Retail Trading in Options and the Rise of the Big Three Wholesalers

Svetlana Bryzgalova,* Anna Pavlova,† and Taisiya Sikorskaya‡

March 26, 2022

Abstract

This paper documents rapid increases in (i) retail investor trading in options and in (ii) payment for order flow (PFOF) for options transactions received by the U.S. retail brokerages. PFOF comes from so-called wholesalers/internalizers – market makers who execute order flow for a retail brokerage. Exploiting new reporting requirements and transaction-level data, we isolate wholesaler trades and propose a novel measure of retail investor trading in options. We find that retail traders prefer cheaper, weekly options, the average quoted bid-ask spread for which is a whopping 12.3%, and lose money on aggregate. The inflow of retail investors also coincides with an increase in call options left suboptimally unexercised. Market makers (and other arbitrageurs) exploit these mistakes via the so-called 'dividend play' trades, producing (virtually) riskless arbitrage profits. Puzzlingly, they forgo 50% of these profits, leaving money on the table for option writers. Our findings suggest that the arbitrageurs behave non-competitively and that the Big Three wholesalers, whose share in PFOF for options surpassed 85%, seem to benefit disproportionately from the growth in retail trading.

JEL Classification: G4, G5, G11, G12, D45

Keywords: Retail, payment for order flow, dividend play, suboptimal option exercise, market concentration, arbitrage, FinTech, WallStreetBets

---

*We thank Brad Barber, Dmitriy Muravyev, Dimitri Vayanos, Ingrid Werner, and seminar participants at Indiana University, London Business School, Stockholm Business School, and the University of Central Florida for their thoughtful comments. We have also benefitted from conversations with Ed Monrad and Shyam Chandramouli of Optiver, and Parker Lim of XTX Markets.

*London Business School and CEPR, sbryzgalova@london.edu

†London Business School and CEPR, apavlova@london.edu

‡London Business School, tsikorskaya@london.edu

1 Introduction

The recent practice of charging zero commissions for trading stocks and other financial securities, pioneered by the mobile app Robinhood, has revolutionized retail brokerage services in the U.S. Since their market entry in 2015, Robinhood and other commission-free brokerages have attracted an unprecedented inflow of retail customers, mainly young and tech-savvy, yet inexperienced investors, who actively trade on these platforms. In 2021, Robinhood alone amassed 22.5 million active users. Such meteoric rise in inexperienced investor trading activity has been subject to an ongoing debate. One feature at the heart of this debate is the controversial practice of payment for order flow (or PFOF), whereby the so-called wholesalers/internalizers – market makers to whom a brokerage routes its orders for execution—pay the brokerage for the orders they receive from it. The brokerage therefore has an incentive to encourage investors to trade more and to trade assets with larger spreads, which could be to their detriment. While the debate has focused primarily on stocks, especially 'meme' stocks, a little appreciated fact is that retail brokerages receive a much higher PFOF for options market transactions. In the second quarter of 2021, brokerages in the U.S. received more than 284.4 million dollars of PFOF for their order flow in stocks and 581.2 million dollars for options.[1]

Given the magnitude of PFOF in options, it is surprising that emerging literature on the growth in retail trading has focused exclusively on equities. Our paper fills this gap. Exploiting a new reporting requirement in the U.S. options exchanges and transaction-level data, we propose a novel measure of retail trades that were executed by wholesalers. We document a significant recent increase in retail investor trading in options, as measured by both wholesaler trades and small trades on options exchanges. Retail traders prefer cheaper, weekly options, the average quoted bid-ask spread for which is as high as 12.3%. A large fraction of retail order flow is serviced by very few wholesalers: The share in PFOF of the top three has grown to over 85% as of the second quarter of 2021.

We explore how the inflow of retail investors has affected the behavior of arbitrageurs. As a laboratory for the study of arbitrage activity, we use the so-called 'dividend play' trades, which are pairs trades that produce (virtually) riskless arbitrage profits for market makers. These profits derive from call options left suboptimally unexercised on cum-dividend dates. We show that an inflow of inexperienced retail investors has boosted potential gains from this strategy. Exploiting the new reporting requirements, we identify all dividend play trades and document that, instead of harvesting all the arbitrage profits from suboptimally unexercised

[1] Our PFOF figures reflect the combined PFOF from the largest U.S. retail brokerages reports under the SEC Rule 606 (routing of orders). See Section 2.1 for the list of brokerages in our sample.

call options, market makers and other sophisticated arbitrageurs leave a half of potential gains to writers of these options (who are potentially other market makers). This puzzling behavior suggests that arbitrageurs behave non-competitively in exploiting the dividend play arbitrage, and the Big Three wholesalers, who often serve as option writers, appear to benefit most.

Unlike stocks that trade on a variety of lit and dark venues, all options trades in the U.S. must go through exchanges. This is perhaps the primary reason why the debate on the consequences of trade internalization by PFOF providers, on their private trading platforms, has been focused solely on equities. We argue that the patterns are similar, if not more acute, in options. Wholesalers are able to cross retail order flow in options without attracting competing quotes—i.e., effectively internalize it—through exchange mechanisms known as 'price improvement auctions'.[2] These market makers typically justify the use of price improvement auctions by their agreement to provide a better execution price than the best quoted price, known as National Best Bid and Offer, or NBBO, to retail brokerages on orders routed to them. Exploiting a reporting requirement introduced by the Options Price Reporting Authority (OPRA) in November 2019, we are able to detect wholesaler trades and study their properties.

Our novel measure of retail trading in the options market are wholesaler trades originating from retail brokerages. In our dataset, those are all trades of OPRA type 'SLAN,' which is a price improvement auction.[3] We correlate our measure with the latest well-publicized retail investor frenzies in GameStop and other meme stocks, as measured by mentions in `WallStreetBets`, an investing forum popular with the new generation of retail investors, Robinhood user counts provided by Robintrack, and the retail frenzies measure of Barber, Huang, Odean, and Schwartz (2022). Our measure is strongly positively related to these retail investor popularity indicators. We also construct an alternative retail investor trading measure based on small trades (up to 10 contracts), commonly used in the industry.[4] Both measures show a marked increase in retail investor trading in our sample. For example, the dollar trading volume in SLAN and small transactions has grown by 143% and 224%,

[2]The term *'internalizer'* typically refers to a particular type of a market maker in *equities*: A firm which executes retail or institutional orders on a private trading platform, off lit exchanges (e.g., Citadel Securities). All the option trades in the U.S. go through exchanges; a market maker that executes customer trades routed to it, is typically called a *'wholesaler'*, since they usually bring already paired buy and sell orders to an exchange.

[3]See Appendix A.4 for a description.

[4]For instance, Deutsche Bank and Bloomberg rely on small trades to proxy retail participation in options, see `https://markets.businessinsider.com/news/stocks/stock-market-outlook-retail-investing-shorts-trading-options-deutsche-bank-2021-1-1030005344`) and `https://www.bloomberg.com/professional/blog/gamestop-highlights-importance-of-option-related-equity-flows/`, respectively.

respectively, from November 2019 to July 2021. This growth closely mimics the growth in PFOF for options in our sample.

The new generation of retail investors is more tech-savvy and more connected to investment forums, but these investors are still financial novices. It is quite striking that they are so active in options markets, despite much higher bid-ask spreads on options relative to stocks. Muravyev and Pearson (2020) report that the average quoted bid-ask spread of options on stocks in the S&P 500 is as high as 17.2%.[5] Despite the market environment of multiple exchanges, pricing mechanisms, and the overall recent improvement in the market liquidity, the average trading costs of retail investors are still quite high. For example, 50% of SLAN (wholesaler) trades in our sample are in ultra short-term options with less than a week to expiration with an average quoted bid-ask spread of 12.3%. However, the true trading costs for options are masked by the zero-commissions (offered by e.g., Robinhood); an opportunity to trade options is displayed prominently on gamified investing apps used by the new generation of investors;[6] and these investors may be attracted by a cheap way of achieving leverage that these options provide.[7]

Our further exploration of SLAN trades reveals that retail investors strongly prefer call options to puts (the volume share in calls is 69%). We also find that retail investors trade mostly at-the-money (72% of trades) or slightly-out-of-the-money (24% of trades) options. The latter involve higher trading costs, with the average quoted bid-ask spread of 28%. 14% of retail trades have a 'micro' size of up to $250, and their average quoted bid-ask spread is 23.4%. Exploring the cross-sectional correlation between SLAN Share and characteristics of the underlying, we document that retail investors prefer options on the underlying with a larger market capitalization and larger trading volume. Consistent with the literature on retail participation in equities, retail share is higher if the price of the underlying is lower and if the underlying is more liquid. We also find that SLAN Share, as well as net SLAN purchases, is positively correlated with stock-based measures of retail activity, such as ticker mentions on `WallStreetBets` forum and Robinhood ownership breadth. We view these cross-sectional relationships as evidence of speculative rather than hedging motives behind retail trades. Finally, we document significant increases in both call and put net purchases during retail investor frenzies, especially in trades of a smaller size.

Do retail investors outperform the market? To address this question, we analyze

[5] For the S&P 500 stocks, this number is 3.55bps (as reported in Hagströmer (2021)). Higher aggregate PFOF for options relative to that for stocks (see Table 12 in the Appendix) indicates that executing order flow in options is a very lucrative business for the market makers.

[6] Chapkovski, Khapko, and Zoican (2021) show that gamification induces risk-taking in novice traders.

[7] See, for example: https://www.nasdaq.com/articles/you-should-be-trading-weekly-options-and-heres-why-2021-01-20.

performance of SLAN trades at the one-, two-, five-, and ten-day horizons. On aggregate, these trades lose money for all horizons. For example, assuming a holding horizon of ten days, we estimate that the aggregate portfolio of retail investors lost $1.14 billion from November 2019 until June 2022. For the same period, all options trades returned a gain of $5.48 billion. The losses are concentrated primarily in short-term call options. However, this calculation does not include trading costs. The aggregate trading costs, measured as a distance from an actual trade price to midquote for all SLAN trades in our sample, amount to a staggering $4.13 billion. This number is much higher than direct trading costs (about $800 million), computed using commissions of retail brokerages in our sample.[8]

Given the recent surge in retail investor activity, it is important to understand its implications for behavior of arbitrageurs in the options market. We focus on one specific mistake that option investors make, for which we can cleanly identify the trading patterns of market makers and other arbitrageurs who exploit it. This mistake is a failure to exercise in-the-money call options before the underlying stock goes ex-dividend when it is optimal to do so.[9] To benefit from it, market makers and other arbitrageurs engage in a 'dividend play,' an arbitrage strategy that diverts windfall gains from the writer of the option that was suboptimally left unexercised. The strategy is normally executed on a physical exchange floor,[10] available to floor market makers and other floor participants. We exploit the new OPRA trade types to accurately classify such arbitrage trades and study the behavior of arbitrageurs. Due to the dividend play, the daily trading volume on last cum-dividend dates in in-the-money call options for which early exercise is optimal often exceeds trading volume on the remaining dates by several orders of magnitude. Even for SPY, the ticker with the most actively traded options in 2021, cum-dividend day volume is typically 14-53 times larger.[11]

Expected profits to floor market makers and other arbitrageurs from the dividend play trades have been growing rapidly during the recent retail investor boom. Most of this profit derives from the sheer increase in open interest due to investor inflow, coupled with a higher fraction of options that are left unexercised on cum-dividend dates. Overall, traders

[8]Robinhood does not charge commissions for options trades but many other brokerages still do.

[9]We note that sometimes call options may be purchased as part of any strategy that involves holding multiple option contracts. In those circumstances, or whenever transactions costs overweigh profits from early exercise, exercising an option may not be optimal.

[10]Some exchanges facilitate these strategies by imposing daily fee caps for floor market makers and other floor traders engaging in them. See e.g., https://listingcenter.nasdaq.com/rulebook/phlx/rules/phlx-options-7, accessed January 12, 2022, for the dividend strategy fee caps imposed by PHLX. Over 2/3 of dividend play transactions in our sample are executed on PHLX.

[11]The lower bound compares the average cum-dividend date dollar trading volume in call options to an average across all days in our sample, while the upper bound compares to the average volume in a week prior to cum-dividend date.

engaging in dividend play behave like unconstrained arbitrageurs in harvesting the windfall gain from failures to exercise options.

There is, however, one striking pattern that emerges from our examination of dividend play transactions. Market makers and other arbitrageurs exploit only 50% of available arbitrage profits, leaving the rest on the table.[12] We show that market makers and other arbitrageurs often exploit profitable opportunities in one contract on a particular stock while leaving another very similar contract unexploited. This is extremely puzzling. Market makers' daily fee on dividend play trades is capped by most exchanges on which dividend play trades take place. Furthermore, other trading costs are very low because such transactions are typically pre-arranged by pairs of market makers and, in our sample, actual transaction prices are close to the midpoint of the bid-ask spread. We discuss the role of transaction costs in detail in Section 4.4.

The only limits-to-arbitrage theory that seems applicable in our context is non-competitive behavior of arbitrageurs. We measure market concentration of arbitrageur firms in two different ways: the (i) payment for order flow (PFOF) share and (ii) share in the internalized volume in underlying equities of the top-3 wholesalers. The first measure is the share of PFOF received from wholesalers, i.e., market makers who bid to execute order flow from retail brokerages. The share of the Big Three—Citadel, Susquehanna, and Wolverine—in PFOF for options has grown to over 85% by the second quarter of 2021. The same Big Three also internalize a large fraction of order flow in equities. The median share of the non-ATS OTC (i.e., internalized) trading volume in equities of the Big Three options wholesalers in our sample is as high as 33%, and this number is even higher for the share of trading volume in 'meme' stocks. We do not have market maker identities in our options data, and so we use the Big Three's share of internalized volume in equities as a proxy. In sum, both measures point to a high market concentration.

We find that market makers and other arbitrageurs avoid engaging in a dividend play strategy in call option contracts that had experienced higher *SLAN order imbalances* in the week preceding the cum-dividend date. This effect is especially large for tickers that have a large share of volume executed by the Big Three PFOF providers in the preceding week. This points to the conclusion that the Big Three wholesalers are the writers of call options purchased by retail investors and hence they are set to receive the windfall gain if retail investors leave their options suboptimally unexercised. It is therefore suboptimal for them to engage in the dividend play trade in those contracts. Intriguingly, other market makers and arbitrageurs appear to avoid those contracts too, effectively leaving windfall gains in

[12] Table 33 in the Appendix quantifies forgone profits of market makers in the top-40 most popular underlying stocks and ETFs for the dividend play strategy in our sample.

those contracts to the option writers, who are likely to be the Big Three wholesalers.

Our paper offers several policy implications. Unlike reporting required by FINRA in equities, there is little transparency on wholesaler activities in the options market. Current highly concentrated market appears to favor leading wholesalers and calls into question the extent of price improvement of retail orders. Additionally, it is important to understand barriers to entry in this market.

Our paper is related to the emerging literature exploring retail investor trading in the age of Robinhood. Welch (2022), Barber, Huang, Odean, and Schwartz (2022), Boehmer, Jones, Zhang, and Zhang (2021), Eaton, Green, Roseman, and Wu (2021), and Fedyk (2021) focus on retail investor equity holdings and trading. This new generation of investors differs from retail investors previously examined in the literature (most notably, by Barber and Odean (2001)) along several important dimensions. While the counts of retail investor equity positions are available from Robintrack, data on their trading in options is not available to researchers. To our knowledge, we are the first to document retail investor preferences and market participation in options, which we infer from transaction-level data that includes newly-introduced OPRA trade types.

We are aware of two papers on retail trading in options. Using account-level data from a brokerage, Bauer, Cosemans, and Eichholtz (2009) document that retail investors' motives for trading appear to be gambling and entertainment and that they incur substantial losses on their options investments. Lakonishok, Lee, Pearson, and Poteshman (2006) argue that speculation is the key driver of retail investors' trading in options and that during the dot-com bubble they favored options on growth stocks. Our paper uses transaction-level data for the entire U.S. options market to document trading patterns of the new generation of retail investors. We show that these investors also have preferences for lotteries and opt for ultra short-term (weekly) options, participate in trading frenzies, and incur large trading costs (possibly masked by zero-commission offers).

Also related to our work are papers on options market structure and liquidity, for example, Battalio, Griffith, and Van Ness (2021), Ramachandran and Tayal (2021), Muravyev and Pearson (2020), Christoffersen, Goyenko, Jacobs, and Karoui (2018), Battalio, Shkilko, and Van Ness (2016), Muravyev (2016), Mayhew (2002). None of these papers, however, constructs measures of retail investor trading and, more generally, examines retail investors.

It has been previously documented that not all American options are exercised rationally (e.g., Poteshman and Serbin (2003)). Battalio, Figlewski, and Neal (2020), Cosma, Galluccio, Pederzoli, and Scaillet (2020), Jensen and Pedersen (2016), and Barraclough and Whaley (2012) focus on early exercise decisions and show in more recent data that a fraction of investors still fail to exercise their options optimally. Hao, Kalay, and Mayhew (2009) and

Pool, Stoll, and Whaley (2008) show how market makers exploit these mistakes by engaging in dividend play trades. Our measure of arbitrageur activity in dividend play, based on the new OPRA codes, is more accurate and it allows us to document surprising reluctance of market makers to harvest arbitrage profits in certain contracts.

Our findings are related to the literature on investor protection (for example, Barbu (2022), Bhattacharya, Illanes, and Padi (2019), Egan (2019), Célérier and Vallée (2017), and Campbell, Jackson, Madrian, and Tufano (2011)). We show how retail brokers and wholesaler-affiliated market makers benefit from the growth of retail trading in the options market, and more so than from retail trading in equities. Furthermore, retail investors' tendencies to trade options contracts with relatively larger spreads and to forgo profits from early exercise directly translate into larger gains to market makers. The complexity of options contracts from the viewpoint of an average retail investor and the potentially misaligned incentives of intermediaries call for enhancements to investor protection on trading platforms.

Finally, there are related studies highlighting potentially non-competitive behavior of market makers in equities. Christie and Schultz (1994) show that NASDAQ market makers collude so as to maintain higher bid-ask spreads. This behavior has stopped after publication of that paper. Our paper uncovers a specific mechanism through which wholesalers are able to avoid competition from other market makers on options exchanges – a price improvement auction. We also document potentially non-competitive behavior of market makers in dividend play trades.

The rest of the paper is organized as follows. Section 2 documents PFOF patterns and examines retail investor trading in options. Section 3 investigates retail investors' failure to exercise options when it is optimal to do so and describes the arbitrageurs' dividend play strategy that exploits these mistakes. Section 4 documents the puzzling behavior of market makers, who leave money on the table, and attempts to rationalize this behavior. Section 5 makes several policy recommendations aimed at retail investor protection and Section 6 concludes. The Appendix presents some technical details and robustness checks.

2 PFOF and rise of retail trading in options market

In this section, we document novel facts about retail trading in the U.S. options market. Leveraging several granular datasets and regulatory filings, we characterize a recent increase in the concentration of retail brokerage markets. We propose a new measure of retail activity in the options market based on transaction-level data, describe its composition and performance, and show how it relates to the existing stock-level retail activity measures and other stock characteristics.

2.1 Dataset

We use option transaction-level data from OPRA LiveVol provided by CBOE. This data covers all trades on 16 US exchanges in index, ETF, and equity options. In our analysis, we focus on ETF and equity options and exclude index options.

Following the literature, we remove the first 15 and last 10 minutes in the day, canceled trades, trades with nonpositive size or price, negative spread (difference between best ask and best bid), and only keep trades for which trade price is above (best bid minus spread) and below (best ask plus spread). We aggregate trades of the same contract with the same quote time, exchange ID, trade price, and trade condition ID into one line. We winsorize trade prices, sizes, and spreads at 99.5^{th} percentile daily. To compute trade imbalances, we follow Savickas and Wilson (2003) and rely on the quote rule, whereby trades with prices above (below) the midpoint are classified as 'buy' ('sell') trades, due to its superior performance for options data. We also confirm that our results hold when using Lee and Ready (1991) algorithm (or tick rule to classify trades at midpoint instead of excluding them).[13]

We use daily option price, volume, and open interest data from OptionMetrics. It comes at a contract level for the period between January 04, 1996, and June 30, 2021. We lag open interest for all the data after November 28, 2000, to have a series of consistent open interest as of the end of day.[14]

All stock-level data comes from the Center for Research in Security Prices (CRSP). This includes dividend history, stock prices and returns, outstanding shares, and rolling monthly volatility of daily returns. To link with OptionMetrics, we rely on the SecId-PERMNO crosswalk provided by WRDS.

Our data on retail investor activity is as follows. We download all comments submitted by users to 'Daily Discussion' and 'What Are Your Moves Tomorrow' threads on `WallStreetBets` subreddit of `reddit.com`. The sample spans October 1, 2019 to June 30, 2021, and is collected via PRAW, which is a Python API toolkit to access `reddit.com`. In particular, we download all the comments (original posts and reactions to them) for each daily DD or MT thread.[15] To count ticker mentions in the downloaded comments, we start from the list of unique historical tickers from CRSP and search for them in all the comments, and then simply sum by date. We only search for capitalized tickers as it is typical for the reddit audience to use those. Since we might omit any lower case mentions, and we do not cover other threads of the forum (such as occasional megathreads), our measure provides a

[13]The resulting ticker-level imbalances have correlation over 99% between the two methods.

[14]The lag is due to the change in the reporting format of OptionMetrics. This implies that end-of-day open interest is measured, therefore, after option exercises.

[15]Some dates are missing due to retrieval limitations on `reddit.com`. We interpolate between the neighboring dates to fill in those values.

lower bound for ticker popularity. For Robinhood breadth of ownership, we use Robintrack data, which is provided in intraday snapshots and covers May 5, 2018, to August 13, 2020. We use the number of users holding a stock as of the last intra-day snapshot.

In addition, we rely on FINRA OTC Transparency data to get stock trading volumes executed off lit exchanges, that is, automated trading system (ATS)[16] and non-ATS OTC trades, where the latter represents internalized trades. Pursuant to the FINRA's Regulatory Notice 15-48,[17] these data is available from April 2016, by security and venue. Securities are split into NMS Tier 1, Tier 2, and OTCE. Details are on the website of FINRA: `https://otctransparency.finra.org/otctransparency/AtsIssueData`.

A recently revised Rule 606[18] requires broker-dealers that route customer orders in stocks and options to report the aggregate data on PFOF, along with its composition across a number of categories. We download these forms (in XML format) for the largest brokers in the U.S. directly from their websites. The list of brokers, largest venues, and their corresponding payments for order flow is reported in Table 12 in the Appendix.

2.2 Zero commissions, PFOF, and market structure

The global retail brokerage industry has changed drastically in recent years. More and more platforms are offering zero-commission trading in equities, and commissions in other asset classes have been reduced. Elimination of commissions has fueled a retail participation boom in financial markets, rise in day trading, and gamification of investing.[19] The success of the zero-commission business model relies on payments for order flow from intermediaries that execute retail orders. Given that the retail order flow is largely uninformed, most of the trades in equities get crossed on private trading platforms, i.e., internalized, by these intermediaries. A similar intermediary in options, known as a wholesaler, typically 'brings' already paired buy and sell orders to the exchange, which, as we explain below, is effectively the same as internalizing them. In response to the changing industry landscape and to promote transparency, the SEC introduced new reporting requirements for brokers.[20] In this section, we use the forms filed in compliance with the new rule (Rule 606 reports) to describe the market for PFOF.

Figure 1 presents evidence on PFOF received by the U.S. retail brokerages in our

[16] ATS are typically referred to as 'dark pools.'

[17] For details, see: `https://www.finra.org/rules-guidance/notices/15-48`.

[18] For details, see `https://www.sec.gov/rules/final/2018/34-84528.pdf`

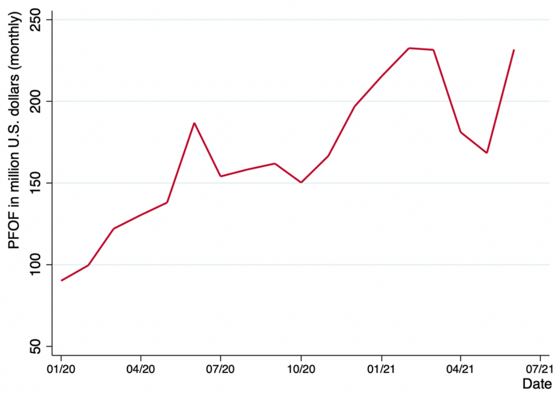
[19] See, for example, the interview with the SEC Chair on brokers' misaligned incentives: `https://www.cnbc.com/amp/2022/01/19/secs-gensler-warns-investors-about-frequent-trades-on-brokerage-apps.html`.

[20] See the SEC rule release at `https://www.sec.gov/rules/final/2018/34-84528.pdf`.

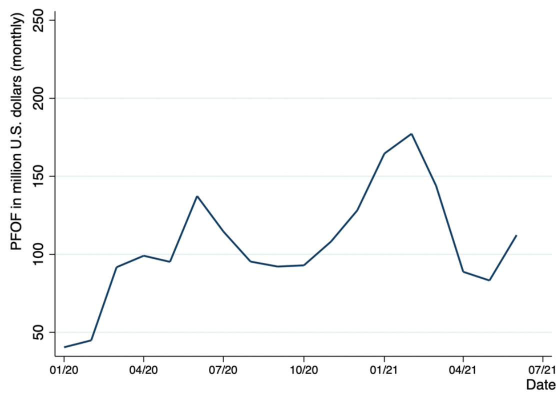
sample since the more detailed reporting of PFOF was made compulsory by the SEC. The amount for PFOF for options transactions significantly exceeds that for stocks, in each quarter in our sample. Generally, bid-ask spreads on options exchanges are considerably higher than those on stock exchanges, and so market makers which receive retail buy and sell orders are likely to benefit more from executing transactions in options.[21] As we show below, there are other ways in which market markers can benefit from interacting with the retail flow in options, in particular, by exploiting the mistakes of young, inexperienced investors who have entered the options market.

Figure 1: Payment for order flow: Options vs stocks

(a) Options



(b) Stocks



This figure plots aggregate monthly payments for order flow received by U.S. retail brokerages.

Growth in retail trading and commercial success of the zero-commission model has prompted the entry of new wholesalers and internalizers into the market,[22] yet it remains quite concentrated, with top-5 PFOF providers accounting for over 90% of the total PFOF received by U.S. brokerages (see Figure 2). Also apparent from Figure 2 is an increasing concentration of PFOF providers in options, with the share of the top-3 providers—Citadel, Susquehanna, and Wolverine —rising from 70% in early 2020 to nearly 85% in the second quarter of 2021. We hereafter refer to these firms as the Big Three wholesalers in options.

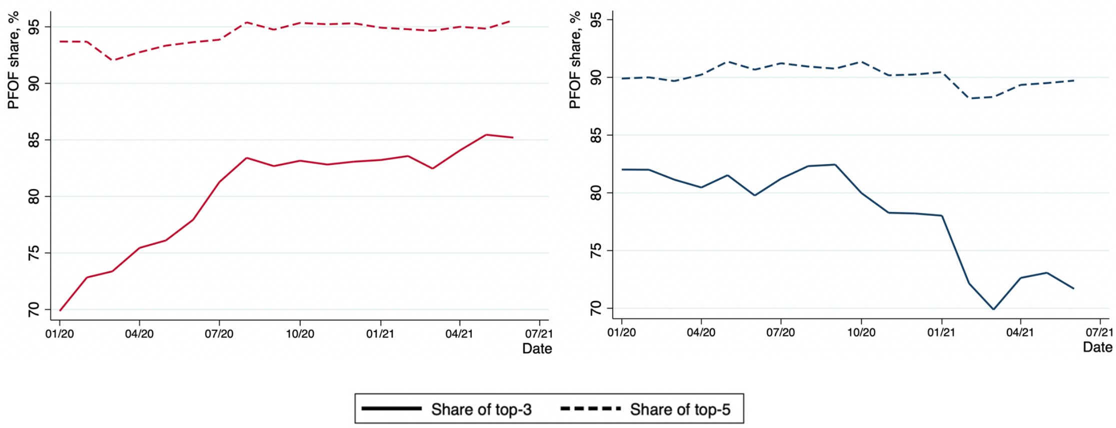
PFOF also tends to be concentrated in a handful of brokerages, as we illustrate in Appendix A.3. This is, however, to a large extent a reflection of their business models: For example, Robinhood is more reliant on the payment from order flow than Fidelity. TD

[21]According to Rule 606 reports, PFOF arrangements differ from broker to broker but the majority are still based on spread, that is, the broker receives a fraction of the spread that the wholesaler/internalizer charges.

[22]For example, Hudson River Trading entered the market for PFOF: https://www.wsj.com/articles/high-frequency-trader-hudson-river-to-execute-retail-stock-trades-11625047200#:~:text=Hudson%20River%20Trading%20LLC%2C%20one,so%2Dcalled%20retail%20wholesaler%20business.

Figure 2: Market concentration in PFOF: Options vs stocks

(a) Options (b) Stocks



This figure plots the share of PFOF received by U.S. retail brokerages from the top-3 and top-5 providers. The top-3 providers in options are Citadel, Susquehanna, and Wolverine while the top-3 providers in stocks are Citadel, Virtu, and Susquehanna.

Ameritrade is by far the largest receiver of payment for order flow in both stocks and options. At the same time, Robinhood's share in the options PFOF has been steadily increasing over our sample period, and is almost as high as that of TD's as of June 2021.

2.3 Retail investor trading in equity and options markets

While recent literature on the ongoing retail investor boom has come up with a number of new retail trading measures, all of them have been focused on equities. These stock-level measures include retail trading imbalances (Boehmer, Jones, Zhang, and Zhang (2021)), breadth of Robinhood user ownership (Welch (2022) and Eaton, Green, Roseman, and Wu (2021)), and counts of `WallStreetBets` ticker mentions (also Eaton, Green, Roseman, and Wu (2021)).[23] Even though these measures are not for options, retail investor frenzies in options and underlying equities tend to occur at the same time, and so we find it useful to include measures of retail investor activity in equities in our dataset.[24]

We add one more measure of retail equity trading to the list: internalized volume, which is the share of non-ATS OTC weekly trading volume in total volume, at a stock level, reported to FINRA.[25] FINRA makes public the identities of the largest market makers

[23] This list is based on the most recent measures with wider coverage, and it omits papers using proprietary data such as NASDAQ TRF data.

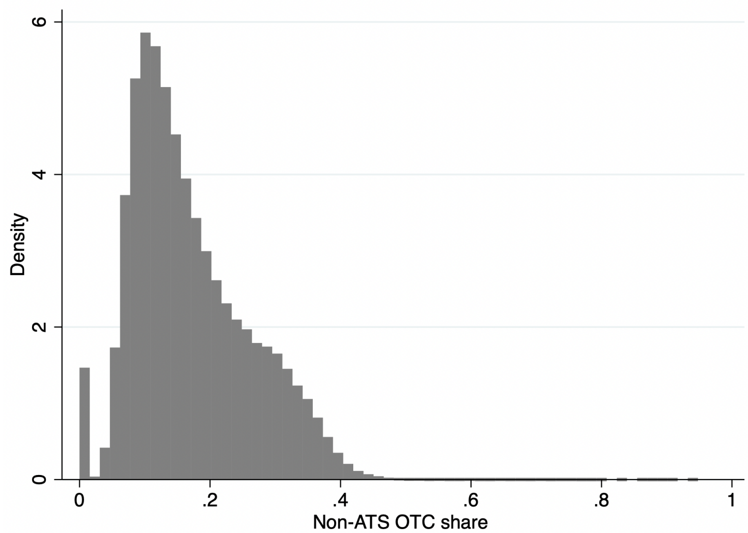
[24] We include the latter two measures because we do not have TAQ data required for constructing the measure of Boehmer, Jones, Zhang, and Zhang (2021).

[25] Not all of these trades originate from retail brokerages (FINRA defines it as "non-ATS electronic trading systems and internalized trades"). However, our results suggest that a significant fraction of these trades

executing non-ATS OTC transactions. Internalized trades for stocks are executed off lit exchanges, but not in 'dark pools' (which are classified as ATS transactions). The non-ATS OTC transactions consist primarily of internalized order flow from retail and institutional customers of wholesalers/internalizers. Table 13 in the Appendix ranks market makers by their non-ATS OTC volume share.[26] This ranking closely resembles that in Panel A of Table 12, in which we sort market makers by their share in PFOF. To the best of our knowledge, this measure has not been used in the extant literature to date.

Figure 3 plots a histogram of weekly non-ATS OTC trading volume (internalized volume) as a share of the total weekly stock trading volume. The average share of internalized volume in the total one is 17% in our sample, and it is trending upwards.

Figure 3: Histogram of non-ATS OTC share



This figure plots the share of non-ATS OTC volume in the total trading volume for all equities and ETFs with options traded in the U.S. in 11/2019-06/2021.

We next move to retail investor trading measures for options. Unlike stocks that trade on a variety of lit and dark venues, all options trades in the U.S. must go through exchanges, and every transaction is recorded. For our measures, we use a transaction-level dataset that includes all options transactions in the U.S. One measure, often used in the industry, is the share of small trades (up to 10 contracts). One could compute it as a frequency share and as a trading volume share. We adopt the latter definition, as it would be more relevant for assessing the influence of retail traders on asset prices. We compute it daily and, in this

do.

[26]Our list of wholesalers/internalizers is very close to that documented in Eaton, Green, Roseman, and Wu (2021) based on Nasdaq data.

section, we aggregate it to a ticker level using traded volumes.

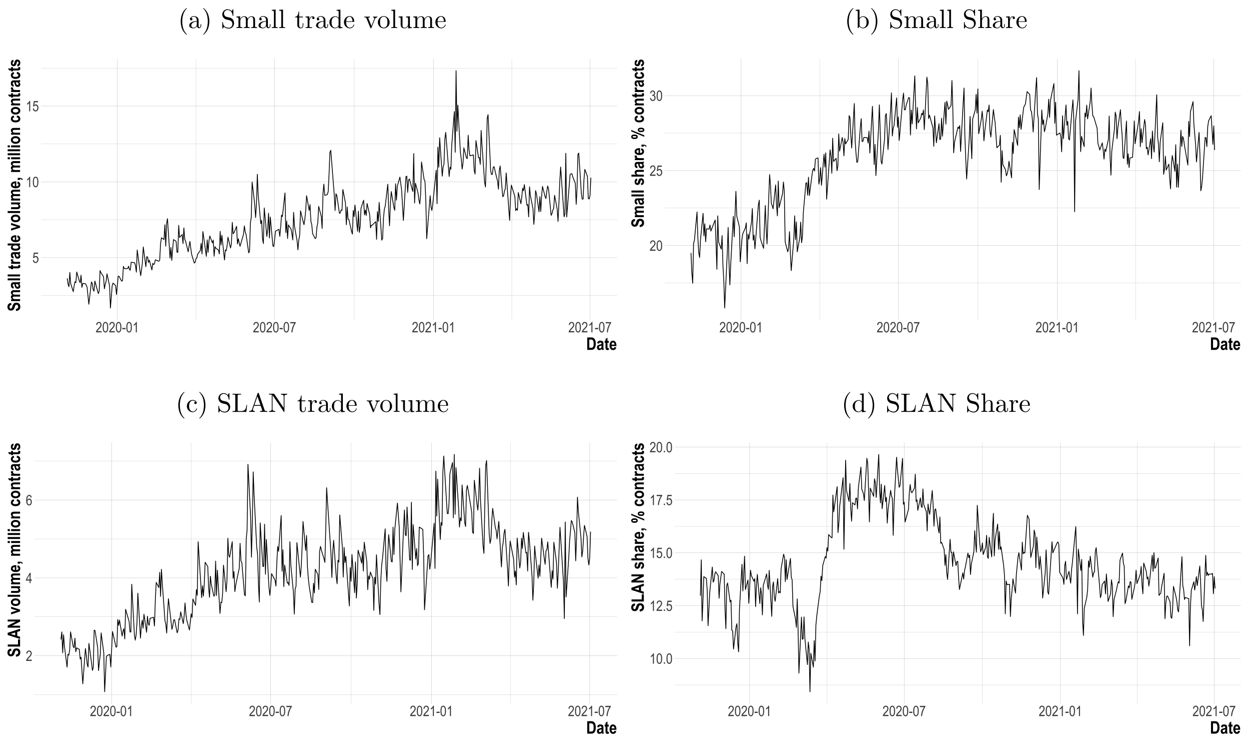
Unique features of our data allow us to construct a novel measure of retail investor trading in options. We take advantage of the new trade type codes, introduced by OPRA on November 4, 2019, which provide a detailed classification of the types of transactions. This reporting requirement is significantly more detailed than its predecessors, and hence we can construct our measure only starting from November 4, 2019.

A highly publicized advantage to investors for having their orders routed to a wholesaler by a retail brokerage, in exchange for PFOF, is that the wholesaler promises a price improvement to the customers, i.e., the execution price that is at least as good or better than the best quoted price, known as National Best Bid and Offer, or NBBO. To meet this commitment, wholesalers execute retail orders through a 'price improvement auction' mechanism offered by options exchanges. The orders entered into an auction typically originate from the same wholesaler. A wholesaler can 'bring' a *paired* order to the exchange to be 'internalized' as long as the order is exposed to other market participants on that exchange. Market participants ('responders') have a window of time to respond with a better price for the agency order (hence, the name 'price improvement auction'), which could lead to the wholesaler losing the trade. In practice, the fees are stacked against responders and it is prohibitively expensive to break up one of these paired trades.[27]

Our novel measure of trades originating from retail brokerages, *SLAN Share*, is a volume share of all the trades with OPRA type 'SLAN', which refers to a single-sided trade that went through a price improvement auction. We plot the two retail investor trading measures in options, Small Share and SLAN Share, in Figure 4, as well as the total volume of small and SLAN trades. Panels (a) and (c) reveal significant growth of and comovement between small and SLAN trading volumes: Retail investor trading shows a marked increase in our sample. For example, the dollar trading volume in SLAN and small transactions has grown by 143% and 224%, respectively, from November 2019 to July 2021. This mimics the growth of PFOF for options, which is 156% over the same period, based on monthly data. The growth in retail trading is especially high from January 2020 until March 2021. This period includes several well-publicized retail investor frenzies in equities and a meteoric rise in the number of Robinhood's active users. This increased participation is also reflected in higher average shares, especially in summer 2020, when the average SLAN Share was almost

[27]For example, on most exchanges, order execution by a wholesaler-affiliated market maker gets charged the contra fee of just $0.05 per contract. In contrast, it would cost another market maker $0.50 to break up/respond to one of these already paired orders during an auction. In the latter case, the wholesaler receives a net rebate of $0.30 per contract just for bringing the order to the exchange. For details, see, for example, the fee schedule for NYSE CUBE: https://www.nyse.com/publicdocs/nyse/markets/american-options/NYSE_American_Options_Fee_Schedule.pdf.

Figure 4: Retail investor trading in options



This figure characterizes retail investor trading in the U.S. options market in 11/2019-06/2021. Panels (a) and (c) plot total daily trading volumes in small trades and SLAN trades, respectively. Panels (b) and (d) plot daily small and SLAN shares, respectively, averaged across all stocks and ETFs in our sample.

as high as 20%.

Table 1 presents various features of SLAN trades and compares them to average trades in the options market. One striking fact is that retail investors prefer to trade options with the shortest maturities: 49.6% of SLAN trades (in terms of their volume share) are in weekly options, compared with 42.8% for the entire universe of trades. This is not surprising: Weekly options have the lowest prices relative to otherwise identical contracts with longer maturities, and retail investors, often being cash-constrained, opt, therefore, for the 'cheapest' alternative. The 'cheapest' alternative, however, is by no means cheap to trade. The average bid-ask spread in options with less than a week to expiration is a whopping 12.3%. At the same time, the average quoted bid-ask spread of retail trades across all the maturity buckets is 13.4%, compared to 11.1% for the overall market trades. This difference is highly significant, both statistically and economically. Lured by recent low- or zero-commission offers, retail investors possibly underestimate the indirect trading costs in the options market.[28] At the same time, a larger share of SLAN trades are executed exactly

[28]The PFOF model and its implications for execution quality and cost transparency have been under scrutiny

at the midpoint, and we find that the effective half-spread, that is, the deviation of trade price from the midpoint, is slightly lower across most categories for these trades.

Table 1: Composition of option trades

Characteristic	Category	SLAN trades: Frequency share, %	SLAN trades: Volume share, %	SLAN trades: Quoted spread, %	SLAN trades: Effective spread, %	All trades: Frequency share, %	All trades: Volume share, %	All trades: Quoted spread, %	All trades: Effective spread, %
Type	Call	71.2	69.0	13.3	6.5	65.0	62.3	10.8	7.7
	Put	28.8	31.0	13.7	6.7	35.0	37.7	11.9	8.0
Trade size (contracts)	1	45.6	6.2	13.7	6.3	44.4	6.0	10.9	7.4
	2-5	30.9	13.3	12.4	6.1	31.6	13.3	10.8	7.5
	6-10	11.8	14.5	13.8	7.1	11.4	13.3	12.0	8.4
	11-100	11.0	53.1	14.6	8.2	11.6	48.7	12.3	9.1
	Above 100	0.6	12.9	14.6	11.6	0.9	18.8	12.6	10.2
Trade size (dollars)	Below 250	41.4	14.3	23.4	11.6	37.9	13.3	19.9	13.7
	250-500	15.4	8.9	8.3	3.7	15.0	8.0	7.7	4.8
	500-1,000	13.6	11.3	7.0	3.0	14.2	10.3	6.5	4.0
	1,000-2,500	13.8	17.3	5.9	2.5	15.0	16.4	5.5	3.3
	2,500-5,000	7.0	13.5	4.9	2.0	7.8	13.3	4.6	2.8
	5,000-10,000	4.6	13.1	4.2	1.8	4.9	12.2	4.0	2.5
	10,000-20,000	2.5	10.1	3.7	2.9	2.7	10.0	3.5	4.8
	20,000-50,000	1.3	7.6	3.3	6.2	1.6	8.9	3.1	10.4
	Above 50,000	0.5	3.9	3.0	11.4	0.8	7.7	2.9	17.6
Trade direction	Sell	46.2	46.4	13.6	7.6	47.5	47.2	10.1	7.6
	Buy	43.4	44.8	13.0	7.1	45.8	46.9	12.0	9.1
	Midpoint	10.4	8.8	14.2	0.0	6.7	6.0	13.3	0.0
Time to expiry	Less than a week	47.2	49.6	12.3	6.5	42.4	42.8	12.5	9.1
	1-2 weeks	13.8	12.8	12.2	6.0	14.4	13.2	9.7	6.6
	2-4 weeks	16.0	15.2	14.9	7.0	17.1	16.5	10.8	7.0
	1-3 months	13.6	13.6	13.7	6.1	15.4	15.9	9.5	6.0
	3-12 months	7.6	7.3	18.3	7.7	8.6	9.4	10.5	7.3
	Over a year	1.4	1.3	17.4	9.1	1.9	1.9	12.2	10.9
Moneyness	Below -2	0.3	0.3	53.3	27.9	0.3	0.3	46.2	30.1
	-2 to -1	0.4	0.4	50.2	25.3	0.4	0.4	42.8	25.7
	-1 to -0.1	23.8	24.1	28.4	13.7	24.2	25.2	21.1	14.0
	At the money	71.2	71.5	8.4	4.1	70.0	69.4	7.8	5.5
	0.1 to 1	4.1	3.6	8.2	4.6	4.8	4.4	5.7	6.4
	1 to 2	0.2	0.1	8.6	7.4	0.2	0.2	6.3	13.8
	Above 2	0.1	0.1	16.3	11.2	0.1	0.1	11.6	24.8
Trade direction and type	Sell - Call	32.7	31.8	13.4	7.4	30.8	29.3	9.6	7.5
	Sell - Put	13.5	14.5	14.3	8.0	16.7	17.9	10.9	7.8
	Buy - Call	31.2	31.2	13.0	7.1	29.9	29.4	11.6	8.9
	Buy - Put	12.2	13.6	13.1	7.1	15.9	17.5	12.7	9.3
	Midpoint - Call	7.3	6.0	14.5	0.0	4.2	3.6	12.9	0.0
	Midpoint - Put	3.1	2.9	13.4	0.0	2.5	2.3	14.0	0.0
ETF	No	81.4	72.4	14.6	7.1	81.6	71.3	11.8	8.2
	Yes	18.6	27.6	8.3	4.4	18.4	28.7	8.4	6.0

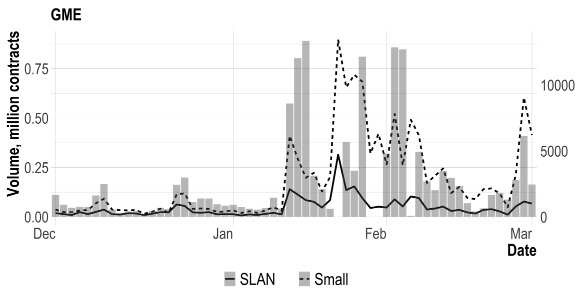
This table reports characteristics of trades by category. (Implied) Trade direction is based on whether the trade price is above (buy), below (sell), or at the midpoint. Quoted spread, % is the spread between the best bid and best ask on the contract (across all exchanges) relative to the midpoint price at the time of the trade. Effective spread is an absolute percentage deviation of the trade price from the midpoint price at the time of the trade, multiplied by 2. For both spreads, we report frequency-weighted averages. Moneyness is measured as $(MidpointPrice - Strike)/Strike$.

Table 1 also reveals that retail investors strongly prefer calls to puts (the volume share in calls is 69%). We see that retail investors trade mostly at-the-money (72% of trades) or slightly-out-of-the-money (24% of trades) options. The latter involves higher trading costs, with the average quoted bid-ask spread of 28.4%. Furthermore, 14.3% of retail trades have a 'micro' size of up to $250, compared to 13.3% in the whole market, and their average quoted

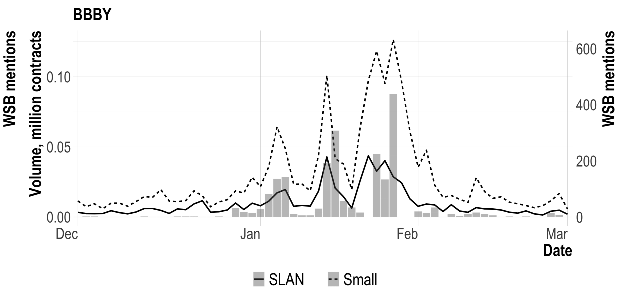
of regulators for years. See, for example, the 2021 U.S. congressional hearing on Robinhood named "Game Stopped? Who Wins and Loses When Short Sellers, Social Media, and Retail Investors Collide.": https://www.nytimes.com/2021/02/19/business/dealbook/robinhood-hearing-congress.html.

Figure 5: ‘Meme’ stocks retail trading and `WallStreetBets` (WSB) mentions in 2021

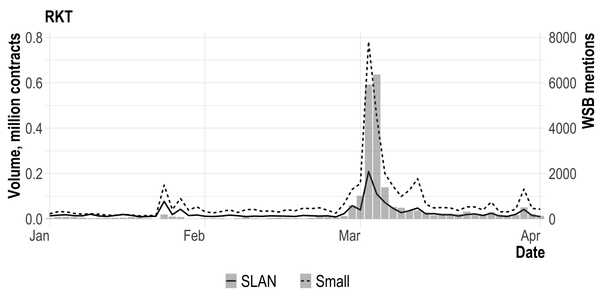
(a) GameStop, GME



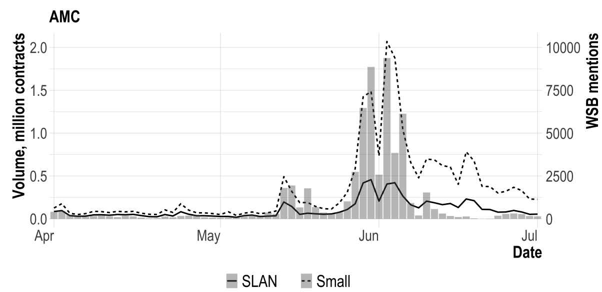
(b) Bed Bath & Beyond, BBBY



(c) Rocket Companies, RKT



(d) AMC Entertainment Holdings, AMC



This figure plots daily WSB mentions (gray bars) and daily volume of SLAN (wholesaler) and Small (i.e., less than 10 contracts) trades.

bid-ask spread is 23.4%.[29] These observations suggest that retail investors are entering the options market with an intent to speculate rather than hedge.[30] Furthermore, there is almost perfectly balanced initiation of buy and sell trades in either call or put options. This is consistent with the idea that retail order flow is symmetric and therefore potentially attractive to wholesalers who earn profits from crossing these trades.[31]

A natural question to ask is how our measures of retail trading in options behave during retail investor frenzies. For illustration, Figure 5 plots SLAN and small trade volumes alongside counts of `WallStreetBets` mentions for four ‘meme’ stocks: GameStop, Bed Bath & Beyond, Rocket Companies, and AMC. We should note that our measure of `WallStreetBets` mentions has some missing dates due to the retrieval limitations on

[29] The literature on retail trading in *equities* typically considers such large trades to be institutional (starting from Lee and Radhakrishna (2000)). Only about 10% of retail trades are above $20,000. Table 15 in the Appendix shows the descriptive statistics of trades below $20,000, which are very similar to those without the size filter.

[30] These observations are consistent with Lakonishok, Lee, Pearson, and Poteshman (2006) and Bauer, Cosemans, and Eichholtz (2009).

[31] Table 17 in the Appendix shows that our conclusions do not change if we we use the fraction of dollar volume in each category instead of frequency or contract volume.

reddit.com, which appear as gaps in the figure.[32] It is apparent from Figure 5 that both measures adequately capture peaks of WallStreetBets mentions of these tickers. In Table 2 below, we establish the cross-sectional relationship between our measures and stock-level retail activity measures formally in a regression framework, for the entire sample.

Having defined our measure of retail activity in the options market, we explore its relation with the characteristics of both options contracts and their underlying. To do that, we first run the following panel regression, separately for call and put options:[33]

$$SLAN\ Trading_{i,t} = \boldsymbol{\gamma}' X_{i,t} + \boldsymbol{\delta}' C_{i,t} + \alpha_{i,t} + \varepsilon_{i,t}, \tag{1}$$

For call or put contracts of each ticker i on date t separately we consider two measures for $SLAN\ Trading_{i,t}$. The first one is $SLAN\ Share_{i,t}$, the volume share of SLAN trades among all the options transactions in ticker i on date t, which reflects the general presence of retail investors. The second measure is $SLAN\ Imbalance_{i,t}$, in both calls and puts, which is the volume difference in buy and sell SLAN trades scaled by the total volume of SLAN trades, corresponding to a buy or sell tilt in retail investor trades.

Our vector of characteristics $X_{i,t}$ includes the following ticker-level variables: log dollar trading volume in options on $t-1$, log price on $t-1$, log total trading volume (lit, ATS, and non-ATS OTC) in the underlying stock or ETF over the previous week, relative spread in the underlying averaged over the previous week, volatility of the underlying returns over the previous week, and log market capitalization value as of $t-1$. Our vector of contract characteristics $C_{i,t}$, equal-weighted at ticker i level, includes: quoted spread, options moneyness, their time to expiration in months, and leverage.[34] We include date and ticker fixed effects, $\alpha_{i,t}$. Finally, we report descriptive statistics for all these variables in Table 18 in the Appendix.

Table 2 presents the results of estimating equation (1). A notable feature of SLAN trades, is that retail investor share is higher in shorter-maturity options on the underlying with a larger market capitalization and a higher trading volume in the previous week. The latter is consistent with higher retail participation in attention-grabbing securities. Furthermore, the lagged price is negatively related to the SLAN share, while last week's returns have a negative coefficient for calls and positive for puts. In addition, retail trading is more prevalent in the options on more liquid stocks and ETFs. Earlier studies have documented

[32] These limitations can only be circumvented with the real-time scraping of reddit.com data.

[33] Splitting the contracts allows us to document differential relationship with the past return on the underlying stock or ETF. All the other results stay very similar if we pool both types of contracts together.

[34] Results are not sensitive to whether we use equal-weighting or volume-weighting for contract characteristics at a ticker level. Furthermore, our results are robust to including implied volatility, trade size, or delta into the list of contract-level controls.

similar relationships for the stock-level imbalances (see Boehmer, Jones, Zhang, and Zhang (2021) and Welch (2022)).

Table 2: Retail trading in options and underlying characteristics

	SLAN Share		**SLAN Imbalance**	
	Call (1)	Put (2)	Call (3)	Put (4)
Option volume, lagged log	-0.016***	-0.042***	0.038***	0.028***
	(-5.34)	(-16.82)	(13.37)	(9.49)
Underlying price, log	-0.278***	-0.221***	-0.035***	-0.064***
	(-16.43)	(-14.69)	(-3.12)	(-6.03)
Underlying return, past week	-0.005***	0.013***	-0.004***	0.006***
	(-4.00)	(9.92)	(-2.90)	(3.80)
Total volume in underlying, past week log	0.045***	0.039***	0.018***	0.037***
	(7.72)	(7.64)	(3.78)	(6.97)
Underlying spread	-0.027***	-0.015***	-0.018***	-0.014***
	(-6.74)	(-4.07)	(-5.17)	(-3.88)
Underlying volatility, past week	0.001	0.001	-0.004**	-0.003
	(0.29)	(0.55)	(-1.97)	(-1.38)
Market cap, lagged log	0.065***	0.041**	-0.079***	-0.001
	(2.70)	(1.98)	(-5.03)	(-0.04)
Option time to expiry	-0.008***	-0.013***	0.001	-0.002
	(-5.69)	(-9.49)	(0.65)	(-1.22)
Option moneyness	-0.016***	-0.016***	-0.001	0.001
	(-8.97)	(-8.85)	(-0.43)	(0.84)
Option spread	-0.023***	-0.026***	-0.006**	-0.007**
	(-11.63)	(-13.78)	(-2.28)	(-2.50)
Option leverage	0.003	0.002	0.001	0.001
	(1.43)	(0.73)	(0.24)	(0.39)
Observations	1,398,642	1,203,829	1,072,741	805,414
Adjusted R-squared	0.110	0.083	0.020	0.023

This table reports the results of estimating (1) on daily data in 11/2019-06/2021. SLAN Share is the ticker-level volume shares of SLAN trades. SLAN Imbalance is the ticker-level volume imbalance for SLAN trades. Underlying price (log) is as of the day before. Underlying return is the total return over the last week. Underlying spread is averaged over the previous week. Underlying volatility is return volatility over the previous week. Option spread is the contract quoted relative spread. Option time to expiry (in months), moneyness, spread, and leverage are equal-weighted across trades at a ticker level. All regressions include date and ticker fixed effects. All variables are standardized within the contract type (call or put). t-statistics are based on standard errors clustered by ticker and date (in parentheses). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Using directional order imbalances in SLAN trades as a dependent variable, we find similar patterns in retail trades (see the last two columns of Table 2). For example, retail investors tend to buy more call options in the securities with higher previously observed trade volume and lower underlying price (and hence, cheaper options price, other things being equal). Interestingly, we see that SLAN Imbalance in calls is likely to be higher in smaller stocks. However, we also see that our chosen set of characteristics has smaller overall explanatory power for imbalances. It suggests that most of the potential price pressure originated from retail investors in the options market seems to be unrelated to fundamentals. This is consistent with the retail flow being fairly balanced and, hence, attractive to market makers.

How are SLAN Share and SLAN Imbalance related to other measures of retail activity? To answer this question, we run a panel regression similar to equation (1) but in addition, consider other measures of retail activity:

$$SLAN\ Trading_{i,t} = \beta Retail_{i,t} + \boldsymbol{\gamma}' X_{i,t} + \boldsymbol{\delta}' C_{i,t} + \alpha_{i,t} + \varepsilon_{i,t}, \tag{2}$$

where $Retail_{i,t}$ is one of the following measures of retail activity at a ticker level: $share^{small}$ is the volume share of trades up to 10 contracts for ticker i on date t (within call and put options), $Internalized\ volume\ in\ underlying_{i,t}$ is the share of non-ATS OTC (i.e., internalized) volume in the total trading volume of ticker i in the week of date t, *Robinhood ownership breadth,* $log_{i,t}$ is the logarithm of the number of Robinhood users holding the ticker i at the end of date t, and *WSB mentions,* $log_{i,t}$ is the logarithm of the number of times ticker i was mentioned on `WallStreetBets` forum on date t. We use the same set of controls for options contracts and their underlying as before.

Table 3 presents the results of estimating equation (2). Our first observation is that all the measures of retail trading are positively correlated with both SLAN Share and SLAN Imbalance in the cross-section, and these coefficients (reflecting partial correlations of the existing measures with our new one) are strongly statistically significant. This provides further validation to our measure of retail trading in options, with additional supporting evidence presented in Section 2.3.1. However, along with the ticker-level X and C characteristics and fixed effects, they explain only 8-12% of the total variation in SLAN Share, showing very limited improvement over the explanatory power documented in Table 2.

It is interesting to note that only WSB mentions seem to exhibit no correlation with SLAN share in calls. We attribute this result to several reasons. First, retrieval limitations of `WallStreetBets` coincided with the days of several retail investor frenzies, which naturally lowers correlation between WSB mentions and our measure. Unfortunately, as we explained earlier, the only way to circumvent this limitation was to scrape the website in real time, making it infeasible for any follow-up analysis. Interestingly, however, WSB mentions are highly significantly correlated with SLAN Imbalance, suggesting that ticker popularity on the investor forum is indeed related to the overall buying pressure in both call and puts, even after conditioning on all the contract and underlying characteristics.

Given that the trading volume in the U.S. options market is highly skewed, one might be concerned that our results hold only for very thinly traded contracts. In Table 20 in the Appendix, we estimate equation (2) for the 341 tickers that constitute the top decile by the total dollar trading volume in our sample. The estimation results are similar to what we document in this section.

Table 3: Retail trading in options and other measures of retail activity

	Retail trading in calls				**Retail trading in puts**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: SLAN Share								
Small share	0.054***				0.052***			
	(23.06)				(24.54)			
Internalized volume in underlying		0.025***				0.019***		
		(9.05)				(6.87)		
Robinhood ownership breadth, log			0.034***				0.059***	
			(3.29)				(5.61)	
WSB mentions, log				-0.001				0.003**
				(-0.30)				(2.15)
Undelying controls X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contract controls C	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,398,642	1,398,642	571,699	1,145,651	1,203,829	1,203,829	496,592	1,020,950
Adjusted R-squared	0.112	0.110	0.104	0.121	0.085	0.084	0.077	0.091
Panel B: SLAN Imbalance								
Small Imbalance	0.533***				0.529***			
	(270.12)				(227.72)			
Internalized volume in underlying		0.015***				0.007**		
		(5.17)				(2.23)		
Robinhood ownership breadth, log			0.047***				0.030***	
			(4.55)				(3.25)	
WSB mentions, log				0.013***				0.009***
				(10.10)				(6.55)
Undelying controls X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contract controls C	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,068,335	1,072,741	423,645	928,569	800,233	805,414	327,645	724,603
Adjusted R-squared	0.198	0.020	0.026	0.019	0.193	0.023	0.025	0.023

This table reports the results of estimating (1) on daily data in 11/2019-06/2021. SLAN and Small Share are the ticker-level volume shares of SLAN and small trades, respectively. SLAN and Small Imbalance are the ticker-level volume imbalance for SLAN and small trades, respectively. Internalized volume in underlying is the share of non-ATS OTC (i.e., internalized) volume in the total trading volume in the underlying stock or ETF. Robinhood ownership breadth, log is the logarithm of the total number of Robinhood users holding the ticker at the end of each day. WSB mentions, log is the logarithm of the number of mentions a ticker gets on `WallStreetBets` during the day. Underlying controls X and contract controls C are described in Section 2.3. All regressions include date and ticker fixed effects. All variables are standardized within the contract type (call or put). t-statistics are based on standard errors clustered by ticker and date (in parentheses). *** p<0.01, ** p<0.05, * p<0.1.

Our measure of retail trading has a few limitations. First, 11.5% of the SLAN volume is concentrated in transactions with over \$20,000 in value (see Table 1), which is considered as a cutoff for retail trades in the related literature (see, e.g., Lee and Radhakrishna (2000)). We therefore exclude trades above this size in our robustness checks. Table 19 in the Appendix confirms that the results are virtually the same. Second, our measure might omit some retail trades executed through trade types other than SLAN. A small fraction of the trades do not go through a wholesaler (most notably, those originating from Interactive Brokers). Furthermore, smaller trades (of size up to 5 contracts) can be routed by wholesaler to an affiliated or other specialist on an exchange depending on the order attractiveness. Finally, our measure omits complex strategies, such as bull spreads, straddles, or butterfly spreads. This happens because complex strategies typically require multi-leg transactions. Therefore, wholesalers looking for price improvement would typically execute them via multi-leg price improvement auctions, as opposed to the single-leg ones. In the OPRA data, these transactions appear as a trade type 'MLAN' (multi-leg auctions). MLAN trades correspond to about 4% of the total market volume, and they are composed primarily of trades of small professional investors and hedge funds, albeit some may be those of retail investors. In addition, in Appendix A.11, we report descriptive statistics and cross-sectional correlations of MLAN with the equity-based measures of retail activity. It further demonstrates that these trades are clearly quite different in nature to those going through single-leg actions. Since we want to capture trading of the new generation of retail investors, we do not include MLAN trades in our analysis.

2.3.1 Additional validation

In this section, we offer additional suggestive evidence that our measure captures *retail* trading in the U.S. options market.

First, we exploit the fact that some U.S. retail brokerages handle expiring options on their clients' accounts in a rule-based manner. For example, Robinhood attempts to exercise in-the-money options (if the account has enough buying power) or sells the contract approximately one hour before the market close (if it does not).[35] This gives us a testable prediction for our measure of retail trading in contracts on their expiration day: We expect to see an imbalance in the direction of sell trades in the last one or two trading hours of the day. To test this prediction, we study volume share of buy and sell trades in each trading hour on option expiration day.

As Table 26 in the Appendix reports, SLAN volume tends to be more imbalanced in the sell direction at the end of an expiration day. At the same time, if anything, there is a buy imbalance in the last trading hour for MLAN trades and other trades that are more likely to be non-retail. We test these differences more formally in Table 25 and find them to

[35]See Robinhood's rules here: https://robinhood.com/us/en/support/articles/expiration-exercise-and-assignment/, accessed on March 21, 2022. This rule was introduced by Robinhood approximately in September 2020.

be statistically significant.

Second, we study directional order imbalances across trade types during the Robinhood herding events (frenzies) uncovered in Barber, Huang, Odean, and Schwartz (2022).[36] In particular, we estimate equation (2) using a dummy for the Robinhood herding event in ticker i on date t instead of $Retail_{i,t}$. This analysis is performed on a subsample of our data (Novermber 4, 2019 to August 10, 2020) due to availability of Robintrack data with which the investor frenzies are identified.

Table 27 in the Appendix documents higher SLAN Imbalance during Robinhood herding events. We also find that the correlation is the highest for SLAN trades sized below \$5,000. Importantly, imbalances in MLAN, all multi-leg, and large trades are not positively related to frenzies. Our results even show negative correlations, suggesting that other types of investors, most likely professional traders or institutions, trade against the retail investors during such events. Overall, we document that during the well-publicized investor frenzies there were directional order imbalances in retail trading in options as well.

2.4 Aggregate performance of retail investors in the U.S. options market

We compute the aggregate retail investor dollar performance over the horizon of h days in the spirit of Barber, Lee, Liu, and Odean (2008):

$$\$Raw\,Perf_h^{SLAN} = \sum_{it} V_{i,t}^{SLAN} \times r_{i,t,t+h}$$

where $V_{i,t}^{SLAN}$ are the net dollar purchases of option contract i through SLAN trade type on day t and $r_{i,t,t+h}$ are the h-day horizon returns on each contract computed as:

$$r_{i,t,t+h} = \frac{Close\ midquote_{i,t+h}}{Average\ trade\ price_{i,t}^{SLAN}} - 1$$

We consider horizons h of one, two, five, and ten days. $Close\,midquote_{i,t+h}$ is the close midquote of contract i on day $t+h$ as reported by OptionMetrics. $Average\,trade\,price_{i,t}^{SLAN}$ is the average trade price of SLAN trades on day $t-1$, which is the average buy price of SLAN trades if $V_{i,t}^{SLAN} > 0$ (retail investors were net buyers of contract i on day t) or the average sell price if $V_{i,t}^{SLAN} < 0$ (retail investors were net sellers of contract i). In the main text of the paper, we report results for equally weighted prices.[37] Further, $r_{i,t,t+h}$ are

[36] We thank Brad Barber for kindly providing us with their data for this exercise.

[37] Results for value-weighted transaction prices are very similar. We report them in Table 28 in the Appendix. Equally weighted prices may be sensitive to outliers while value-weighted prices might be affected by price

winsorized at 0.25^{th} and 99.75^{th} percentiles each day.

Table 4 summarizes the performance of retail investor options trades and compares it to that of all the trades in the market. Under the assumption of a ten-day holding period, retail investors lost \$1.14 billion on their options trades between November 2019 and June 2021. During the same time period, all options trades returned a gain of \$5.48 billion. Interestingly, Table 4 also shows that retail investor losses were concentrated in at-the-money or slightly in-the-money calls with a very short time to expiration (less than a week).

Table 4: SLAN trade performance, aggregate and by contract characteristics

	SLAN Raw performance, \$ billion				**Market Raw performance, \$ billion**			
	1 day	2 days	5 days	10 days	1 day	2 days	5 days	10 days
Panel A: All contracts								
	-0.425	-0.970	-1.095	-1.135	3.748	4.440	4.728	5.477
Panel B: By contract type								
Call	-0.187	-0.754	-0.918	-1.062	3.637	3.842	3.870	3.368
Put	-0.237	-0.215	-0.177	-0.072	0.106	0.592	0.856	2.101
Panel C: By moneyness								
Below -2	-0.003	-0.003	-0.004	-0.002	-0.059	-0.071	-0.074	-0.092
-2 to -1	-0.004	-0.003	0.000	0.006	-0.073	-0.079	-0.087	-0.097
-1 to -0.1	0.003	0.044	0.183	0.295	0.720	1.030	1.411	2.026
At the money	0.217	-0.177	-0.458	-0.559	2.522	3.192	2.767	2.604
0.1 to 1	-0.384	-0.572	-0.557	-0.598	1.348	1.077	1.540	1.947
1 to 2	-0.114	-0.118	-0.130	-0.137	-0.187	-0.167	-0.185	-0.172
Above 2	-0.137	-0.137	-0.127	-0.137	-0.525	-0.589	-0.646	-0.754
Panel D: By time to expiry								
Less than a week	-0.309	-0.728	-1.062	-1.064	5.447	5.546	5.278	5.394
1-2 weeks	-0.059	-0.119	-0.141	-0.258	0.847	0.984	1.564	1.588
2-4 weeks	-0.022	-0.093	-0.089	-0.188	0.911	1.087	1.203	1.487
1-3 months	-0.003	-0.020	0.068	0.136	0.815	0.994	0.753	0.616
3-12 months	0.060	0.071	0.179	0.274	-1.357	-1.275	-1.156	-0.680
Over a year	-0.091	-0.079	-0.049	-0.034	-2.920	-2.901	-2.915	-2.936

This table reports the performance of SLAN trades in November 2019 to June 2021. Raw performance at each horizon is computed as explained in Section 2.4.

In Table 29 in the Appendix, we report the overall trade performance by month and day of the week. Retail investor losses are not concentrated in any particular month, while, at the same time, January and February 2021 are the worst months in our sample, corresponding to losses of \$681 and \$282 million, respectively. The same table reveals that, on average, performance seems to be lower on Thursdays and Fridays, potentially due to weekly options' expiry on Fridays. Table 30 in the Appendix also reports the top-10 and

impact of large trades. We winsorize trade prices, sizes, and spreads as in our earlier analysis at 99.50^{th} percentile each day.

bottom-10 tickers, based on performance of retail trades and the whole market. Similar to the latter, retail investors on on average, realized a gain on such large-cap names as Amazon (AMZN) and Apple (AAPL). Interestingly, however, in contrast to the market, they also lost on trading in 'meme' stocks, such as GameStop (GME) and AMC Entertainment (AMC).

Our analysis so far has not taken transaction costs into account. Some of the brokerages in our sample, such as Robinhood, offer commission-free options trading. However, the majority of brokerages still charge around \$0.65 per contract.[38] Using the fraction of PFOF in options paid to Robinhood as the upper bound of their share in the retail options trading, we can therefore estimate the aggregate *direct* transaction costs paid by retail investors. Using 1.59 million contracts as the aggregate SLAN volume and 25% as Robinhood's average share in PFOF for options, the direct transaction costs of retail trades in our sample period amount to $\$0.65 \times 1.59 \times 10^6 \times 0.75 \approx \$773$ million.

Importantly, we also evaluate *indirect* transaction costs at a contract level, aggregated across all the contracts. They are computed by summing up the products of effective half-spread and trade size across all SLAN trades in our sample, resulting in \$4.13 billion.[39] These costs are not as transparent as brokerage fees, and are likely to be overlooked by retail investors. Furthermore, they become revenue for market makers and exchanges executing retail orders (rather than for retail brokerages). These costs are economically large, being five times the direct costs of retail trading, and more than three times larger than the actual trading loss estimate in Table 4. Our calculation approach captures the actual gains and losses of retail trading, and does not require any assumptions regarding their opportunity costs. Finally, the overall magnitude of trading costs (relative to the raw trading performance of retail investors) is also consistent with the findings in Barber, Lee, Liu, and Odean (2008) on retail trading in stocks.

One limitation of our data is that some trades might come from multi-leg strategies involving options as well as underlying equities (e.g., a covered call), and we do not observe equity legs of these transactions. However, since the retail investor boom in our sample is largely driven by novice investors, we believe that only a small fraction of them use such sophisticated strategies. Therefore, it has little impact on our aggregate retail performance estimates.

[38] As of March 2022, TD Ameritrade, Charles Schwab, Interactive Brokers, E*TRADE, and Fidelity all charge \$0.65 per contract, according to their websites. Some of the brokers provide commission discounts for frequent traders or for large transactions. However, given the stylized features of retail trading highlighted in Table 1, these discounts are unlikely to have a material impact on our estimates.

[39] The corresponding figure for the whole market stands at \$39.41 billion.

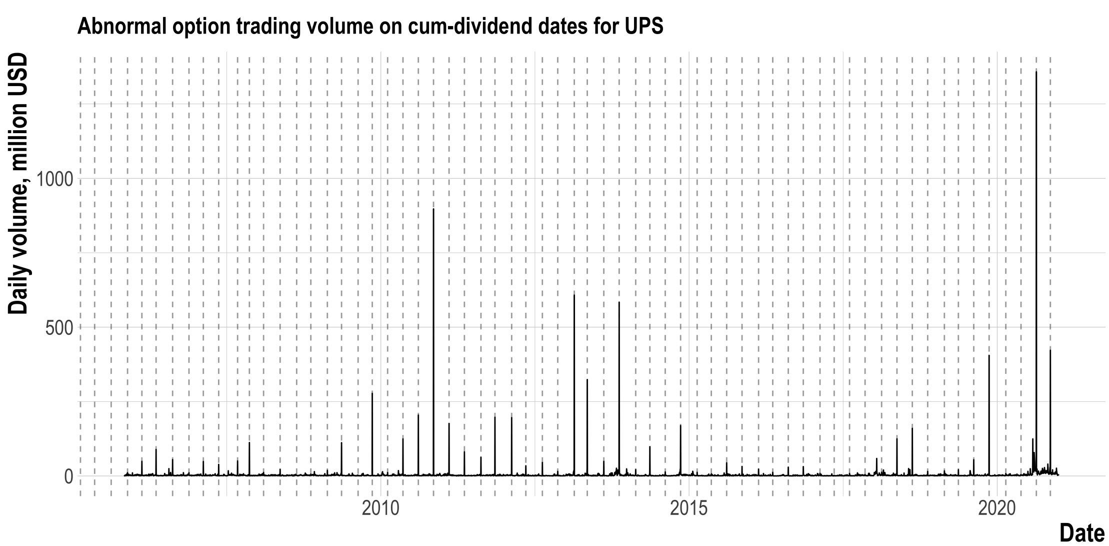
3 Retail investor presence and failure to exercise

Our aim is to study how the inflow of retail investors has affected the behavior of arbitrageurs in the options market. To this end, we focus on a particular arbitrage strategy, known as a dividend play, in which we can accurately identify trades of arbitrageurs. We present our measure of arbitrageur activity in the dividend play trade and discuss channels through which the inflow of retail investors has made this strategy more profitable for the arbitrageurs.

3.1 Resurgence of dividend play

Daily trading volume in options on high-dividend stocks in the U.S. exhibits an intriguing seasonality, illustrated in Figure 6 for the case of UPS. The spikes in trading volume apparent from the figure occur every quarter, on the last cum-dividend date, i.e., the day before UPS pays a dividend. The average daily traded notional for UPS is \$125.3 million on cum-dividend dates and only \$2.5 million on the remaining dates. This pattern is common for options on high dividend paying stocks; Appendix A.19 presents more examples.

Figure 6: Abnormal trading volume on cum-dividend dates for UPS



This figure plots daily trading volume for all call option contracts on UPS, in millions of U.S. dollars, as reported in OptionMetrics. The dashed lines indicate cum-dividend dates.

On cum-dividend dates, market makers (and other arbitrageurs) engage in an arbitrage trade known as the dividend play.[40] This strategy is available only for transactions

[40] Pool, Stoll, and Whaley (2008) and Hao, Kalay, and Mayhew (2009) also study this trade.

originating from the floor of the exchange,[41] or, in other words, only to the market participants who must be physically located on the trading floor. The strategy involves establishing long and short options positions that are so large that an operational error may potentially destabilize the market. Concerned about the impact of dividend play trades on the orderly functioning of the market, in 2014 the SEC issued a new rule designed to make the strategy impractical,[42] which resulted in much lower trading volumes on cum-dividend dates. However, the recent dramatic increase in options trading by inexperienced retail investors appears to have led to a resurgence of the strategy, despite the barriers created by the SEC rule.

The goal of the dividend play strategy is to take advantage of inattentive investors who fail to exercise their call options on dividend paying stocks when it is optimal to do so. It is optimal to exercise a call option if the value of exercising it on a cum-dividend date and collecting a dividend exceeds the value of the call the next day when the stock goes ex-dividend. Computing option values involves an application of the Black-Scholes-Merton formula or a more sophisticated option pricing method, which is typically difficult for novice retail investors. Alternatively, some retail investors may be unaware of the possibility of early exercise or are simply inattentive.[43] Since a fraction of in-the-money call options remains suboptimally unexercised, the writers of these options would not be asked to deliver the stock and would profit from this inattention. It is a zero-sum game.

If all in-the-money call option contracts on a stock have been exercised on the cum-dividend date, all holders of short positions in the same contracts receive a request to deliver the stock. If some contracts are left unexercised, however, the U.S. Options Clearing Corporation (OCC) randomly 'assigns' short positions that must deliver the stock. The unassigned holders simply hold on to their options and profit from a capital gain. Market makers (and other arbitrageurs) can divert this capital gain to themselves by simultaneously buying and selling a large number of in-the-money call options on the same ticker.[44] They exercise all long positions and deliver on all assigned short positions. Since some fraction remains (suboptimally) not assigned, they capture dividends on their net long stock positions while staying fully hedged. Usually, two arbitrageurs agree on a dividend play trade in advance

[41] In fact, dividend play could be organized off the exchange floor but it would then not qualify for transaction fee caps. In our data, most abnormal volume on cum-dividend dates goes through floor trades on two exchanges, PHLX and BOX, as we discuss below.

[42] See `https://www.sec.gov/rules/sro/occ/2014/34-73438.pdf`.

[43] There might be other reasons why investors do not exercise, such as costs of unwinding more complex strategies. Hao, Kalay, and Mayhew (2009) show that dividend play profits outweigh such costs in most cases.

[44] The current SEC rule, presented in footnote 42, prohibits simultaneous buying and selling of the same contract.

and serve as counterparty to each other on their arbitrage positions.

Table 5 illustrates the mechanics of the dividend play strategy by means of an example. Suppose there is 1 call option contract outstanding and it is optimal to exercise it.[45] Case 1 corresponds to the case when the option is exercised, the holder of the short position get assigned to deliver the underlying, and so there is no profit for a dividend play strategy to harvest. Case 2 describes what happens if the contract is left unexercised. Without arbitrageur involvement, the short position in the contract does not get assigned, and the option writer received a windfall gain of $500 for sure. Now consider the entry of a market maker. The market maker attempts to pocket most of the potentially harvestable profit of $500. To do so, the market maker buys and simultaneously sells 100 contracts and exercises all their long positions. The probability of assignment increases, but, because of the OCC's random assignment, with probability 100/101, the market marker holds the short position that does not get assigned and hence yields a gain. For the original option writer, this probability is now only 1/101. Hence, the expected gain of the market maker is $495 out of the total gain of $500 and that of the original option writer drops to $5. A dividend play strategy, therefore, dilutes the share of the gain that accrues to the original option writer.

Table 5: Dividend play: An Example

	OI_{t-1} (A)	New positions(t) (B)	Available for ex. (C)	No. exercised (D)	Prob. non-assign. orig. option writer (E)	Prob. non-assign. market maker (F)	Gain per share (G)	Expected gain orig. option writer (E*G*100)	Expected gain market maker (F*G*100)
Case 1. Optimal exercise									
Customer	1	0	1	1	0		5	0	
Case 2. Suboptimal exercise									
Case 2.1. Without dividend play									
Customer	1	0	1	0	1		5	**500**	
Case 2.2. With dividend play									
Customer	1	0	1	0	1/101		5	**5**	
Market maker	0	100	100	100		100/101	5		**495**
Total	1	100	101	100					

This table illustrates the dividend play strategy. Date t refers to the cum-dividend date and OI_t stands for the open interest on date t.

In the next section, we detect dividend play activity at a contract level in the full sample and characterize its importance relative to the overall trading volume on cum-dividend dates.

[45] Appendix A.20 provides another example, in which there are multiple contracts outstanding, some of which are exercised optimally and some are not.

3.2 Arbitrageur activity in dividend play strategy

We first present our measure of arbitrageur activity in the dividend play strategy. Through fee caps, exchanges incentivize cum-dividend day arbitrage strategies to originate from the physical floor. We therefore again exploit OPRA trade types to isolate option transactions that are executed on the floor. The trade types that cover most of the dividend play transactions are SLFT and MLFT, which are single-leg and multi-leg floor trades, respectively (see Appendix A.4 for a more detailed description). Other floor trade types, used infrequently in our sample are MLCT, MSFL, SLCN, TLFT, and TLFT. To our knowledge, this is the most precise measure of arbitrageur activity in the dividend play strategy in the literature, which typically uses trading volume on cum-dividend date in excess of the past average volume.

Table 6: Characteristics of activity on cum-dividend dates

	Average ticker dollar volume ($ million) on		**Total market dollar volume share (%) on**	
	cum-dividend date	any other date	cum-dividend date	any other date
	(1)	(2)	(3)	(4)
Panel A. Option type				
Call	24.9	1.5	92.8	54.3
Put	2.1	1.4	7.2	45.7
Panel B. Moneyness				
In the money	26.8	0.8	80.6	18.3
At the money	4.8	2.0	17.6	70.2
Out of the money	0.6	0.4	1.8	11.4
Panel C. Trade size				
Small	1.5	0.7	5.6	27.5
Large	31.0	2.8	94.4	72.5
Panel D. Floor trade				
Yes	48.2	0.9	76.9	6.5
No	6.1	2.4	23.1	93.5
Panel E. Exchange				
PHLX or BOX	24.6	0.5	79.7	15.0
Any other	5.5	2.3	20.3	85.0

This table compares option trading activity for dividend-paying tickers (2,153 stocks and ETFs) on cum-dividend date with any other date. The average volume in columns (1) and (2) is computed at ticker-day level, and the volume share in columns (3) and (4) is for the entire market. In Panel B, we define 'in the money' as $(Midpoint\,Price - Strike)/Strike > 0.1$ for call options and $(Midpoint\,Price - Strike)/Strike < -0.1$ for put options. 'At the money' are contracts for which this value is between -0.1 and 0.1, and 'out of the money' are all other contracts. In Panel C, we define trade as 'small' if the trade size is at or below 10 contracts. In Panel D, we define floor trades as trades with SLFT and MLFT OPRA trade types.

In our data, we see bursts of simultaneous buy and sell activity in neighboring-strike call option contracts, executed normally within several seconds, all coming from the floor. In an effort to reduce operationally risky dividend play trades, since 2014 the SEC forbids the market makers (and other arbitrageurs) to simultaneously buy and sell the same contract. Market participants have adjusted their trading strategies and they now simultaneously buy and sell neighboring contracts, which ultimately achieves the same objective. The trades are typically pre-arranged by pairs of market makers (or other arbitrageurs). We see no similar bursts of simultaneous buy and sell activity in call option contracts in any other OPRA trade types, which assures us that our measure very accurately captures arbitrageur activity in the dividend play strategy.

Table 6 presents some descriptive statistics of trading activity on cum-dividend vs. any other dates for dividend-paying stocks and ETFs. We see an enormous difference in floor trading volume and volume of large trades on cum-dividend dates relative to other dates. Moreover, on cum-dividend dates we see a colossal spike in volume on two exchanges that cap fees for the dividend play strategy: PHLX and BOX. Breaking the trades by moneyness, we see that the primary increase in volume comes from trading deep-in-the-money calls (that are more likely to be optimal to exercise). This pattern is a signature of the dividend play strategy. The sheer size of the dividend play positions is astonishing, especially after the SEC passed a rule intended to clamp down on this strategy.[46]

3.3 Failure to exercise and dividend play profits

In this section, we compute exploitable profits from a dividend play strategy. Some of these profits come from an increase in the open interest, some from investors' failure to exercise, and some from the value of early exercise of each contract. With an inflow of inexperienced investors in the options market, we expect the first two components to increase. We therefore find it useful to decompose the exploitable profit from a contract into three parts: the (i) open interest, (ii) fraction unexercised, and (iii) early exercise value.

The exploitable dividend play profit on all the interest for each contract is defined as

$$\pi_t = OI_{t-1} \times f_t \times EEV_t, \tag{3}$$

where $t-1$ is the day before the cum-dividend date, OI_{t-1} denotes open interest on that date (measured after all trades, exercises, and assignments on that date), $f_t \equiv OI_t/OI_{t-1}$ is the fraction unexercised, and EEV_t the early exercise value, computed below. Note that the fraction unexercised reflects the fraction of open interest in an option contract that remains

[46]See https://www.sec.gov/rules/sro/occ/2014/34-73438.pdf.

outstanding after the cum-dividend date (after all trades, exercises, and assignments on that date). Both EEV_t and f_t are estimated quantities. Open interest as of the day before the cum-dividend day (OI_{t-1}) and fraction not exercised (f_t) are available from OptionMetrics. In rational and frictionless markets, we expect $f_t = 0$ if $EEV > 0$.

The early exercise value is model based, and we rely on the Black-Scholes-Merton option pricing formula to compute it.[47] Denote the expected ex-dividend price of an option by c_{ex}, its strike by K, and the current (cum-dividend) underlying stock price by S. The expected option ex-dividend price represents the expected time value of the option. *Early exercise value (EEV)* is therefore the difference between the current stock price, strike, and this expected time value of the option: $S - K - c_{ex}$.[48] The details of the computation of c_{ex} are in Appendix A.21.

In the following analyses, we restrict our sample to call option contracts that are optimal to exercise on cum-dates and refer to it as the *dividend play sample*. Further details related to its construction are provided in Appendix A.22, and Table 32 in the Appendix presents the descriptive statistics for our dividend play sample.

How do retail trading trends relate to cum-dividend date exercise rates? To answer this question, we run the following regression.

$$Y_{c,t} = \beta_1 \times share^{SLAN}_{c,t} + \beta_2 \times share^{small}_{c,t} + \boldsymbol{\gamma}' X_{c,t} + \alpha_{i,t} + \varepsilon_{c,t} \tag{4}$$

where, for each contract c on cum-date t, we consider two dependent variables, $Y_{c,t}$: Fraction of open interest not exercised by ex-dividend date and potential profits from dividend play strategy as defined in Equation (3). $share^{SLAN}_{c,t}$ is the average dollar volume share in OPRA trade type SLAN over one trading week before the cum-dividend date t, and $share^{small}_{c,t-h}$ is the average dollar volume share of trades up to 10 contracts over over one trading week before the cum-dividend date t. In some specifications we also use ticker-level measures of retail investor popularity such as *Internalized volume in underlying*, which is the share of non-ATS OTC (i.e., internalized) volume in the total trading volume of the underlying stock or ETF i in the week of date t, and *WSB mentions, log*, the logarithm of the number of times ticker i was mentioned on `WallStreetBets` forum on date t. Our vector of controls $X_{c,t}$ includes the following contract-level variables: log OI, EEV, log dollar trading volume, relative spread,

[47] To make sure our results are robust to the choice of the underlying pricing model, we considered the sample of broad-index ETFs and computed their corresponding option prices with the Merton and Bates models, following Bakshi, Cao, and Chen (1997) and Cosma, Galluccio, Pederzoli, and Scaillet (2020). Options on these ETFs represent over 10% of contracts in our dividend play sample and 55% of potential dividend play profits. All our results go through in that sample and are available upon request.

[48] Note that this definition is from Pool, Stoll, and Whaley (2008) and it is equivalent to the definition in Hao, Kalay, and Mayhew (2009). The latter uses dividend instead: $Dividend - c_{ex} + S_{ex} - K$.

implied volatility, moneyness, days to expiry.[49] Our specification also includes the ticker by date fixed effects $\alpha_{i,t}$. Our measures of retail investor trading are computed over one trading week before the cum-date because the new generation of retail investors have a strong preference for options expiring within a week (see Section 2.3). Table 34 in Appendix A.24 presents an alternative specification in which we measure retail trading over two weeks preceeding a cum-dividend date.

Panel A of Table 7 reports the results of the regression in (4), with the fraction of open interest unexercised as the outcome variable. We find that there is a strong positive relationship between retail investor trading and the fraction of options that were suboptimally not exercised on the cum-dividend day. We measure retail investor trading in two different ways—by the share of volume executed by wholesalers (OPRA trade type SLAN) over the past week and by the share of small trades—and both variables come out as strong predictors of failures to exercise the option. A one standard deviation increase in the share of SLAN or small trades in the contract in the week preceding the cum-date raises the fraction unexercised by about one percentage point, depending on the specification. This result is robust and the magnitudes of the coefficients of interest do not change much as we relax the specification of fixed effects and switch on ticker-level controls instead (columns (1)–(3) and (5)-(6)).

Another measure of retail investor activity that strongly predicts the fraction left unexercised is the share of internalized volume in the total trade volume in the underlying stock or ETF, measured over the preceding trading week. We introduced this measure in Section 2.3. While this is a measure of retail investor trading in the underlying stock, it is correlated with retail investor trading in options on that stock. A call option is a leveraged position in a stock, and we hypothesize that some risk-loving retail investors may wish to trade options rather than the stock. We find that a one standard deviation increase in the internalized volume increases the fraction of options left unexercised by 1.6 percentage points (column (6)). Using another measure of retail trading discussed in Section 2.3, *WallStreetBets* count, we also find a positive and significant relationship between retail investor interest and the fraction unexercised.

One may argue that an alternative explanation for our findings is that the failures to exercise the options may be driven by transactions costs that make exercise impractical. To rule out this explanation, we restrict the sample to the top EEV tercile, the most profitable contracts to exercise (column (5)). We find that the size of the effect goes up significantly relative to our base case, implying that investor mistakes are a more likely driver of our

[49] Since log OI and EEV are components of potential dividend play profits, we do not include them in the specification in Panel B below.

findings. Another possible alternative explanation is that investors hold the call options in our sample as part of a sophisticated strategy, and exercising the option breaks one leg of the strategy. While this is possible and we do see mentions of a number of options strategies on `WallStreetBets`, we believe that the new generation of retail investors that drive our results are financial novices and relatively few of them engage in options strategies. Furthermore, to engage in such strategies, investors must qualify for a certain level of investment proficiency, required by investing platforms.

Table 7: Suboptimal exercise and retail investor popularity

	Dividend play profitability feature					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Fraction of OI not exercised, %						
SLAN Share	5.946***		5.791***	6.724***	5.792***	6.447***
	(3.48)		(3.37)	(2.66)	(3.38)	(3.58)
Small Share		4.438**	4.137**	10.365***	4.768**	4.569**
		(2.17)	(2.02)	(3.69)	(2.46)	(2.05)
Internalized volume in underlying						31.763*
						(1.88)
WSB mentions, log						0.528*
						(1.85)
Observations	21,105	21,105	21,105	6,942	21,105	19,134
Adjusted R-squared	0.237	0.237	0.238	0.345	0.214	0.213
Panel B. Potential profits, log U.S. dollar						
SLAN Share	1.465***		1.435***	1.701***	1.531***	1.564***
	(7.24)		(7.09)	(5.22)	(7.54)	(7.58)
Small Share		0.944***	0.873***	2.394***	0.887***	0.870***
		(3.27)	(3.08)	(4.09)	(3.27)	(3.04)
Internalized volume in underlying						2.991*
						(1.72)
WSB mentions, log						0.027
						(0.73)
Observations	21,105	21,105	21,105	6,942	21,105	19,134
Adjusted R-squared	0.307	0.304	0.308	0.314	0.286	0.292
Sample	All	All	All	Top EEV tercile	All	All
FE	Ticker*Date	Ticker*Date	Ticker*Date	Ticker*Date	Ticker and Date	Ticker and Date
Contract controls	Y	Y	Y	Y	Y	Y
Ticker controls	N	N	N	N	Y	Y

This table reports estimates of (4) in our dividend play sample. SLAN Share and Small Share are the contract-level volume shares of SLAN and small trades, respectively, averaged over one trading week before the cum-dividend date. Internalized volume in underlying is the share of non-ATS OTC (i.e., internalized) volume in the total trading volume in the underlying stock or ETF, averaged over one trading week before the cum-dividend date. WSB mentions, log, is the logarithm of total mentions of the ticker on *WallStreetBets* forum. In Panel B, contract controls include: log dollar trading volume, relative spread, IV, moneyness, days to expiry. In Panel A, they additionally include log OI and EEV. Ticker controls include: underlying price, underlying volatility, underlying relative bid-ask spread, underlying market cap. S.E. are clustered by ticker and date. Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Panel B of Table 7 considers the regression in (4), but with potential profits as the outcome variable. It reveals that the coefficients on all measures of retail activity are positive and significant: A one standard deviation increase in SLAN Share corresponds to around $4,200 higher profit *per contract*. In other words, the higher the retail activity in a contract in a week preceding the cum-dividend date, the more profitable it is for arbitrageurs to engage in a dividend play in the contract. Higher profits come from both (i) higher fraction unexercised (documented in Panel (a)) and (ii) higher open interest in the contracts popular with retail investors.

4 Money left on the table: A puzzle

In this section, we show that arbitrageurs engaging in the dividend play strategy leave money on the table by failing to capture arbitrage profits in some call option contracts. We explore the determinants of this puzzling behavior and present suggestive evidence that arbitrageurs may behave non-competitively.

4.1 Case study

November 11, 2020, was a cum-dividend date for UPS, a high-dividend paying stock, and a number of call options on UPS were deeply in-the-money and optimal to exercise on that day. Table 8 zooms in on a pair of such contracts, both expiring on November 20, 2020. We first compare the trading volume in the contracts on November 11, 2020. Notice that the

Table 8: Case study of arbitrageur activity: Two UPS call options on cum-dividend date

	Strike	EEV	OI (t-1)	Moneyness	Spread	Fraction not exercised	Cum-date volume	Floor share
Contract 1	160	0.29	1,945	3.15	0.045	0.76	45	0.000
Contract 2	155	0.43	2,487	4.62	0.039	0.47	3,255	0.998

trading volume in Contract 2 exceeds that in Contract 1 by two orders of magnitude. Notice also that Contract 2 has a very high share of orders from the trading floor on that day, while Contract 1 has zero. We also see characteristic bursts of floor orders in the transaction-level data for Contract 2. This means that market makers (or other arbitrageurs) engaging in a dividend play trade entered Contract 2 but not in Contract 1.

Why did the arbitrageurs leave money on the table in Contract 1? The contract had a high EEV and a large fraction unexercised. Using equation (3) to compute the arbitrageur's forgone profits from not entering Contract 1, we arrive at $1,945 \times 0.76 \times 0.29 \times 100 \approx 42,900$ dollars, a significant sum.[50]

Trading costs do not explain the market participants' reluctance to trade Contract 1. First, exchanges offer daily fee caps for the dividend play strategy, and so if market

[50]Each options contract in our sample is for 100 shares of the underlying stock or ETF.

makers (or other arbitrageurs) entered Contract 2, they should have also entered Contract 1. Second, contract bid-ask spreads in Table 4 are very similar. In the regression framework that follows, we further control for the options contract liquidity and show that trading costs do not explain why arbitrageurs forgo profitable opportunities.

It is very puzzling why arbitrageurs fully exploited the arbitrage opportunity in Contract 2 but not Contract 1. In the following section, we show that this pattern is general in our sample. The unexploited profit in Contract 1 accrued to the writer of this contract, which could be a market maker or perhaps a retail investor. The latter is less likely because retail brokerages take an automated action to close short positions that have dividend risk on behalf of their clients.[51] Appendix A.25 presents an excerpt from Robinhood's Terms and Conditions to provide an example of such automated action. It is therefore more likely that the writer of the contract who received the windfall gain was a market maker. The market maker who is a writer of the contract of course has no incentive to engage in a dividend play strategy in this contract because this would mean sacrificing own profit. But it is puzzling why other market makers or arbitrageurs would not wish to enter Contract 1 and reap arbitrage profits.

Table 33 in the Appendix generalizes this case study and reports forgone profits by ticker for the top-40 underlying stocks and ETFs sorted by the total size of forgone profits in our sample. We aggregate our data to and report the number of profitable individual contracts per ticker. The total amount of harvested profits in top-40 tickers in our sample is around \$51 million, whereas the total amount of forgone profit stands at \$67 million. For a virtually riskless arbitrage strategy, the amount of money left on the table is striking!

Furthermore, Table 33 does not reveal any particular pattern in harvested vs forgone profits: There is a large variation in arbitrageur participation across and within tickers. In what follows, we examine possible explanations for the puzzling reluctance of market participants to harvest arbitrage profits in some contracts.

4.2 Drivers of arbitrageur activity

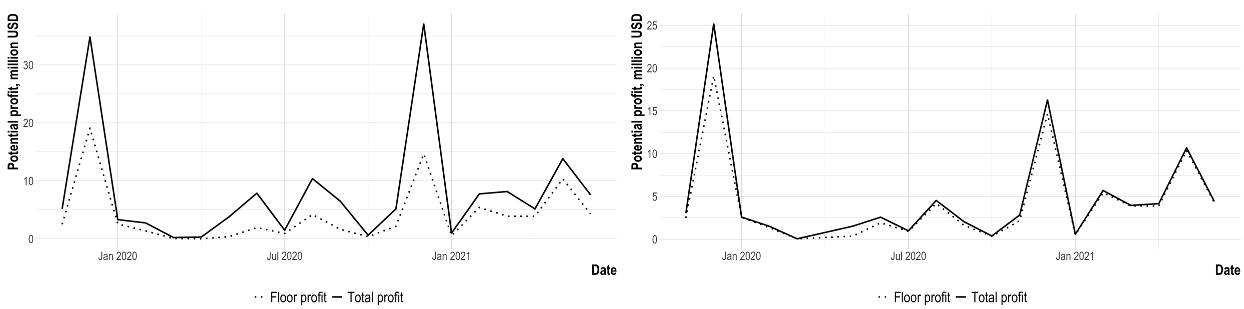
To examine drivers of arbitrage activity, we start by contrasting potential and harvested profits from the dividend play strategy on cum-dividend dates in our sample. Figure 7 presents potential profit of the dividend play strategy in all outstanding contracts, computed using equation (3), and profits harvested by floor traders. It emerges from panel (a), that a

[51] Since each options contract is for delivery of 100 shares of the underlying, for small retail investors the cash outlay needed for purchases of the underlying stock and delivering it could be quite significant. A brokerage would therefore close a short position if there are not enough funds in the account to buy and deliver the underlying.

large fraction of potential profit, about 50%, remains unharvested. If we restrict the sample, however, to the contracts with non-zero floor trading volume—i.e., contracts in which we detect dividend play activity—most of the potential profit resulting from the failure of investors to exercise their options on cum-dividend dates is harvested. In other words, market makers (and other arbitrageurs) selectively enter profitable contracts, capturing almost 100% of exploitable gains, but forgo arbitrage profits in contracts they do not enter.

Figure 7: Total and floor trader profit from dividend play strategy

(a) All profitable contracts (b) Contracts with non-zero floor share



This figure illustrates the implied share of potential dividend play profits captured by arbitrageurs on the trading floor. The solid plot is for the potential profit from the dividend play strategy and the dashed plot is for the profit harvested by floor traders (arbitrageurs).

The tendency of market makers (and other arbitrageurs) to leave money on the table in some profitable arbitrage opportunities is puzzling. In what follows, we try to understand the features on the contracts into which market participants are likely to enter.

The total exploitable profit is a calculated quantity, not known for sure on cum-dividend dates. The determinants of potential profit from a contract, however, are well-understood (see equation (3)), and projected fraction suboptimally unexercised is one of them. As we know from Table 7, this fraction is increasing in retail investor popularity. We therefore examine whether market makers (and other arbitrageurs) exploit increased investor inattention in contracts popular with retail investors. We estimate the following regression in the sample of contracts that should optimally be exercised on cum-date:

$$share_{c,t}^{floor} = \beta_1 \times share_{c,t}^{SLAN} + \beta_2 \times share_{c,t}^{Small} + \boldsymbol{\gamma}' X_{c,t} + \alpha_{i,t} + \varepsilon_{c,t}, \quad (5)$$

where the regressors are as in our previous specification (4) and the outcome variable is now the share of floor trades, which are predominantly market maker dividend play trades, in contract c on date t. Table 9 reports the results of the regression.

Table 9 reveals that arbitrageur activity, as measured by floor trading share, is pos-

Table 9: Arbitrageur activity and retail investor popularity

	Floor trading share on cum-date				**D(floor share > 0)**	**Floor trading volume, log**
	(1)	(2)	(3)	(4)	(5)	(6)
SLAN share	0.048***		0.037**	0.057*	0.036*	0.547***
	(2.59)		(2.11)	(1.87)	(1.89)	(3.43)
Small share		0.288***	0.286***	0.123*	0.301***	4.025***
		(6.41)	(6.39)	(1.66)	(6.53)	(8.53)
Observations	21,105	21,105	21,105	6,942	21,105	21,105
Adjusted R-squared	0.403	0.408	0.408	0.473	0.397	0.475
Sample	All	All	All	Top EEV tercile	All	All
Contract controls	Y	Y	Y	Y	Y	Y

This tables reports estimates of (5) in our dividend play sample. Floor trading share on cum-date is the contract-level volume share of trades executed on the traded floor in the total traded volume on the cum-dividend date. SLAN Share and Small Share are the contract-level volume shares of SLAN and small trades, respectively, averaged over one trading week before the cum-dividend date. Contract controls include: log OI, EEV, log dollar trading volume, relative spread, IV, moneyness, days to expiry. All regressions include ticker by date fixed effects. t-statistics are based on standard errors clustered by ticker and date (in parentheses). *** p<0.01, ** p<0.05, * p<0.1.

itively related to both measures of retail investor trading over the preceding week. This means that market makers (or other arbitrageurs) are aware of and do exploit suboptimal exercise strategies of retail investors by engaging in more dividend play trades in contracts that have experienced elevated retail investor activity. The effects of SLAN and Small Shares are statistically significant across all specifications. The magnitudes of the effect can be interpreted as follows. A one standard deviation increase in SLAN (Small) Share increases the share of floor trading by about 0.048*100*0.16=1 (0.288*100*0.18=5) percentage point(s). The magnitude is similar for the extensive margin (column (5)): A one standard deviation increase in SLAN (Small) Share increases the probability of floor entry by around a half (five) percentage point(s). If, instead of the floor trading volume share, we use floor trading volume in a contract on the cum-dividend date as an outcome variable, we again see strong effects of retail investor participation in the contract (column (6)). All of these effects are highly statistically significant.

If we restrict our sample to the most profitable contracts (column (4)), the relationship between retail trading and arbitrageur activity weakens and becomes marginally significant. This is surprising, given that our earlier analysis documents that retail investors do not exercise more profitable contracts more (column (4) in Table 7), and therefore market makers (or other arbitrageurs) seem to leave money on the table in more profitable dividend play trades.

To ascertain the robustness of our results, we pursue an alternative empirical strategy, based on propensity score matching. Such empirical exercise brings us closer to the case study explored in Table 8. Matching is a natural strategy in our setup because the set of characteristics on which one should match options to keep the expected profitability constant is well understood. We, again, study the relationship between floor trading share on cum-dividend date and retail popularity. However, here we isolate contracts with high retail popularity (top decile of SLAN share or Small share of 100%)[52] and construct the control group of contracts matched on profitability characteristics from contracts with low retail popularity. In the basic set of characteristics, we use open interest, early exercise value, and moneyness. We also report results with the characteristics extended to relative spread and underlying price. The corresponding covariate balance plots are presented in Appendix A.26.

Table 10: Arbitrageur activity and retail popularity: Matched contracts

	Floor trading share on cum-date							
		Matched		OLS		Matched		OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D(SLAN share top decile)	0.035***	0.039***	0.032***	0.019*				
	(2.82)	(3.67)	(3.10)	(1.79)				
D(Small share = 1)					0.033***	0.035***	0.042***	0.033***
					(3.99)	(5.03)	(6.19)	(4.82)
Observations	21,105	21,105	21,105	21,105	21,105	21,105	21,105	21,105
No. neighbors	1	10	10		1	10	10	
Short controls	Y	Y	Y	Y	Y	Y	Y	Y
Extended controls	N	N	Y	Y	N	N	Y	Y

This table reports the results of propensity score matching estimation and OLS estimates for the same set of contract characteristics in our dividend play sample. Columns (1)-(3) and (5)-(7) report ATE. SLAN and small share are the contract-level volume shares of SLAN and small trades, respectively, averaged over one trading week before the cum-dividend date. Short controls include: log OI, EEV, moneyness. Extended controls include relative spread and underlying price. Robust z-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 10 generalizes our case study in Table 8 to the sample of all matched contracts from our dividend play sample.[53] We again see that contracts that had experienced a larger volume of retail trading in the week preceding the cum-dividend date are more targeted by the arbitrageurs. The magnitudes are very similar to those in Table 9. We vary controls and numbers of neighbors across our specifications in Table 10, and the magnitudes of the coefficients do not vary much. The coefficients are also statistically indistinguishable from the OLS estimates from the same specification (columns (4) and (8)), which offers further

[52] Results are robust to the choice of percentiles. Small share of 100% corresponds to the 57^{th} percentile in our sample.

[53] In fact, we can use matching on the same profitability characteristics to study arbitrageur entry. In Appendix A.27, we show that across the propensity score spectrum, there exist contracts with both zero and positive floor volume (as we showed before, in the latter case floor traders represent almost 100% of trading so they seem to exhaust most of the potential profits). This result suggests that profitability characteristics do not predict entry very well, hence emphasizing the puzzle.

evidence that our results are robust.

4.3 Wholesaler concentration and limits to arbitrage

In Section 2.2, we have reached a striking conclusion that PFOF in options is highly concentrated. The share of the Big Three—Citadel, Susquehanna, and Wolverine—stands at nearly 85% in the second quarter of 2021. We unfortunately do not have trader identities in our options dataset. However, FINRA now provides a breakdown by firm of the non-ATS OTC (i.e., internalized) volume in underlying equities and ETFs.[54] These data are available at a weekly frequency. Since retail investor frenzies in equities and options tend to occur at the same time, we use the trading volume share in equities internalized by the Big Three as our proxy for the trading volume share is options. Moreover, two of the Big Three in options, Citadel and Susquehanna, also belong to the top-3 providers of PFOF in equities in our sample (together with Virtu).[55]

We first examine the Big Three's combined share in stock trading volume in our dividend play dataset. Figure 8 plots a histogram of this share in our dataset. It is striking that internalized volume in equities has become so high that we see stocks and ETFs in which the Big Three had a share in *total* trading volume of around 15% (panel (a)). Those stocks include many usual suspects from highly publicized retail investor frenzies (ticker and the mean Big Three share in brackets): Rocket Companies (RKT, 16%) and Nvidia Corporation (NVDA, 14%), Pfizer (PFE, 12%) and AstoZeneca (AZN, 11%), Metro-Goldwyn-Mayer Studios (MGM, 16%) and AMC Entertainment Holdings (AMC, 11%), and even Apple (AAPL, 13%) and Microsoft (MSFT, 11%). The Big Three's share in the *OTC volume* is very high (panel (b)), with the sample median of 33%. For the names we just listed, this share ranges between 42 and 57%. It seems that the market for OTC transactions in equities has become very concentrated in the recent past, during which the overall trading volume, as well as retail investor trading, has increased substantially. Given that the FINRA data we use is in the public domain, market participants must be well aware of the fact that they are operating in a highly concentrated market.

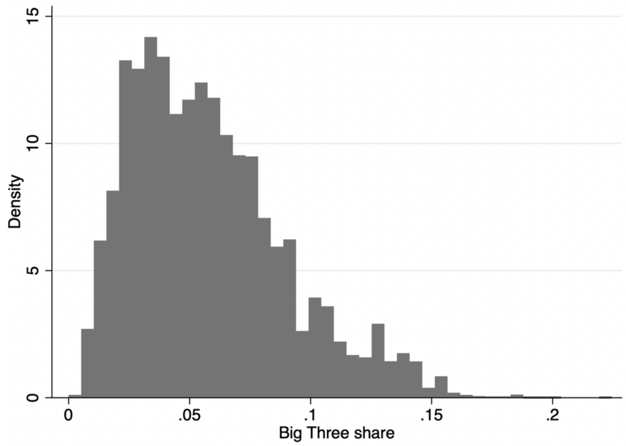
Why could market concentration matter for the surprising reluctance of market participnts to engage in dividend play in certain contracts? One potential hypothesis is that a large market maker, to which retail orders to buy call options are routed by retail investing platforms, serves as a counterparty to these transactions and writes these options. Then this market maker is the one to receive windfall gains from investors' failures to exercise these options. There is no incentive to this market maker to engage in dividend play in contracts

[54] We summarize it in Appendix A.2.

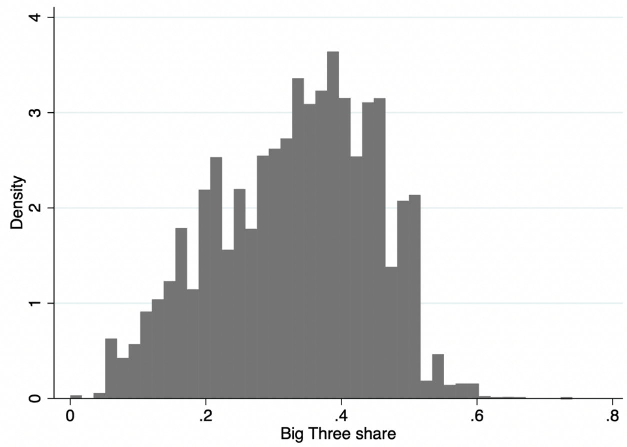
[55] See Appendix A.1 for a detailed description.

Figure 8: The Big Three: Share of trading volume in underlying equities and ETFs

(a) Share of total volume



(b) Share of non-ATS OTC (internalized) volume



This figure plots histograms of weekly non-ATS OTC volume share in underlying equities and ETFs by the Big Three PFOF providers in options: Citadel, Susquehanna, and Wolverine. Panel (a) uses share in total trading volume (on lit exchanges, ATS and non-ATS OTC), panel (b) uses share in total non-ATS OTC volume. Based on FINRA OTC Transparency and CRSP data for the underlying securities in our dividend play sample.

in which it is short. But this does not explain why *other* arbitrageurs do not wish to trade these contracts and divert windfall gains from the market maker holding a short position in them.

Table 11 attempts to shed some light on the economic mechanism behind the puzzling decision of arbitrageurs to leave money on the table. We capture arbitrageur concentration in two different ways: (i) using the Big Three internalizers' non-ATS OTC volume share in the underlying stock, a continuous variable and (in Panel A); and (ii) a dummy indicating that the Big Three share is in the top decile (in Panel B). It is likely that the effect of arbitrageur concentration is highly non-linear, which is why we have included Panel B.

Columns (1) and (2) demonstrate an important role of SLAN order imbalance—i.e., more buy relative to sell orders from retail investors and vice versa—in the week preceding the cum-dividend day. SLAN order imbalance has a negative effect on arbitrageurs' decision to engage in a dividend play strategy. Recent SLAN imbalance in a contract indicates that wholesalers are likely to have a (long or short) position in that contract. This feature of the contract appears to reduce the appetite for arbitrage in that contract.

The Big Three share has a negative effect on arbitrageur entry in a contract, although it is significant only when the Big Three share is high (Panel B). That is, market makers (and other arbitrageurs) seem more reluctant to enter a contract that has experienced exceptionally high trading volume in the underlying that was internalized by the Big

Three.

Table 11: Arbitrageur activity and market concentration

	Floor trading share on cum-date					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Continuous Big Three share						
D(SLAN imbalance)	-0.120***			-0.093***		
	(-13.23)			(-5.97)		
D(SLAN buy imbalance)		-0.089***			-0.061***	
		(-7.68)			(-2.92)	
Big Three share			-0.814	-0.665	-0.734	-0.688
			(-1.50)	(-1.21)	(-1.34)	(-1.27)
Big Three share × D(SLAN trade imbalance)				-0.428*		
				(-1.96)		
Big Three share × D(SLAN buy trade imbalance)					-0.407*	
					(-1.69)	
Big Three share × Log change in OI						-0.611**
						(-2.24)
Log change in OI						-0.073***
						(-3.97)
Observations	21,105	21,105	20,682	20,682	20,682	20,677
Adjusted R-squared	0.417	0.411	0.370	0.379	0.373	0.375
Panel B. Big Three share in the top decile						
D(SLAN imbalance)				-0.111***		
				(-12.52)		
D(SLAN buy imbalance)					-0.080***	
					(-6.56)	
D(Big Three share > 10%)			-0.049*	-0.034	-0.044	-0.045
			(-1.69)	(-1.12)	(-1.51)	(-1.53)
D(Big Three share > 10%) × D(SLAN trade imbalance)				-0.051***		
				(-2.66)		
D(Big Three share > 10%) × D(SLAN buy trade imbalance)					-0.043**	
					(-2.35)	
D(Big Three share > 10%) × Log change in OI						-0.081***
						(-3.13)
Log change in OI						-0.096***
						(-7.99)
Observations			21,105	21,105	21,105	21,100
Adjusted R-squared			0.372	0.380	0.374	0.377
FE	Ticker*Date	Ticker*Date	Ticker and Date	Ticker and Date	Ticker and Date	Ticker and Date
Contract controls	Y	Y	Y	Y	Y	Y
Ticker controls	N	N	Y	Y	Y	Y

This table further explains floor trader activity in our dividend play sample. Big Three share is the total share of the Big Three internalizers' non-ATS OTC volume in the total stock trading volume over the past trading week. $D(SLAN\,imbalance) = 1$ if there was an order imbalance in SLAN trades over the past trading week. $D(SLAN\,buy\,imbalance) = 1$ if there was a positive order imbalance in SLAN trades over the past trading week. 'Log change in OI' is the log ratio of cum-dividend date open interest to the average open interest in the past trading week. Contract controls include: SLAN volume share, Small volume share, log OI, EEV, log dollar trading volume, relative spread, IV, moneyness, days to expiry. Ticker controls include: underlying price, underlying volatility, underlying relative bid-ask spread, underlying market cap. t-statistics are based on standard errors clustered by ticker and date (in parentheses). *** p<0.01, ** p<0.05, * p<0.1

Some of the strongest effects documented in Table 11 are for the interactions of the Big Three share with SLAN buy imbalances, as well as with log changes in the open interest, and the signs on the interactions are all negative. A possible interpretation of these results is as follows. Due to a large imbalance in options,[56] the Big Three have written a large number

[56] See e.g., Barber, Huang, Odean, and Schwartz (2022) document such frenzies during our sample period.

of call options on that stock in the preceding week and are holding a large short position in these calls in their inventory on a cum-dividend date. They are then set to benefit from retail investor suboptimal option exercise behavior, documented in Section 3.1, and collect a windfall gain. The entry of other market participants in these contracts would dilute the share of the windfall gain the Big Three are set to receive. All arbitrageurs who are able to engage in the dividend play strategy in the contract are located physically on the exchange floor (of primarily PHLX and also BOX options exchanges). They are reluctant to openly trade against the Big Three, and so competition in the dividend play strategy breaks down and the Big Three benefit disproportionally from this.

For robustness, we use matching as an alternative approach for studying the effects of arbitrageur concentration in the dividend play strategy. The set of characteristics on which we match option contracts is the same as in Table 10. The results of this analysis are reported in Appendix A.29. They are fully consistent with the results presented in Table 11.

Finally, we attempt to quantify the number of arbitrageurs simultaneously engaging in a dividend play strategy in a particular contract and show that this number is typically quite small. A signature of the dividend play strategy is the bursts (several trades within milliseconds of one another) of simultaneous buy and sell activity in neighboring-strike call option contracts, originating from the floor. Sizes of trades within each burst are always the same in our sample, but they differ across bursts. We use the number of unique trade sizes as a proxy for the number of arbitrageurs engaging in a dividend play trade in a call option contract. Figure 17 in the Appendix plots a percentage split of dividend play trades by unique trade sizes. The figure reveals that the most common trade size by far is one, which means that only one arbitrageur enters the (long side) of the contract. We also observe two or three unique trade sizes, but the occurrence of trade sizes higher than three is fairly rare. This provides suggestive evidence that the number of arbitrageurs entering a dividend play strategy in a given contract is very low, which could be an additional reason why they behave non-competitively.

4.4 Alternative explanations

First, there exist dividend-play specific fee caps on Philadelphia and, more recently, Boston options exchanges (PHLX and BOX.)[57] Those fee caps limit the total costs paid by the market maker on a particular day at the options class level: Harvesting the profit from an additional contract would not increase payments to the exchanges once the limit is hit. Second, given that dividend play usually requires two participating parties, it is highly

[57]See PHLX pricing schedule: https://listingcenter.nasdaq.com/rulebook/phlx/rules/phlx-options-7 and BOX fee schedule: https://boxoptions.com/regulatory/fee-schedule/.

likely they agree on the transaction price that allows for mutually beneficial profit sharing. There is no clear reason why they would omit any particular contract from their agreement due to its otherwise lower liquidity. Finally, in the above analysis, we always control for contract liquidity or match on contract relative spread. It is therefore unlikely that the contracts in which market makers (or other arbitrageurs) do not engage in dividend play are systematically less liquid.

One might be concerned that our choice to exclude trades at exchange open and close might affect the results. To rule this out, consider the following thought experiment. Assume that all the difference in volumes between our sample and the total daily volume (available through OptionMetrics) is due to dividend play activity. On can then compute all the corresponding profits from the arbitrage strategy. Even under this fairly unrealistic assumption, the harvested profit share increases by only 5 percentage points (from 50% to 55%). Therefore, omitting trades at the open and close cannot explain our findings about the money arbitrageurs leave on the table.

Another potential explanation is that arbitrageurs' capital constraints bind. However, most regulatory requirements typically involve netted positions, which are relatively low given the symmetric and fully hedged nature of the strategy. So it is not clear why capital constraints may bind unless they bind due to the arbitrageurs' internal risk management guidelines. Relatedly, such large trades are associated with high operational risks. According to SIFMA, Bank of America Merrill Lynch incurred a $10 million loss due to a human error when executing the dividend play strategy.[58] Still, such explanations cannot produce the variation in floor trader activity within and across tickers that we document: Table 33 illustrates that there are many profitable contracts in which floor traders do not enter at all.

Finally, it has been documented that even sophisticated market players exhibit limits to attention (Kacperczyk, Nieuwerburgh, and Veldkamp (2016)). Indeed, there may be hundreds of potentially profitable contracts available to dividend play on each cum-dividend day (thousands in case of SPY). Perhaps, traders simply cannot evaluate all relevant pricing parameters, enter into an agreement with each other, and process the necessary number of trades? First, it is not clear why other exchange members do not enter to reap arbitrage profits if such limits exist. However, we went on to test this hypothesis more formally. To do so, we used the number of stock-level EPS (Hirshleifer, Lim, and Teoh (2009)) and macroeconomic announcements (Savor and Wilson (2014)) as proxies for limits to attention and did not find that those mattered for floor trader activity. These additional results are available upon request.

One alternative explanation that we cannot rule out is that some profits are left

[58]See https://www.reuters.com/article/us-usa-options-apple-idUSKBN0IQ2FA20141106.

unexploited because of the stigma and reputational costs associated with the dividend play strategy. The SEC has clearly signaled its disapproval of the strategy in its 2014 Rule aimed at making the strategy impractical (see footnote 42). Reputational costs could explain the lack of entry of new arbitrageurs. However, they cannot explain why arbitrageurs who regularly engage in this strategy, and hence are willing to incur reputation costs, still leave money on the table.

5 Discussion and policy implications

Our paper calls for more transparency in reporting wholesaler activities in the options market, consistent with the current practice by FINRA in equities. In particular, it would be useful to know how often market makers affiliated with wholesalers get order allocations through price improvement auctions. The current highly concentrated market significantly favors leading wholesalers and calls into question the extent of price improvement of retail orders. One particularly fruitful avenue for future research is uncovering the barriers to entry in this market and developing a structural framework to analyze the effects of market power on prices and welfare.

We would not be the first ones calling for more transparency in trading costs in zero-commission offers of retail brokerages.[59] However, most prior calls were related to equities. Trading costs in options are orders of magnitude higher, and so a regulatory requirement to disclose these costs to investors would be a welcome first step.

It is apparent from our analysis that the new generation of investors, while tech-savvy and active on investing forums, is still lacking in financial education that is required to trade options. For example, retail investors trade options frequently, opting for options with very short maturities. This behavior creates significant trading costs, which are masked by zero-commission option trading offers by investing platforms (e.g., Robinhood). Another example of mistakes is that retail investors fail to exercise their options when it is optimal to do so.

It is not clear whether retail investing platforms have the right incentives to prevent their customers from making trading mistakes. Frequent trading produces large order flow and hence large revenue from PFOF for retail investing platforms. Trading less liquid assets, such as options, enhances these profits further. Investor option exercise mistakes also

[59]Regulators have long been interested in various aspects of the system of payment for order flow, and in particular, whether internalization of orders really provides price improvement for the clients. In 2017 SEC found that some of the algorithms used by Citadel Securities to route retails orders, did not seek to obtain the best price on the marketplace, leading to a settlement fee of $22.6 mln (see https://www.sec.gov/news/pressrelease/2017-11.html). Furthermore, in June 2021 Gary Gensler, chair of the SEC, announced an upcoming comprehensive review of the current microstructure rules, including the system of payment for order flow, see https://www.ft.com/content/83dff8fc-14ac-4e67-a969-20b358c349e8.

generate additional revenue for market makers, via e.g., dividend play trades, which in turn may lead to a larger PFOF for retail brokerages. The question of optimal options exercise requires knowledge of option pricing models, which retail investors are likely to be lacking. One possibility would be to require retail brokerages to report options' early exercise values to investors. The early exercise value could be computed from the Black-Scholes model. Another possibility is to make *automatic* early exercise on cum-dividend dates when it is optimal to do so a default option for investors, from which they can opt out if they wish.

6 Conclusion

This paper focuses on the recent boom in retail investor trading in options. The new generation of retail investors are young and tech-savvy, yet amateur investors. Exploiting a new OPRA reporting requirement, we develop a novel measure of retail investor trading in options and document a rapid rise in retail investor trading in our sample. We argue that retail investors enter the options market for speculative reasons. Lured by recent low- or zero-commission offers, they prefer options with very short maturities, primarily calls. These options have a 60% higher bid-ask spread than the average in the options market, making the options business a very lucrative one for wholesalers which execute retail order flow. The ballooning PFOF for options received by retail brokerages is an indication of a strong wholesaler interest in executing retail order flow in options.

PFOF for options is very concentrated, with the share of the top-3 providers exceeding 85% toward the end of our sample. We are interested in exploring whether such a high concentration of wholesaler volume affects arbitrage activity in the options market. To this end, we isolate one specific arbitrage strategy executed predominantly by market makers: the dividend play. This strategy produces (virtually) riskless arbitrage profits for market makers. However, the entry of additional arbitrageurs dilutes the incumbent market makers' share of the gain from the strategy. We document that market makers do not engage in dividend play in some contracts, forgoing large profits. Our exploration of the reasons behind this puzzling behavior points to the unwillingness of market makers to dilute the share of gains received by the Big Three PFOF providers. This hypothesis requires further investigation. If indeed market concentration has become so high that market makers choose to behave non-competitively, this has significant implications for investor protection.

References

Bakshi, Gurdip, Charles Cao, and Zhiwu Chen, 1997, Empirical performance of alternative option pricing models, *The Journal of Finance* 52, 2003–2049.

Barber, Brad M., Xing Huang, Terrance Odean, and Chris Schwartz, 2022, Attention-Induced Trading and Returns: Evidence from Robinhood Users, *The Journal of Finance* forthcoming.

Barber, Brad M., Yi-Tsung Lee, Yu-Jane Liu, and Terrance Odean, 2008, Just How Much Do Individual Investors Lose by Trading?, *The Review of Financial Studies* 22, 609–632.

Barber, Brad M., and Terrance Odean, 2001, Boys will be Boys: Gender, Overconfidence, and Common Stock Investment, *Quarterly Journal of Economics* 116, 261–292.

Barbu, Alexandru, 2022, Ex-Post Loss Sharing In Consumer Financial Markets, working paper.

Barraclough, Kathryn, and Robert E. Whaley, 2012, Early Exercise of Put Options on Stocks, *The Journal of Finance* 67, 1423–1456.

Battalio, Robert, Stephen Figlewski, and Robert Neal, 2020, Option Investor Rationality Revisited: The Role of Exercise Boundary Violations, *Financial Analyst Journal* 76, 82–99.

Battalio, Robert, Todd Griffith, and Robert Van Ness, 2021, Do (Should) Brokers Route Limit Orders to Options Exchanges That Purchase Order Flow?, *Journal of Financial and Quantitative Analysis* 56, 183–211.

Battalio, Robert, Andriy Shkilko, and Robert Van Ness, 2016, To Pay or Be Paid? The Impact of Taker Fees and Order Flow Inducements on Trading Costs in U.S. Options Markets, *Journal of Financial and Quantitative Analysis* 51, 1637–1662.

Bauer, Rob, Mathijs Cosemans, and Piet Eichholtz, 2009, Option trading and individual investor performance, *Journal of Banking & Finance* 33, 731–746.

Bhattacharya, Vivek, Gaston Illanes, and Manisha Padi, 2019, Fiduciary Duty and the Market for Financial Advice, working paper, available at SSRN: https://ssrn.com/abstract=3281345.

Boehmer, Ekkehart, Charles M. Jones, Xiaoyan Zhang, and Xinran Zhang, 2021, Tracking Retail Investor Activity, *The Journal of Finance* 76, 2249–2305.

Campbell, John Y, Howell E Jackson, Brigitte C Madrian, and Peter Tufano, 2011, Consumer Financial Protection, *Journal of Economic Perspectives* 25, 91–114.

Célérier, Claire, and Boris Vallée, 2017, Catering to Investors Through Security Design: Headline Rate and Complexity, *The Quarterly Journal of Economics* 132, 1469–1508.

Chapkovski, Philipp, Mariana Khapko, and Marius Zoican, 2021, Does Gamified Trading Stimulate Risk Taking?, working paper, available at SSRN: https://ssrn.com/abstract_id=3971868.

Christie, William G., and Paul H. Schultz, 1994, Why do NASDAQ Market Makers Avoid Odd-Eighth Quotes?, *The Journal of Finance* 49, 1813–1840.

Christoffersen, Peter, Ruslan Goyenko, Kris Jacobs, and Mehdi Karoui, 2018, Illiquidity Premia in the Equity Options Market, *The Review of Financial Studies* 31, 811–851.

Cosma, Antonio, Stefano Galluccio, Paola Pederzoli, and Olivier Scaillet, 2020, Early Exercise Decision in American Options with Dividends, Stochastic Volatility, and Jumps, *Journal of Financial and Quantitative Analysis* 55, 331–356.

Eaton, Gregory W., T. Clifton Green, Brian Roseman, and Yanbin Wu, 2021, Zero-Commission Individual Investors, High Frequency Traders, and Stock Market Quality, working paper, available at SSRN: https://ssrn.com/abstract_id=3776874.

Egan, Mark, 2019, Brokers versus Retail Investors: Conflicting Interests and Dominated Products, *The Journal of Finance* 74, 1217–1260.

Fedyk, Valeria, 2021, This Time is Different: Investing in the Age of Robinhood, working paper.

Hagströmer, Björn, 2021, Bias in the Effective Bid-Ask Spread, *Journal of Financial Economics* 142, 314–337.

Hao, Jia, Avner Kalay, and Stewart Mayhew, 2009, Ex-dividend Arbitrage in Option Markets, *The Review of Financial Studies* 23, 271–303.

Hirshleifer, David, Sonya Seongyeon Lim, and Siew Hong Teoh, 2009, Driven to Distraction: Extraneous Events and Underreaction to Earnings News, *The Journal of Finance* 64, 2289–2325.

Jensen, Mads V., and Lasse H. Pedersen, 2016, Early Option Exercise: Never Say Never, *Journal of Financial Economics* 121, 278–299.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2016, A Rational Theory of Mutual Funds' Attention Allocation, *Econometrica* 84, 571–626.

Lakonishok, Josef, Inmoo Lee, Neil D. Pearson, and Allen M. Poteshman, 2006, Option Market Activity, *The Review of Financial Studies* 20, 813–857.

Lee, Charles M.C., and Balkrishna Radhakrishna, 2000, Inferring investor behavior: Evidence from TORQ data, *Journal of Financial Markets* 3, 83–111.

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

Mayhew, Stewart, 2002, Competition, Market Structure, and Bid-Ask Spreads in Stock Option Markets, *The Journal of Finance* 57, 931–958.

Muravyev, Dmitriy, 2016, Order Flow and Expected Option Returns, *The Journal of Finance*

71, 673–708.

——— , and Neil D Pearson, 2020, Options Trading Costs Are Lower than You Think, *The Review of Financial Studies* 33, 4973–5014.

Pool, Veronika Krepely, Hans R. Stoll, and Robert E. Whaley, 2008, Failure to Exercise Call Options: An Anomaly and a Trading Game, *Journal of Financial Markets* 11, 1–35.

Poteshman, Allen M., and Vitaly Serbin, 2003, Clearly Irrational Financial Market Behavior: Evidence from the Early Exercise of Exchange Traded Stock Options, *The Journal of Finance* 58, 37–70.

Ramachandran, Lakshmi S., and Jitendra Tayal, 2021, Mispricing, Short-sale Constraints, and the Cross-Section of Option Returns, *Journal of Financial Economics* 141, 297–321.

Savickas, Robert, and Arthur J. Wilson, 2003, On inferring the direction of option trades, *The Journal of Financial and Quantitative Analysis* 38, 881.

Savor, Pavel, and Mungo Wilson, 2014, Asset Pricing: A Tale of Two Days, *Journal of Financial Economics* 113, 171–201.

Welch, Ivo, 2022, The Wisdom of the Robinhood Crowd, *The Journal of Finance* forthcoming.

A Appendix

A.1 Payment for order flow, by broker and firm

Table 12: Payment for order flow - Data description

Firms	Broker													**Total paid, $ mln.**	**Total paid, %**
	TD Ameritrade	Robinhood	E*TRADE	Charles Schwab	Webull	Fidelity	tastyworks	Tradestation	Apex	Ally	Interactive Brokers	Bank of America	Wealthfront		
Panel A: Stocks															
CITADEL	329.2	175.2	90.0	54.4	46.5	0.0	0.7	8.5	8.6	4.0	3.2		0.0	**720.3**	37.7
SUSQUEHANNA	92.9	63.8	54.4	32.3		0.0	0.4	0.0	3.6	2.4				**249.8**	13.1
VIRTU	243.4	110.2	80.1	45.3	4.3	-0.3		15.7	7.4	2.6	1.9		0.0	**510.6**	26.7
WOLVERINE		29.3	0.0				0.0			0.1				**29.4**	1.5
DASH					0.0				0.0					**0.0**	0.0
MORGAN STANLEY						-0.4								**-0.4**	0.0
TWO SIGMA	58.7	53.6	9.8	6.2	3.8	0.0		4.3	0.2					**136.6**	7.2
NASDAQ			5.5	0.8	0.1	33.2		0.0		0.0	37.1			**76.7**	4.0
UBS	64.5		10.2	24.0		-0.1		3.7			0.0			**102.3**	5.4
CBOE			10.5	0.3		38.0		0.7			-13.8			**35.7**	1.9
OTHER	0.3		0.7		22.3	-0.7		9.8	11.2	0.0	5.1			**48.7**	2.6
Total received, $ mln.	**789.0**	**432.1**	**261.2**	**163.3**	**77.0**	**69.7**	**1.1**	**42.7**	**31.0**	**9.1**	**33.5**	**0.0**	**0.0**	**1909.7**	
Total received, %	41.3	22.6	13.7	8.6	4.0	3.6	0.1	2.2	1.6	0.5	1.8	0.0	0.0		
Panel B: Options															
CITADEL	511.2	370.6	137	66.4	38.9	67.5	31.8	7.7	1.6	5.2		0		**1237.9**	41.0
SUSQUEHANNA	378.7	224.5	101.7	64.3	23.3	25.8	13.7	0.5	3.8	5.3				**841.6**	27.9
VIRTU														**0.0**	0.0
WOLVERINE	83	154.7	41.3	43.8	4.2	31.4	0	7.1	1.8	3.4				**370.7**	12.3
DASH	76.9		60.6	17.1	12.3	11.1	20.7	7.5	3.8			0		**210.0**	7.0
MORGAN STANLEY	66.1	52.9	36.9	20.7		7.1		7.1						**190.8**	6.3
TWO SIGMA			5.1		0.3				0					**5.4**	0.2
NASDAQ						0					38.7	2.6		**41.3**	1.4
UBS														**0.0**	0.0
CBOE						0					47.3	4.7		**52.0**	1.7
OTHER	2.1			3.4	0	6.6	7.6		2.3	0	39.3	5		**66.3**	2.2
Total received, $ mln.	**1118.0**	**802.7**	**382.6**	**215.7**	**79.0**	**149.5**	**73.8**	**29.9**	**13.3**	**13.9**	**125.3**	**12.3**	**0.0**	**3016.0**	
Total received, %	37.1	26.6	12.7	7.2	2.6	5.0	2.4	1.0	0.4	0.5	4.2	0.4	0.0		

This table reports the total payment for order flow in stocks (Panel A) and options (Panel B) for each broker-firm pair in Q1/2020-Q2/2021. Relationships with missing values do not exist. PFOFs with zero values are rounded to zero. All data is from SEC Rule 606 reports.

A.2 OTC trading volume, by venue

Table 13: Top-15 internalizers in the U.S.

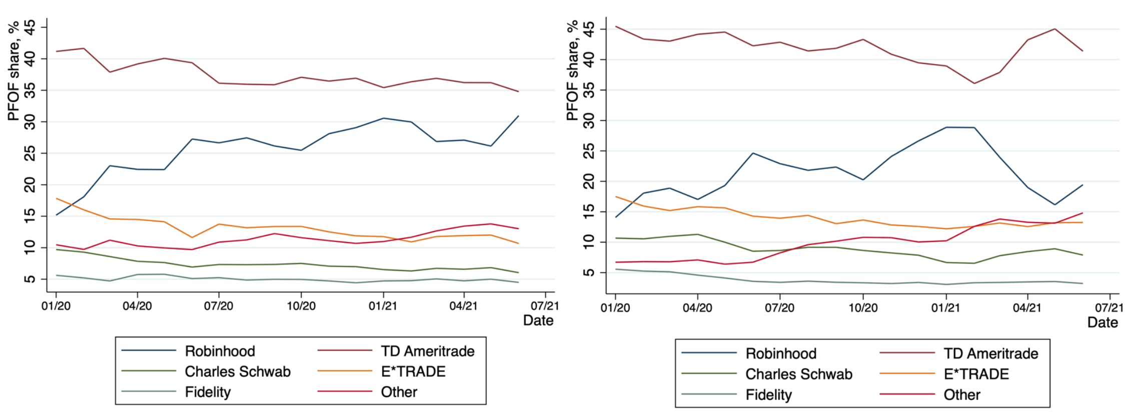
Firm	OTC volume, billion shares	Venue share in total volume, %	Cumulative share, %
CITADEL SECURITIES	477.82	44.31	44.31
VIRTU	357.61	33.16	77.47
SUSQUEHANNA	119.10	11.04	88.52
TWO SIGMA	48.50	4.50	93.01
JANE STREET CAPITAL	28.49	2.64	95.66
UBS	25.35	2.35	98.01
WOLVERINE	7.29	0.68	98.68
COMHAR CAPITAL MARKETS	3.84	0.36	99.04
HRT EXECUTION SERVICES	3.46	0.32	99.36
LEK SECURITIES CORPORATION	2.27	0.21	99.57
GOLDMAN	2.20	0.20	99.77
ACS EXECUTION SERVICES	0.44	0.04	99.81
IMC	0.32	0.03	99.84
MORGAN STANLEY	0.29	0.03	99.87
COWEN	0.28	0.03	99.90

This table reports the top-15 firms in terms of their total OTC non-ATS (i.e., internalized) volume in 11/2019-06/2021. Based on FINRA OTC Transparency data.

A.3 PFOF trends, by broker

Figure 9: Share in the payment for order flow: Options vs stocks, by brokerage

(a) Options (b) Stocks



This figure plots the share in monthly payments for order flow of the largest U.S. retail brokerages.

A.4 OPRA trade types

The table below presents OPRA trade types, together with their descriptions, implemented on November 4, 2019. We also include the corresponding Trade Condition IDs from LiveVol, our data provider.

Table 14: OPRA trade types for transactions in U.S. options exchanges

OPRA Type Description	**OPRA Message Type**	**LiveVol Trade Condition ID**	**OPRA Condition Description**
AUTO		18	Transaction was executed electronically. Prefix appears solely for information; process as a regular transaction.
CANC		40	Transaction previously reported (other than as the last or opening report for the particular option contract) is now to be cancelled.
CBMO	Multi Leg Floor Trade of Proprietary Products	133	Transaction represents execution of a proprietary product non-electronic multi leg order with at least 3 legs. The trade price may be outside the current NBBO.
CNCL		41	Transaction is the last reported for the particular option contract and is now cancelled.
CNCO		42	Transaction was the first one (opening) reported this day for the particular option contract. Although later transactions have been reported, this transaction is now to be cancelled.
CNOL		43	Transaction was the only one reported this day for the particular option contract and is now to be cancelled.
ISOI		95	Transaction was the execution of an order identified as an Intermarket Sweep Order. Process like normal transaction.
LATE		13	Transaction is being reported late, but is in the correct sequence; i.e., no later transactions have been reported for the particular option contract.
MASL	Multi Leg Auction against single leg(s)	125	Transaction was the execution of an electronic multi leg order which was "stopped" at a price and traded in a two sided auction mechanism that goes through an exposure period and trades against single leg orders/ quotes. Such auctions mechanisms include and not limited to Price Improvement, Facilitation or Solicitation Mechanism.
MESL	Multi Leg auto-electronic trade against single leg(s)	123	Transaction represents an electronic execution of a multi Leg order traded against single leg orders/ quotes.
MLAT	Multi Leg Auction	120	Transaction was the execution of an electronic multi leg order which was "stopped" at a price and traded in a two sided auction mechanism that goes through an exposure period in a complex order book. Such auctions mechanisms include and not limited to Price Improvement, Facilitation or Solicitation Mechanism.
MLET	Multi Leg auto-electronic trade	119	Transaction represents an electronic execution of a multi leg order traded in a complex order book.

continuation on the next page

Table 14: OPRA trade types for transactions in U.S. options exchanges (cont.)

MLCT	Multi Leg Cross	121	Transaction was the execution of an electronic multi leg order which was “stopped” at a price and traded in a two sided crossing mechanism that does not go through an exposure period. Such crossing mechanisms include and not limited to Customer to Customer Cross and QCC with two or more options legs.
MLFT	Multi Leg floor trade	122	Transaction represents a non-electronic multi leg order trade executed against other multi-leg order(s) on a trading floor. Execution of Paired and Non-Paired Auctions and Cross orders on an exchange floor are also included in this category.
MSFL	Multi Leg floor trade against single leg(s)	126	Transaction represents a non-electronic multi leg order trade executed on a trading floor against single leg orders/ quotes. Execution of Paired and Non-Paired Auctions on an exchange floor are also included in this category.
OPEN		6	Transaction is a late report of the opening trade and is out of sequence; i.e., other transactions have been reported for the particular option contract.
OPNL		7	Transaction is a late report of the opening trade, but is in the correct sequence; i.e., no other transactions have been reported for the particular option contract.
OSEQ		2	Transaction is being reported late and is out of sequence; i.e., later transactions have been reported for the particular option contract.
REOP		21	Transaction is a reopening of an option contract in which trading has been previously halted. Prefix appears solely for information; process as a regular transaction.
SCLI	Single Leg Cross ISO	117	Transaction was the execution of an Intermarket Sweep electronic order which was “stopped” at a price and traded in a two sided crossing mechanism that does not go through an exposure period. Such crossing mechanisms include and not limited to Customer to Customer Cross.
SLAI	Single Leg Auction ISO	115	Transaction was the execution of an Intermarket Sweep electronic order which was “stopped” at a price and traded in a two sided auction mechanism that goes through an exposure period. Such auctions mechanisms include and not limited to Price Improvement, Facilitation or Solicitation Mechanism marked as ISO.
SLAN	Single Leg Auction Non ISO	114	Transaction was the execution of an electronic order which was “stopped” at a price and traded in a two sided auction mechanism that goes through an exposure period. Such auctions mechanisms include and not limited to Price Improvement, Facilitation or Soliciation Mechanism.
SLCN	Single Leg Cross Non ISO	116	Transaction was the execution of an electronic order which was “stopped” at a price and traded in a two sided crossing mechanism that does not go through an exposure period. Such crossing mechanisms include and not limited to Customer to Customer Cross and QCC with a single option leg.

continuation on the next page

Table 14: OPRA trade types for transactions in U.S. options exchanges (cont.)

MLCT	Multi Leg Cross	121	Transaction was the execution of an electronic multi leg order which was “stopped” at a price and traded in a two sided crossing mechanism that does not go through an exposure period. Such crossing mechanisms include and not limited to Customer to Customer Cross and QCC with two or more options legs.
SLFT	Single Leg Floor Trade	118	Transaction represents a non-electronic trade executed on a trading floor. Execution of Paired and Non-Paired Auctions and Cross orders on an exchange floor are also included in this category.
TASL	Stock Options Auction against single leg(s)	131	Transaction was the execution of an electronic multi leg stock/options order which was “stopped” at a price and traded in a two sided auction mechanism that goes through an exposure period and trades against single leg orders/ quotes. Such auctions mechanisms include and not limited to Price Improvement, Facilitation or Solicitation Mechanism.
TESL	Stock Options auto-electronic trade against single leg(s)	130	Transaction represents an electronic execution of a multi Leg stock/options order traded against single leg orders/ quotes.
TFSL	Stock Options floor trade against single leg(s)	132	Transaction represents a non-electronic multi leg stock/options order trade executed on a trading floor against single leg orders/ quotes. Execution of Paired and Non-Paired Auctions on an exchange floor are also included in this category.
TLAT	Stock Options Auction	124	Transaction was the execution of an electronic multi leg stock/options order which was “stopped” at a price and traded in a two sided auction mechanism that goes through an exposure period in a complex order book. Such auctions mechanisms include and not limited to Price Improvement, Facilitation or Solicitation Mechanism.
TLCT	Stock Options Cross	128	Transaction was the execution of an electronic multi leg stock/options order which was “stopped” at a price and traded in a two sided crossing mechanism that does not go through an exposure period. Such crossing mechanisms include and not limited to Customer to Customer Cross.
TLET	Stock Options auto-electronic trade	127	Transaction represents an electronic execution of a multi leg stock/options order traded in a complex order book.
TLFT	Stock Options floor trade	129	Transaction represents a non-electronic multi leg order stock/options trade executed on a trading floor in a Complex order book. Execution of Paired and Non-Paired Auctions and Cross orders on an exchange floor are also included in this category.

This table reports OPRA trade types and their descriptions. The type of each transaction in U.S. options exchanges has to be classified using a type description from the table and reported to OPRA. This reporting requirement was implemented on November 4, 2019.

A.5 Descriptive statistics on SLAN trades below $20,000 by category

Table 15: Composition of SLAN trades below $20,000 in size

Characteristic	**Category**	Frequency share, %	Volume share, %	Quoted spread, %	Effective spread, %
Type	Call	71.2	69.8	13.5	6.5
	Put	28.8	30.2	13.9	6.8
Trade size (contracts)	1	46.3	7.0	13.7	6.2
	2-5	31.3	14.9	12.4	6.0
	6-10	11.7	15.9	14.1	7.2
	11-100	10.2	52.7	15.7	8.9
	Above 100	0.4	9.5	19.8	16.2
Trade size (dollars)	Below 250	42.1	16.2	23.4	11.6
	250-500	15.7	10.0	8.3	3.7
	500-1,000	13.9	12.8	7.0	3.0
	1,000-2,500	14.0	19.5	5.9	2.5
	2,500-5,000	7.1	15.2	4.9	2.0
	5,000-10,000	4.7	14.8	4.2	1.8
	10,000-20,000	2.5	11.5	3.7	2.9
	20,000-50,000				
	Above 50,000				
Trade direction	Sell	46.1	46.2	13.9	7.5
	Buy	43.4	44.6	13.2	7.2
	Midpoint	10.5	9.2	14.3	0.0
Time to expiry	Less than a week	47.4	50.8	12.4	6.5
	1-2 weeks	13.9	12.9	12.3	6.0
	2-4 weeks	16.1	15.2	15.1	7.0
	1-3 months	13.5	13.0	14.0	6.0
	3-12 months	7.5	6.7	18.9	7.4
	Over a year	1.4	1.1	18.4	7.8
Moneyness	Below -2	0.3	0.3	53.3	27.9
	-2 to -1	0.4	0.4	50.3	25.3
	-1 to -0.1	24.1	25.9	28.5	13.8
	At the money	71.2	70.5	8.5	4.1
	0.1 to 1	3.9	2.8	8.5	3.7
	1 to 2	0.1	0.1	9.2	4.3
	Above 2	0.1	0.1	17.6	7.7
Trade direction and type	Sell - Call	32.6	32.0	13.6	7.3
	Sell - Put	13.5	14.1	14.5	8.0
	Buy - Call	31.2	31.4	13.1	7.2
	Buy - Put	12.2	13.2	13.3	7.2
	Midpoint - Call	7.4	6.3	14.6	0.0
	Midpoint - Put	3.1	2.9	13.5	0.0
ETF	No	81.4	74.0	14.8	7.0
	Yes	18.6	26.0	8.5	4.5

This table reports characteristics of SLAN trades (single-leg price improvement auctions) that are smaller than $20,000 in size by category. (Implied) Trade direction is based on whether the trade price is above (buy), below (sell), or at the midpoint. Quoted spread, % is the spread between the best bid and best ask on the contract (across all exchanges) relative to the midpoint price at the time of the trade. Effective spread is an absolute percentage deviation of the trade price from the midpoint price at the time of the trade, multiplied by 2. For both spreads, we report frequency-weighted averages. Moneyness is measured as $(Midpoint\,Price - Strike)/Strike$.

A.6 Descriptive statistics on small option trades by category

Table 16: Composition of small trades

Characteristic	**Category**	Frequency share, %	Volume share, %	Quoted spread, %	Effective spread, %
Type	Call	65.3	64.8	10.5	7.4
	Put	34.7	35.2	11.9	7.8
Trade size (contracts)	1	50.8	18.5	10.9	7.4
	2-5	36.1	40.8	10.8	7.5
	6-10	13.1	40.7	12.0	8.4
	11-100				
	Above 100				
Trade size (dollars)	Below 250	42.1	29.9	18.9	12.8
	250-500	16.2	14.7	6.8	4.1
	500-1,000	14.7	16.2	5.7	3.4
	1,000-2,500	14.3	18.9	4.7	2.7
	2,500-5,000	6.6	9.9	3.9	2.2
	5,000-10,000	3.5	5.7	3.2	1.9
	10,000-20,000	1.6	2.8	2.7	6.6
	20,000-50,000	0.7	1.4	2.2	22.8
	Above 50,000	0.2	0.5	1.4	61.6
Trade direction	Sell	47.6	47.4	9.9	7.5
	Buy	45.6	46.1	11.8	8.8
	Midpoint	6.8	6.6	13.2	0.0
Time to expiry	Less than a week	42.4	42.7	12.3	8.8
	1-2 weeks	14.6	14.3	9.5	6.4
	2-4 weeks	17.2	16.9	10.6	6.7
	1-3 months	15.3	15.3	9.4	5.9
	3-12 months	8.4	8.7	10.5	7.3
	Over a year	1.9	1.9	12.1	11.2
Moneyness	Below -2	0.3	0.3	44.8	28.6
	-2 to -1	0.4	0.4	42.3	24.9
	-1 to -0.1	23.9	24.2	20.9	13.7
	At the money	70.2	70.1	7.7	5.3
	0.1 to 1	4.9	4.8	5.6	6.5
	1 to 2	0.2	0.2	6.2	14.2
	Above 2	0.1	0.1	10.8	25.0
Trade direction and type	Sell - Call	31.1	30.7	9.5	7.4
	Sell - Put	16.6	16.7	10.9	7.8
	Buy - Call	30.0	30.0	11.4	8.6
	Buy - Put	15.6	16.0	12.6	9.1
	Midpoint - Call	4.3	4.1	12.7	0.0
	Midpoint - Put	2.5	2.5	14.0	0.0
ETF	No	83.0	81.1	11.5	7.9
	Yes	17.0	18.9	8.4	5.9

This table reports characteristics of small option trades (below 10 contracts) by category. (Implied) Trade direction is based on whether the trade price is above (buy), below (sell), or at the midpoint. Quoted spread, % is the spread between the best bid and best ask on the contract (across all exchanges) relative to the midpoint price at the time of the trade. Effective spread is an absolute percentage deviation of the trade price from the midpoint price at the time of the trade, multiplied by 2. For both spreads, we report frequency-weighted averages. Moneyness is measured as $(\mathit{Midpoint\,Price} - \mathit{Strike})/\mathit{Strike}$.

A.7 Additional descriptive statistics on SLAN trades

Table 17: Composition of SLAN trades, additional statistics

Characteristic	**Category**	Dollar volume share, %	Dollar spread, $	Implied volatility	Trade price, $
Type	Call	69.92	0.20	0.84	4.88
	Put	30.08	0.19	0.80	4.51
Trade size (contracts)	1	13.24	0.23	0.84	5.62
	2-5	22.32	0.20	0.82	4.73
	6-10	18.49	0.16	0.83	3.69
	11-100	39.77	0.12	0.80	2.78
	Above 100	6.17	0.06	0.67	1.33
Trade size (dollars)	Below 250	2.01	0.08	0.93	0.73
	250-500	2.73	0.14	0.80	2.29
	500-1,000	4.78	0.19	0.78	3.70
	1,000-2,500	10.71	0.28	0.76	6.48
	2,500-5,000	12.15	0.41	0.73	11.50
	5,000-10,000	15.74	0.50	0.71	16.68
	10,000-20,000	16.90	0.62	0.69	22.77
	20,000-50,000	19.73	0.77	0.67	29.25
	Above 50,000	15.24	1.06	0.68	42.44
Trade direction	Sell	48.97	0.23	0.82	5.22
	Buy	45.04	0.20	0.86	4.89
	Midpoint	6.00	0.10	0.74	2.34
Time to expiry	Less than a week	40.14	0.16	0.88	4.10
	1-2 weeks	12.15	0.17	0.83	4.42
	2-4 weeks	14.27	0.20	0.85	4.21
	1-3 months	16.60	0.24	0.73	5.56
	3-12 months	12.66	0.37	0.69	7.72
	Over a year	4.02	0.80	0.60	13.94
Moneyness	Below -2	0.06	0.20	2.45	0.94
	-2 to -1	0.12	0.26	1.89	1.42
	-1 to -0.1	10.98	0.17	1.23	2.08
	At the money	77.00	0.19	0.66	5.19
	0.1 to 1	10.98	0.52	1.14	13.03
	1 to 2	0.58	0.76	1.49	18.48
	Above 2	0.28	0.82	1.60	18.10
Trade direction and type	Sell - Call	34.34	0.23	0.83	5.37
	Sell - Put	14.62	0.22	0.80	4.84
	Buy - Call	31.50	0.20	0.87	4.96
	Buy - Put	13.53	0.19	0.82	4.72
	Midpoint - Call	4.07	0.10	0.75	2.37
	Midpoint - Put	1.93	0.09	0.73	2.26
ETF	No	78.97	0.23	0.91	5.20
	Yes	21.03	0.09	0.46	2.92

This table reports characteristics of trades by category. (Implied) Trade direction is based on whether the trade price is above (buy), below (sell), or at the midpoint. Dollar spread, $ is the spread between the best bid and best ask on the contract (across all exchanges) in U.S. dollars at the time of the trade. Implied volatility is trade implied volatility reported by OPRA. For all measures, we report frequency-weighted averages. Moneyness is measured as $(MidpointPrice - Strike)/Strike$.

A.8 Descriptive statistics for the ticker-level sample

Table 18: Descriptive statistics for the ticker-level variables

	Call options					Put options				
	Mean	Median	St. Dev.	p1	p99	Mean	Median	St. Dev.	p1	p99
SLAN Share	0.21	0.14	0.24	0.00	1.00	0.17	0.07	0.25	0.00	1.00
SLAN < $5k$ Share	0.18	0.10	0.23	0.00	1.00	0.15	0.05	0.24	0.00	1.00
SLAN < $20k$ Share	0.21	0.13	0.24	0.00	1.00	0.17	0.06	0.25	0.00	1.00
SLAN > $20k$ Share	0.01	0.00	0.04	0.00	0.17	0.01	0.00	0.04	0.00	0.13
MLAN Share	0.03	0.00	0.09	0.00	0.48	0.04	0.00	0.12	0.00	0.67
Complex Share	0.12	0.01	0.20	0.00	0.98	0.17	0.03	0.26	0.00	1.00
Large Share	0.06	0.00	0.14	0.00	0.71	0.05	0.00	0.15	0.00	0.77
> $50k$ Share	0.02	0.00	0.08	0.00	0.41	0.02	0.00	0.08	0.00	0.42
SLAN Imbalance	-0.11	-0.11	0.66	-1.00	1.00	-0.17	-0.23	0.71	-1.00	1.00
SLAN < $5k$ Imbalance	-0.12	-0.11	0.65	-1.00	1.00	-0.17	-0.23	0.70	-1.00	1.00
SLAN < $20k$ Imbalance	-0.11	-0.11	0.65	-1.00	1.00	-0.17	-0.23	0.71	-1.00	1.00
SLAN > $20k$ Imbalance	-0.03	-0.04	0.80	-1.00	1.00	-0.09	-0.13	0.84	-1.00	1.00
MLAN Imbalance	-0.08	0.00	0.53	-1.00	1.00	-0.11	-0.03	0.55	-1.00	1.00
Complex Imbalance	-0.04	0.00	0.49	-1.00	1.00	-0.06	0.00	0.52	-1.00	1.00
Large Imbalance	-0.03	0.00	0.74	-1.00	1.00	-0.05	-0.01	0.75	-1.00	1.00
> $50k$ Imbalance	-0.01	0.00	0.75	-1.00	1.00	-0.06	-0.05	0.77	-1.00	1.00
Small Share	0.60	0.55	0.34	0.00	1.00	0.64	0.65	0.35	0.00	1.00
Small Imbalance	-0.05	-0.04	0.53	-1.00	1.00	-0.02	-0.01	0.59	-1.00	1.00
Internalized volume in underlying	0.17	0.15	0.09	0.00	0.39	0.17	0.15	0.08	0.00	0.38
Robinhood ownership breadth, log	6.94	6.83	1.75	3.33	11.81	7.06	6.95	1.78	3.40	11.98
WSB mentions, log	0.18	0.00	0.56	0.00	3.22	0.20	0.00	0.58	0.00	3.26
Option trading volume, lagged log	5.30	5.19	2.85	0.18	12.17	4.82	4.62	2.81	0.18	11.66
Underlying price, log	3.30	3.37	1.29	0.34	6.04	3.40	3.46	1.26	0.46	6.11
Underlying return, past week	0.01	0.00	0.09	-0.25	0.32	0.01	0.00	0.09	-0.25	0.33
Total volume in underlying, log	15.46	15.42	1.49	11.97	19.20	15.64	15.60	1.46	12.21	19.29
Underlying spread	0.05	0.04	0.03	0.00	0.18	0.05	0.04	0.03	0.00	0.18
Underlying volatility	0.48	0.36	0.42	0.04	2.40	0.49	0.36	0.43	0.04	2.46
Market cap, log	7.59	7.59	1.94	3.28	12.14	7.80	7.79	1.90	3.50	12.22
D(is ETF)	0.14	0.00	0.35	0.00	1.00	0.14	0.00	0.35	0.00	1.00
Option spread	0.46	0.33	0.40	0.05	2.00	0.44	0.30	0.41	0.05	2.00
Option moneyness	-0.04	-0.04	0.13	-0.47	0.44	-0.09	-0.07	0.18	-0.92	0.37
Option time to expiry	0.08	0.06	0.06	0.00	0.28	0.07	0.05	0.06	0.00	0.30
Option leverage	14.44	10.61	12.54	2.36	74.32	13.58	10.04	12.26	0.98	70.35

This table reports the descriptive statistics for the daily ticker-level sample in 11/2019-06/2021, separately for call and put options. The sample includes all stock and ETF tickers with lagged price above $1. SLAN and Small Share are the ticker-level volume shares of SLAN and small trades, respectively. SLAN and Small Imbalance are the ticker-level volume imbalance for SLAN and small trades, respectively. Share and imbalance are constucted similarly for SLAN < $\$5,000$, SLAN < $\$20,000$, SLAN $\$5,000 - 20,000$, MLAN, complex (all multi-leg), large (above 100 contracts) trades and trades above $\$50,000$. Internalized volume in underlying is the share of non-ATS OTC (i.e., internalized) volume in the total trading volume in the underlying stock or ETF. Robinhood ownership breadth, log is the logarithm of the total number of Robinhood users holding the ticker at the end of each day. WSB mentions, log is the logarithm of the number of mentions a ticker gets on `WallStreetBets` during the day. Underlying price (log) is as of the day before. Underlying return is the total return over the last week. Total volume in underlying, log is the logarithm of the total trading volume (lit, ATS and non-ATS OTC) in underlying ticker over the previous week. Underlying spread is averaged over the previous week. Underlying volatility is return volatility over the previous week. Option spread is the contract quoted relative spread. Option time to expiry (in months), moneyness, spread, and leverage are equal-weighted across trades at a ticker level. WSB mentions, Robinhood ownership breadth, underlying volatilit, and spread as well as option spread, time to expiry, and lambda are winsorized at 99^{th} percentile. Underlying return and option moneyness are winsorized at 0.5^{th} and 99.5^{th} percentiles.

A.9 SLAN trades below $20,000 and other measures of retail activity

Table 19: Share of SLAN option trades below $20,000 in size and other measures of retail activity

	SLAN $< \$20k$ trades in calls				**SLAN $< \$20k$ trades in puts**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: MLAN Share								
Small Share	0.076***				0.071***			
	(32.51)				(33.69)			
Internalized volume in underlying		0.024***				0.019***		
		(8.60)				(7.09)		
Robinhood ownership breadth, log			0.028***				0.058***	
			(2.75)				(5.58)	
WSB mentions, log				-0.001				0.003**
				(-0.56)				(2.21)
Undelying controls X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contract controls C	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,398,642	1,398,642	571,699	1,145,651	1,203,829	1,203,829	496,592	1,020,950
Adjusted R-squared	0.122	0.118	0.111	0.132	0.092	0.089	0.083	0.097
Panel B: MLAN Imbalance								
Small Imbalance	0.539***				0.534***			
	(275.16)				(228.18)			
Internalized volume in underlying		0.015***				0.007**		
		(5.21)				(2.25)		
Robinhood ownership breadth, log			0.047***				0.028***	
			(4.60)				(3.07)	
WSB mentions, log				0.013***				0.010***
				(10.16)				(7.00)
Undelying controls X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contract controls C	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,066,981	1,071,276	423,057	927,439	798,787	803,815	326,784	723,249
Adjusted R-squared	0.202	0.020	0.026	0.020	0.196	0.024	0.025	0.023

This table reports the results of estimating (1) on daily data in 11/2019-06/2021. SLAN $< \$20k$ and Small Share are the ticker-level volume shares of SLAN (below $20,000) and small trades, respectively. SLAN $< \$20k$ and Small Imbalance are the ticker-level volume imbalance for SLAN (below $20,000) and small trades, respectively. Internalized volume in underlying is the share of non-ATS OTC (i.e., internalized) volume in the total trading volume in the underlying stock or ETF. Robinhood ownership breadth, log is the logarithm of the total number of Robinhood users holding the ticker at the end of each day. WSB mentions, log is the logarithm of the number of mentions a ticker gets on `WallStreetBets`during the day. Underlying controls X and contract controls C are described in Section 2.3. All regressions include date and ticker fixed effects. All variables are standardized within the contract type (call or put). t-statistics are based on standard errors clustered by ticker and date (in parentheses). *** p<0.01, ** p<0.05, * p<0.1.

A.10 SLAN trades and other measures of retail activity, most traded tickers only

Table 20: Retail trading in options and other measures of retail activity, most traded tickers only

	Retail trading in calls				Retail trading in puts			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: SLAN Share								
Small Share	0.339***				0.245***			
	(16.27)				(17.65)			
Internalized volume in underlying		0.100***				0.075***		
		(3.76)				(4.29)		
Robinhood ownership breadth, log			0.091				0.028	
			(1.54)				(0.52)	
WSB mentions, log				-0.012				0.024***
				(-1.09)				(2.87)
Undelying controls X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contract controls C	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	130,271	130,271	55,855	121,829	129,942	129,942	55,591	121,500
Adjusted R-squared	0.426	0.391	0.366	0.385	0.364	0.340	0.294	0.337
Panel B: SLAN Imbalance								
Small Imbalance	0.291***				0.217***			
	(37.20)				(32.86)			
Internalized volume in underlying		0.029***				0.005		
		(2.86)				(0.49)		
Robinhood ownership breadth, log			0.082**				0.042	
			(2.57)				(1.61)	
WSB mentions, log				0.045***				0.023***
				(8.58)				(4.49)
Undelying controls X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contract controls C	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	129,800	129,806	55,499	121,364	128,965	128,975	54,822	120,533
Adjusted R-squared	0.116	0.044	0.041	0.044	0.094	0.055	0.050	0.054

This table reports the results of estimating (1) on daily data for the underlying securities in the top decile by their total option dollar traded volume in 11/2019-06/2021 (341 tickers). SLAN and Small Share are the ticker-level volume shares of SLAN and small trades, respectively. SLAN and Small Imbalance are the ticker-level volume imbalance for SLAN and small trades, respectively. Internalized volume in underlying is the share of non-ATS OTC (i.e., internalized) volume in the total trading volume in the underlying stock or ETF. Robinhood ownership breadth, log is the logarithm of the total number of Robinhood users holding the ticker at the end of each day. WSB mentions, log is the logarithm of the number of mentions a ticker gets on `WallStreetBets` during the day. Underlying controls X and contract controls C are described in Section 2.3. All regressions include date and ticker fixed effects. All variables are standardized within the contract type (call or put). t-statistics are based on standard errors clustered by ticker and date (in parentheses). *** p<0.01, ** p<0.05, * p<0.1.

A.11 Characteristics of MLAN trades

In this section, we describe trades that are multi-leg and went through price improvement auctions. These trades are on average larger than SLAN trades, more balanced by option type, and are negatively correlated with equity-based measures of retail activity. Furthermore, a larger fraction of these trades are executed at midpoint.

Table 21: MLAN trades in options and other measures of retail activity

	MLAN trades in calls				**MLAN trades in puts**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: MLAN Share								
Small Share	0.046***				0.044***			
	(16.38)				(21.75)			
Internalized volume in underlying		-0.003				0.007***		
		(-1.17)				(2.85)		
Robinhood ownership breadth, log			0.005				0.014	
			(0.70)				(1.49)	
WSB mentions, log				-0.008***				-0.002
				(-5.68)				(-0.99)
Undelying controls X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contract controls C	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,398,642	1,398,642	571,699	1,145,651	1,203,829	1,203,829	496,592	1,020,950
Adjusted R-squared	0.060	0.059	0.046	0.061	0.049	0.047	0.043	0.049
Panel B: MLAN Imbalance								
Small Imbalance	0.299***				0.391***			
	(65.17)				(78.41)			
Internalized volume in underlying		-0.002				-0.000		
		(-0.41)				(-0.05)		
Robinhood ownership breadth, log			-0.029**				-0.029*	
			(-2.31)				(-1.81)	
WSB mentions, log				-0.002				-0.006***
				(-1.47)				(-3.62)
Undelying controls X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contract controls C	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	520,921	521,155	190,664	482,981	437,361	437,954	167,774	408,759
Adjusted R-squared	0.038	0.016	0.017	0.015	0.065	0.023	0.025	0.020

This table reports the results of estimating (1) on daily data in 11/2019-06/2021. MLAN and Small Share are the ticker-level volume shares of MLAN and small trades, respectively. MLAN and Small Imbalance are the ticker-level volume imbalance for MLAN and small trades, respectively. Internalized volume in underlying is the share of non-ATS OTC (i.e., internalized) volume in the total trading volume in the underlying stock or ETF. Robinhood ownership breadth, log is the logarithm of the total number of Robinhood users holding the ticker at the end of each day. WSB mentions, log is the logarithm of the number of mentions a ticker gets on `WallStreetBets` during the day. Underlying controls X and contract controls C are described in Section 2.3. All regressions include date and ticker fixed effects. All variables are standardized within the contract type (call or put). t-statistics are based on standard errors clustered by ticker and date (in parentheses). *** p<0.01, ** p<0.05, * p<0.1.

Table 22: Composition of MLAN trades

Characteristic	**Category**	Frequency share, %	Volume share, %	Quoted spread, %	Effective spread, %
Type	Call	54.0	53.5	15.0	7.0
	Put	46.0	46.5	17.8	8.5
Trade size (contracts)	1	54.4	10.4	17.3	8.7
	2-5	28.7	17.1	14.9	6.3
	6-10	9.8	16.8	15.6	6.6
	11-100	6.6	40.4	15.7	7.2
	Above 100	0.5	15.4	14.4	7.6
Trade size (dollars)	Below 250	40.6	15.5	29.5	15.1
	250-500	14.5	8.1	9.5	2.8
	500-1,000	13.9	10.2	7.9	2.1
	1,000-2,500	14.4	15.4	6.8	1.7
	2,500-5,000	7.2	11.9	5.9	1.4
	5,000-10,000	4.5	10.9	5.2	1.3
	10,000-20,000	2.5	9.1	4.5	4.9
	20,000-50,000	1.5	8.8	4.0	12.4
	Above 50,000	0.8	10.1	3.5	18.9
Trade direction	Sell	50.0	49.2	13.0	6.5
	Buy	35.6	37.1	20.9	12.4
	Midpoint	14.4	13.7	16.4	0.0
Time to expiry	Less than a week	35.4	39.4	23.0	13.0
	1-2 weeks	14.9	14.8	14.4	6.2
	2-4 weeks	22.0	19.2	13.2	4.4
	1-3 months	20.5	17.8	9.9	3.1
	3-12 months	5.8	7.1	15.1	7.3
	Over a year	1.2	1.5	14.2	9.4
Moneyness	Below -2	0.1	0.3	79.7	37.6
	-2 to -1	0.3	0.3	67.6	24.6
	-1 to -0.1	25.2	23.3	32.3	14.9
	At the money	69.4	70.7	11.0	5.1
	0.1 to 1	4.9	5.3	5.6	5.1
	1 to 2	0.1	0.1	6.3	15.4
	Above 2	0.0	0.1	11.5	22.9
Trade direction and type	Sell - Call	26.6	26.1	12.2	6.2
	Sell - Put	23.4	23.1	14.0	7.0
	Buy - Call	19.8	20.2	18.7	10.7
	Buy - Put	15.8	16.8	23.6	14.5
	Midpoint - Call	7.5	7.1	15.3	0.0
	Midpoint - Put	6.8	6.5	17.5	0.0
ETF	No	74.9	71.4	17.0	7.1
	Yes	25.1	28.6	14.1	9.5

This table reports characteristics of MLAN trades (multi-leg price improvement auctions) by category. (Implied) Trade direction is based on whether the trade price is above (buy), below (sell), or at the midpoint. Quoted spread, % is the spread between the best bid and best ask on the contract (across all exchanges) relative to the midpoint price at the time of the trade. Effective spread is an absolute percentage deviation of the trade price from the midpoint price at the time of the trade, multiplied by 2. For both spreads, we report frequency-weighted averages. Moneyness is measured as $(Midpoint\,Price - Strike)/Strike$.

A.12 Complex strategy trades and measures of retail activity

Table 23: Complex strategy trades in options and measures of retail activity

	Trades in calls				Trades in puts			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Complex Share								
Small Share	-0.016***				0.013***			
	(-6.13)				(4.93)			
Internalized volume in underlying		-0.006**				0.001		
		(-2.28)				(0.49)		
Robinhood ownership breadth, log			-0.010				0.005	
			(-1.10)				(0.41)	
WSB mentions, log				-0.009***				0.000
				(-4.70)				(0.19)
Undelying controls X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contract controls C	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,398,642	1,398,642	571,699	1,145,651	1,203,829	1,203,829	496,592	1,020,950
Adjusted R-squared	0.128	0.128	0.124	0.136	0.108	0.108	0.103	0.114
Panel B: Complex Imbalance								
Small Imbalance	0.418***				0.525***			
	(111.02)				(130.84)			
Internalized volume in underlying		-0.003				0.003		
		(-0.94)				(1.09)		
Robinhood ownership breadth, log			-0.022**				-0.009	
			(-2.04)				(-0.86)	
WSB mentions, log				-0.002*				-0.003*
				(-1.89)				(-1.87)
Undelying controls X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contract controls C	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	767,858	769,022	292,638	691,809	667,209	669,619	263,116	608,363
Adjusted R-squared	0.074	0.010	0.013	0.010	0.127	0.013	0.016	0.011

This table reports the results of estimating (1) on daily data in 11/2019-06/2021. Complex and Small Share are the ticker-level volume shares of all multi-leg strategy and small trades, respectively. Complex and Small Imbalance are the ticker-level volume imbalance for all multi-leg and small trades, respectively. Internalized volume in underlying is the share of non-ATS OTC (i.e., internalized) volume in the total trading volume in the underlying stock or ETF. Robinhood ownership breadth, log is the logarithm of the total number of Robinhood users holding the ticker at the end of each day. WSB mentions, log is the logarithm of the number of mentions a ticker gets on `WallStreetBets` during the day. Underlying controls X and contract controls C are described in Section 2.3. All regressions include date and ticker fixed effects. All variables are standardized within the contract type (call or put). t-statistics are based on standard errors clustered by ticker and date (in parentheses). *** p<0.01, ** p<0.05, * p<0.1.

A.13 Trades above $50,000 and measures of retail activity

Table 24: Trades in size above $50,000 in options and measures of retail activity

	Trades in calls				Trades in puts			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Share of trades sized above $50,000								
Small Share	-0.205***				-0.191***			
	(-32.72)				(-28.02)			
Internalized volume in underlying		0.016***				-0.006*		
		(6.85)				(-1.87)		
Robinhood ownership breadth, log			0.044***				-0.007	
			(4.40)				(-0.46)	
WSB mentions, log				0.010***				0.001
				(4.06)				(0.31)
Undelying controls X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contract controls C	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,398,642	1,398,642	571,699	1,145,651	1,203,829	1,203,829	496,592	1,020,950
Adjusted R-squared	0.150	0.121	0.116	0.122	0.129	0.102	0.104	0.102
Panel B: Imbalance in trades sized above $50,000								
Small Imbalance	0.254***				0.239***			
	(30.36)				(25.04)			
Internalized volume in underlying		-0.003				0.007		
		(-0.40)				(0.79)		
Robinhood ownership breadth, log			-0.056**				-0.028	
			(-2.19)				(-1.03)	
WSB mentions, log				0.003				0.006**
				(1.48)				(2.37)
Undelying controls X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contract controls C	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	166,518	166,746	59,467	155,334	121,597	121,991	50,690	113,256
Adjusted R-squared	0.026	0.016	0.021	0.015	0.029	0.021	0.026	0.020

This table reports the results of estimating (1) on daily data in 11/2019-06/2021. Small Share is the ticker-level volume shares of small trades. Small Imbalance is the ticker-level volume imbalance for small trades. Internalized volume in underlying is the share of non-ATS OTC (i.e., internalized) volume in the total trading volume in the underlying stock or ETF. Robinhood ownership breadth, log is the logarithm of the total number of Robinhood users holding the ticker at the end of each day. WSB mentions, log is the logarithm of the number of mentions a ticker gets on `WallStreetBets` during the day. Underlying controls X and contract controls C are described in Section 2.3. All regressions include date and ticker fixed effects. All variables are standardized within the contract type (call or put). t-statistics are based on standard errors clustered by ticker and date (in parentheses). *** p<0.01, ** p<0.05, * p<0.1.

A.14 Characteristics of SLAN and MLAN trades on option expiration day

To shed light on statistical significance of observations in Section 2.3.1, we regress the daily series of differences between buy and sell shares onto dummies for each trading hour. Table 25 reports the results for SLAN, MLAN, complex trades as well as trades larger than \$50k. We see that SLAN trades exhibit a statistically significant intraday pattern: There is a larger sell volume share at the end of the trading day. Furthermore, the magnitude of the buy-sell share difference significantly increases after 30/09/2020, when Robinhood introduced the rule-based selling of expiring options. MLAN, complex or larger trades do not reveal anything similar.

Table 25: Intra-day buy-sell patterns for expiring options

	Sell-buy volume share by trade type:							
	SLAN		MLAN		Complex		Sized above \$50k	
Variable	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.
D(Expiration -1 hour)	-0.92***	-8.62	0.63	1.22	1.10***	5.65	1.38**	2.13
× D(after 30/09/2020)	-0.33*	-1.80	-0.69	-1.20	0.25	0.86	1.24	1.03
D(Expiration -2 hours)	-0.05	-0.51	-0.78***	-3.97	0.01	0.11	0.86	1.39
× D(after 30/09/2020)	-0.23	-1.49	0.08	0.26	0.25	1.31	1.56	1.21
D(Expiration -3 hours)	0.20**	2.43	-0.48***	-2.93	-0.12	-0.89	1.83***	4.09
× D(after 30/09/2020)	0.01	0.08	-0.32	-1.08	0.18	0.99	1.26	1.16
D(Expiration -4 hours)	0.39***	4.11	-0.91***	-6.23	-0.48***	-3.22	-0.26	-0.34
× D(after 30/09/2020)	0.09	0.69	0.16	0.55	0.56***	2.82	2.38*	1.74
D(Expiration -5 hours)	0.67***	6.76	-1.27***	-6.62	-0.15	-1.10	0.60	0.80
× D(after 30/09/2020)	-0.01	-0.08	0.27	0.88	-0.01	-0.06	0.67	0.55
D(Expiration -6 hours)	0.95***	6.91	-1.98***	-7.71	-0.28*	-1.76	0.19	0.20
× D(after 30/09/2020)	0.17	1.07	0.59	1.56	0.15	0.75	3.19**	1.97
D(Expiration -7 hours)	0.42***	7.10	-0.77***	-3.94	-0.12*	-1.85	1.09	1.60
D(after 30/09/2020)	-0.11	-1.51	0.50**	2.24	0.13	1.43	-0.67	-0.75

This table reports estimation results from regressing hourly volume share difference between buy and sell trades on hourly dummies in 11/2019-06/2021. D(Expiration -X hours) equals 1 for Xth hour to expiration. D(after 30/09/2020) equals 1 after 30/09/2020. Complex trades are all multi-leg trades. t-statistics are based on Newey-West standard errors (lag 5). *** p<0.01, ** p<0.05, * p<0.1.

Table 26: Composition of option trades on expiration day

Hour to expiry	Trade direction	SLAN Frequency share, %	SLAN Volume share, %	MLAN Frequency share, %	MLAN Volume share, %
Panel A: Before September 2020					
7	Sell	4.05	3.24	3.60	3.54
7	Buy	3.93	3.42	3.00	3.09
7	Midpoint	0.85	0.66	0.95	0.86
6	Sell	11.62	10.01	10.31	10.08
6	Buy	11.06	10.35	8.70	8.88
6	Midpoint	2.53	2.07	2.98	2.71
5	Sell	7.57	7.23	6.90	7.00
5	Buy	7.15	7.50	6.05	6.32
5	Midpoint	1.65	1.45	2.10	1.96
4	Sell	5.93	6.08	5.54	5.51
4	Buy	5.53	6.18	4.91	5.08
4	Midpoint	1.27	1.18	1.70	1.57
3	Sell	5.12	5.54	4.98	4.99
3	Buy	4.73	5.51	4.55	4.92
3	Midpoint	1.09	1.04	1.53	1.42
2	Sell	6.22	6.60	6.47	6.35
2	Buy	5.18	6.15	6.07	6.20
2	Midpoint	1.21	1.16	1.98	1.81
1	Sell	6.72	7.43	7.50	7.25
1	Buy	5.41	6.04	7.87	8.23
1	Midpoint	1.18	1.17	2.30	2.24
Panel B: After September 2020					
7	Sell	4.41	3.53	3.77	3.31
7	Buy	4.32	3.65	3.08	2.96
7	Midpoint	0.93	0.71	1.00	0.85
6	Sell	12.43	10.70	10.48	9.79
6	Buy	11.73	10.96	8.83	8.97
6	Midpoint	2.67	2.15	2.92	2.62
5	Sell	7.32	7.20	6.87	6.55
5	Buy	6.81	7.27	6.04	6.18
5	Midpoint	1.55	1.37	2.02	1.86
4	Sell	5.69	6.13	5.53	5.38
4	Buy	5.27	6.09	5.04	5.29
4	Midpoint	1.18	1.12	1.68	1.62
3	Sell	4.91	5.67	5.10	5.05
3	Buy	4.45	5.44	4.73	5.09
3	Midpoint	1.02	1.01	1.55	1.49
2	Sell	4.93	5.98	5.64	5.66
2	Buy	4.23	5.27	5.41	5.99
2	Midpoint	1.03	1.04	1.74	1.72
1	Sell	8.10	7.74	7.86	8.09
1	Buy	5.68	5.80	8.39	9.21
1	Midpoint	1.34	1.17	2.31	2.33

This table reports characteristics of trades by category for options on their expiration day. SLAN is a single-leg price improvement auction, our measure of retail activity. MLAN is a multi-leg price improvement auction. (Implied) Trade direction is based on whether the trade price is above (buy), below (sell), or at the midpoint.

A.15 SLAN volume and Robinhood herding events (frenzies)

Table 27: Options trade imbalances and herding events

	Imbalance in trades of type:							
	SLAN		**SLAN** $< \$5k$		**SLAN** $< \$20k$		**SLAN** $> \$20k$	
	Call	Put	Call	Put	Call	Put	Call	Put
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D(Robinhood frenzy)	0.066***	0.095***	0.084***	0.130***	0.065***	0.102***	-0.068	-0.093
	(2.68)	(2.90)	(3.43)	(4.05)	(2.63)	(3.16)	(-0.89)	(-0.78)
Observations	437,428	337,841	433,274	332,861	436,830	336,955	61,379	39,374
Adjusted R-squared	0.025	0.025	0.027	0.025	0.026	0.025	0.023	0.031
	MLAN		**All complex**		**All** $> \$50k$		**All** > 100 **contracts**	
	Call	Put	Call	Put	Call	Put	Call	Put
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
D(Robinhood frenzy)	-0.151***	-0.024	-0.079**	0.003	0.115	-0.176**	-0.045	-0.052
	(-3.41)	(-0.41)	(-2.45)	(0.08)	(1.27)	(-2.07)	(-1.14)	(-0.82)
Observations	196,457	172,732	302,335	271,650	61,265	52,036	121,160	91,698
Adjusted R-squared	0.016	0.025	0.013	0.016	0.021	0.026	0.022	0.032

This table reports the results of estimating (1) on daily data in 04/11/2019-10/08/2020, separately for call and put options. The sample includes all stock and ETF tickers with lagged price above \$1. As a dependent variable, we use imbalance of contract volume traded via the indicated trade type, aggregated at the ticker level. SLAN is a single-leg price improvement auction, through which we measure retail activity. SLAN $< \$5k$, $< \$20k$, and $> \$20k$ correspond to SLAN trades of the respective dollar size. MLAN is a multi-leg price improvement auction. D(Robinhood frenzy) equals 1 if the ticker experienced a Robinhood herding event using the data of Barber, Huang, Odean, and Schwartz (2022). All regressions include X and C controls, as described in Section 2.3, as well as date and ticker fixed effects. t-statistics are based on standard errors clustered by ticker and date (in parentheses). *** p<0.01, ** p<0.05, * p<0.1.

A.16 Aggregate SLAN performance, value-weighted prices

Table 28: SLAN trade performance, aggregate and by contract characteristics, using value-weighted prices

	SLAN Raw performance, $ billion				Market Raw performance, $ billion			
	1 day	2 days	5 days	10 days	1 day	2 days	5 days	10 days
Panel A: All contracts								
	-0.532	-1.076	-1.201	-1.242	2.808	3.501	3.789	4.538
Panel B: By contract type								
Call	-0.303	-0.869	-1.034	-1.178	2.873	3.078	3.107	2.604
Put	-0.228	-0.206	-0.168	-0.063	-0.071	0.415	0.679	1.925
Panel C: By moneyness								
Below -2	-0.003	-0.003	-0.004	-0.002	-0.056	-0.068	-0.071	-0.089
-2 to -1	-0.005	-0.004	0.000	0.006	-0.074	-0.080	-0.088	-0.097
-1 to -0.1	-0.018	0.023	0.161	0.273	0.694	1.003	1.385	2.000
At the money	0.136	-0.258	-0.540	-0.640	1.846	2.516	2.091	1.928
0.1 to 1	-0.386	-0.575	-0.560	-0.600	1.147	0.875	1.339	1.745
1 to 2	-0.114	-0.118	-0.129	-0.137	-0.198	-0.178	-0.196	-0.182
Above 2	-0.138	-0.138	-0.128	-0.138	-0.549	-0.613	-0.670	-0.778
Panel D: By time to expiry								
Less than a week	-0.346	-0.765	-1.099	-1.101	4.614	4.713	4.446	4.562
1-2 weeks	-0.079	-0.139	-0.161	-0.278	0.807	0.943	1.523	1.547
2-4 weeks	-0.045	-0.115	-0.112	-0.210	0.890	1.067	1.182	1.466
1-3 months	-0.021	-0.038	0.050	0.118	0.765	0.944	0.703	0.566
3-12 months	0.052	0.063	0.171	0.265	-1.367	-1.285	-1.165	-0.689
Over a year	-0.092	-0.080	-0.051	-0.036	-2.908	-2.889	-2.903	-2.924

This table reports the performance of SLAN trades in November 2019 to June 2021. Raw performance at each horizon is computed as explained in Section 2.4. We use value-weighted average transaction prices.

A.17 Aggregate SLAN performance, by month and weekday

Table 29: SLAN trade performance, by month and weekday

	SLAN Raw performance, \$ billion				Market Raw performance, \$ billion			
Horizon h	1 day	2 days	5 days	10 days	1 day	2 days	5 days	10 days
Panel A: By month								
Nov-19	-0.017	-0.010	-0.009	-0.016	-0.009	-0.015	-0.061	-0.073
Dec-19	-0.014	-0.015	-0.029	-0.025	-0.122	-0.115	-0.101	-0.102
Jan-20	0.022	0.016	0.055	0.109	0.695	0.658	0.832	0.749
Feb-20	-0.083	-0.108	-0.145	-0.082	0.677	0.692	0.704	1.296
Mar-20	0.081	0.063	0.146	0.192	0.403	0.594	0.553	0.285
Apr-20	0.009	-0.025	-0.037	-0.059	0.334	0.408	0.406	0.236
May-20	0.003	0.020	0.012	0.001	0.215	0.258	0.179	0.254
Jun-20	-0.078	-0.139	-0.017	-0.066	-0.673	-0.618	-0.909	-1.063
Jul-20	0.010	-0.005	0.031	0.085	0.172	0.371	0.589	0.625
Aug-20	0.076	0.067	0.043	0.016	1.138	0.927	0.993	0.525
Sep-20	0.041	0.006	-0.012	-0.021	0.365	0.161	0.059	0.202
Oct-20	0.033	0.003	0.016	0.043	0.469	0.712	0.605	0.850
Nov-20	0.009	0.002	-0.006	0.044	0.425	0.227	-0.068	-0.066
Dec-20	0.050	0.046	-0.005	-0.047	0.825	0.867	0.967	1.050
Jan-21	-0.134	-0.442	-0.735	-0.681	1.033	1.172	0.495	0.921
Feb-21	-0.026	-0.123	-0.157	-0.282	0.681	0.329	1.391	1.892
Mar-21	-0.098	-0.063	-0.038	-0.071	-0.259	-0.094	-0.820	-0.756
Apr-21	-0.148	-0.164	-0.180	-0.218	-0.373	-0.126	0.007	0.198
May-21	-0.048	-0.006	0.000	-0.021	-0.501	-0.181	0.304	0.133
Jun-21	-0.111	-0.093	-0.029	-0.037	-1.749	-1.787	-1.398	-1.678
Panel B: By weekday								
Mon	0.032	0.048	-0.108	-0.222	-0.320	-0.606	-1.300	-1.952
Tue	0.008	-0.217	-0.131	-0.087	1.070	1.030	0.824	1.867
Wed	0.089	-0.116	-0.177	-0.135	1.839	1.775	1.667	2.242
Thu	-0.205	-0.304	-0.261	-0.302	1.214	1.607	1.697	0.875
Fri	-0.350	-0.380	-0.418	-0.388	-0.056	0.634	1.841	2.444

This table reports the performance of SLAN trades in November 2019 to June 2021. Raw performance at each horizon is computed as explained in Section 2.4.

A.18 Aggregate SLAN performance, best and worst tickers

Table 30: Best and worst tickers by trade performance, SLAN and the whole market

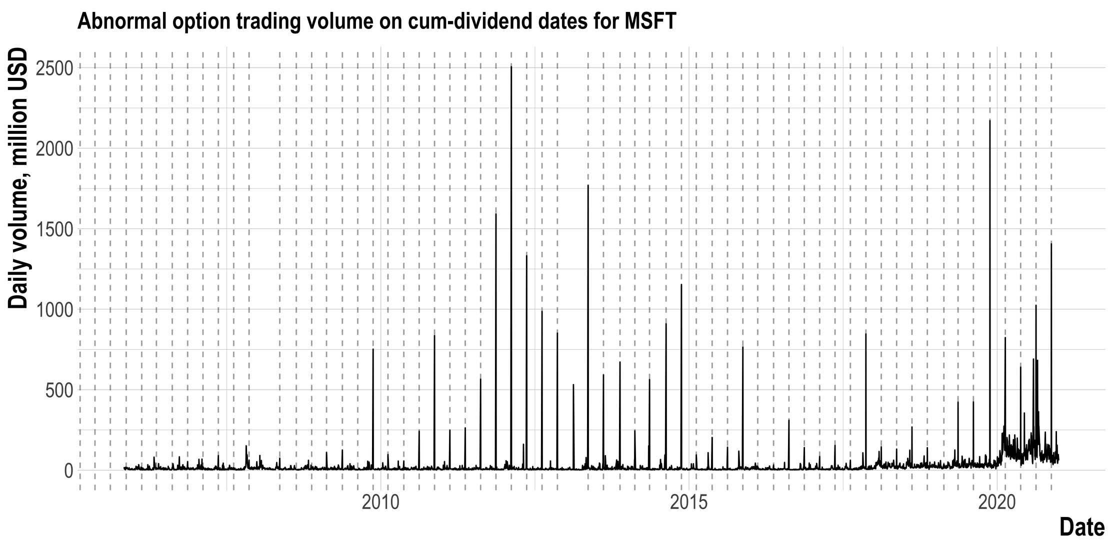
Ticker	Name	SLAN Raw performance, $ billion 1 day	2 days	5 days	10 days	Indirect TC, $ billion	Ticker	Name	Market Raw performance, $ billion 1 day	2 days	5 days	10 days	Indirect TC, $ billion
Top 10 tickers for SLAN trades							**Top 10 tickers for all market trades**						
AMZN	Amazon.com Inc	0.371	0.373	0.310	0.271	-0.203	AMZN	Amazon.com Inc	1.489	1.656	1.583	1.661	-1.756
NVDA	NVIDIA Corp	0.033	0.052	0.092	0.101	-0.056	GME	GameStop Corp	0.909	0.968	1.244	1.051	-0.774
AAPL	Apple Inc	-0.003	0.011	0.027	0.094	-0.106	SHOP	Shopify Inc	0.514	0.560	0.629	0.556	-0.452
SHOP	Shopify Inc	0.063	0.062	0.067	0.070	-0.043	AMC	AMC Entertainment Holdings Inc	0.043	-0.016	0.308	0.363	-0.323
MRNA	Moderna Inc	0.012	0.017	0.027	0.027	-0.022	AAPL	Apple Inc	0.482	0.524	0.388	0.345	-0.779
BABA	Alibaba Group Holding Ltd	0.017	0.028	0.032	0.022	-0.030	NVDA	NVIDIA Corp	0.133	0.221	0.212	0.317	-0.522
DIS	Walt Disney Co	0.000	0.004	0.007	0.020	-0.016	BA	Boeing Co	0.312	0.387	0.265	0.296	-0.478
GOOGL	Alphabet Inc	0.007	0.010	0.032	0.020	-0.032	ZM	Zoom Video Communication Inc	0.279	0.236	0.233	0.231	-0.393
MSFT	Microsoft Corp	-0.006	-0.011	-0.005	0.013	-0.035	ROKU	Roku Inc	0.121	0.191	0.241	0.213	-0.243
CMG	Chipotle Mexican Grill Inc	0.009	0.008	0.008	0.013	-0.014	MRNA	Moderna Inc	0.108	0.080	0.183	0.189	-0.256
Bottom 10 tickers for SLAN trades							**Bottom 10 tickers for all market trades**						
TSLA	Tesla Inc	-0.430	-0.598	-0.882	-0.917	-0.314	TSLA	Tesla Inc	-1.921	-1.820	-1.780	-2.002	-2.935
SPY	SPDR S&P 500 ETF	-0.198	-0.363	-0.377	-0.333	-0.201	SPY	SPDR S&P 500 ETF	-0.423	-0.825	-1.006	-0.676	-1.408
QQQ	Invesco Nasdaq-100 ETF	-0.043	-0.088	-0.077	-0.115	-0.056	TLRY	Tilray Brands Inc	-0.360	-0.358	-0.303	-0.340	-0.102
GME	GameStop Corp	0.030	-0.121	-0.106	-0.109	-0.058	QQQ	Invesco Nasdaq-100 ETF	0.029	-0.042	-0.106	-0.151	-0.500
AMC	AMC Entertainment Holdings Inc	-0.050	-0.074	-0.080	-0.079	-0.027	LQD	iShares iBoxx $ Inv. Grade Corp. Bond ETF	0.013	0.008	-0.040	-0.133	-0.020
MARA	Marathon Digital Holdings Inc	-0.001	-0.002	-0.006	-0.031	-0.006	NFLX	Netflix Inc	-0.054	-0.062	-0.066	-0.063	-0.402
IWM	iShares Russell 2000 ETF	-0.003	-0.005	-0.008	-0.029	-0.017	MARA	Marathon Digital Holdings Inc	0.014	0.006	0.006	-0.061	-0.047
RIOT	Riot Blockchain Inc	-0.004	-0.003	-0.015	-0.027	-0.014	UVXY	ProShares Ultra VIX Short-Term Futures ETF	-0.004	0.030	-0.012	-0.061	-0.152
MSTR	MicroStrategy Inc	-0.023	-0.020	-0.022	-0.027	-0.007	ITB	iShares US Home Construction ETF	-0.002	0.002	-0.002	-0.053	-0.006
PLUG	Plug Power Inc	-0.006	-0.012	-0.020	-0.022	-0.009	NKE	NIKE Inc	-0.012	-0.020	-0.031	-0.050	-0.044

This table reports the performance of top-10 and bottom-10 ticker by their aggregate 10-day SLAN (Market) performance in November 2019 to June 2021. Raw performance at each horizon and the indirect transaction costs (TC) are computed as explained in Section 2.4.

A.19 Abnormal trading volume on cum-dividend dates: Further examples

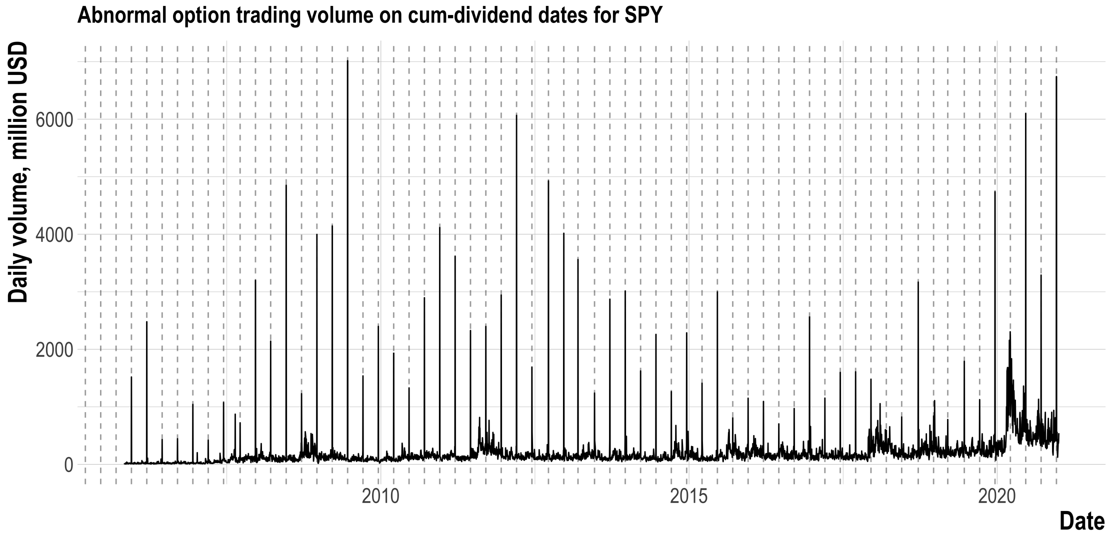
This appendix contains two further examples of abnormal trading volume on cum-dividend dates. The figures below plot daily trading volume of options on Microsoft, MSFT, and on the largest S&P 500 ETF, SPY.

Figure 10: Abnormal trading volume on cum-dividend dates for Microsoft



This figure plots daily trading volume for all call option contracts on MSFT, in millions of U.S. dollars, as reported in OptionMetrics. The dashed lines indicate cum-dividend dates.

Figure 11: Abnormal trading volume on cum-dividend dates for Microsoft



This figure plots daily trading volume for all call option contracts on SPY, in millions of U.S. dollars, as reported in OptionMetrics. The dashed lines indicate cum-dividend dates.

A.20 Dividend play: Another example

Table 31 provides an additional example illustrating the mechanics of the dividend play strategy. Case 1 corresponds to the case when all 1,000 outstanding contracts are exercised, all 1,000 short positions get assigned and so there is no profit for a dividend play strategy to harvest. Case 2 describes what happens if 500 of 1,000 outstanding contracts are left unexercised. Without arbitrageur involvement, half of the short positions in the contract get assigned; the remaining positions deliver a gain of \$0.5 per share and \$25,000 in total for the unassigned short positions, a gain to the original customers with short positions. Now consider the entry of market makers. The market makers attempt to recover most of the potentially harvestable profit of \$25,000. To do so, they buy and simultaneously sell 5,000 contracts and exercise all their long positions. The probability of assignment increases, but, because of the OCC's random assignment, some of the short positions of the market makers remain unassigned and hence yield a gain. In our example, market makers harvest \$20,850 out of the total gain of \$25,000. To divert a larger fraction of the total gain from the original customers with short positions, market makers simply increase the number of contracts they buy and sell.

Table 31: Dividend play: Another Example

	OI_{t-1}	New positions(t)	Available for ex.	No. exercised	Prob. Assign.	No. assign.	No. not assign.	Gain per share	Total gain on unassign. positions	OI_t	Fraction unex.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Case 1. Optimal exercise											
Customer	1000	0	1000	1000	100%	1000	0.00	0.5	0	0.00	
Case 2. Suboptimal exercise											
Case 2.1. Without dividend play											
Customer	1000	0	1000	500	50%	500	500	0.5	25000	500	0.5
Case 2.2. With dividend play											
Customer	1000	0	1000	500		916.7	83.33	0.5	4166.7		
Market makers	0	5000	5000	5000		4583.3	416.67	0.5	20833.3		
Total	1000	5000	6000	5500	92%	5500	500		25000	500	0.5

This table illustrates the dividend play strategy. Date t refers to the cum-dividend date and OI_t stands for the open interest on date t. This table is similar to Table 1 in Pool, Stoll, and Whaley (2008).

A.21 Dividend play: Technical details

We compute the expected call option ex-dividend price using Black-Scholes-Merton formula:

$$\begin{aligned} c_{ex} &= S_{ex} e^{-y(T-t)} N(d_1) - K e^{-r(T-t)} N(d_2) \\ d_1 &= \frac{1}{\sigma\sqrt{T-t}} ln\left(\frac{S_{ex}}{K} + \left[r - y + \frac{\sigma^2}{2}\right](T-t)\right) \\ d_2 &= d_1 - \sigma\sqrt{T-t} \\ y &= Dividend_{ex}/S_{ex} \end{aligned}$$

where S_{ex} is the expected price after the stock goes ex-dividend, i.e., price at close on the cum-dividend day minus expected dividend, $T-t$ is time to maturity in years, i.e., difference in the expiration date and the current date in days divided by 360, K is the contract strike, σ^2 is the annualized implied volatility,[60] and r is the interpolated maturity-specific interest rate provided by OptionMetrics (annualized %), $Dividend_{ex}$ is the expected dividend after the ex-date.[61]

[60] We use the daily contract-level implied volatility from OptionMetrics. If it is missing, we interpolate it from the neighboring strikes.

[61] We assume that its size is equal to the current dividend if the stock pays one more dividend after the current dividend until the option expires and 0 otherwise.

A.22 Dividend play sample: Data filters and calculated variables

We use our dataset described in 2.1 together with the following filters to arrive at the final dividend play sample. We include all call option contracts with $EEV > 0$. Furthermore, since our valuation might be imperfect, we add a market-based filter of the optimality of exercise: We only keep contracts with a decline in open interest on the cum-dividend date.[62] By implication, we only have contracts with non-zero open interest on the cum-dividend date and the date before that.

Following the early papers on dividend play, we remove contracts with no trading volume on cum-dividend date. Additionally, we remove contracts expring immediately after the ex-dividend.[63]

To measure arbitrageur activity, we use floor trading share, defined as the total volume in transactions of OPRA types SLFT and MLFT, divided by the total volume on the cum-dividend date.[64] For both SLAN and Small Share, we compute a one-week moving average and use its lagged value on the cum-dividend date. We use the same rolling measures for the retail activity variables described in the main text, as well as volume, spread, and implied volatility controls.

We compute relative spread quoted at the time of each option trade as $2(best\,ask - best\,bid)/(best\,ask + best\,bid)$ (relative to the midpoint price). We compute moneyness of the trade as $0.5(underlying\,bid + underlying\,ask)/strike - 1$.[65]

[62] This is consitent with Hao, Kalay, and Mayhew (2009).
[63] The last filter does not change results significantly.
[64] In unreported tests, we confirm that using dollar volume based measures instead yields similar results.
[65] In the absence of TAQ data, we use underlying bid-ask midpoint as a high-frequency price.

Table 32: Dividend play sample descriptive statistics

	Mean	Median	St. Dev.	p1	p99
Fraction of OI not exercised, %	23.38	6.98	31.02	0.00	99.32
Floor trades volume share on cum-date	0.56	0.85	0.47	0.00	1.00
D(floor share > 0)	0.61	1.00	0.49	0.00	1.00
SLAN Share	0.15	0.15	0.16	0.00	0.65
Small Share	0.85	0.88	0.18	0.27	1.00
Internalized volume in equities	0.17	0.16	0.05	0.07	0.30
WSB mentions	8.20	0.33	36.15	0.00	232.67
OI, log	4.97	4.89	1.99	0.69	9.80
Early exercise value (EEV), $	0.49	0.30	0.59	0.00	2.79
Market EEV, $	0.08	0.02	0.39	-0.41	1.12
Dollar potential profit	7,059.9	215.4	57,949.3	0.00	121,549.1
Dollar volume, log	4.16	4.09	1.40	1.17	7.85
Relative spread	0.07	0.04	0.09	0.00	0.46
Implied volatility, annualized	0.41	0.36	0.28	0.07	1.29
Moneyness	5.56	3.99	5.57	0.47	29.27
Days to expiry	67	21	124	4	625

This table reports descriptive statistics for all contracts in the dividend play sample (21,997 observations). SLAN and small share are the contract-level volume shares of SLAN and small trades, respectively, averaged over one trading week before the cum-dividend date. Internalized volume in equities is the ticker-level share of volume executed in the non-ATS OTC space relative to the total trading volume, averaged over one trading week before the cum-dividend date. WSB mentions is the number of underlying ticker mentions on `WallStreetBets` forum, averaged over one trading week before the cum-dividend date. Relative spread is options contract quoted spread at the time of the trade relative to the midpoint price. Implied volatility is as reported in LiveVol, interpolated using nearest strikes if missing. Moneyness is measured as $(Midpoint\, Price - Strike)/Strike$.

A.23 Dividend play profits by ticker

Table 33: Dividend play profits by ticker

Ticker	Profit, USD Harvested	Forgone	No. contracts Fully harvested	Partly harvested	Forgone	Traded volume (contracts)
Ticker	(1)	(2)	(3)	(4)	(5)	(6)
SPY	2,395,306.0	23,600,000.0	228	38	991	722,404
AAPL	3,464,514.0	8,591,433.0	302	129	238	849,464
EEM	10,800,000.0	4,907,643.0	142	5	46	5,266,442
IWM	1,484,462.0	3,528,205.0	51	3	162	512,940
EFA	2,436,712.0	3,324,776.0	97	7	25	1,087,908
XLE	1,794,247.0	2,844,799.0	171	7	91	447,875
VALE	1,913,317.0	2,737,224.0	66	6	11	1,876,400
QQQ	105,585.1	2,039,488.0	23	2	208	27,750
EWZ	3,374,398.0	1,330,285.0	87	1	31	1,327,432
KO	440,711.1	1,097,762.0	76	23	66	322,120
HYG	36,411.0	912,840.1	11	4	57	63,710
SAN	-	753,484.1	0	0	11	-
HD	936,386.6	674,275.4	95	21	96	197,887
COST	1,703.4	658,710.4	11	4	43	1,207
XLF	478,434.8	620,190.3	57	5	57	344,130
IBM	684,029.4	567,198.3	116	116	36	383,594
BHP	176,163.5	553,367.0	33	4	12	57,055
DIA	156,767.9	539,148.9	60	8	146	17,401
ET	620,329.4	529,149.1	51	12	48	574,990
QCOM	749,520.4	497,014.1	68	16	49	426,659
GOLD	313,229.0	449,155.9	27	4	32	68,580
VIAC	1,770,252.0	420,720.0	97	1	51	437,395
XOM	8,123,625.0	404,302.2	242	82	62	1,734,910
XLI	10,843.2	401,816.1	12	1	15	17,123
RIO	27,621.1	375,591.7	16	4	5	57,782
XLP	116,856.2	370,016.7	16	0	32	15,990
T	2,700,908.0	369,322.1	155	35	47	2,381,173
JPM	883,655.1	365,889.3	80	32	37	1,096,394
CVX	623,103.9	320,866.8	234	86	91	419,757
FXI	877,237.6	309,056.5	77	4	18	1,242,431
GILD	355,846.2	308,605.1	65	23	43	280,310
MRO	-	307,556.9	0	0	23	-
NVDA	-	283,368.6	0	0	57	-
BP	339,841.9	277,599.5	99	28	52	209,456
DIS	836,205.3	273,563.9	41	5	1	503,899
PGR	661,473.3	263,863.8	12	5	20	56,496
MPC	561,419.5	251,516.0	105	47	64	418,852
TGT	90,064.5	241,550.0	77	29	66	65,810
DOW	138,056.3	231,155.0	32	13	61	92,369
PRU	68,932.0	224,236.2	33	29	18	85,664
Total	**50,548,168.7**	**66,756,745.0**	**3,165**	**839**	**3,219**	**23,691,759**

This table reports the top-40 tickers in terms of dividend play profits forgone by floor traders in our sample. Values are aggregated across all contracts within a ticker in 11/2019-06/2021. Total dividend play potential profits are computed as in Equation (3). To compute 'harvested' profits, we multiply the total profits by the floor volume share on cum-dividend date, and attribute the residual to 'forgone' profits. No. of 'fully harvested' contracts in column (3) is the number of contracts with floor share above 90%, and in column (5) – with zero floor share.[a] Traded volume in column (6) is the total floor trading volume in all contracts.

[a]The average floor share is over 99% in 'fully harvested' contracts and 69% in 'partly harvested' contracts.

A.24 Retail trading measured over a longer window

In this appendix, we redefine our retail trading measures. Instead of measuing shares of retail trading in the dollar trading volume in options over one trading week preceeding a cum-dividend date, we measure them over two trading weeks. Tables 34 and 35 are the analogs of Tables 7 and 9, respectiviely, but with the redefined measures of retail trading.

Table 34: Suboptimal exercise and retail investor popularity

	Dividend play profitability feature					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Fraction of OI not exercised, %						
SLAN share	7.460***		7.280***	7.689**	7.633***	7.628***
	(3.46)		(3.36)	(2.34)	(3.57)	(3.41)
Small share		5.120**	4.767**	12.893***	5.818***	5.041**
		(2.23)	(2.06)	(3.94)	(2.62)	(2.00)
Non-ATS OTC share						25.559
						(1.19)
WSB mentions, log						0.557*
						(1.69)
Observations	21,105	21,105	21,105	6,942	21,105	19,134
Adjusted R-squared	0.238	0.237	0.238	0.345	0.214	0.214
Panel B. Potential profits, log U.S. dollar						
SLAN share	1.970***		1.954***	2.238***	2.078***	2.040***
	(7.56)		(7.52)	(5.46)	(8.10)	(7.59)
Small share		0.568	0.485	2.504***	0.520	0.531
		(1.61)	(1.40)	(3.49)	(1.62)	(1.47)
Non-ATS OTC share						1.388
						(0.69)
WSB mentions, log						0.036
						(0.87)
Observations	21,105	21,105	21,105	6,942	21,105	19,134
Adjusted R-squared	0.315	0.311	0.315	0.318	0.293	0.299
Sample	All	All	All	Top EEV tercile	All	All
FE	Ticker*Date	Ticker*Date	Ticker*Date	Ticker*Date	Ticker and Date	Ticker and Date
Contract controls	Y	Y	Y	Y	Y	Y
Ticker controls	N	N	N	N	Y	Y

This table reports estimates of (4) in our dividend play sample. SLAN Share and Small Share are the contract-level volume shares of SLAN and small trades, respectively, averaged over two trading weeks before the cum-dividend date. Non-ATS OTC share is the ticker-level share of volume executed in the non-ATS OTC space relative to the total trading volume, averaged over two trading weeks before the cum-dividend date. WSB mentions, log, is the logarithm of total mentions of the ticker on *WallStreetBets* forum. In Panel B, contract controls include: log dollar trading volume, relative spread, IV, moneyness, days to expiry. In Panel A, they additionally include log OI and EEV. Ticker controls include: underlying price, underlying volatility, underlying relative bid-ask spread, underlying market cap. S.E. are clustered by ticker and date. Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 35: Arbitrageur activity and retail investor popularity: Retail trading measured over a longer window

	Floor trading share on cum-date				**D(floor share > 0)**	**Floor trading volume, log**
	(1)	(2)	(3)	(4)	(5)	(6)
SLAN share	0.051*		0.038	0.057	0.033	0.604**
	(1.82)		(1.45)	(1.21)	(1.22)	(2.57)
Small share		0.338***	0.336***	0.126	0.351***	4.848***
		(6.05)	(6.07)	(1.22)	(6.24)	(8.26)
Observations	21,105	21,105	21,105	6,942	21,105	21,105
Adjusted R-squared	0.404	0.410	0.410	0.473	0.399	0.479
Sample	All	All	All	Top EEV tercile	All	All
Contract controls	Y	Y	Y	Y	Y	Y

This tables reports estimates of (5) in our dividend play sample. SLAN Share and Small Share are the contract-level volume shares of SLAN and small trades, respectively, averaged over two trading weeks before the cum-dividend date. Contract controls include: log OI, EEV, log dollar trading volume, relative spread, IV, moneyness, days to expiry. All regressions include ticker by date fixed effects. Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1

A.25 Dividend risk and automatic actions of retail brokerages

This appendix presents an example of an automatic action to close short positions exposed to dividend risk on cum-dividend dates undertaken by retail brokerages. The example is from the Terms and Conditions of Robinhood.

Figure 12: Excerpt from Robinhood's Terms and Conditions

Options Dividend Risk

Dividend risk is the risk that you'll get assigned on any short call position (either as part of a covered call or spread) the trading day before the underlying security's ex-dividend date. If this happens, you'll open the ex-date with a short stock position and actually be responsible for paying that dividend yourself. You can potentially avoid this by closing any position that includes a short call option at any time before the end of the regular-hours trading session the day before the ex-date.

Robinhood may take action in your account to close any positions that have dividend risk the day before an ex-dividend date. Generally, we'll only take action if your account wouldn't be able to cover the dividend that would be owed after an assignment. This is done on a best-efforts basis.

A.26 Covariate balance for matching

Figure 13: Covariate balance for SLAN share in Table 10

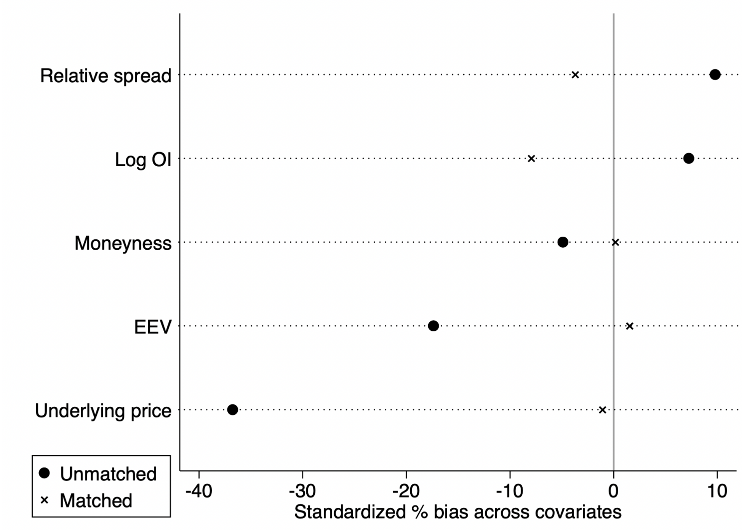
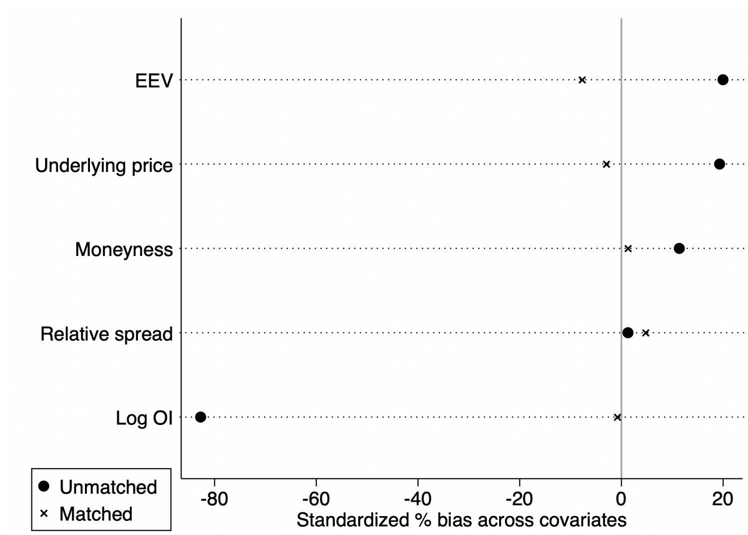
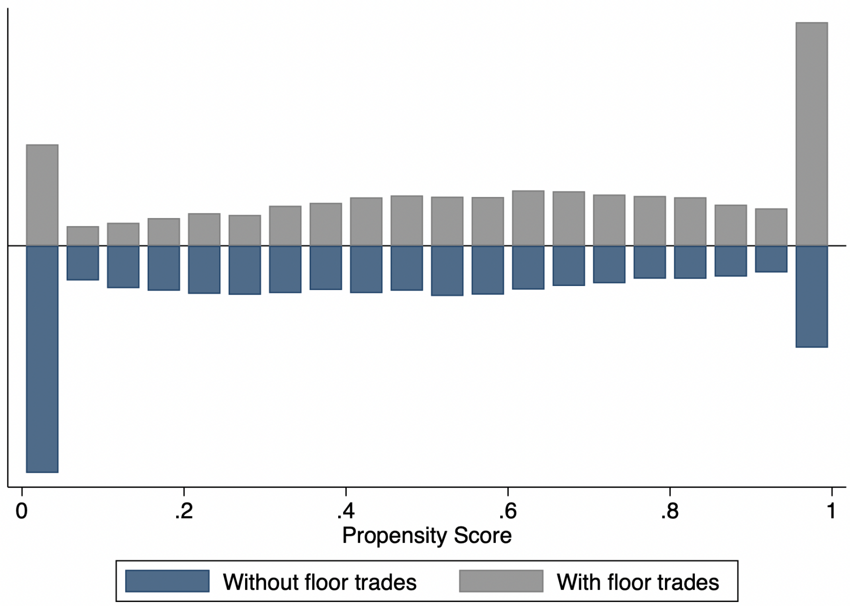


Figure 14: Covariate balance for Small share in Table 10



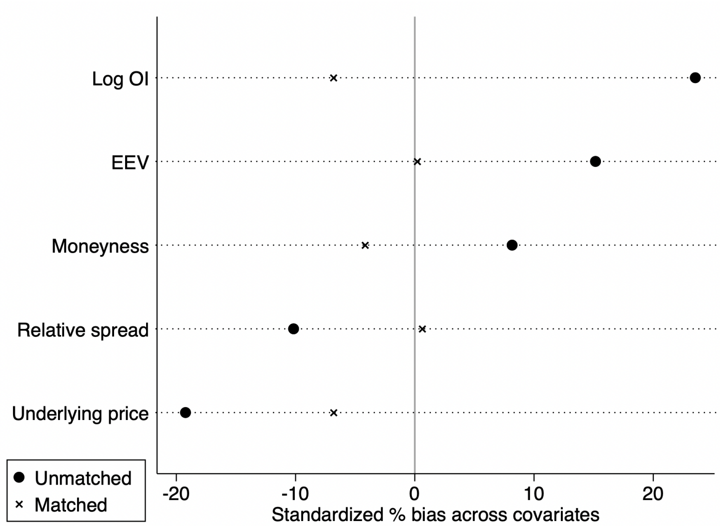
A.27 Dividend play puzzle in matched contracts

Figure 15: Floor traders' entry across propensity score levels



This figure depicts the number of contracts with and without floor trades across the scores of propensity to have floor trades. The propensity scores are based on the full set of controls: log OI, EEV, log trading volume, relative spread, IV, moneyness, days to expiry, underlying price, underlying volatility, underlying relative bid-ask spread, underlying market cap. We report the balance tests in Appendix A.26.

Figure 16: Covariate balance for Floor share in Figure 15

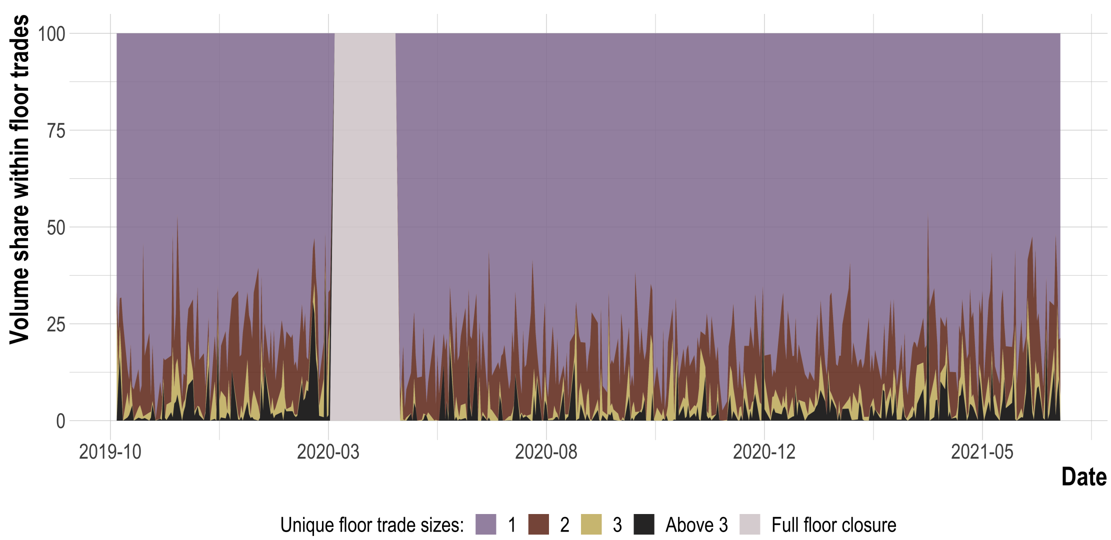


A.28 Few arbitrageurs engaging in dividend play

This appendix provides suggestive evidence for the number of arbitrageurs simultaneously engaging in a dividend play strategy in a particular contract. Figure 17 plots a percentage split of dividend play trades by unique trade sizes, which is our proxy for a number of arbitrageurs engaging in dividend play in each contract.

The gray shaded area in Figure 17 corresponds to the closure of all exchange floors in the U.S. due to the COVID-19 pandemic. Our measure of floor trading is indeed zero over this period. Furthermore, the total trading volume on cum-dividend dates during the closures is the same as on any other day, which provides additional validation of the measure. Even when PHLX floor was closed but ARCA and BOX floors were open, the mean trading volume on cum-dates was an order of magnitude lower.

Figure 17: Floor trading by number of floor trade sizes



This figures depicts percentage split of trades executed on exchange floor by the number of unique trade sizes. We only include contracts in our dividend play sample. The gray shaded area corresponds to the period of floor closures on all exchanges.

A.29 Big Three share and floor trading, matching approach

Similar to Table 10 in the main text, we employ a matching approach to study the importance of concentration in PFOF market for the floor trading share on cum-dividend date. For matching, we use the same characteristics as in the main text. The corresponding covariate balance plot is presented below.

Table 36: Arbitrageur activity and market concentration: Matched contracts

	Floor trading share on cum-date			
	Matched			OLS
	(1)	(2)	(3)	(4)
D(Big Three share > 10%)	-0.031***	-0.032***	-0.034***	-0.038***
	(-2.71)	(-2.98)	(-3.06)	(-3.84)
Observations	21,105	21,105	21,105	21,105
No. neighbors	1	10	10	
Short controls	Y	Y	Y	Y
Extended controls	N	N	Y	Y

Columns (1)-(3) report ATE. Big Three share is the total share of the Big Three internalizers' non-ATS OTC volume in the total stock trading volume over the past trading week. Short controls include: log OI, EEV, moneyness. Extended controls include relative spread and underlying price. Robust z-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 18: Covariate balance for the top decile of Big-Three share dummy in Table 36