UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Investigating the Nature of Retail Investor Activity

A thesis presented for the degree of

Bachelor of Financial Economics

Special thanks to Marcin Zamojski for the tireless guidance and support throughout the process of writing this thesis. We are very happy to have had you as our supervisor!

Authors:	Johan Birgersson
	Samuel Carlén
Supervisor:	Marcin Zamojski
Submission Date:	June 7, 2021

Contents

Abstract

We study the nature of retail investor activity and how this differs when looking at periods in connection with quarterly reports and when looking at a special period on the market like Covid-19. Further, we investigate how this differs between S&P500 stocks and meme stocks. We find that retail order imbalances can predict positive future returns for S&P500 stocks and find suggestive evidence that this doesn't differ significantly for periods in connection with reports. We find that this doesn't differ significantly during the initial impact of Covid-19. We do however find evidence that retail order imbalances predict negative returns for meme stocks, contrary to the behavior exhibited in the S&P500 stocks and in previous literature.

1

Introduction

We expand on the literature surrounding the nature of retail investor activity by looking at patterns of retail trade between 2016 to 2021. We look at how the predictability of returns for retail order imbalances differs in in periods in connection with quarterly reports and in connection with the initial impact of Covid-19. We also investigate how this differs between S&P500 stocks and "meme stocks" like GameStop or AMC Entertainment.

Retail trading has been widely debated among academics and regulators for a long time, but even more so recently as an emergence of e-trading platforms like Robinhood or Fidelity have enabled easier access to capital markets for retail traders (Martin and Wigglesworth 2021). Low interest rates, high volatility creating potential for large gains in a short time and "lock-down boredom", are all factors fueling an increase in retail investor activity (Shalchi 2021). As retail investor activity has been increasing, we have witnessed several cases of unusual activity in so-called "meme stocks". For example, in 2020 Tesla rose 700% and Gamestop rose from $18.84 at the end of 2020 to $325 on Jan. 29 2021 to then decline to $90 within the first two trading days in February (Lazzara 2021).[1] According to Gobler (2021) meme stocks often become overvalued and see drastic price increases in a short amount of time. The price behaviour witnessed in

[1]A meme stock is a stock that sees an increase in volume due to the hype on social media and online forums like Reddit rather than due to the company's performance. They are generally popular among retail traders.

these stocks aligns with previous findings on retail trading from Barber & Odean (2000), Barber, Odean and Zhu (2008) and Han and Kumar (2013) who show retail trading to be speculative by nature and that stocks with high retail proportion are more likely to have "lottery features", attracting retail traders with a propensity for gambling. This behaviour can be seen frequently on online forums such as Wall-Street Bets, where users post their speculative investments while encouraging others to do so as well through "YOLOs". YOLO stands for you only live once and in this case signals gambling big by for example going all in on one stock. (Keshner 2021)

Some literature also claims that retail trading is contrarian by nature (Boehmer et al. 2020) and that it leads to retail investors unwillingly providing liquidity and therefore losing money (Kaniel, Saar, and Titman 1999). However, they also find that retail activity can accurately predict future returns, further validated by findings from both Kelley and Tetlock (2013), and Boehmer, Jones, Zhang and Zhang (2020)[2]. BJZZ also find evidence that stocks with a high retail order imbalance [3] outperform stocks with negative imbalances over the following week. Moreover, they find suggestive evidence that retail marketable orders contain firm-level information that is not yet included in the stock price, implying that retail trades are informed or that our definition of market efficiency is incorrect.

Contradictions in previous literature coupled with recent developments on the market in for example meme stocks makes retail investor activity an important and interesting subject to investigate with many unanswered questions.

The research area on understanding of how retail traders interact with and cause events on the market is of interest from a range of standpoints and for multiple applications. From an economic sustainability standpoint, it provides a basis for decisions on whether retail investors need to be further regulated for their own protection or, on the contrary, if a democratization of the capital markets is purely positive as it can contribute to

[2]Henceforth referenced as BJZZ

[3]A stock ending a given day with retail buys exceeding retail sells.

increased equality. Market participants may be interested in finding out if patterns in retail trading can help predict market movements and if this should be accounted for in their financial models.

To contribute to this area of research, this study aims to investigate whether the predictability of returns through retail order imbalances differs between large caps and meme stocks. Furthermore, it will look at how the effect differs for periods in connection with quarterly reports compared to non quarterly report periods as well for the period during the initial impact of Covid-19 compared to non-Covid-19 periods. Thus, the following hypotheses will be investigated:

1. Retail order imbalances can predict future returns.
2. When decomposing order imbalance into the three components OIB contrarian, OIB persistence and Other, these will all contribute to the prediction of returns but to a varying degree.
3. The predictability of returns will differ when looking at:
   (a) S&P500 large caps compared to meme stocks
   (b) Trading days in connection with reports compared to non-report trading days
   (c) Trading days in connection with the initial impact of Covid-19 compared to other trading days

Previous research both on order imbalances in general (Chordia, Roll, and Avanidhar Subrahmanyam 2000) and retail order imbalances (Boehmer et al. 2020) & (Farrell et al. 2020) indicates that retail order imbalances can be used to predict future returns and thus we hypothesize that this will be the case in our paper as well. (Boehmer et al. 2020) find evidence that both retail order imbalance persistence and retail order imbalance other contribute to the predictive power of order imbalances while not being able confirm whether contrarian trading contributes significantly. We believe that we will find similar results when decomposing the order imbalance into the three components.

Further, we expect differing results when accounting for the three given circumstances above. We expect that the patterns of retail trading and the ensuing predictability of future returns will be significantly different for meme stocks when compared to S&P500 large caps due to the gambling like features features they possess. We expect different results in connection with reports as previous research finds that the informativeness of retail order flow increases in connection with the release of new information (Farrell et al. 2020). Finally, we hypothesize that the initial impact of Covid-19 and the ensuing fear and panic will have shifted the patterns of retail traders to such a degree that the predictability of returns will also have become significantly different.

Our thesis uses retail order imbalances to look at the nature of retail investor activity and how it can be used to predict future returns. Our findings suggest that retail order imbalances can be used to predict future returns, although this differs when accounting for particular circumstances. Specifically, when we look at S&P500 large caps compared to meme stocks, we find that the large caps exhibit similar characteristics as seen in previous research while the meme stocks predict returns a week ahead negatively. This is interesting as it indicates that when retail traders buy meme stocks, they are generally losing money at a one week-ahead period. These findings could support claims of a need for better financial literacy for retail investors. Though we do not rule it out, we find no significant evidence supporting that either Covid-19 or periods in connection with quarterly reports affect the informativeness of retail order imbalances on future returns any differently than "normal" periods.

The paper is structured as follows. We look at previous literature related to our subject in section 2. In section 3 we specify our method of data collection, our retail order imbalance identification and data analysis. In section 4 we show and analyse our results. In 5 we discuss our results further while comparing them to previous research and discuss what they indicate. We conclude the paper in section 6 and in section A we include an appendix of additional regressions and results.

2

Literature Review

2.1 Trading Strategies

Contrarian and momentum trading are two common strategies employed by investors, founded on the premise of large under- or overreactions to news or research due to behavioral biases (Daniel, Hirshleifer, and Avindhar Subrahmanyam 2002; Barberis, Shleifer, and Vishny 1999).

Contrarian trading can be explained by the overreaction hypothesis; the hypothesis states that investors tend to react excessively to information and cause disproportionate stock movements. The contrarian trader tries to capture the reversed movement as the overreaction is corrected. The profitability of contrarian trading strategies has been subject to debate among academics; (Lehmann 1990) finds that in the short term, contrarian strategies yield significant positive returns even when correcting for transaction costs while Conrad, Gultekin, and Kaul (1997) contest this and claim that the positive returns are rather caused by bid-ask bounce in transaction prices. When accounting for bid-ask bounce, they find that all profits from price reversals are eliminated for NASDAQ-NMS stocks and most of the profits for NYSE/AMEX stocks are eliminated.

Momentum trading on the other hand, can be explained by the under-reaction hypothesis meaning that investors react to information to a lesser degree than they should, leading to prices not adjusting enough. The momentum trader tries to capture the continuous positive or negative effect, the momentum, following news announcements or the releases of research. It is widely agreed upon within the literature that momentum trading is a profitable trading method but the reason for this is more up for debate (Galariotis 2013).

2.2 Retrieving information from trade data

A common issue when working with trade data is identifying the origin of a trade as this can't be retrieved directly from the data. Therefore, researchers looking at patterns of a specific group of traders need to use estimations or proxies for these instead. For example when identifying retail order flow, researchers commonly use the trade size as a proxy. However, with the increased prevalence of algorithmic trading and with institutional investors, to a larger degree, placing more frequent but smaller orders, this proxy becomes ineffective. (BJZZ) In order to counter these issues and to get a more accurate measurement of retail order flow, BJZZ propose using retail order imbalances. An order imbalance occurs when the distribution of buy and sell orders is slanted heavily in one direction. As shown by Chordia, Roll and Subrahmanyam (2000), order imbalances can, when aggregated on a daily basis, be used as a means to explain future returns. They find that order imbalances increase following market declines and reverse following a market increase, indicating that traders are on average contrarian in their trading behaviour. Further, they find that order imbalances in either direction, excess buy or sell orders, reduce liquidity.

In order to calculate order imbalances, researchers must first determine whether a trade is a market buy or sell. Similarly to the origin of the trade, this can not be identified directly from the trade data. Two of the most common methods to solve this issue are called the Tick Rule (TR) and the Lee-Ready algorithm (LR) (Chakrabarty, Pascual, and Shkilko 2015). TR classifies a trade as buyer-initiated when the trade price is above the previously traded price and as seller-initiated when the traded price is below the previously trade price. If the trade price is the same as the previous trade price, the rule looks for the closest previous price difference from the current trade price and classifies it based on the same premise. LR uses both quotes and price changes to sign trades.[1] Both TR and LR often require proprietary data as well as being very time consuming for the researcher.[2]. A more recent method called Bulk Volume Classification (BVC) by Easley, Prado, and O'Hara (2012) claims to provide similar accuracy while solving these issues. Chakrabarty, Pascual, and Shkilko (2015) test the validity of these claims by comparing the accuracy of BVC with that of both TR and LR.

In doing so, they find that both TR and LR are significantly more accurate in trade classification for equities than BVC for most of the common research issues. For identifying order imbalances, accuracy using TR/LR ranges from 52% to 92.3% while BVC ranges from 13.9% to 76.6%.

2.3 Retail Investor Activity

Boehmer et al. (2020) propose a method for identifying trades initiated by retail investors and their direction using US transaction data. In this method, they take advantage of the fact that most marketable retail order flows in U.S. equity markets are internalized or sold to wholesalers, and are per SEC 612 regulation subjected to price improvement, making them easily identifiable. This price improvement is observable as a small fraction on the penny of a share. Institutional orders almost never receive this kind of price improvement. Thus it becomes possible to use sub-penny trade prices to identify marketable retail order flows.

The trading direction is similarly easily identified. Purchases can be distinguished by

[1] Our algorithm uses the quote rule to sign trades that occur anywhere other than the midpoint, and the TR to sign trades that occur at the midpoint. The quote rule is applied as follows: if a trade occurs above (below) the midpoint, it is classified as buyer-(seller-) initiated.

[2] This paper has access to a limited set of proprietary data and can thus only use this trade signing algorithm for 2020-01-02 to 2021-04-27

having prices below a round penny, while sales have prices just above a round penny.

By cross-validation with NASDAQ TRF DATA, BJZZ show that the sub-penny approach matches the NASDAQ TRF's correct buy/sell sign 98.2% of the time. The standard Lee and Ready (1991) trade-signing algorithm has a similar, but lower, accuracy rate of 96.7%.[3]

BJZZ examine the return predictability using marketable retail order imbalances by a decomposition the measure to represent price pressure, liquidity provision and, informed trading. BJZZ find that all three dimensions of OI are highly significant and that marketable retail order flows are contrarian for horizons ranging between one week and six months. The control variables indicate that investors tend to buy more aggressively in relatively larger firms, growth firms, and firms with higher turnover, as well as firms with higher volatility. They also find that if retail investors buy more than they sell in a given week, the return on that stock in the following week is significantly higher than days with similar buy and sell volumes.

Finally, their decomposition shows that approximately half of the predictive power of the marketable retail order imbalance comes from the persistence of the order imbalance measures. Majority of the remainder comes from the residual component after they eliminate order persistence and the contrarian trading pattern. As this residual component turns out to predict future stock returns, it is consistent with the hypothesis that marketable retail investor trading contains valuable information about future stock price movements.

Farrell et al. (2020) further the research conducted by BJZZ by investigating the effect of social media and specifically that of SeekingAlpha on retail investor activity.

Their results indicate that there is a causal relation between Seeking Alpha articles and

[3]These accuracy rates are true for stocks with a share price below $100.

retail investor trading and that the articles have a strong influence on both the intensity and direction of retail trading. According to their results, retail trading in the half-hour after the publication of a Seeking Alpha article is 7.73% higher than in the half-hour before.

Furthermore, they also investigate whether the ability to predict future returns based on retail order imbalances differs on days with Seeking Alpha articles, media articles, or firm news compared to normal days without such activity. Their findings align with those of BJZZ in that retail order imbalances can be informative about returns a week ahead. However, they also find that the informativeness doubles on days with Seeking Alpha research.

3

Method

3.1 Data collection and processing

We randomly select 11 stocks that are included in the S&P500 as well as GameStop, AMC Entertainment and The Eastman Kodak Company to perform our analysis on.[1] We limit the number of firms due to time constraints.[2] For these stocks we extract quarterly data on market capitalisation, EBITDA, monthly turnover, operating income before depreciation, total outstanding shares, and earning announcement dates for the years 2016-2021 from Compustat.

We use data retrieved from Finnhub.io that includes the national best bid offers of the individual stocks as well as transaction data. Each set of firm data is retrieved at a millisecond interval using an API, Application Programming interface, which is a software intermediary that allows two applications to talk to each other and exchange data. However the data retrieval rate is limited, leading to a smaller selection of stocks is compared to that of previous research.

We use the method developed by BJZZ to identify trades initiated by retail investors. The

[1] The randomly selected stocks where extracted from the population of S&P500 stocks using the RAND function in excel.

[2] Each ticker takes approximately 12 hours to retrieve, process and manually prepare for Stata

method uses the fact that retail order flow can receive minor price improvements down to a cent per share, contrary to institutional order flow. Most of the orders that execute with price-improvements down to a cent a share take place off-exchange and are then reported to a Trade Reporting Facility (TRF)[3]. Data from a group of wholesale brokers in January 2020 shows that 85% of all market orders are price-improved (Mackintosh 2020). With access to this data, it is possible to identify both retail sales and buys. Retail market sales are assumed where the transaction price is slightly above the round penny, while retail buys are assumed by the transaction price being slightly below the round penny.

In the initial step of the processing retail trades are identified in three steps: (i) Trades executed through FINRA which are labeled as "D" in the tick data are extracted. (ii) We utilize the sub-penny rule described in the previous section to identify retail trades. We apply the following condition to each price datapoint:

$$[Price_{i,\tau} \times 100] \; mod(1) \neq 0, \tag{1}$$

where i denotes firm and τ denotes timestamp. If this condition is fulfilled, a trade is categorized as a retail trade. (iii) We identify whether the retail trade is to be identified as a buy or a sell on the trades that are assumed to be retail initiated

If the following condition is fulfilled, the trade is identified as aggressive buy:

$$[Price_{i,\tau} \times 100] \; mod(1) > 0.5 \tag{2}$$

Similarly, if the following condition is fulfilled, the trade is identified as an aggressive sell:

$$[Price_{i,\tau} \times 100] \; mod(1) < 0.5 \tag{3}$$

[3]This is because orders executed by wholesalers or through internalisation must be publicly reported; they are usually reported to a FINRA Trade Reporting Facility (TRF), which provides broker-dealers with a mechanism through which to report transactions that take place off-exchange. These TRF executions subsequently included in the TAQ "consolidated tape" of all reported transactions with exchange code "D"

The result of this process is a dataset with marketable retail trades and their corresponding trade direction. This then enables us to extract firm specific metrics from this data as well. We determine the signed volume of individual trades in the following manner:

$$Price_{i,\tau} \times ShareVolume_{i,\tau} \times TradeDirection_{i,\tau} = SignedVolume \tag{4}$$

Since our method relies on the measurements of order imbalance shown in the next section we also summarize each day's volume, irrespective of the trade direction. This produces the individual stock order imbalance on the given day.

Institutional signed volume is calculated in the same manner, but uses the LR trade signing algorithm to determine trade direction (BJZZ).

Next, we extract the realized intraday volatility. Firstly we re-sample the data at a frequency of 5 minutes in order to eliminate the effect of bid-ask bounce, which is a situation where the price bounces within the very limited range between the bid price and ask price but no real movement in price occur. We then proceed to evaluate the log returns.

$$log(return)_{i,\tau} = log(price_{i,\tau}) - log(price_{i,\tau+1}) \tag{5}$$

Next each data point of difference is squared and the data is summarized as following equation shows:

$$RealizedVolatility_i, t = \sum_{i=1}^{N} \sqrt{log(return)^2_{i,\tau}} \tag{6}$$

Where i denotes firm, t denotes day and τ denotes intraday. This produces the realized volatility of the stock for the given day. Similarly to BJZZ, with access to this data, we can now compute and measure the daily retail order imbalances, and thus retail investors' directional trades using:

$$RetailOIBvol_{i,t} = \frac{RetailBvol(i,t) - RetailSvol}{RetailBvol(i,t) + RetailSvol(i,t)} \tag{7}$$

where *Retail* denotes marketable retail trade, *OIB* denotes order imbalance, *vol* stands for share volume, *B* and *S* denote buy and sell respectively.

3.2 The model

We examine the informativeness of retail order imbalances at the one week, two weeks and 4 weeks in the future by estimating:

$$\begin{aligned} Y_{i(w+1,w+2,w+4)} = \alpha &+ \beta_1 Retail\_OIB_{i,w} + \beta_3 Retail\_OIB_{i,w} \times Log(Size)_{i,q} \\ &+ \beta_4 Inst\_OIB_{i,w} + \beta_5 Inst\_OIB_{i,w} \times Log(Size)_{i,q} \\ &+ \beta_7 meme \times Retail\_OIB_{i,w} + \beta_8 report\_5 \times Retail\_OIB_{i,w} \\ &+ \beta_9 covid \times Retail\_OIB_{i,w} + \beta_{10} Char_{i,t} + Month_i + Firm_i + \epsilon_{i,t}, \end{aligned} \tag{8}$$

where $Y_{i(w+1,w+2,w+4})$ is stock i's return from the close of day compounded to the close of day compounded into weekly observations in three intervals one, two, and, four weeks in the future and q denotes quarterly observations. $Retail\_OIB_{i,w}$ is the daily retail buy volume minus daily retail sell volume scaled by daily retail trading volume and compounded to weekly observations. We include $Retail\_OIB \times Size$ as an interaction term between profitability of retail trading and firm size. $Inst\_OIB_{i,w}$ is the total non-retail buy volume minus total non-retail sell volume, scaled by total non-retail trading volume and $Inst\_OIB_{i,w} \times Log(Size)_{i,q}$ is an interaction term between firm size and institutional order imbalance. $Char_{i,q}$ is a vector of firm characteristics and includes past returns estimated over the prior week, market capitalization, monthly turnover, volatility of daily returns, and operating income before depreciation. $report\_5 \times Retail\_OIB_{i,w}$ is an interaction term between a dummy variable for 5 days around reporting and retail order imbalance. $covid \times Retail\_OIB_{i,w}$ is an interaction term between a dummy variable for the Covid-19 for the effect on March and April of 2020, i.e the months where most governments enforced lockdowns (Sandford 2020) and retail order imbalance. Finally, $meme \times Retail\_OIB_{i,w}$ is an interaction term between a dummy variable for meme stocks and retail order imbalance.

We use returns that are similar to BJZZ's "CRSP returns" as they are computed based on close price, that is latest close in trade. This can trigger an upward bias due to bid-ask bounce according to Blume and Stambaugh (1983). For this reason BJZZ report most of their results and returns in the average bid-ask price at the end of the day. When comparing BJZZ's results for CRSP returns and closing bid-ask price, we note that the differences are marginal. Furthermore, we mainly compute our results on S&P500 stocks, meaning that the bid-ask spread should be relatively small and therefore the bid-ask bounce should only have a marginal effect on our results. However, for the meme-stocks we admit that there may be a larger spread and that it has a larger significance. For these reasons, we opt to use the latest closing price for our returns but in case of additional time at hand we would include this in our report as well.

We then investigate whether the correlation between retail imbalances and future returns depends on price pressure, liquidity provision or informed trading. We use the daily return decomposition of BJZZ; we decompose the order imbalance into three components: OIB persistence representing price pressure, OIB contrarian to represent liquidity provision and OIB Other to represent informed trading.[4] The three components are estimated using the following panel regression:

$$Retail\_OIB_w = \alpha + \beta_1 Retail\_OIB_{i,w-1} + \beta_2 Ret_{i,w-1} + \epsilon_{i,t}, \tag{9}$$

[4] This is not included in our regression

where OIB persistence is $\beta_1 Retail\_OIB_{i,w-1}$, OIB contrarian is $\beta_2 Ret_{i,w-1}$and OIB other is $\hat{\epsilon}_{i,t}$.

Through this decomposition we can differentiate between the different effects' contribution towards the predictive power. This is done through a regression with return as the dependent variable and the decomposed effects as independent variables along with the residual from the first stage regression as an independent variable.

$$Return_w = \alpha + \beta_1 Retail\_OIB_{i,w-1} + \beta 2Ret_{i,w-1} + \widehat{\mu}i,t + \epsilon i,t \tag{10}$$

In our full-scale model above, we utilize the Lee-Ready trade signing algorithm on data on national best bid and ask prices to also be able to determine the trade direction of institutional trade direction and the resulting institutional order imbalance. However, we rely on finnhub.io for our data collection and they currently only have national best bid and ask price data dating back to 2020-01-02. Due to limited time and literature showing signs of low accuracy in using Bulk-Value Classification for calculating order imbalances (Chakrabarty, Pascual, and Shkilko 2015), we instead opt to use a simplified version of our model excluding institutional order imbalance for the full period between 2016-05-27 and 2021-04-27.

3.3 The simplified model

Our simplified model is used in the same way as the full-scale model, with the exception that we omit $Inst\_OIB_{i,w}$ and $Inst\_OIB \times Log(Size)_{i,q}$. As the model is simplified, the results may carry omitted variable bias. However we use it as a robustness check to our full scale model to control for a longer time period.

$$\begin{aligned} Y_{i(w+1,w+2,w+4)} = \alpha &+ \beta_1 Retail\_OIB_{i,w} + \beta_3 Retail\_OIB_{i,w} \times Log(Size)_{i,q} \\ &+ \beta_4 meme \times Retail\_OIB_{i,w} + \beta_5 report\_5 \times Retail\_OIB_{i,w} \\ &+ \beta_6 covid \times Retail\_OIB_{i,w} + \beta_7 Char_{i,t} + Month_i + Firm_i + \epsilon_{i,t}, \end{aligned} \tag{11}$$

The simplified model is applied on the longer time scope and we then investigate whether the correlation between retail imbalances and future returns depends on price pressure, liquidity provision or informed trading in the same manner as for the full-scale model.

4

Results and Analysis

4.1 Summary Statistics

Our sample, with summary statistics shown in Table 4.1 through 4.5, ranges back from 25-05-2016 up until 25-03-2021 with statistics computed on 358 million unique trades. We consider 11 randomly selected S&P500 stocks and an additional two manually selected meme stocks: GameStop and AMC Entertainment.

The stocks in our sample have a mean market value of USD 44.9 billion and a mean quarterly turnover of USD 10,6 billion. The mean market size and quarterly turnover together with the fact that we only use S&P500 constituents, means that our sample is tilted towards towards larger firms. As such we see lower spreads due to higher trade volumes (Palmer 2021), which can also be seen in Table 4.1 as the mean realized spread is zero. Larger firms will also mean that retail order flow is slightly more persistent according to previous literature (Boehmer et al. 2020). Moreover, in Table 4.5 we can see that there is a notable general uptrend in the proportion of trade volume being executed by retail traders. This is especially true for two of the smaller meme stocks, i.e, AMC, GME.

The mean weekly aggregated Retail OIB ratio is 1.9 through 2016-2021 and 2.1 through

2020 to mid 2021 indicating that, within our dataset, retail traders have been net buyers. The mean weekly aggregated Institutional OIB ratio is -0.012 for the period 2020 to mid 2021, indicating that institutional traders on average buy and sell the stocks approximately equally much in our sample. When grouping the meme stocks we find a significantly lower weekly order imbalance in all instances, 1.07 for the period 2016-2021, 1.7 for 2020 through mid 2021. We also see a lower mean institutional order imbalance ratio of -0.003 indicating that on average institutional investors on average buy and sell GME, AMC Entertainment approximately equally during the period 2020 to mid 2021.

The mean one week return is 0.4% which is 23.4% aggregated for a full year or 186% for the full sample period while the mean 5 year return for the sample period is 127%. These can be compared to the 10% average return of the S&P500 since its inception or the 101% return of S&P500 between May 2016 and April 2021 (representing a 15.2% annualized return). This can be interpreted as our sample being relatively representative of the returns achieved by the S&P500 during the sample period.

Table 4.1: Summary statistics for S&P500 stocks, 2020-2021

	Sum	Mean	SD	Min	Max	N
Return_w	21.15	0.006	0.085	-0.71	1.35	3,564.00
Retail_OIB_w	7,553.30	2.106	0.823	-1.14	4.40	3,586.00
Inst_OIB_w	-44.68	-0.012	0.828	-3.41	2.94	3,586.00
RealizedVol_d	55.67	0.015	0.012	0.00	0.20	3,630.00
RealizedSpread	0.13	0.000	0.000	-0.01	0.00	3,630.00
VolShareOfRetail	563.10	0.155	0.064	0.02	0.55	3,630.00
Retailtrades	7,901,239.00	2,176.650	4,509.589	68.00	50,788.00	3,630.00
Market Value - Total	167,110,694.98	46,036.004	60,211.803	1,120.36	283,508.00	3,630.00
Operating Income Before Depreciation - Quarterly	7,007,823.69	1,930.530	3,912.925	-1,581.00	15,079.00	3,630.00
Sales/Turnover (Net)	38,575,943.73	10,626.982	14,991.408	126.47	46,821.00	3,630.00

Table 4.2: Summary statistics for S&P500 stocks, 2016-2021

	Sum	Mean	SD	Min	Max	N
Return_w	51.89	0.004	0.054	-0.71	1.35	13,529.00
Retail_OIB_w	25,464.72	1.879	0.921	-1.29	4.77	13,550.00
RealizedVol_d	146.98	0.011	0.009	0.00	0.20	13,595.00
VolShareOfRetail	2,195.53	0.161	0.075	0.01	0.72	13,595.00
Retailtrades	21,044,769.00	1,547.979	3,356.410	13.00	78,501.00	13,595.00
Market Value - Total	610,225,059.75	44,885.992	63,533.882	537.03	283,508.00	13,595.00
Operating Income Before Depreciation - Quarterly	24,211,587.64	1,959.025	3,927.186	-2,018.00	15,871.00	12,359.00

Table 4.3: Summary statistics for meme stocks, 2020-2021

	Sum	Mean	SD	Min	Max	N
Return_w	58.97	0.091	0.549	-0.80	7.88	648.00
Retail_OIB_w	696.33	1.068	0.503	0.05	2.68	652.00
Inst_OIB_w	-1.65	-0.003	0.385	-1.10	2.12	652.00
RealizedVol_d	43.05	0.065	0.064	0.01	0.91	660.00
RealizedSpread	0.31	0.000	0.001	-0.01	0.01	660.00
VolShareOfRetail	204.79	0.310	0.080	0.07	0.68	660.00
Retailtrades	17,129,881.00	25,954.365	66,424.659	258.00	644,976.00	660.00
Market Value - Total	828,034.46	1,254.598	3,887.877	246.91	21,222.50	660.00
Operating Income Before Depreciation - Quarterly	-47,592.80	-72.110	198.211	-355.00	244.90	660.00

Table 4.4: Summary statistics for meme stocks, 2016-2021

	Sum	Mean	SD	Min	Max	N
Return_w	48.20	0.020	0.290	-0.80	7.88	2,460.00
Retail_OIB_w	4,221.49	1.715	1.018	-0.97	4.75	2,462.00
RealizedVol_d	74.68	0.030	0.040	0.00	0.91	2,472.00
VolShareOfRetail	559.06	0.226	0.095	0.01	0.68	2,472.00
Retailtrades	18,321,634.00	7,411.664	36,092.413	7.00	644,976.00	2,472.00
Market Value - Total	4,582,106.33	1,853.603	2,173.999	246.91	21,222.50	2,472.00
Operating Income Before Depreciation - Quarterly	210,947.26	85.335	155.174	-355.00	342.70	2,472.00

Table 4.5: Mean retail volume share of trades in percentages

	2016	2017	2018	2019	2020	2021	Total
AFL	14.5	12.8	13.1	12.0	13.7	10.3	12.9
BBY	17.4	13.0	16.9	17.3	16.2	12.5	15.8
CI	18.7	12.7	14.4	22.5	14.1	9.7	15.9
CTSH	12.8	15.1	14.7	16.3	13.7	15.8	14.7
CZR	20.0	24.7	23.2	18.4	16.7	17.6	20.5
ECL	12.1	15.6	18.6	13.7	12.5	11.1	14.5
ED	15.7	11.5	13.9	12.3	15.9	19.0	14.0
HES	10.1	11.6	12.3	14.6	14.2	15.6	13.0
OMC	12.2	14.3	18.8	18.4	14.2	17.1	16.0
PTC	15.0	12.7	14.4	16.6	14.8	18.7	14.9
T	23.3	25.2	27.2	27.0	24.7	23.3	25.5
AMC	14.1	20.2	23.7	24.3	34.1	34.3	24.7
GME	15.1	16.6	17.5	20.9	28.2	27.0	20.5
All	15.5	15.8	17.6	18.0	17.9	17.8	17.1

4.2 Results

Our main variables of interest are marketable retail order imbalance and return on individual stocks, denoted $Retail\_OIB_w$ and $Return_{w+t}$ respectively in our results. In our regressions we use weekly data on both returns and order imbalance.

We start out by looking at the results from the sample of 11 S&P500 stocks with our full model regressing over the shorter time period followed by the results from our simplified model on the longer time period. We then repeat this process but for the sample of meme stocks.

We run our simplified model on the full time period but as we also have access to additional data allowing us to use our full scale model for the time period 2020-2021, we run these regressions for this period separately as a form of robustness check.

We note that the return and marketable retail order imbalance residual is likely to be correlated within firms, so we cluster standard errors by firm in all regressions. Both time periods have been subjected to month and firm controls individually and are presented in the appendix Tables A.2 through A.5.

4.3 Results from the full model on short period

We find no statistical significance when regressing without the interaction between order imbalance and firm size. However, when the interaction term is included we find a slightly positive coefficient of 0.028 for retail order imbalance when predicting returns one week ahead. An interpretation of this is that we cannot predict future returns in large firms. This finding is in line with the findings of BJZZ and Farells et Al, for the one week period they also find a slight positive coefficient. We use the same calculations for our proxy of order imbalance, based on share volume, however the returns are calculated as the compounded 5 day difference between close price. A method of calculation, shown to produce slightly lower coefficients, albeit with stronger predictive power in BJZZs, findings. Our findings suggest that if retail investors buy more than they sell (in aggregate) of a stock in a given week, the returns tend to increase, in effect, retail investors tend to sell losers and buy winners in the short horizon of one week.

Table 4.6: Predicting future return short period 2020-2021 - full model

	Return_{w+1}	Return_{w+1}	Return_{w+1}	Return_{w+1}	Return_{w+2}	Return_{w+4}
Retail_OIB_w	−0.024	−0.019	−0.001	0.028***	−0.004	−0.020
	(0.015)	(0.014)	(0.002)	(0.008)	(0.020)	(0.022)
Inst_OIB_w	0.010*	0.010*	0.010	0.127	0.009	−0.053
	(0.006)	(0.005)	(0.006)	(0.102)	(0.085)	(0.064)
Return_w		−0.035	−0.027	−0.027	−0.034	0.058
		(0.028)	(0.041)	(0.041)	(0.024)	(0.063)
OIBDq_q		−0.000	−0.000**	−0.000*	−0.000*	−0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sales/Turnover_q		0.000	0.000	0.000***	0.000	−0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
RealizedVol		0.553	−0.090	−0.136	0.145	0.631**
		(0.319)	(0.449)	(0.466)	(0.240)	(0.244)
report_5=1			−0.004	−0.003	−0.032	−0.042
			(0.052)	(0.051)	(0.044)	(0.047)
meme=1			0.268***	0.279***	0.215***	0.132
			(0.056)	(0.061)	(0.044)	(0.120)
report_5=1 × Retail_OIB_w			−0.003	−0.003	0.013	0.026
			(0.021)	(0.021)	(0.019)	(0.022)
meme=1 × Retail_OIB_w			−0.167***	−0.184***	−0.131***	−0.072
			(0.015)	(0.020)	(0.023)	(0.080)
$\text{Retail\_OIB}_w\_X\_logSize$				−0.003***	−0.000	0.001
				(0.001)	(0.002)	(0.002)
$\text{Inst\_OIB}_w\_X\_logSize$				−0.011	−0.001	0.005
				(0.010)	(0.008)	(0.006)
Constant	0.066*	0.047	0.012	0.010	0.021	0.010
	(0.037)	(0.038)	(0.010)	(0.009)	(0.015)	(0.014)
Observations	4212	4147	4147	4147	4082	3952

Standard errors in parentheses
Dependent variables: Future return, in week 1,2 and,4.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

For the control variables, we observe negative coefficients on the previous week's return, which indicates weekly return reversals, size, operating income before depreciation, turnover, and realized volatility all carry the expected signs, and most are not statistically significant. This result also confirms that the predictability we find is not due to size, turnover, or volatility. The average adjusted R2's from the first stage cross-sectional estimation are mostly around 4.4%, apart from the first regression without control variables which has an R2 of 0.18%. This shows that the amount of variance the realized volatility and previous 5 day return and other controls impact the model.

4.4 Decomposition results for short period

In table 4.7 we find that the coefficient is significant at an alpha of 0.1 on our proxy for persistence, i.e marketable retail order imbalance with a five day lag. The positive coefficient of 0.73 on persistence is in line with previous research by BJZZ and Farell et al. and shows a high degree of persistence on the weekly level. Contrarian trading is found to be significant, at an alpha of 0.01 which is in line with previous research. In the second

Table 4.7: Decomposition short period 2020-2021

	Retail_OIB$_w$	Return$_w$
Retail_OIB$_{w-1}$	0.730***	−0.027*
	(0.031)	(0.014)
Return$_{w-1}$	−0.125***	0.006
	(0.033)	(0.018)
$\widehat{\mu} i, t$		−0.009
		(0.011)
Constant	0.526***	0.071*
	(0.086)	(0.036)
Observations	4160	4147

Standard errors in parentheses
Dependent variables: Retail order imbalance and return.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

step we do not find we find suggestive evidence that persistence is explanatory of return in the short period. We do not find the proxy for contrarian trading to be significant, which is in line with previous research by BJZZ. Contrary to previous research we do not find the residual, i.e the proxy for informational trading to be significant.

4.5 Results from the simplified model on long period

Similarly to the shorter period we do not find statistical significance that retail order imbalance is a predictor of future return without controlling for firm size. The longer time period, reported in table 4.8, supports the findings of the full featured model, with a slightly positive predictive coefficient of 0.008 on marketable retail order imbalance, at an alpha of 0.01.

This alludes to the importance of the features in the full model and this simplified albeit longer time period model not being able to find any coefficients to be statistically significant. It could also be the the fact the time periods are very different, as is evident when looking at Figure 4.1 and 4.2.

In the longer period we find that realized volatility is a highly significant contributor to future return. This finding is different to that of the shorter period where realized volatility was shown to be largely insignificant apart from future returns four weeks ahead. This suggests that the risk in stocks from our sample becomes less important in explaining the future return.

Table 4.8: Predicting future return long period 2016-2021 - full model

	Return_{w+1}	Return_{w+1}	Return_{w+1}	Return_{w+1}	Return_{w+2}	Return_{w+4}
Retail_OIB_w	−0.004	−0.004	0.001	0.008***	0.004	0.001
	(0.004)	(0.003)	(0.001)	(0.003)	(0.005)	(0.008)
Return_w		−0.034*	−0.033	−0.033	−0.036	0.050
		(0.018)	(0.020)	(0.020)	(0.029)	(0.058)
OIBDq_q		−0.000*	−0.000	−0.000*	−0.000*	−0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sales/Turnover_q		0.000	0.000	0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
RealizedVol		0.688***	0.566***	0.553***	0.674*	0.956***
		(0.210)	(0.181)	(0.177)	(0.340)	(0.164)
report_5=1			−0.004	−0.004	−0.012	−0.016
			(0.017)	(0.017)	(0.013)	(0.016)
meme=1			0.044**	0.043*	0.031*	0.013
			(0.020)	(0.020)	(0.016)	(0.022)
report_5=1 × Retail_OIB_w			−0.003	−0.003	0.004	0.008
			(0.007)	(0.007)	(0.006)	(0.007)
meme=1 × Retail_OIB_w			−0.023**	−0.024**	−0.018*	−0.010
			(0.010)	(0.010)	(0.009)	(0.010)
$\text{Retail\_OIB}_w\_X\_logSize$				−0.001**	−0.001	−0.000
				(0.000)	(0.001)	(0.001)
Constant	0.014	0.005	−0.003	−0.002	0.000	−0.003
	(0.009)	(0.007)	(0.002)	(0.002)	(0.004)	(0.002)
Observations	15989	14699	14699	14699	14638	14518

Standard errors in parentheses
Dependent variables: Future return, in week 1,2 and,4.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.6 Decomposition results for long period

Results are reported in table 4.9. Similarly to the shorter time period we find a statistical significance at an alpha of 0.01 for persistence, with a coefficient of 0.7. The insignificant contrarian trading is in line with previous research.

From the second step of the decomposition we find that the persistence, contrarian trading as well as informational trading is insignificant.

Table 4.9: Decomposition short period 2020-2021

	Retail_OIB$_w$	Return$_w$
Retail_OIB$_{w-1}$	0.704***	−0.005
	(0.018)	(0.003)
Return$_{w-1}$	−0.176**	0.013
	(0.059)	(0.020)
$\widehat{\mu} i, t$		−0.001
		(0.003)
Constant	0.551***	0.015*
	(0.044)	(0.008)
Observations	15937	15924

Standard errors in parentheses
Dependent variables: Retail order imbalance and return.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.7 Differences between the models and time samples

When comparing the decomposition findings in the two data sets, we notice that persistence is statistically significant at an alpha of 0.01 in both regressions, though we find a higher coefficient in the period 2020-2021.

Disregarding the predictive power, it is interesting that the coefficient for marketable retail order imbalances increases in magnitude in the period 2020 to mid 2021, compared to the period 2016-2021. As shown by Figure 4.1 and 4.2 the nature of the samples looks very different in 2020-2021 compared to 2016-2021 while marketable retail order flow remains at a similar level.

Figure 4.1: Weekly return and retail order imbalance for S&P500 sample over the period mid 2016 to mid 2021

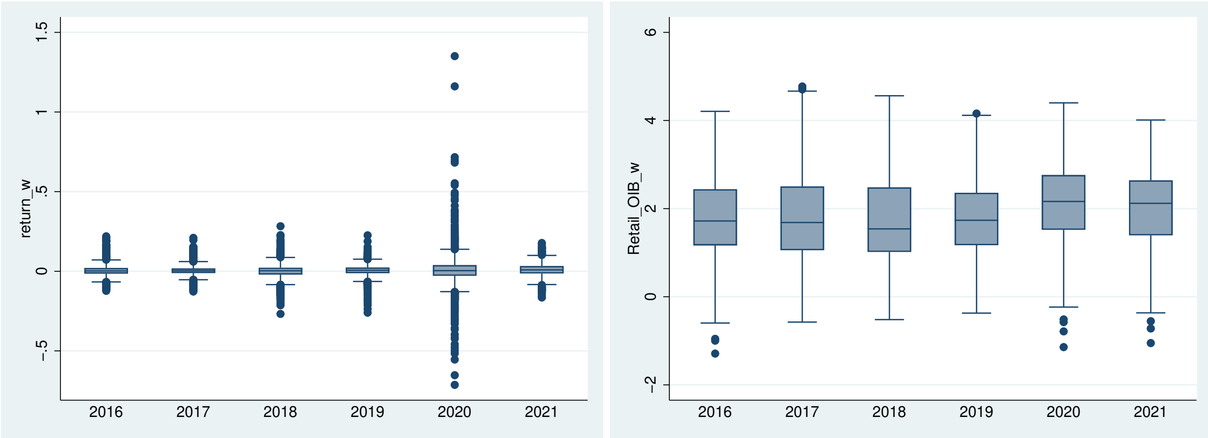
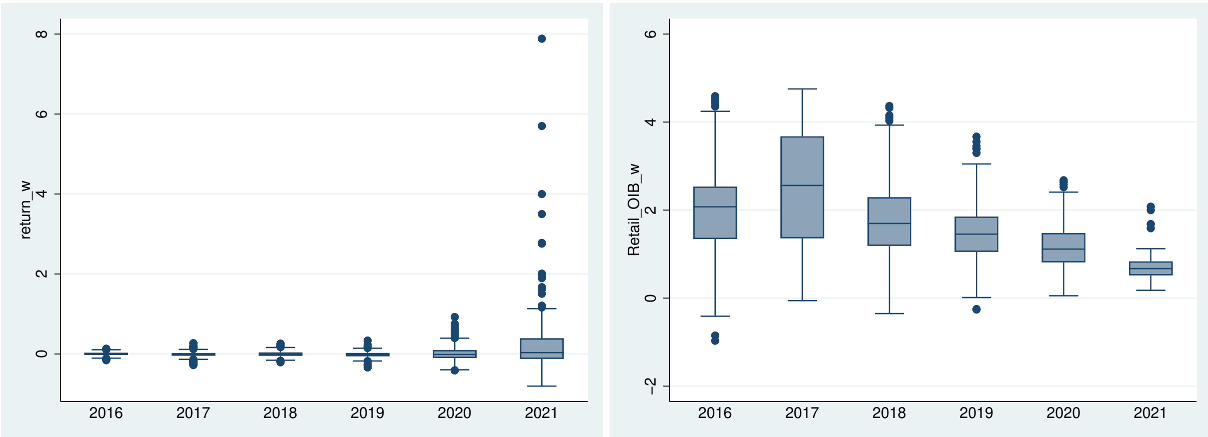


Figure 4.2: Weekly return and retail order imbalance for meme stock sample over the period mid 2016 to mid 2021



4.8 Reporting days

We report our findings for reporting dates throughout all time periods and sub-groups of all 11 stocks and only meme stocks. We do not find any significance for 5-days around reporting, which might not be surprising considering the size of the data becoming so much smaller when only accounting for days in connection with reports. The insignificant coefficients all share the same signs as the overall marketable retail order imbalance coefficients, though they are, in all cases, close to zero. On this basis, we cannot reject the null hypothesis in this case. Though it could be the case that there is no contribution, it seems unlikely since there is an overall correlation between return and marketable retail order flow in previous literature as well as in most of our findings.

4.9 The meme stocks

When regressing on only meme stocks in the shorter period we find that on average retail investor order imbalance tend to negatively correlate with returns one and two weeks ahead, with a coefficient of -0.18 and -0.13, as seen in table 4.6.

When regressing on the full period with the meme stocks which can be seen in table 4.8, we also find negative coefficients for all time horizons although none of them are statistically significant. For the full period however, the coefficients are lower than for the shorter time period. For example, coefficient for one week returns is -0.023 in the longer period compared -0.18 for the shorter period. This indicates that the recent surge in interest in these specific stocks may be the cause for our findings.

We also note that the returns for meme stocks are much higher than for the rest of the sample. Despite this, we find that when retail investors are buying meme stocks to a higher degree, the following weekly returns are negative. This implies that retail investors are generally losing money in these stocks.

4.10 Covid effect

When looking at the shorter time period, we note differing results compared to those observed for the longer time period. A natural explanation for this is Covid-19 starting to affect the financial markets in the beginning of 2020. To see if there is a significant difference in the retail order flow and the predictability of returns specifically around the outbreak of Covid-19 and the resulting intense lockdowns, we introduce a Covid-dummy variable for the months of March and April in 2020 when lockdowns were at the highest level. As can be seen in table 4.10, we do not find statistical evidence that there is a difference due to the initial lockdowns. We cannot reject the null hypothesis that the initial outbreak of Covid-19 has an effect on retail order flow and the predictability of returns. Possibly due the fact that the initial market decline in March was regained over

April. However, despite not being able to say with certainty that the initial impact of Covid-19 is a root of our differing results, our results in general point towards Covid-19 having a notable impact.

Table 4.10: Predicting future return short period 2020-2021 - full model

	Return_{w+1}	Return_{w+1}	Return_{w+1}	Return_{w+1}	Return_{w+2}	Return_{w+4}
Retail_OIB_w	-0.024	-0.019	-0.002	0.024^{***}	-0.006	-0.020
	(0.015)	(0.014)	(0.004)	(0.007)	(0.021)	(0.022)
Inst_OIB_w	0.010^{*}	0.010^{*}	0.010^{*}	0.133	0.015	-0.050
	(0.006)	(0.005)	(0.005)	(0.103)	(0.079)	(0.065)
Return_w		-0.035	-0.033	-0.033	-0.034	0.062
		(0.028)	(0.044)	(0.044)	(0.028)	(0.059)
OIBDq_q		-0.000	-0.000^{**}	-0.000^{*}	-0.000^{*}	-0.000^{*}
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sales/Turnover (Net)		0.000	0.000	0.000^{***}	0.000	-0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
RealizedVol		0.553	-0.015	-0.061	0.135	0.561^{**}
		(0.319)	(0.497)	(0.508)	(0.278)	(0.194)
report_5=1			-0.006	-0.005	-0.034	-0.043
			(0.051)	(0.051)	(0.044)	(0.049)
meme=1			0.262^{***}	0.273^{***}	0.215^{***}	0.137
			(0.058)	(0.062)	(0.039)	(0.119)
Covid=1			-0.034	-0.042	-0.037	-0.017
			(0.031)	(0.032)	(0.049)	(0.032)
report_5=1 × Retail_OIB_w			-0.002	-0.003	0.014	0.027
			(0.021)	(0.021)	(0.019)	(0.022)
meme=1 × Retail_OIB_w			-0.164^{***}	-0.181^{***}	-0.131^{***}	-0.075
			(0.015)	(0.020)	(0.021)	(0.080)
Covid=1 × Retail_OIB_w			0.008	0.012	0.019	0.016
			(0.014)	(0.015)	(0.020)	(0.014)
$\text{Retail\_OIB}_w\_logSize$				-0.003^{**}	-0.000	0.001
				(0.001)	(0.002)	(0.002)
Inst_OIB_w_Size				-0.012	-0.002	0.005
				(0.010)	(0.007)	(0.006)
Constant	0.066^{*}	0.047	0.014	0.014	0.026	0.015
	(0.037)	(0.038)	(0.012)	(0.012)	(0.017)	(0.016)
Observations	4212	4147	4147	4147	4082	3952

Standard errors in parentheses
Dependent variables: Future return, in week 1,2 and,4.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5

Discussion

5.1 Results compared to previous research

Using a method developed by BJZZ, we identify trades executed by retail traders on a sample of stocks and use this data to calculate a retail order imbalance measure which tells us about the nature and direction of retail trades. We further the research of BJZZ, albeit on a smaller sample size, by running regressions on specific time periods in connection with quarterly reports and by including a sample of meme stocks to investigate whether this leads to different findings.

Similarly to the conclusions of BJZZ and other previous literature, our results indicate that retail order imbalances can be used to predict future returns. Using our full model on S&P500 stocks between 2020-2021, we find statistically significant results showing that retail order imbalance correlates positively with one-week returns. However, contradictory to BJZZ's findings, when we look at longer horizons such as two-week returns or four-week returns, we find a negative correlation although not statistically significant. Using our simplified model on 2016-2021, we find similar results as for the shorter period but with much lower coefficients. Retail order imbalance has a statistically significant positive correlation with one-week returns and the correlation turns negative during week two and forward.

There are several reasons as to why we have different findings compared to previous literature. Firstly, it could be due to the sample size; BJZZ look at 3000 stocks for a 6 year time period whereas we look at 12 stocks for a five year time period. This could also in part explain why many of the results aren't statistically significant.

Another explanation connected to our sample could be that it is biased towards larger firms as, disregarding the meme stock sample, we only look at stocks within the S&P500 while BJZZ's sample of 3000 stocks naturally includes stocks outside of the S&P500 as well. However, according to BJZZ's results, larger firms tend to show higher positive coefficients, which is the opposite of what we observe, thus we can disregard this as an explanation.

Another reason could be connected to bid-ask bounce, as our method of calculating returns isn't the most optimal for eliminating these errors. However, as noted in our method we believe this should only be a marginal error due to the size of the companies in our sample and the minimal effect witnessed due to bid-ask bounce in BJZZ's results.

In contradiction to previous research on which components drive retail order imbalance and return, we find no statistical evidence that our proxy for informed trading contributes to return. We do however find that retail order imbalance can be explained by our proxy for persistence as well as contrarian trading. However, we can only find suggestive evidence that the persistence is explanatory to future return, and no statistical evidence that contrarian trading has effect on future return.

For our meme stock sample, our findings on what direction retail order imbalances predict returns stand in direct contradiction to BJZZ; we find that retail order imbalances have a negative correlation with returns for all examined horizons with varying degrees of statistical significance. This goes for both the shorter period where we apply our full model and for the longer period with the simplified model. Interestingly, the coefficients are considerably higher for the shorter time period and we specifically find a statistically significant, strong negative correlation between retail order imbalances and one week returns. We believe that the high negative correlation could be a direct effect of the high risk, gambling-like nature, that meme stocks tend to posses. Similarly to what is witnessed in the S&P500 sample, we see that one-week returns become much more volatile in 2020 and in 2021 for the meme stock sample. (Figure 4.2) We also note a large amount of outliers in one-week returns, especially in 2021.

We are unable to find any statistical evidence of retail order imbalances being more or less explanatory of future returns in connection with the release of quarterly reports. Therefore, we can not confirm our hypothesis that predictability of returns differs in connection with quarterly reports. We do however note that it seems unlikely that there is no real contribution from increased information since previous research suggests that information is explanatory to retail trading behaviour (Boehmer et al. 2020; Farrell et al. 2020). Thus, we conclude that our sample size is a more likely reason for the insignificant coefficients.

Finally, we hypothesized that the initial effect of Covid-19 would cause differing results. We were unable to find statistical evidence of this. However, we find close to zero coefficients before 2020 and 2021 with low statistical significance and confidence intervals going above zero. When we instead run regressions solely on 2020-2021 we find larger coefficients with similar statistical significance. Looking at a comparison of the distribution of one week returns (Figure 4.1) between different years may in part provide an explanation for this difference. Between 2016 and 2019 we mostly see one week returns distributed evenly around the 0%-mark while the returns become much more volatile with a positive tilt during both 2020 and 2021. With a larger amount of extreme negative or positive returns within the investigated time horizons, we get an increased effect on our observed coefficient. While this difference could be interpreted as an effect of Covid-19 we find no statistical evidence which can attribute this to Covid-19.

5.2 The trend of increased volume of retail trades

A need for better financial literacy?

Table A.1 in appendix shows that realized volatility can explain an increased marketable retail order imbalance. This could imply that the increased volatility in the market during Covid-19 is the main explanation for the trend of increase in retail trades as seen in Figure 4.1. However, when looking at the correlation between realized volatility and marketable retail order imbalance for meme stocks during 2020 to mid 2021, we find suggestive evidence, that an increase in volatility has a negative effect on marketable retail order imbalance which contradicts the previous finding. We interpret this as retail traders not choosing meme stocks purely on the grounds of them having a high volatility but rather due to other factors catering to their gambling propensity. Without further evidence we can only speculate and it is hard to draw any more conclusions, but one reason for this could be the rise of forums such as Wall-Street Bets and an increased coverage of meme stocks capturing the eye of more retail traders leading to the increased activity.

When looking at our summary statistics, it becomes evident that the share of volume traded by retail investors is increasing significantly. If you also couple this with the negative correlation between increased marketable retail order imbalance and future returns within meme stocks between 2020 - to mid 2020, it highlights a further need and interest in studying this area of research more extensively and with bigger scopes. Previous research on liquidity provision and price pressures find that retail investors generally supply liquidity, i.e pay prices at premium in transitory effects.(Barrot, Kaniel and, Sraere 2015) While we do not find this in our current smaller data set it is fair to assume this is the case in a general setting. We do however find that in especially meme stocks, retail investors exhibit a contrarian-like trading behaviour where they buy losers and sell winners, which in part indicates that they are losing money.

Even if several of our results are low in significance individually, the aggregate of the findings suggest an increased role of retail investors in the financial markets. Given the increase in retail trades, and the gambling-like features of meme stocks it is becoming increasingly important to supply information to retail investors. For many people, their first encounters with the financial markets might come through these meme stocks and without prior knowledge, misguided decision-making caused by hausse in social media forums could lead to the loss of large sums of money. However, the increased level of retail investor activity should also, if coupled with an increased level of financial literacy, be seen as something positive through an equality perspective as more people start taking advantage of the opportunities that the financial markets can provide .

6

Conclusion

We contribute to the research area surrounding the nature of retail trading activity by investigating how future returns correlate with retail order imbalances and how this differs for periods in connection with quarterly reports, for periods following the outbreak of Covid-19 and for meme stocks. For the period of 2016 to 2021 with a sample of randomly selected stocks from the S&P500 we find evidence supporting the findings of BJZZ that retail investors order flows correlate positively with future returns, we do however only find this to be statistically significant for returns one week in the future. Regarding the periods of reporting we are unable to find statistical significance likely as a result of the limited size of our sample. However, without rejecting the null hypothesis, we find that the data suggests similar results as the longer period due to the sign of the coefficient, albeit very close to zero. Interestingly, we find that for our group of meme stocks there is a significant and negative correlation between one-week returns and retail order imbalances. Although the sample is small, we find suggestive evidence that retail investors lose money when investing in stocks with gambling characteristics. We investigate the possibility that there is a correlation between volatility and retail order imbalances within meme stocks and hence that the investors choose these stocks on the basis of volatility but find that this isn't specifically the case. Finally, while our results suggest an effect of Covid-19 in the shorter time frame, we are unable to support this with statistical evidence when looking at the initial lockdown effect.

We suggest further research in stocks with gambling characteristics with larger samples to find to further evidence of uninformed retail investors losing money in them. An interesting route connected to meme stocks could be to perform sentiment analysis' on forums such as Wall-street Bets and performing a similar analysis as the one done by Farrell et. Al (2020) on SeekingAlpha research. Although our method for identifying retail trades relies on a circumstance specific for the US market, we believe an interesting and important future subject would be to extend the research to other international markets. In doing so, we would find whether these results and findings are exclusive for the US market or if the nature of retail trading follows the same pattern irrespective of geography and market.

In general, we believe that the nature of retail investing is an important subject to investigate further as it can provide an answer to whether retail traders need to be protected, either through legislation or through increasing the general level of financial literacy through education. A capital market where people have the tools to make informed decisions and where they don't lose money to others due to unawareness is, in our opinion, absolutely vital in creating a financially sustainable society.

Bibliography

[1] Nicholas Barberis, Andrei Shleifer, and Robert Vishny. "A model of investor sentiment". In: *Journal of financial economics* (1999). URL: https://doi.org/10.1016/S0304-405X(98)00027-0 (visited on 05/26/2021).

[2] Marshall Blume and Robert Stambaugh. "Biases in computed returns: An application to the size effect". In: *Journal of financial economics* 12 (1983). URL: https://doi.org/10.1016/0304-405X(83)90056-9 (visited on 05/22/2021).

[3] Ekkehart Boehmer et al. "Tracking Retail Investor Activity". In: *Journal of Finance* (2020). URL: https://doi.org/10.1111/jofi.130 (visited on 05/21/2021).

[4] Bidisha Chakrabarty, Roberto Pascual, and Andriy Shkilko. "Evaluating Trade Classification Algorithms: Bulk Volume Classification Versus the Tick Rule and the Lee-Ready Algorithm". In: *Journal of Financial Markets* 25 (2015).

[5] Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. "Order imbalance, liquidity and market returns". In: *Journal of financial economics* 65 (2000).

[6] Jennifer Conrad, Mustafa N. Gultekin, and Gautam Kaul. "Profitability of Short-Term Contrarian Strategies: Implications for Market Efficiency". In: *Journal of Business and Economic Statistics)* 15 (1997).

[7] Kent Daniel, David Hirshleifer, and Avindhar Subrahmanyam. "Investor Psychology and Security Market Under- and Overreactions". In: *The Journal of Finance* 6 (2002).

[8] David Easley, Marcos Lopez de Prado, and Maureen O'Hara. "Flow Toxicity and Liquidity in a High Frequency World". In: *Review of Financial Studies* (2012). URL: https://dx.doi.org/10.2139/ssrn.1695596 (visited on 05/14/2021).

[9] Michael Farrell et al. "The Democratization of Investment Research and the Informativeness of Retail Investor Trading". In: *Revise and Resubmit (Journal of Financial Economics)* (2020). URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3222841 (visited on 05/21/2021).

[10] Emilios C. Galariotis. "Contrarian and momentumtrading: a review of the literature". In: *Review of Behavioral Finance* (2013). URL: https://doi.org/10.1108/RBF-12-2013-0043 (visited on 05/26/2021).

[11] Erin Gobler. "What is a Meme Stock". In: *The Balance* (2021). URL: https://www.thebalance.com/what-is-a-meme-stock-5118074 (visited on 05/19/2021).

[12] Ron Kaniel, Gideon Saar, and Sheridan Titman. "Individual Investor Trading and Stock Returns". In: *The Journal of the American Finance Association* (1999). URL: https://doi.org/10.1111/j.1540-6261.2008.01316.x (visited on 05/25/2021).

[13] Eric K. Kelley and Paul C. Tetlock. "How Wise Are Crowds? Insights from Retail Orders and Stock Returns". In: *Journal of Finance* 68 (2013).

[14] Andrew Keshner. *Tendies? Diamond hands? Your guide to the lingo on WallStreetBets, the Reddit forum fueling Gamestops wild rise.* 2021. URL: https://www.marketwatch.com/story/tendies-diamond-hands-your-guide-to-the-lingo-on-wallstreetbets-the-reddit-forum-fueling-gamestops-rise-11611780829 (visited on 05/22/2021).

[15] Craig Lazzara. "Efficient Markets and Irrational Exuberance". In: *S&P Global Research Insights* 49 (2021).

[16] Charles M. C. Lee and Mark J. Ready. "Inferring Trade Direction from Intraday Data". In: *The Journal of the American Finance Association* (1991). URL: https://doi.org/10.1111/j.1540-6261.1991.tb02683.x (visited on 05/14/2021).

[17] Bruce N. Lehmann. "Fads, Martingales, and Market Efficiency". In: *The Quarterly Journal of Economics* (1990). URL: `https://doi.org/10.2307/2937816` (visited on 05/26/2021).

[18] Phil Mackintosh. "Who Counts as a Retail Investor?" In: *Nasdaq* (2020). URL: `https://www.nasdaq.com/articles/who-counts-as-a-retail-investor-2020-12-17` (visited on 05/22/2021).

[19] Katie Martin and Robin Wigglesworth. "Rise of the retail army: the amateur traders transforming markets". In: *Financial Times* (2021). URL: `https://www.ft.com/content/7a91e3ea-b9ec-4611-9a03-a8dd3b8bddb5` (visited on 05/28/2021).

[20] Barclay Palmer. *What determines a Stocks's Bid-Ask Spread?* 2021. URL: `https://www.investopedia.com/ask/answers/06/bidaskspread.asp` (visited on 05/14/2021).

[21] Alasdair Sandford. *Coronavirus: Half of humanity now on lockdown as 90 countries call for confinement.* 2020. URL: `https://www.euronews.com/2020/04/02/coronavirus-in-europe-spain-s-death-toll-hits-10-000-after-record-950-new-deaths-in-24-hou` (visited on 05/28/2021).

[22] Ali Shalchi. "The rise of armchair retail trading: Risks and regulation". In: *UK House of Commons Library* (2021).

Appendix A

Table A.1: Realized volatility as an explanatory variable for retail orderimbalance in memestocks during 2020-2021.

	Retail_OIB$_w$	Retail_OIB$_w$
RealizedVol	-1.296^{*}	-1.374^{*}
	(0.596)	(0.685)
Return$_{w-1}$	0.029	0.037
	(0.034)	(0.038)
Retail_OIB$_{w-1}$	0.340^{***}	0.275^{***}
	(0.068)	(0.075)
OIBDq$_q$	-0.001^{**}	-0.001^{**}
	(0.000)	(0.000)
Sales/Turnover$_q$	0.000^{***}	0.000^{***}
	(0.000)	(0.000)
Month=1		0.000
		(.)
Month=2		0.295^{***}
		(0.016)
Month=3		0.105^{***}
		(0.020)
Month=4		-0.182^{***}
		(0.039)
Month=5		-0.189^{**}
		(0.073)
Month=6		0.012
		(0.028)
Month=7		0.032
		(0.032)
Month=8		0.166^{***}
		(0.026)
Month=9		0.184^{***}
		(0.033)
Month=10		0.218^{***}
		(0.030)
Month=11		0.182^{***}
		(0.032)
Month=12		0.093^{***}
		(0.028)
Constant	0.354^{***}	0.283^{***}
	(0.076)	(0.084)
Observations	960	960

Standard errors in parentheses

Dependent variables: Retail order imbalance and return.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Predicting future return long period 2016-2021 firm specific

	Return_{w+1}	Return_{w+1}	Return_{w+1}	Return_{w+1}	Return_{w+2}	Return_{w+4}
Retail_OIB_w	-0.005	-0.004	0.000	0.005	-0.002	-0.002
	(0.004)	(0.003)	(0.001)	(0.006)	(0.006)	(0.014)
AFL	0.000					
	(.)					
AMC	0.007^{***}	0.000	0.000	0.000	0.000	0.000
	(0.000)	(.)	(.)	(.)	(.)	(.)
BBY	0.003^{**}	0.008	-0.024	-0.025	-0.029	-0.021
	(0.001)	(0.004)	(0.032)	(0.031)	(0.033)	(0.027)
CI	0.003^{**}	0.011	-0.024	-0.024	-0.027	-0.019
	(0.001)	(0.007)	(0.030)	(0.029)	(0.031)	(0.026)
CTSH	0.000	0.008^{*}	-0.026	-0.026	-0.029	-0.020
	(0.001)	(0.004)	(0.031)	(0.031)	(0.032)	(0.026)
CZR	0.014^{***}	0.013^{***}	-0.021	-0.024	-0.025	-0.020
	(0.002)	(0.002)	(0.032)	(0.030)	(0.034)	(0.028)
ECL	0.000	0.009^{*}	-0.024	-0.024	-0.028	-0.019
	(0.000)	(0.005)	(0.031)	(0.031)	(0.032)	(0.027)
ED	-0.006^{*}	0.005	-0.026	-0.026	-0.031	-0.021
	(0.003)	(0.005)	(0.030)	(0.030)	(0.031)	(0.026)
GME	0.026^{***}	0.019^{***}	0.012^{**}	0.011^{**}	0.014^{***}	0.016^{**}
	(0.001)	(0.002)	(0.005)	(0.005)	(0.004)	(0.006)
HES	0.001	0.003	-0.031	-0.031	-0.035	-0.028
	(0.001)	(0.003)	(0.032)	(0.031)	(0.034)	(0.027)
OMC	-0.002^{***}	0.005	-0.029	-0.029	-0.032	-0.023
	(0.000)	(0.004)	(0.032)	(0.031)	(0.033)	(0.027)
PTC	0.008^{**}	0.011^{***}	-0.025	-0.026	-0.027	-0.019
	(0.003)	(0.004)	(0.033)	(0.031)	(0.034)	(0.028)
T	-0.005^{***}	0.039	0.000	0.000	0.000	0.000
	(0.001)	(0.037)	(.)	(.)	(.)	(.)
Return_w		-0.034	-0.034	-0.034	-0.037	0.048
		(0.019)	(0.020)	(0.020)	(0.029)	(0.058)
OIBDq_q		-0.000	-0.000	-0.000	-0.000	-0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sales/Turnover_q		0.000	0.000	0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
RealizedVol		0.657^{***}	0.565^{***}	0.565^{***}	0.686^{*}	0.978^{***}
		(0.199)	(0.177)	(0.178)	(0.347)	(0.159)
report_5=1			-0.005	-0.004	-0.013	-0.017
			(0.017)	(0.017)	(0.013)	(0.016)
meme=1			0.009	0.008	-0.009	-0.021
			(0.038)	(0.037)	(0.038)	(0.038)
report_5=1 × Retail_OIB_w			-0.002	-0.002	0.004	0.008
			(0.007)	(0.007)	(0.006)	(0.007)
meme=1 × Retail_OIB_w			-0.021^{*}	-0.022^{*}	-0.015	-0.007
			(0.010)	(0.011)	(0.009)	(0.009)
$\text{Retail\_OIB}_w\_X\_logSize$				-0.000	0.000	0.000
				(0.001)	(0.001)	(0.001)
Constant	0.012	-0.002	0.025	0.025	0.032	0.019
	(0.007)	(0.009)	(0.033)	(0.032)	(0.034)	(0.029)
Observations	15989	14699	14699	14699	14638	14518

Standard errors in parentheses
Dependent variables: Future return, in week 1,2 and,4.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Predicting future return long period 2016-2021 month specific

	Return_{w+1}	Return_{w+1}	Return_{w+1}	Return_{w+1}	Return_{w+2}	Return_{w+4}
Retail_OIB_w	−0.004	−0.003	0.001	0.009***	0.005	−0.000
	(0.004)	(0.003)	(0.001)	(0.002)	(0.006)	(0.009)
Month=1	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Month=2	−0.026	−0.030*	−0.028	−0.028	−0.022**	−0.006
	(0.015)	(0.016)	(0.016)	(0.016)	(0.009)	(0.011)
Month=3	−0.037*	−0.045**	−0.044**	−0.044**	−0.026	0.017***
	(0.017)	(0.019)	(0.018)	(0.018)	(0.020)	(0.005)
Month=4	−0.014	−0.019	−0.018	−0.018	−0.022	0.000
	(0.017)	(0.018)	(0.018)	(0.018)	(0.016)	(0.004)
Month=5	−0.025	−0.029	−0.027	−0.027	−0.018	0.005
	(0.022)	(0.023)	(0.022)	(0.022)	(0.019)	(0.008)
Month=6	−0.032	−0.035	−0.034	−0.034	−0.027	0.006
	(0.019)	(0.020)	(0.019)	(0.019)	(0.017)	(0.007)
Month=7	−0.032	−0.034	−0.033	−0.033	−0.026	0.008
	(0.019)	(0.019)	(0.019)	(0.019)	(0.016)	(0.008)
Month=8	−0.025	−0.026	−0.025	−0.025	−0.019	0.010
	(0.016)	(0.016)	(0.015)	(0.015)	(0.013)	(0.011)
Month=9	−0.029	−0.031	−0.031	−0.031	−0.028	−0.001
	(0.018)	(0.018)	(0.018)	(0.018)	(0.016)	(0.008)
Month=10	−0.036*	−0.039*	−0.038*	−0.038*	−0.024	0.017**
	(0.020)	(0.020)	(0.020)	(0.020)	(0.018)	(0.007)
Month=11	−0.015	−0.017	−0.017	−0.017	−0.015	0.002
	(0.018)	(0.018)	(0.018)	(0.018)	(0.015)	(0.007)
Month=12	−0.032	−0.034	−0.035	−0.035	−0.022	0.045
	(0.020)	(0.021)	(0.021)	(0.021)	(0.017)	(0.028)
Return_w		−0.045**	−0.044*	−0.043*	−0.042	0.055
		(0.019)	(0.021)	(0.021)	(0.031)	(0.056)
OIBDq_q		−0.000**	−0.000*	−0.000**	−0.000**	−0.000*
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sales/Turnover_q		0.000	0.000	0.000*	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
RealizedVol		0.730***	0.617***	0.603***	0.700*	0.956***
		(0.220)	(0.194)	(0.190)	(0.357)	(0.179)
report_5=1			−0.004	−0.004	−0.012	−0.013
			(0.016)	(0.016)	(0.012)	(0.017)
meme=1			0.042*	0.041*	0.030*	0.012
			(0.020)	(0.020)	(0.015)	(0.022)
report_5=1 × Retail_OIB_w			−0.004	−0.004	0.004	0.008
			(0.007)	(0.007)	(0.005)	(0.007)
meme=1 × Retail_OIB_w			−0.022*	−0.023**	−0.017*	−0.009
			(0.010)	(0.010)	(0.009)	(0.010)
$\text{Retail\_OIB}_w\_X\_logSize$				−0.001***	−0.001	−0.000
				(0.000)	(0.001)	(0.001)
Constant	0.039	0.032	0.024	0.025	0.021	−0.013
	(0.025)	(0.023)	(0.018)	(0.018)	(0.014)	(0.009)
Observations	15989	14699	14699	14699	14638	14518

Standard errors in parentheses
Dependent variables: Future return, in week 1,2 and,4.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Predicting future return short period 2020-2021 firm specific

	Return_{w+1}	Return_{w+1}	Return_{w+1}	Return_{w+1}	Return_{w+2}	Return_{w+4}
Retail_OIB$_w$	−0.013	−0.014	−0.001	−0.006	−0.087	−0.133**
	(0.009)	(0.010)	(0.002)	(0.023)	(0.052)	(0.052)
Inst_OIB$_w$	0.009	0.009	0.011	0.135	0.011	−0.055
	(0.006)	(0.006)	(0.007)	(0.110)	(0.088)	(0.060)
AFL	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
AMC	0.034***	0.006	−0.080***	−0.083***	−0.088***	−0.089***
	(0.009)	(0.020)	(0.004)	(0.004)	(0.005)	(0.010)
BBY	0.003***	0.039	0.034	0.032	0.042	0.043
	(0.001)	(0.027)	(0.025)	(0.024)	(0.030)	(0.033)
CI	0.009	0.177	0.151	0.127	0.143	0.147
	(0.006)	(0.138)	(0.129)	(0.111)	(0.150)	(0.162)
CTSH	−0.001	0.006	0.004	0.008	0.013**	0.004
	(0.002)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
CZR	0.019***	0.003	0.006	0.009	0.041*	0.044*
	(0.003)	(0.016)	(0.016)	(0.012)	(0.021)	(0.021)
ECL	0.002	0.003	−0.010	0.001	0.000	−0.023
	(0.005)	(0.009)	(0.013)	(0.011)	(0.013)	(0.015)
ED	−0.014*	−0.005	0.000	0.006	0.007	0.001
	(0.007)	(0.011)	(0.009)	(0.010)	(0.011)	(0.011)
GME	0.115***	0.098***	0.000	0.000	0.000	0.000
	(0.009)	(0.018)	(.)	(.)	(.)	(.)
HES	0.004	−0.010	−0.012	−0.005	0.006	0.003
	(0.003)	(0.012)	(0.014)	(0.011)	(0.019)	(0.017)
OMC	0.006	0.007	−0.000	0.007	0.022	0.021
	(0.005)	(0.008)	(0.006)	(0.010)	(0.016)	(0.012)
PTC	0.012**	−0.006	−0.011	−0.007	0.021	0.025
	(0.004)	(0.016)	(0.019)	(0.013)	(0.021)	(0.020)
T	−0.007***	0.356	0.304	0.299	0.335	0.270
	(0.001)	(0.219)	(0.197)	(0.188)	(0.224)	(0.245)
Return$_w$		−0.023	−0.033	−0.033	−0.040	0.053
		(0.036)	(0.040)	(0.041)	(0.027)	(0.064)
OIBDq$_q$		−0.000*	−0.000*	−0.000*	−0.000**	−0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sales/Turnover$_q$		−0.000	−0.000	−0.000	−0.000	−0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
RealizedVol		0.154	−0.093	−0.116	0.145	0.626**
		(0.351)	(0.426)	(0.456)	(0.233)	(0.225)
report_5=1			−0.005	−0.004	−0.035	−0.045
			(0.052)	(0.051)	(0.045)	(0.049)
meme=1			0.287***	0.301***	0.246***	0.165*
			(0.029)	(0.035)	(0.021)	(0.086)
report_5=1 × Retail_OIB$_w$			−0.002	−0.003	0.014	0.027
			(0.021)	(0.021)	(0.020)	(0.022)
meme=1 × Retail_OIB$_w$			−0.161***	−0.165***	−0.092***	−0.024
			(0.015)	(0.015)	(0.029)	(0.072)
Retail_OIB$_w$_$X_logSize$				0.001	0.008	0.013**
				(0.002)	(0.005)	(0.005)
Inst_OIB$_w$_$X_logSize$				−0.012	−0.001	0.005
				(0.010)	(0.008)	(0.006)
Constant	0.031	0.050	0.029	0.024	0.030	0.022
	(0.019)	(0.039)	(0.026)	(0.022)	(0.027)	(0.030)
Observations	4212	4147	4147	4147	4082	3952

Standard errors in parentheses
Dependent variables: Future return, in week 1,2 and,4.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Predicting future return short period 2020-2021 month specific

	Return_{w+1}	Return_{w+1}	Return_{w+1}	Return_{w+1}	Return_{w+2}	Return_{w+4}
Retail_OIB_w	-0.024	-0.019	-0.001	0.028^{***}	-0.003	-0.011
	(0.015)	(0.014)	(0.002)	(0.009)	(0.022)	(0.016)
Inst_OIB_w	0.009	0.008	0.009	0.135	0.011	-0.056
	(0.006)	(0.005)	(0.005)	(0.102)	(0.076)	(0.065)
Month=1	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Month=2	-0.078	-0.094	-0.089	-0.088	-0.083^{**}	0.001
	(0.045)	(0.056)	(0.053)	(0.052)	(0.033)	(0.031)
Month=3	-0.093^{*}	-0.118^{*}	-0.111^{*}	-0.111^{*}	-0.078	0.041^{**}
	(0.050)	(0.056)	(0.055)	(0.057)	(0.058)	(0.015)
Month=4	-0.043	-0.060	-0.063	-0.061	-0.070	0.020^{*}
	(0.051)	(0.059)	(0.062)	(0.063)	(0.053)	(0.011)
Month=5	-0.051	-0.067	-0.070	-0.070	-0.050	0.013
	(0.063)	(0.070)	(0.073)	(0.072)	(0.061)	(0.012)
Month=6	-0.090	-0.106	-0.109	-0.110	-0.104	0.014
	(0.057)	(0.065)	(0.068)	(0.070)	(0.061)	(0.013)
Month=7	-0.075	-0.089	-0.092	-0.092	-0.068	0.033
	(0.055)	(0.063)	(0.066)	(0.066)	(0.056)	(0.019)
Month=8	-0.054	-0.066	-0.068	-0.070	-0.072	0.021
	(0.042)	(0.051)	(0.053)	(0.055)	(0.046)	(0.028)
Month=9	-0.084	-0.099	-0.099	-0.101	-0.075	0.019
	(0.054)	(0.062)	(0.063)	(0.064)	(0.052)	(0.022)
Month=10	-0.082	-0.097	-0.094	-0.093	-0.069	0.058^{***}
	(0.056)	(0.065)	(0.065)	(0.064)	(0.057)	(0.016)
Month=11	-0.027	-0.041	-0.043	-0.040	-0.054	0.007
	(0.046)	(0.054)	(0.057)	(0.055)	(0.050)	(0.020)
Month=12	-0.082	-0.095	-0.096	-0.095	-0.067	0.174
	(0.057)	(0.067)	(0.067)	(0.067)	(0.048)	(0.115)
Return_w		-0.062^{**}	-0.054	-0.054	-0.051	0.065
		(0.028)	(0.042)	(0.042)	(0.033)	(0.057)
OIBDq_q		-0.000	-0.000^{**}	-0.000^{*}	-0.000^{**}	-0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sales/Turnover_q		0.000	0.000	0.000^{**}	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
RealizedVol		0.687^{*}	0.065	0.015	0.226	0.652^{***}
		(0.353)	(0.493)	(0.509)	(0.278)	(0.150)
report_5=1			-0.001	0.001	-0.028	-0.035
			(0.046)	(0.046)	(0.040)	(0.057)
meme=1			0.257^{***}	0.269^{***}	0.211^{***}	0.136
			(0.058)	(0.064)	(0.042)	(0.103)
report_5=1 × Retail_OIB_w			-0.007	-0.007	0.010	0.029
			(0.019)	(0.019)	(0.016)	(0.026)
meme=1 × Retail_OIB_w			-0.162^{***}	-0.181^{***}	-0.131^{***}	-0.081
			(0.016)	(0.023)	(0.025)	(0.069)
$\text{Retail\_OIB}_w\_X\_logSize$				-0.003^{**}	-0.001	0.000
				(0.001)	(0.002)	(0.001)
$\text{Inst\_OIB}_w\_X\_logSize$				-0.012	-0.001	0.006
				(0.010)	(0.007)	(0.006)
Constant	0.129	0.121	0.088	0.086	0.086	-0.022
	(0.080)	(0.091)	(0.062)	(0.062)	(0.049)	(0.020)
Observations	4212	4147	4147	4147	4082	3952

Standard errors in parentheses
Dependent variables: Future return, in week 1,2 and,4.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$