The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXVI, NO. 5 • OCTOBER 2021

Many researchers have concluded that retail equity investors are generally uninformed and make systematic mistakes when selecting equity investments (see, for example, Barber and Odean (2000, 2008)). More recent studies, however, suggest otherwise (Kaniel, Saar, and Titman (2008), Barber, Odean, and Zhu (2009), Kaniel et al. (2012), Kelley and Tetlock (2013), Fong, Gallagher, and Lee (2014), Barrot, Kaniel, and Sraer (2016)). These studies show that retail investors' trading can predict future stock returns. Unfortunately, most existing studies of retail order flow are based on proprietary data sets with relatively small subsets of overall retail order flow. For example, Barber and Odean (2000) use data from a single U.S. retail brokerage firm, while Barber and Odean (2008) examine individual investor trading data from a total of three different retail or discount brokerage firms. Kelley and Tetlock (2013) use data from a single U.S. wholesaler, Fong, Gallagher, and Lee (2014) analyze data from the Australian Securities Exchange (ASX), and Barrot, Kaniel, and Sraer (2016) use data from a single French brokerage firm. Kaniel, Saar, and Titman ((2008), Kaniel et al. (2012), and Boehmer, Jones, and Zhang (2008) use proprietary account-type data from the NYSE during the early 2000s. During that period, only a small number of brokerages sent their retail order flow to the NYSE, and thus, the NYSE's market share of overall retail order activity was (and has remained) quite small.

In existing work, many researchers use trade size as a proxy for retail order flow. Before the spread of computer algorithms that "slice and dice" large institutional parent orders into a sequence of small child orders, small trades were much more likely to come from retail customers, while institutions were likely behind the larger reported trades. For example, Lee and Radhakrishna (2000) use a $20,000 cutoff to separate smaller individual trades from larger institutional trades. More recently, Campbell, Ramadorai, and Schwartz (2009) allow the cutoff to vary using through a regression approach that is calibrated to observed quarterly changes in institutional ownership, but they maintain the same basic assumption that small trades are more likely to arise from individual trading. However, with algorithms becoming an important feature of institutional order executions in the early 2000s, a trade-size partition has become far less useful as a proxy for retail order flow. Indeed, the tendency of algorithms to slice orders into smaller and smaller pieces has progressed so far that during our recent sample period, the retail order flow that we identify actually has a slightly larger average trade size than other flow.

Given the current automated and algorithm-driven market structure, researchers need an alternative measure to isolate retail order flow. We introduce such a measure in this paper. As one of our main contributions, we show that our measure can identify a broad swath of marketable retail order flow. Our measure builds on the fact that, due to regulatory restrictions in the United States and the resulting institutional arrangements, retail order flow—but not institutional order flow—can receive price improvement, measured in small fractions of a cent per share. We use this observation to identify marketable retail price-improved orders from the Trade and Quote (TAQ) data, a publicly available data set that contains all transactions for stocks listed on a national

exchange in the United States. Specifically, we identify trades that execute at share prices with fractional pennies. Most such price-improved transactions take place off-exchange and are reported to a Trade Reporting Facility (TRF). Using these TRF data, we identify transactions as retail buys if the transaction price is slightly below the round penny and as retail sells if the transaction price is slightly above the round penny. This approach separates retail investors' marketable orders from those of institutions because institutional trades generally cannot receive this type of fractional penny price improvement.[1] We discuss our approach in greater detail in Section I.B. Notice that our retail order flow measure only includes marketable orders, that is, it does not include limit orders. Overall, we believe that our method of retail trade identification is conservative, and we cross-validate the accuracy of our approach using a small sample of NASDAQ TRF audit trail data.

We analyze retail marketable order flow from the U.S. equity market for the six years between January 2010 and December 2015. We find that retail investors are slightly contrarian at a weekly horizon, and that the cross-section of weekly marketable retail order imbalances predicts the cross-section of returns over the next several weeks, consistent with the findings of Kaniel, Saar, and Titman (2008), Kaniel et al. (2012), Kelley and Tetlock (2013), Fong, Gallagher, and Lee (2014), and Barrot, Kaniel, and Sraer (2016) but inconsistent with the findings of many others.

The predictability of marketable retail order flow for future returns is consistent with persistence in retail order flow, liquidity provision, and informed trading. We conduct a decomposition exercise and separate the marketable retail order imbalance into proxies for these three components. The empirical findings show that persistence in order flow and order flow driven by return reversals (our proxy for liquidity provision) account for about half of the predictive power of the marketable retail order imbalance for future returns. We attribute the other half of this predictive power to potential informed trading.

We also investigate the nature of the information embedded in retail trading. Our results show that the marketable retail order imbalance is positively correlated with some firm-level surprises in public news, and that marketable retail order flow has predictive power beyond public news, which suggests (but only suggests) that retail investors may possess firm-level information that is not yet incorporated into prices.

Finally, we conduct a battery of robustness checks and provide further discussion. We find that our results are robust, and we provide additional evidence that, despite the predictive power of marketable retail order flow in the cross-section, aggregate marketable retail flows cannot predict future market returns.

[1] In contrast, institutional trades often occur at the midpoint of the prevailing bid and ask prices. If the bid-ask spread is an odd number of cents, the resulting midpoint trade price ends in a half-penny. Many of these midpoint trades take place on crossing networks and are reported to the TRF. Thus, trades at or near a half-penny are likely to be from institutions and are not assigned to the retail category.

Given the nature of our data, our work is also related to recent studies of off-exchange trading in the United States. For instance, Kwan, Masulis, and McInish (2015) study the competition between traditional stock exchanges and new dark-trading venues and find that the minimum pricing increment regulation (typically one penny) drives orders to dark pools and limits the competitiveness of the exchanges. Battalio, Corwin, and Jennings (2016) examine make-take fees and how brokers route order flow, and suggest that current order-routing practices may not maximize the quality of limit order execution. Menkveld, Yueshen, and Zhu (2017) directly investigate the pecking order of trading venues in dark pools and document that investors strategically put low-cost, low-immediacy orders in front of high-cost, high-immediacy orders.

Compared to the earlier literature on retail orders and studies of off-exchange trades, we make three main contributions. First and most importantly, we propose a novel methodology for identifying and signing marketable retail trades using publicly available data with substantial coverage. Second, we show that the marketable retail trades that we identify can predict the cross-section of future stock returns. Third, we analyze the nature of the predictive power of marketable retail order flow and show that half of its predictability is likely driven by order imbalance persistence and liquidity provision, while the other half is consistent with informed trading. We also track potential informed trading to different types of news and provide suggestive evidence on the nature of the information possessed by these retail investors.

Two studies, Kaniel, Saar, and Titman (2008) and Kelley and Tetlock (2013), study similar questions and are closely related to our research, but they employ different data and draw different interpretations. For instance, using proprietary data from the NYSE between January 2000 and December 2003, Kaniel, Saar, and Titman (2008) document that retail order flows can predict stock returns. Kaniel, Saar, and Titman (2008) examine the contemporaneous relation between their retail order flows and stock returns. They find that the contemporaneous return is significantly positive for stocks that retail investors sell and negative for stocks that they buy, which is consistent with a liquidity provision interpretation and inconsistent with the information story. We follow their approach using our new marketable retail order flow variables. We are able to replicate the predictive relation between retail order flow and future stock returns, but our results for the contemporaneous relation are different: the contemporaneous return is significantly negative for stocks that retail investors sell and positive for stocks that they buy when they use marketable orders. Our findings are more in line with an information interpretation than a liquidity provision interpretation.

Kelley and Tetlock (2013) obtain data from a major retail wholesaler between February 2003 and December 2007. Their data allow them to separate retail orders into market orders and limit orders. They find that both retail market orders and limit orders can predict future stock returns, but for different reasons. The aggressive market orders can correctly predict future news, suggesting that these trades are informed, while the passive limit orders are contrarian, consistent with liquidity provision. Our marketable retail order

flow measure only identifies market orders, and for these marketable orders, we follow their tests and replicate their results. In addition, we decompose our marketable retail order imbalance into components related to order flow persistence, contrarian trading, public news, and a residual, which potentially contains nonpublic information. The decomposition exercise shows that public news contributes little to the predictive power of marketable retail trades, whereas the residual part is more important. With more recent data and wider coverage, our study provides interesting new findings that complement the studies by Kaniel, Saar, and Titman (2008) and Kelley and Tetlock (2013).

The remainder of this paper is organized as follows. We describe the data and our identification method in Section I. Section II presents our main empirical results. We provide further discussion of the results and perform robustness and plausibility checks in Section III. Finally, Section IV concludes the paper.

I. Identifying Retail Order Flows

As we note in the introduction, our most important contribution is to provide a simple, new method to identify a wide swath of marketable retail order flow using publicly available equity transaction data. We introduce our data sources in Section I.A. In Section I.B, we provide the institutional background. Sections I.C and I.D report summary statistics and results of cross-validation tests, respectively.

A. Data Sources

From TAQ trade data, we start with only trades that occur off-exchange, designated with exchange code "D." We merge these TAQ data with stock returns and accounting data from CRSP and Compustat, respectively. We include only the common stocks with share code 10 or 11 (which mainly excludes ETFs, ADRs, and REITs) listed on the NYSE, NYSE MKT (formerly Amex), and NASDAQ. We remove low-priced stocks by requiring that the minimum stock price be $1 at the previous month-end.

Our sample spans the period January 3, 2010 to December 31, 2015. Data on subpenny price improvement actually extend back to 2005. In Internet Appendix Figure IA.1[2], we plot the time series from January 2005 (the start of Regulation National Market System, or Reg NMS, which established the current regulatory framework for subpenny price improvement in the United States) to December 2017. We choose to study the 2010 to 2015 period for two reasons. First, during the first few years under Reg NMS, there is a strong upward trend in the number of subpenny trades, possibly because an increasing number of brokerage firms were adopting the practice of providing fractional cents of price improvement to retail investors via internalization or wholesalers. The upward trend disappears and stabilizes after 2009. Second, from

[2] The Internet Appendix is available in the online version of this article on *The Journal of Finance* website.