, the Bloomberg abnormal attention measure can reflect the cross-sectional attention distribution. For each stock-quarter, we average the number of days for which the Bloomberg attention measure is greater than 0.

We consider (log) market capitalization, book-to-market, research and development expenditure scaled by total sales, and institutional ownership in each quarter. In one of the specifications, we also interact institutional ownership with the event dummy to reflect the shift in investor composition starting in 2020. We also include quarterly average daily stock turnover, firm age (the number of years since first listed on CRSP), the quarterly stock return and total absolute quarterly return, and quarterly stock volatility (the standard deviation of daily returns).

Our analysis is conducted at a quarterly frequency because we observe the number of analysts at that frequency. The sample period is from January 1, 2014 to December 31, 2020 for a total of 4978 unique firms. We start our analysis from 2014 because before 2014, the TAQ database that we employ to compute the share of retail trading did not include trades with less than 100 shares, i.e., odd-lots (see O'Hara et al., 2014).

Panel A in Table 1 reports descriptive statistics for firm-level variables computed quarterly. The average number of analysts covering a firm (N. analyst) is 4.21. For stocks with at least one analyst, the median number is 5.24. The average share of retail dollar volume is 9.85%. For stocks with no analyst coverage, the average share of retail trading is 15.93% and 5.11% for stocks

[5] For more detail on the construction of the measure, we refer the reader to Ben-Rephael et al. (2017) and to https://www3.nd.edu/~zda/AIA_App.pdf on how to retrieve the data.

Table 1
Firm-level summary statistics.

	Full sample					Analyst coverage group			
	Mean	P25	P50	P75	Std.	Zero	Low	Medium	High
						Mean			
N. analyst	4.21	0	2.00	6.00	5.70	0	1.88	5.24	13.69
Log market capitalization ($)	20.36	18.80	20.34	21.83	2.17	18.76	20.12	21.14	22.72
Share of retail volume (%)	9.85	3.43	5.68	13.48	9.33	15.93	8.35	5.59	5.11
Attention	5.79	0	0	7.00	11.25	1.70	2.68	5.74	17.07
Turnover (%)	1.01	0.33	0.64	1.13	1.34	0.98	0.83	1.01	1.27
Book-to-Market	1.38	0.46	0.85	1.43	4.03	1.56	1.42	1.27	1.12
R&D as fraction of sales (%)	0.10	0	0	0.06	0.26	0.12	0.11	0.10	0.06
IO (%)	51.91	14.46	59.97	85.20	35.87	32.30	55.41	65.37	68.71
Age (years)	19.78	5.00	16.00	28.00	18.25	15.69	19.03	20.65	27.13

This table presents the summary statistics of firm-specific variables, computed quarterly. N. analyst is the number of quarterly unique analyst providing quarterly earnings forecast for each firm. Log market capitalization is the natural logarithm of the market value of the firm on quarter t. Share of retail volume is the proportion of retail dollar volume over the total dollar volume. Attention is the total number of days with high abnormal attention from Bloomberg Terminal. Turnover is the average daily turnover. Book-to-Market corresponds to the quaterly book-to-market ratio. R&D is the research and development scaled by total sale on quarter t. IO is the institutional ownership coverage on quarter t. Age is the number of years since first listed on CRSP. The table reports the mean, the 25th, 50th, and 75th percentiles, and standard deviation (Std.) for the full sample of stocks as well as the mean for four groups: zero, low, medium, and high analyst coverage. The sample period is from 2014Q1 to 2020Q4.

with high analyst coverage. This confirms previous findings that the proportion of retail trading is greater in small stocks than large stocks (Barber and Odean, 2000). Unsurprisingly, high-analyst firms receive greater institutional investor attention and have the largest market capitalization, greater institutional ownership, and lower return volatility.

4.2. Retail trading and analyst coverage

We first examine the relationship between the number of analyst coverage and retail trading. We estimate the following regression specification:

$$\text{N. analyst}_{i,t} = \beta_1 \text{Retail share}_{i,t} + \Gamma'_{i,t}\delta_1 + \alpha_k + \alpha_t + \text{error}_{i,t}, \tag{22}$$

for stock i and quarter t. Γ' represents a vector of control variables and α_k and α_t corresponds to industry and year-quarter fixed-effects. We report the results in Table 2. We find that the coefficient on retail trading varies between 0.049 and 0.096 across all model specifications, and all coefficients are significant at the 1% level. In the last model specification that includes all control variables, a 1% increase in retail share is associated with 0.068 more analysts covering the stock. From Table 1, the standard deviation of the retail trading share is 9.33 percentage points. Our results imply that a one standard deviation increase in retail trading leads to 0.63 more analysts covering a particular stock. The magnitude of the effect is economically large, given that the median stock in our sample receives only two analyst reports each quarter.

The results in Table 2 also highlight other important determinants of analyst coverage: In particular, analysts are more likely to cover large stocks, stocks with high investor attention, or large institutional ownership.

Fig. 6 illustrates the result in Table 2. We compute residual analyst coverage and residual retail share by regressing the number of analysts and retail share on all other control variables in regression (22), including fixed effects. In the left panel, we control for market capitalization and fixed effects, and we showcase that residual analyst coverage monotonically increases in the residual retail share quintiles. The right panel includes all control variables in the regressions for Table 2, notably investor attention and institutional ownership. Analyst coverage is particularly clustered in high-retail stocks (i.e., Q5): there is a coverage gap of 0.6 analyst reports between the average high-retail and the average low-retail stock – almost one-third of the analyst coverage for the median U.S. stock.

We provide two additional robustness checks in Appendix D. First, we acknowledge that analyst coverage might respond slowly to changing market conditions. In Table D.1, we report the results of estimating model (22) by regressing analyst coverage on lagged retail trading share and controls: the effect is qualitatively and quantitatively unchanged. Second, we exclude from our analysis a small subset of "meme" stocks that might have attracted abnormal retail attention in 2020. In Table D.2, we report our main analysis excluding the following stocks: GameStop, AMC, Nikola, Hertz, United States Oil Fund LP, Novavax, Kodak, Bed Bath & Beyond Inc., Koss Corp., and Vinco Ventures.[6]

4.3. Instrumental variable analysis: Stock splits

The analysis in Section 4.2 documents a correlation between retail trading and analyst coverage. It is plausible that retail traders are drawn to stocks with more analyst coverage, rather than analysts covering stocks with high retail interest. To establish evidence

[6] The list is taken from Bloomberg, YearoftheMemeStock:Hertz,KodakTopListof2020Highlights, December 26, 2020. Accessed January 25, 2023. URL: https://www.bloomberg.com/news/articles/2020-12-26/year-of-the-meme-stock-hertz-kodak-top-list-of-2020-highlights.

Table 2
Analyst coverage and retail trading.

	Number of analysts				
	(1)	(2)	(3)	(4)	(5)
Retail share	0.096***	0.049***	0.069***	0.068***	0.068***
	(10.016)	(5.082)	(6.464)	(6.448)	(6.435)
Ln MCAP	1.987***	1.520***	1.471***	1.465***	1.462***
	(15.256)	(13.154)	(12.841)	(13.072)	(13.259)
Turnover	52.017***	41.970***	38.169***	38.253***	39.635***
	(4.184)	(3.534)	(3.329)	(3.299)	(3.079)
Attention		0.104***	0.106***	0.106***	0.106***
		(9.956)	(10.127)	(10.174)	(10.181)
IO			1.158***	1.173***	1.162***
			(4.364)	(4.442)	(4.347)
Book-to-market				0.009	0.008
				(0.918)	(0.921)
R&D expenses				0.158	0.146
				(1.659)	(1.619)
Firm age				0.003	0.003
				(0.858)	(0.857)
Return					−0.381***
					(−3.398)
$\|Return\|$					0.176
					(0.843)
Volatility					−1.898
					(−0.771)
Industry FE	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes
Observations	100,213	100,213	100,213	100,213	100,213
R^2	0.513	0.535	0.537	0.537	0.537
Adjusted R^2	0.513	0.534	0.536	0.536	0.537

This table reports the coefficients of the following regression:

$$\text{N. analyst}_{i,t} = \beta_1 \text{Retail share}_{i,t} + \Gamma'_{i,t}\delta_1 + \alpha_k + \alpha_t + \text{error}_{i,t},$$

for stock i and quarter t. Γ' represents a vector of control variables and α_k and α_t corresponds to industry (four-digit GIC codes) and year-quarter fixed-effects. The dependent variable is the number of analysts (*N. analyst*) for a given stock-quarter. The *Retail share* is computed as the proportion of retail dollar volume for stock i in total retail dollar volume across all stocks for quarter t, in percent. *Ln MCAP* is the natural logarithm of the market capitalization at the beginning of the quarter. *Turnover* is the quarterly stock turnover. *Attention* is the average number of days with high institutional investor attention from Bloomberg Terminal in quarter t. *IO* is the institutional ownership, in percent. *Book-to-market* is the quarterly book-to-market ratio. *R&D expenses* is the quarterly research and development expenses scaled by total sale. *Age* is the firm age as the number of years since first listed on CRSP. *Return* and $|Return|$ is the quarterly stock return and absolute return on quarter t, respectively. *Volatility* is the stock volatility in quarter t. Robust standard errors clustered by industry and year-quarter are presented in parenthesis. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. The sample period is from 2014Q1 to 2020Q4.

of a causal link from retail trading to analyst coverage, we aim to find a source of plausibly exogenous variation in retail volume share.

In this subsection, we instrument the retail trading share using stock split events as a plausibly exogenous shock. Brandt et al. (2010) and Cox et al. (2022) find a positive link between forward stock splits and retail trading. The economic channel is appealing: a stock split reduces the stock price, making it more affordable to retail traders who (i) cannot trade fractional shares and (ii) are usually capital constrained. A reverse stock split, where a company decreases the number of its outstanding shares and consequently increases the stock price, has the opposite effect.

To sharply capture variation in analyst coverage around 931 stock splits, we restrict our sample to a window of two years (i.e., 8 quarters) before and after each forward and reverse split event. We compute a "split factor" variable starting from the FACSHR variable in the CRSP data set, which stands for the *factor to adjust shares outstanding*. In particular, we have:

$$Splitfactor_{i,t} = 1 + \text{FACSHR}_{i,t} = \frac{Sharesoutstanding_{i,t}}{Sharesoutstanding_{i,t-1}}, \tag{23}$$

where i indexes stocks and t runs over quarters. If there is no split in a given quarter, the expression in Eq. (23) equals one. A forward split implies an increase in the number of stocks and therefore $Splitfactor > 1$, whereas a reverse split translates to $Splitfactor < 1$.

To implement the instrumental variables approach, we estimate the following first-stage regression:

$$Retailshare_{i,t} = \beta_1 Splitfactor_{i,t} + \Gamma'_{i,t}\delta_1 + \alpha_k + \alpha_t + e_{i,t}. \tag{24}$$

For the instrument to be valid, it must strongly affect retail trading. We first examine how abnormal retail trading volume changes around split events. We plot the average retail trading volume in Fig. 7 around forward and reverse split events. Consistent

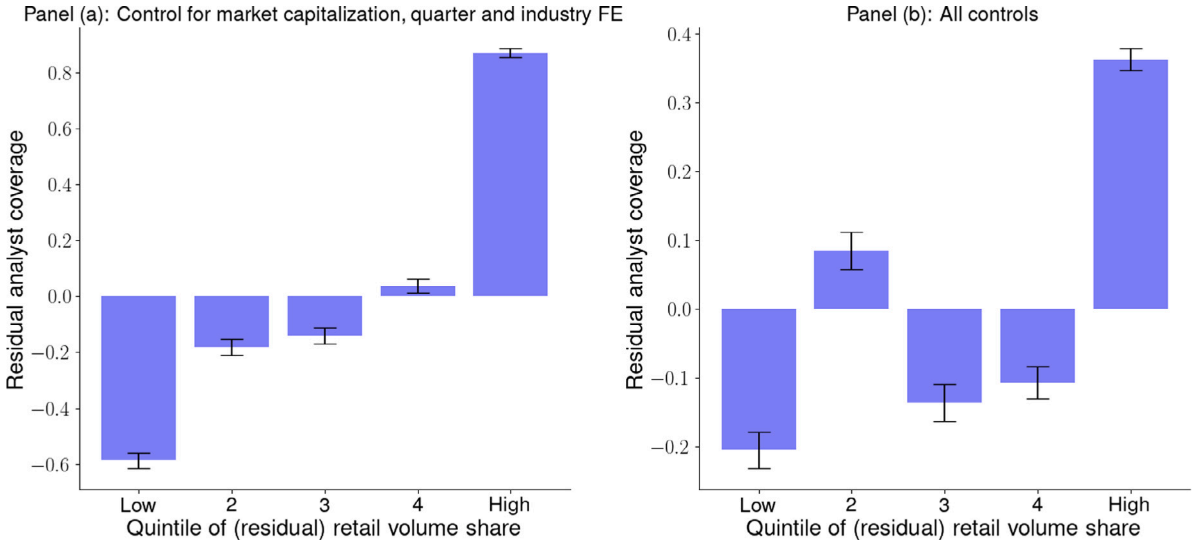

Fig. 6. Retail trading and analyst coverage.
This figure shows the average residual analyst coverage between 2014Q1 and 2020Q4 for U.S. equities by the residual retail volume share. The retail share for stock i is computed as the proportion of retail volume for stock i in total dollar volume for stock i and quarter t. Residual analyst coverage and residual retail share are obtained after controlling for market capitalization and industry fixed effects in Panel (a) and market capitalization, institutional ownership, turnover, book-to-market ratio, R&D expenses, firm age, return, absolute return, volatility, and industry fixed effects in Panel (b).

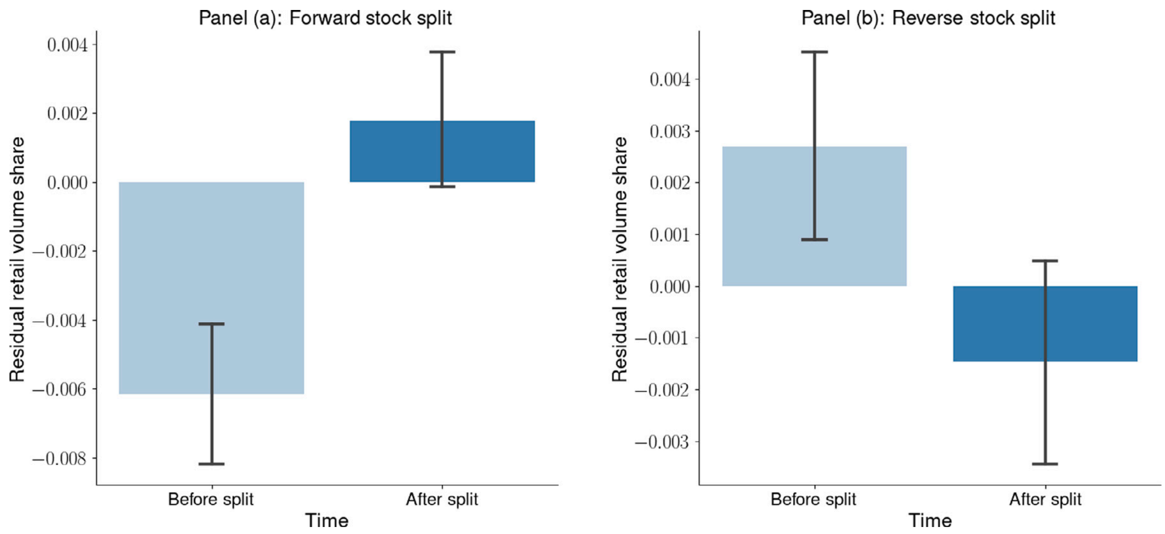

Fig. 7. Abnormal changes in retail trading following splits.
This figure shows the average residual analyst coverage between 2014Q1 and 2020Q4 for U.S. equities by forward and reverse stock splits. Residual retail share is obtained after controlling for market capitalization, institutional ownership, turnover, book-to-market ratio, R&D expenses, firm age, return, absolute return, volatility, and industry fixed effects.

with Brandt et al. (2010), we find an increase in retail trading following forward splits and a decline in retail trading following reverse splits.

The second-stage regression (2SLS) estimates the impact of retail trading on analyst coverage:

$$\text{N. Analyst}_{i,t} = \beta_2 \widehat{\text{Retail share}}_{i,t} + \Gamma'_{i,t}\delta_2 + \alpha_k + \alpha_t + \varepsilon_{i,t}, \tag{25}$$

where $\widehat{\text{Split factor}}$ are the predicted values from the first-stage regression. If the conditions for a valid instrumental variable are met, β_2 captures the causal effect of retail trading on analyst coverage.

Table 3
Instrumental variables analysis using stock splits.

	Retail share (1)	*N. analysts* (2)	*Retail share* (3)	*N. analysts* (4)	*Retail share* (5)	*N. analysts* (6)
Split factor	1.503***		1.082***		1.064***	
	(3.847)		(3.684)		(3.507)	
$\widehat{Retail share}$		0.436***		0.488***		0.485***
		(5.024)		(4.882)		(4.735)
Ln MCAP	−3.877***	2.987***	−3.276***	2.618***	−3.263***	2.612***
	(−23.657)	(8.088)	(−19.874)	(6.290)	(−16.323)	(6.327)
Turnover	−1.602	29.163***	−7.117	24.629***	−0.177	22.682***
	(−0.147)	(5.748)	(−0.646)	(3.723)	(−0.023)	(2.897)
Attention			0.125**	0.028	0.126**	0.028
			(2.657)	(1.112)	(2.659)	(1.169)
IO			−8.289***	4.899***	−8.179***	4.829***
			(−8.592)	(4.246)	(−8.303)	(4.110)
Book-to-market					−0.004*	0.003
					(−1.843)	(1.138)
R&D expenses					0.837	−0.386*
					(1.691)	(−1.985)
Firm age					−0.025	0.002
					(−1.124)	(0.235)
Return					−0.637**	0.093
					(−2.268)	(0.479)
$\|Return\|$					0.205	−0.159
					(0.713)	(−1.230)
Volatility					−10.088***	3.132
					(−2.878)	(1.136)
1st-stage F-statistic	14.803		13.568		12.299	
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,320	12,320	12,320	12,320	12,320	12,320
R^2	0.695	0.234	0.728	0.203	0.730	0.210
Adjusted R^2	0.694	0.231	0.727	0.199	0.729	0.206

This table reports the coefficients of the following regression:

$$\text{Retail share}_{i,t} = \beta_1 \text{Split factor}_{i,t} + \Gamma'_{i,t}\delta_1 + \alpha_k + \alpha_t + e_{i,t}.$$

The second-stage equation estimates the impact of retail trading on analyst coverage:

$$\text{N. Analyst}_{i,t} = \beta_2 \widehat{\text{Retail share}}_{i,t} + \Gamma'_{i,t}\delta_2 + \alpha_k + \alpha_t + \epsilon_{i,t},$$

where the dependent variable is the number of unique analysts providing earnings forecasts for quarter t. Γ' represents a vector of control variables and α_k and α_t corresponds to industry (four-digit GIC codes) and year-quarter fixed-effects. The *Retail share* is computed as the proportion of retail dollar volume for stock i in total retail dollar volume across all stocks for quarter t, in percent. *Split factor* is factor to adjust shares outstanding; see Eq. (23). $\widehat{Retailshare}$ is the predicted retail share from the first-stage regression. *Ln MCAP* is the natural logarithm of the market capitalization at the beginning of the quarter. *Turnover* is the quarterly stock turnover. *Attention* is the average number of days with high institutional investor attention from Bloomberg Terminal in quarter t. *IO* is the institutional ownership, in percent. *Book-to-market* is the quarterly book-to-market ratio. *R&D expenses* is the quarterly research and development expenses scaled by total sale. *Age* is the firm age as the number of years since first listed on CRSP. *Return* and $|Return|$ is the quarterly stock return and absolute return on quarter t, respectively. *Volatility* is the stock volatility in quarter t. Robust standard errors clustered by firm and year-quarter are presented in parenthesis, and ***, **, and * denote the statistical significance at the 1, 5, and 10% level, respectively. The sample period is from 2014Q1 to 2020Q4. Standard errors clustered at industry and year-quarter level.

Columns (1), (3), and (5) in Table 3 report the first-stage regression results for different model specifications. Across all specifications, the split factor is positively associated with retail trading and statistically significant at the 1% level. A two-for-one forward split increases the retail trading share by 1.064 percentage points in column (5), or by more than 10% given the unconditional average retail share of 9.85%.

The F-statistic across all model specifications exceeds the threshold of $F = 10$, which suggests that the instrument is strong and unlikely to be biased towards the OLS estimates (Bound et al., 1995). We report the two-stage least squares estimates in columns (2), (4), and (6). We find a strong and significant impact of retail trading activity on information supply, in that the number of analysts covering a stock increases in the predicted value of $\widehat{Retailshare}$. The magnitude of the effect is several times larger than for the OLS estimation results in Table 2. The discrepancy is not uncommon in empirical finance studies [see the discussion in Jiang (2017)]. One potential reason is that the IV coefficient measures a "local average treatment effect" (LATE) and might not result in uniform assignment between treatment and non-treatment stocks. In our setup, it is plausible, for example that treatment stocks that undergo forward splits have ex ante low retail trading interest, which prompted the managers to consider a split to begin with.

Overall, our empirical analysis supports the main prediction of the model that analyst are more likely to cover stocks with high retail trading activity.

Table 4
Decomposition of analyst coverage variance.

Rank	Variable	Shapley value	Rank	Variable	Shapley value
1	*Ln MCAP*	45.72%	6	*Turnover*	3.15%
2	*Attention*	30.40%	7	*Firm age*	2.15%
3	*Retail share*	7.28%	8	$\lvert Return \rvert$	0.59%
4	*IO*	6.61%	9	*R&D expenses*	0.32%
5	*Volatility*	3.50%	10	*Book-to-market*	0.18%

This table reports the Shapley-Owen (Owen, 1977) relative contributions to analyst coverage, computed as in Eq. (26). *Retail share* is computed as the proportion of retail dollar volume for stock i in total dollar volume for quarter t, in percent. The control variables include *Attention* that corresponds to the average number of days with high institutional investor attention from Bloomberg Terminal over quarter t. *Ln MCAP* is the natural logarithm of the market capitalization at the beginning of the quarter, quarterly research and development expenses scaled by total sale (*R&D*), institutional ownership (in percent) (*IO*), firm age as the number of years since first listed on CRSP (*Age*), the quarterly stock return (*Return*) and absolute return ($\lvert Return \rvert$) on quarter t, and the stock volatility on quarter t (*Volatility*).

4.4. A variance decomposition

Our previous results indicate that retail trading activity is a key determinant of analyst coverage. To establish the marginal contribution of retail trading to explain the variability in analyst coverage, we decompose the R^2 of regression model (5) in Table 2 using the Shapley-Owen methodology (Owen, 1977). To implement the Shapley-Owen methodology on a model with k independent variables $\Omega = \{x_1, x_2, \ldots, x_k\}$, we estimate 2^k regressions: for each possible subset of independent variables. Next, we compute the marginal contribution of independent variable j (i.e., its Shapley value) to the regression R^2 as:

$$R_j^2 = \sum_{S \subseteq \Omega - \{x_j\}} \frac{R^2\left(S \cup \{x_j\}\right) - R^2(S)}{k\binom{k-1}{|S|}}, \tag{26}$$

where $R^2(S)$ is the R^2 of a regression on the independent variables in S and $|S|$ is the number of elements in S.

We document in Table 4 that retail trading activity has a Shapley value of 7.28%. That is, retail trading is the third most important determinant of analyst coverage after market capitalization (Shapley value of 45.72%) and investor attention (30.40%), but ahead of institutional ownership, investor turnover, volatility, R&D activity, or firm age.

5. Conclusion

In this paper, we highlight the link between retail trading and the supply of information in financial markets in the cross-section of U.S. stocks. We find that a one standard deviation increase in retail trading leads to 0.6 more sell-side analyst reports for the stock after controlling for market capitalization, industry, turnover, or institutional investor attention. The effect magnitude is economically large, as the median U.S. stock receives only two analyst reports each quarter.

We build a noisy rational expectation equilibrium model to flesh out the economic channel linking retail trading to analyst coverage. Analysts provide informative signals to institutional investors, with the goal to maximize institutional trading volume. The signals are most valuable in stocks with high retail trading interest, in which institutional investors can trade more aggressively without revealing information. As a result, since information production is costly, analysts provide more precise forecasts in stocks with a large mass of retail investors. One implication of the model is that the un-bundling of research and trading services, as mandated in Europe by MiFID II regulation, should weaken the link between retail trading and analyst coverage and reduce clustering.

A key implication of the model is that improving market access to retail traders should boost information production by sell-side analysts, leading to higher market liquidity and more informative prices. Second, since retail traders are more active in small stocks, they help even out the information supply in the cross-section of stocks. Historically, analyst coverage is skewed towards large firms. The impact of a skewed information supply is not always benign: greater analyst coverage is often associated with a lower financing cost (Easley and O'Hara, 2004), which can lead to a re-allocation of capital in the economy. For example, Begenau et al. (2018) find that the increase in big data disproportionately benefits big firms help these firms to reduce the cost of capital and enabling them to grow more through investment. We argue that increasing access to financial markets to individual investors could lead to a more balanced analyst coverage in the cross-section of stocks, particularly for previously neglected securities.

Data availability

Data will be made available on request

Appendix A. Notation summary

Exogenous parameters and their interpretation.	
Parameter	Definition
I and R	Pertaining to institutional and retail traders, respectively.
v, τ_v	Asset value and the unconditional precision of asset payoff.
u, τ_u	Noise trader demand and the inverse of noise trader variance.
μ	Mass of retail traders, with $\mu \leq 1$.
c	Cost parameter for analyst signal production.
γ	Investors' CARA risk-aversion.
f	Volume-based fee for brokerage houses.
Endogenous quantities and their interpretation.	
Variable	Definition
y, τ_y	Analyst signal and its precision.
$\phi < 1$	Share of analyst signal that is endogenously revealed in prices.
p	Market-clearing price at $t = 2$.
$\mathcal{D}_i$	Demand for investor type j at $t = 1$.

Appendix B. Numerical solution for assets in positive net supply

We relax the assumption that the asset is in zero net supply, and assume instead a positive supply Q to (numerically) study the model implications when there is a positive risk premium. In this case, we conjecture a linear price function:

$$p = p_y \tilde{y} + p_u \tilde{u} + p_Q Q. \tag{B.1}$$

The institutional and retail trader demand functions remain the same, since the conditional expectation and variance of the asset do not change. However, the market clearing condition in (10) becomes:

$$\int_0^1 \mathcal{D}_{\mathbf{I}} \mathrm{d}i + \int_0^\mu \mathcal{D}_{\mathbf{R}} \mathrm{d}i + \tilde{u} = Q. \tag{B.2}$$

We solve for the market clearing price and obtain:

$$p = \frac{\tau_y (1+\mu\phi)}{\tau_v (1+\mu) + \tau_y (1+\mu\phi)} \tilde{y} + \frac{\gamma (1+\mu\phi)}{\tau_v (1+\mu) + \tau_y (1+\mu\phi)} \tilde{u} - \frac{\gamma (1+\mu\phi)}{\tau_v (1+\mu) + \tau_y (1+\mu\phi)} Q, \tag{B.3}$$

where we note that $p_Q = -p_u$.

$$\begin{aligned} \mathcal{D}_{\mathbf{I}}(p) &= \frac{1}{\gamma} \left[\left(\tau_y - \left(\tau_y + \tau_v\right) p_y\right) y - \left(\tau_y + \tau_v\right) p_u u + \left(\tau_y + \tau_v\right) p_u Q \right] \\ &\sim \mathcal{N}\left(\frac{1}{\gamma} \left(\tau_y + \tau_v\right) p_u Q, \frac{1}{\gamma^2} \left[\frac{\left(\tau_y - \left(\tau_y + \tau_v\right) p_y\right)^2 \left(\tau_y + \tau_v\right)}{\tau_y \tau_v} + \frac{\left(\left(\tau_y + \tau_v\right) p_u\right)^2}{\tau_u} \right] \right), \end{aligned} \tag{B.4}$$

where we use that $\mathrm{var}(y) = \frac{\tau_y + \tau_v}{\tau_y \tau_v}$ and $\mathrm{var}(u) = \tau_u^{-1}$. From the properties of the folded normal distribution, we obtain that the expected volume (conditional on informed traders starting out with no position in the asset) is:

$$\mathbb{E}\mathrm{Volume}_{\mathbf{I}} = \mathbb{E}\left|\mathcal{D}_{\mathbf{I}}\right| = \sigma \sqrt{\frac{2}{\pi}} e^{-\frac{\mu^2}{2\sigma^2}} + \mu \operatorname{erf}\left(\frac{\mu}{\sqrt{2\sigma^2}} \right), \tag{B.5}$$

where $\mu = \frac{1}{\gamma}\left(\tau_y + \tau_v\right) p_u Q$, $\sigma^2 = \frac{1}{\gamma^2} \left[\frac{\left(\tau_y - \left(\tau_y + \tau_v\right) p_y\right)^2 \left(\tau_y + \tau_v\right)}{\tau_y \tau_v} + \frac{\left(\left(\tau_y + \tau_v\right) p_u\right)^2}{\tau_u} \right]$, and

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\left(-t^2\right) \mathrm{d}t \tag{B.6}$$

is the Gauss error function. Fig. B.1 illustrates numerically that the main result of the paper is robust to allowing a positive net supply for the asset. That is, the equilibrium analyst signal precision increases in the mass of retail traders – we plot the relationship for $Q \in \{0, 2, 4, 6\}$. Another salient pattern in Fig. B.1 is that the volume tends to increase in Q: a higher asset supply leads to a larger risk premium, depressing the clearing price and increasing the demand from informed traders.

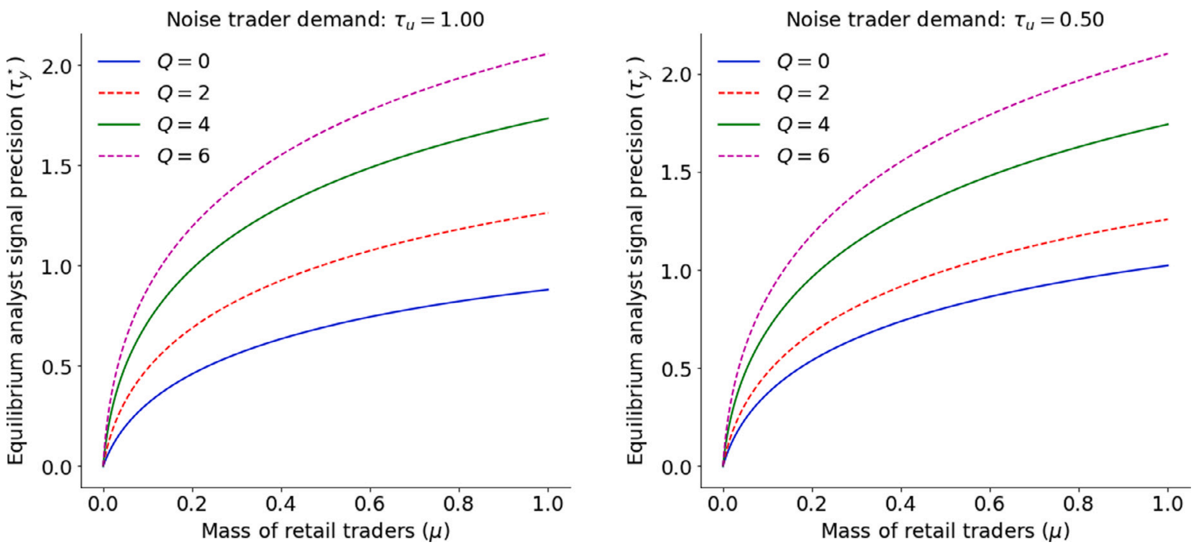

Fig. B.1. Equilibrium analyst signal precision with positive asset supply.
This figure illustrates the equilibrium analyst signal precision $\tau_y^\star$ as a function of the mass of retail traders μ. Parameter values: $\tau_v = 1$, $\gamma = 5$, $\tau_u = 0.5$, $c = 0.2$.

Appendix C. Proofs

Proposition 1

Proof. We replace p_y and p_u in the analysts' problem (19) to obtain a closed form expression for volume V (scaled by the $\sqrt{\frac{2}{\pi}}$ constant):

$$V = \sqrt{\frac{(\tau_v + \tau_y)\left(\gamma^2(\mu+1)^2\tau_u^2\tau_y^2(\tau_v+\tau_y) + \gamma^4\tau_u\tau_y((\mu(\mu+2)+2)\tau_v + 2(\mu+1)\tau_y) + \gamma^6(\tau_v+\tau_y)\right)}{\tau_u\left(\gamma^2(\mu\tau_v+\tau_v+\tau_y) + (\mu+1)\tau_u\tau_y(\tau_v+\tau_y)\right)^2}}. \tag{C.1}$$

The analysts' problem is to maximize:

$$U\left(\tau_y\right) = V\left(\tau_y\right) - \frac{c\tau_y^2}{2}. \tag{C.2}$$

We need to show that:

$$\frac{\partial \tau_y^\star}{\partial \mu} \geq 0, \tag{C.3}$$

and therefore the precision of analyst signals increases in the mass of uninformed traders.

Let $f\left(\tau_y, \mu\right) \equiv \frac{\partial U}{\partial \tau_y}$. Then $\tau_y^\star$ solves:

$$f\left(\tau_y^\star, \mu\right) = 0. \tag{C.4}$$

From the implicit function theorem, we have that:

$$\frac{\partial \tau_y^\star}{\partial \mu} = -\left(\frac{\partial f}{\partial \tau_y}\right)^{-1}\frac{\partial f}{\partial \mu}. \tag{C.5}$$

We start with $\frac{\partial f}{\partial \mu}$, which can be written as:

$$\frac{\partial f}{\partial \mu} = \frac{g_1}{g_2}, \tag{C.6}$$

where:

$$g_2 = 2\tau_u^2\left(\gamma^2(\mu\tau_v+\tau_v+\tau_y) + (\mu+1)\tau_u\tau_y(\tau_v+\tau_y)\right)^6 \times \\ \times\left(\frac{\gamma^2(\tau_v+\tau_y)\left(\gamma^2\tau_u\tau_y((\mu(\mu+2)+2)\tau_v + 2(\mu+1)\tau_y) + \gamma^4(\tau_v+\tau_y) + (\mu+1)^2\tau_u^2\tau_y^2(\tau_v+\tau_y)\right)}{\tau_u\left(\gamma^2(\mu\tau_v+\tau_v+\tau_y) + (\mu+1)\tau_u\tau_y(\tau_v+\tau_y)\right)^2}\right)^{3/2} \tag{C.7}$$

and:

$$g_1 = \gamma^6\tau_v(\tau_v+\tau_y)^2(\gamma^2+\tau_u\tau_y)(\gamma^4\tau_u^2\tau_y((\mu(\mu(7\mu+36)+58)+30)\tau_v^2\tau_y+ \\ +2(\mu+1)(2\mu(\mu+2)+3)\tau_v^3+2(\mu(3\mu(\mu+7)+40)+21)\tau_v\tau_y^2+18(\mu+1)^2\tau_y^3)+ \\ +\gamma^2(\mu+1)\tau_u^3\tau_y^2(\tau_v+\tau_y)((5\mu(\mu+2)+6)\tau_v^2+2(3\mu(2\mu+5)+10)\tau_v\tau_y+2(\mu+1)(4\mu+7)\tau_y^2)+ \\ +\gamma^6\tau_u(((7-2\mu)\mu(\mu+2)+14)\tau_v^2\tau_y+(\mu+1)(\mu(\mu+2)+2)\tau_v^3+22(\mu+1)\tau_v\tau_y^2+2(6\mu+5)\tau_y^3) \\ +2\gamma^8(\tau_v+\tau_y)((1-\mu)\tau_v+\tau_y)+2(\mu+1)^3\tau_u^4\tau_y^3(\tau_v+\tau_y)^2(\tau_v+2\tau_y)), \tag{C.8}$$

and both g_2 and g_1 are positive for $\mu \leq 1$. Therefore, it follows that $\frac{\partial f}{\partial \mu} > 0$.

It remains to show that $\frac{\partial f}{\partial \tau_y}\left(\tau_y^\star\right) < 0$. We can write:

$$\frac{\partial f}{\partial \tau_y}\left(\tau_y^\star\right) = -\frac{g_3}{g_4}, \tag{C.9}$$

where:

$$g_3 = 2(\gamma^6\tau_u^2(3(5\mu(7\mu(\mu+4)+40)+96)\tau_v^3\tau_y^2+4(\mu(\mu(29\mu+160)+258)+128)\tau_v^2\tau_y^3+ \\ +6(\mu+1)(\mu(\mu+4)^2+12)\tau_v^4\tau_y+(\mu+1)^2(\mu+2)(\mu(\mu+2)+4)\tau_v^5+ \\ +24(\mu+1)(\mu(2\mu+17)+17)\tau_v\tau_y^4+120(\mu+1)^2\tau y^5)+ \\ +2\gamma^4(\mu+1)\tau_u^3\tau_y(\tau_v+\tau_y)(2(\mu(\mu(3\mu(\mu+11)+125)+184)+92)\tau_v^2\tau_y^2+ \\ +3(\mu+2)^3(2\mu+3)\tau_v^3\tau_y+3(\mu+1)(\mu+2)(\mu(\mu+2)+2)\tau_v^4+4(\mu+1)(\mu(9\mu+47)+50)\tau_v\tau_y^3 \\ +2(\mu+1)(35\mu+38)\tau_y^4)+\gamma^2(\mu+1)^2\tau_u^4\tau_y^2(\tau_v+\tau_y)^2(8(3\mu(\mu(\mu+5)+8)+13)\tau_v^2\tau_y+ \\ +3(\mu+2)(3\mu(\mu+2)+4)\tau_v^3+8(3\mu(\mu(\mu+7)+13)+22)\tau_v\tau_y^2+24(\mu+1)(3\mu+4)\tau_y^3) \\ +4\gamma^8\tau_u(\tau_v+\tau_y)^2((2-\mu)(\mu+1)\tau_v^2+(\mu(2\mu+13)+14)\tau_v\tau_y+12(\mu+1)\tau_y^2)+ \\ +8\gamma^{10}(\tau_v+\tau_y)^3+4(\mu+1)^4(\mu+2)\tau_u^5\tau_y^3(\tau_v+\tau_y)^3(\tau_v^2+3\tau_v\tau_y+3\tau_y^2)) > 0 \tag{C.10}$$

for any $\mu < 2$, and

$$g_4 = (\tau_v+\tau_y)(\gamma^2(\mu\tau_v+\tau_v+\tau_y)+(\mu+1)\tau_u\tau_y(\tau_v+\tau_y))(\gamma^2\tau_u\tau_y((\mu(\mu+2)+2)\tau_v+2(\mu+1)\tau_y)+ \\ +\gamma^4(\tau_v+\tau_y)+(\mu+1)^2\tau_u^2\tau_y^2(\tau_v+\tau_y))(\gamma^2\tau_u((\mu+1)(\mu+2)\tau_v^2+(\mu(2\mu+9)+8)\tau_v\tau_y+6(\mu+1)\tau_y^2)+ \\ +2\gamma^4(\tau_v+\tau_y)+(\mu+1)(\mu+2)\tau_u^2\tau_y(\tau_v+\tau_y)(\tau_v+2\tau_y)) > 0, \tag{C.11}$$

for any μ. It follows that $\frac{\partial f}{\partial \tau_y}\left(\tau_y^\star\right) < 0$ and finally that $\frac{\partial \tau_y^\star}{\partial \mu} \geq 0$, which concludes the proof. □

Corollary 1

Proof. We define price efficiency as in Goldstein and Yang (2017), that is:

$$\tau_{v|p} \equiv \text{var}\left[v \mid p\right]^{-1} = \tau_v + \frac{\tau_u}{\tau_u + \frac{\gamma^2}{\tau_y}}\tau_y. \tag{C.12}$$

It is straightforward to show that the asset payoff precision, conditional on price, increases in the analyst signal precision since:

$$\frac{\partial \tau_{v|p}}{\partial \tau_y} = 1 - \frac{\gamma^4}{\left(\gamma^2+\tau_u\tau_y\right)^2} > 0 \tag{C.13}$$

since $\left(\gamma^2+\tau_u\tau_y\right)^2 < \gamma^4$. From Proposition 1, τ_y increases in μ and therefore by the chain rule:

$$\frac{\partial \tau_{v|p}}{\partial \mu} = \frac{\partial \tau_{v|p}}{\partial \tau_y}\frac{\partial \tau_y}{\partial \mu} > 0, \tag{C.14}$$

which concludes the proof. □

Corollary 2

Proof. Plug the equilibrium signal precision, i.e., $\tau_y^\star$, in the formula for p_u, we have:

$$p_u = \gamma\frac{1+\mu\phi}{\tau_v\left(1+\mu\right)+\tau_y^\star\left(1+\mu\phi\right)}.$$

Therefore, liquidity is defined as:

$$p_u^{-1} = \frac{\tau_v\left(1+\mu\right)+\tau_y^\star\left(1+\mu\phi\right)}{\gamma\left(1+\mu\phi\right)}.$$

Hence, we have:

$$\begin{aligned}\frac{\partial p_u^{-1}}{\partial \mu} &= \frac{\left(\tau_v + \frac{\partial \tau_y^\star}{\partial \mu}(1+\mu\phi) + \tau_y^\star \phi\right)\gamma(1+\mu\phi) - \gamma\phi\left(\tau_v + \tau_v\mu + \tau_y^\star + \tau_y^\star\mu\phi\right)}{\gamma^2(1+\mu\phi)^2} \\ &= \frac{\left(\tau_v + \frac{\partial \tau_y^\star}{\partial \mu}(1+\mu\phi) + \tau_y^\star \phi\right)(1+\mu\phi) - \phi\left(\tau_v + \tau_v\mu + \tau_y^\star + \tau_y^\star\mu\phi\right)}{\gamma(1+\mu\phi)^2} \\ &= \frac{\frac{\partial \tau_y^\star}{\partial \mu}(1+\mu\phi)^2 + \tau_v + \tau_v\mu\phi + \tau_y^\star\phi + \tau_y^\star\mu\phi^2 - \tau_v\phi - \tau_v\mu\phi - \tau_y^\star\phi - \tau_y^\star\mu\phi^2}{\gamma(1+\mu\phi)^2} \\ &= \frac{\frac{\partial \tau_y^\star}{\partial \mu}(1+\mu\phi)^2 + (1-\phi)\tau_v}{\gamma(1+\mu\phi)^2} > 0,\end{aligned}$$

as $\frac{\partial \tau_y^\star}{\partial \mu} \geq 0$ by Proposition 1. □

Appendix D. Additional empirical results

See Tables D.1 and D.2.

Table D.1
Analyst coverage and retail trading (lagged independent variables).

	Number of analysts				
	(1)	(2)	(3)	(4)	(5)
Retail share	0.095***	0.047***	0.070***	0.070***	0.070***
	(9.678)	(4.790)	(7.150)	(7.237)	(7.322)
Ln MCAP	2.009***	1.519***	1.457***	1.461***	1.453***
	(15.382)	(13.184)	(12.848)	(13.142)	(13.575)
Turnover	57.708***	46.673***	42.205***	41.893***	44.510***
	(4.454)	(3.813)	(3.566)	(3.521)	(3.263)
Attention		0.108***	0.110***	0.110***	0.110***
		(10.188)	(10.499)	(10.514)	(10.497)
IO			1.402***	1.402***	1.377***
			(5.982)	(6.083)	(5.977)
Book-to-market				0.004	0.004
				(0.631)	(0.600)
R&D expenses				0.162	0.162
				(1.412)	(1.397)
Firm age				−0.001	−0.001
				(−0.330)	(−0.371)
Return					−0.261**
					(−2.336)
|Return|					0.003
					(0.014)
Volatility					−2.708
					(−1.029)
Industry FE	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes
Observations	98,744	98,744	98,744	98,744	98,744
Adjusted R^2	0.523	0.545	0.548	0.548	0.548

This table reports the coefficients of the following regression:

$$\text{N. analyst}_{i,t} = \beta_1 \text{Retail share}_{i,t-1} + \Gamma'_{i,t-1}\delta_1 + \alpha_k + \alpha_t + \text{error}_{i,t},$$

for stock i and quarter t. Γ' represent a vector of control variables and α_k and α_t corresponds to industry (four-digit GIC codes) and year-quarter fixed-effects. The dependent variable is the number of analysts (*N. analyst*) for a given stock-quarter. The *Retail share* is computed as the proportion of retail dollar volume for stock i in total retail dollar volume across all stocks for quarter t, in percent. *Ln MCAP* is the natural logarithm of the market capitalization at the beginning of the quarter. *Turnover* is the quarterly stock turnover. *Attention* is the average number of days with high institutional investor attention from Bloomberg Terminal in quarter t. *IO* is the institutional ownership, in percent. *Book-to-market* is the quarterly book-to-market ratio. *R&D expenses* is the quarterly research and development expenses scaled by total sale. *Age* is the firm age as the number of years since first listed on CRSP. *Return* and *|Return|* is the quarterly stock return and absolute return on quarter t, respectively. *Volatility* is the stock volatility in quarter t. Robust standard errors clustered by industry and year-quarter are presented in parenthesis. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. The sample period is from 2014Q1 to 2020Q4.

Table D.2
Analyst coverage and retail trading (exclude meme stocks).

	Number of analysts				
	(1)	(2)	(3)	(4)	(5)
Retail share	0.096***	0.050***	0.069***	0.068***	0.068***
	(10.018)	(5.038)	(6.418)	(6.407)	(6.641)
Ln MCAP	1.988***	1.521***	1.473***	1.467***	1.470***
	(15.262)	(13.117)	(12.806)	(13.043)	(12.599)
Turnover	52.167***	42.180***	38.459***	38.525***	38.780***
	(4.183)	(3.543)	(3.343)	(3.312)	(3.174)
Attention		0.104***	0.106***	0.106***	0.106***
		(9.833)	(10.006)	(10.048)	(10.053)
IO			1.139***	1.153***	1.157***
			(4.305)	(4.381)	(4.294)
Book-to-market				0.009	0.008
				(0.921)	(0.926)
R&D expenses				0.157	0.142*
				(1.672)	(1.717)
Firm age				0.003	0.003
				(0.801)	(0.798)
Return					−0.388***
					(−3.198)
$\lvert Return \rvert$					0.135
					(0.637)
Volatility					−0.601
					(−0.107)
Industry FE	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes
Observations	100,071	100,071	100,071	100,071	100,071
R^2	0.514	0.535	0.537	0.537	0.537
Adjusted R^2	0.514	0.535	0.537	0.537	0.537

This table reports the coefficients of the following regression:

$$\text{N. analyst}_{i,t} = \beta_1 \text{Retail share}_{i,t} + \Gamma'_{i,t}\delta_1 + \alpha_k + \alpha_t + \text{error}_{i,t},$$

for stock i and quarter t. Γ' represents a vector of control variables and α_k and α_t corresponds to industry (four-digit GIC codes) and year-quarter fixed-effects. The dependent variable is the number of analysts (*N. analyst*) for a given stock-quarter. The *Retail share* is computed as the proportion of retail dollar volume for stock i in total retail dollar volume across all stocks for quarter t, in percent. *Ln MCAP* is the natural logarithm of the market capitalization at the beginning of the quarter. *Turnover* is the quarterly stock turnover. *Attention* is the average number of days with high institutional investor attention from Bloomberg Terminal in quarter t. *IO* is the institutional ownership, in percent. *Book-to-market* is the quarterly book-to-market ratio. *R&D expenses* is the quarterly research and development expenses scaled by total sale. *Age* is the firm age as the number of years since first listed on CRSP. *Return* and $\lvert Return \rvert$ is the quarterly stock return and absolute return on quarter t, respectively. *Volatility* is the stock volatility in quarter t. Robust standard errors clustered by industry and year-quarter are presented in parenthesis. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. The sample period is from 2014Q1 to 2020Q4. We exclude the following "meme" stocks: GameStop, AMC, Nikola, Hertz, United States Oil Fund LP, Novavax, Kodak, Bed Bath & Beyond Inc., Koss Corp., and Vinco Ventures. Standard errors clustered at industry and year-quarter level.

References

Admati, A.R., Pfleiderer, P., 1988. A theory of intraday patterns: Volume and price variability. Rev. Financ. Stud. 1 (1), 3–40.
Barber, B.M., Huang, X., Odean, T., Schwarz, C., 2022. Attention-induced trading and returns: Evidence from robinhood users. J. Finance 77 (6), 3141–3190.
Barber, B.M., Lee, Y.-T., Liu, Y.-J., Odean, T., 2008. Just how much do individual investors lose by trading? Rev. Financ. Stud. 22 (2), 609–632.
Barber, B.M., Odean, T., 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. J. Finance 55 (2), 773–806.
Barber, B.M., Odean, T., 2007. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. Rev. Financ. Stud. 21 (2), 785–818.
Barber, B.M., Odean, T., 2013. Chapter 22 - the behavior of individual investors. In: Constantinides, G.M., Harris, M., Stulz, R.M. (Eds.), In: Handbook of the Economics of Finance, vol. 2, Elsevier, pp. 1533–1570.
Barrot, J.-N., Kaniel, R., Sraer, D., 2016. Are retail traders compensated for providing liquidity? J. Financ. Econ. 120 (1), 146–168.
Barth, M.E., Kasznik, R., McNichols, M.F., 2001. Analyst coverage and intangible assets. J. Account. Res. 39 (1), 1–34.
Begenau, J., Farboodi, M., Veldkamp, L., 2018. Big data in finance and the growth of large firms. J. Monetary Econ. 97, 71–87.
Ben-Rephael, A., Da, Z., Israelsen, R.D., 2017. It depends on where you search: Institutional investor attention and underreaction to news. Rev. Financ. Stud. 30 (9), 3009–3047.
Bhushan, R., 1989. Firm characteristics and analyst following. J. Account. Econ. 11 (2–3), 255–274.
Boehmer, E., Jones, C.M., Zhang, X., Zhang, X., 2021. Tracking retail investor activity. J. Finance 76, 2249–2305.
Bound, J., Jaeger, D.A., Baker, R.M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. J. Amer. Statist. Assoc. 90 (430), 443–450.

Brandt, M.W., Brav, A., Graham, J.R., Kumar, A., 2010. The idiosyncratic volatility puzzle: Time trend or speculative episodes? Rev. Financ. Stud. 23 (2), 863–899.
Brennan, M.J., Jegadeesh, N., Swaminathan, B., 1993. Investment analysis and the adjustment of stock prices to common information. Rev. Financ. Stud. 6 (4), 799–824.
Chinco, A., Fos, V., 2021. The sound of many funds rebalancing. Rev. Asset Pricing Stud. 11 (3), 502–551.
Cox, J., Van Ness, B., Van Ness, R., 2022. Stock splits and retail trading. Financ. Rev. 57 (4), 731–750.
Easley, D., O'Hara, M., 2004. Information and the cost of capital. J. Finance 59 (4), 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. J. Financ. Econ. 146 (2), 502–528.
Frankel, R., Kothari, S.P., Weber, J., 2006. Determinants of the informativeness of analyst research. J. Account. Econ.
Gleason, C.A., Lee, C.M.C., 2003. Analyst forecast revisions and market price discovery. Account. Rev. 78 (1), 193–225.
Goldstein, I., Yang, L., 2017. Information disclosure in financial markets. Annu. Rev. Finan. Econ. 9, 101–125.
Grossman, S., 1976. On the efficiency of competitive stock markets where trades have diverse information. J. Finance 31 (2), 573–585.
Grossman, S.J., Stiglitz, J.E., 1980. On the impossibility of informationally efficient markets. Am. Econ. Rev. 70 (3), 393–408.
Groysberg, B., Healy, P.M., Maber, D.A., 2011. What drives sell-side analyst compensation at high-status investment banks? J. Account. Res. 49 (4), 969–999.
Hameed, A., Morck, R., Shen, J., Yeung, B., 2015. Information, analysts, and stock return comovement. Rev. Financ. Stud. 28 (11), 3153–3187.
Holmström, B., 1999. Managerial incentive problems: A dynamic perspective. Rev. Econom. Stud. 66 (1), 169–182.
Hong, H., Kubik, J.D., 2003. Analyzing the analysts: Career concerns and biased earnings forecasts. J. Finance 58 (1), 313–351.
Jiang, W., 2017. Have instrumental variables brought us closer to the truth. Rev. Corp. Finance Stud. 6 (2), 127–140.
Kaniel, R., Saar, G., Titman, S., 2008. Individual investor trading and stock returns. J. Finance 63 (1), 273–310.
Livnat, J., Mendenhall, R.R., 2006. Comparing the post-earnings announcement drift for surprises calculated from analyst and time series forecasts. J. Account. Res. 44 (1), 177–205.
Lush, M., Fontes, A., Zhu, M.Z., Valdes, O.V., Mottola, G.M., 2021. Investing 2020: New Accounts and the People Who Opened Them. Tech. rep., FINRA.
McCrank, J., 2021. Factbox: The U.S. retail trading frenzy in numbers. https://www.reuters.com/article/us-retail-trading-numbers-idUSKBN29Y2PW. (Accessed 4 May 2021).
Mondria, J., Vives, X., Yang, L., 2022. Costly interpretation of asset prices. Manage. Sci. 68 (1), 52–74.
O'Hara, M., Yao, C., Ye, M., 2014. What's not there: Odd lots and market data. J. Finance 69 (5), 2199–2236.
Osipovich, A., 2020. Individual-investor boom reshapes U.S. stock market. https://www.wsj.com/articles/individual-investor-boom-reshapes-u-s-stock-market-11598866200. (Accessed 4 May 2021).
Owen, G., 1977. Values of games with a priori unions. In: Henn, R., Moeschlin, O. (Eds.), Mathematical Economics and Game Theory. Springer, Berlin, Heidelberg, pp. 76–88.
Ozik, G., Sadka, R., Shen, S., 2021. Flattening the illiquidity curve: Retail trading during the COVID-19 lockdown. J. Financ. Quant. Anal. 56 (7), 2356–2388.
Verrecchia, R.E., 1982. Information acquisition in a noisy rational expectations economy. Econometrica 50 (6), 1415–1430.
Wang, J., 1994. A model of competitive stock trading volume. J. Polit. Econ. 102 (1), 127–168.
Welch, I., 2021. The wisdom of the robinhood crowd. J. Finance 77, 1489–1527.