Unravelling the Influence of Retail Investors on Liquidity and Volatility in UK Markets

James Symons-Hicks

April 2023

Abstract

Retail investor presence has been growing in UK markets over the past decade and does not seem likely to fade anytime soon. With the changes in the investing landscape and the boom in investors during the pandemic, it seems necessary to re-evaluate their impact on UK markets. This paper analyses the impact of retail investors on liquidity and volatility using pricing data on the UK FTSE 100 index, website traffic analysis to trading platforms, and the Google Search Volume Index. Overall, I find that retail investing activity contributes to increased stock volatility, particularly for stocks that are more well-known, such as consumer staples and consumer discretionary; they have a positive, but lower than average, impact on financial stocks. I find that these investors had an increased impact on volatility during the pandemic, with this effect remaining in place post-COVID. There is some evidence to suggest that increased investor attention increases liquidity the following week.

Keywords: *Retail investors, Equity markets, Volatility, Liquidity, Google SVI, COVID-19*

1 Introduction

The number of retail investors has increased over the past decade with the introduction of commission-free trading, fintech innovations, and ease of access to markets. However, the COVID-19 pandemic saw an unprecedented increase in retail investing activity, fuelled by the 'stay at home' mandates issued in many countries. The increased free time during the pandemic led people to fill their days with investing, and as many households cut spending and instead increased the level of saving over this period, (Li et al., 2020), this increased the capital available with which individuals could invest. The combination of these factors contributed to the increased stock market participation by retail investors and dramatic changes in the retail investing landscape. These investors do not seem to be going anywhere anytime soon, and so, we need to ask the question: *what impact do they have on markets?* In this paper, I aim to broaden the set of stocks considered to the equities that constitute the FTSE 100 index, rather than analysing individual assets. I investigate how retail investing activity impacts the volatility and liquidity of UK stocks, both across time and across stock characteristics.

There currently exists a gap in the literature on this topic in two regards. Firstly, there is an inconclusive understanding of their modern impacts in academic literature, with much of the older literature becoming less relevant when accounting for the fintech innovations and changes to the retail investing landscape over the last decade (Welch, 2022). The influence of retail investors over time is "subject to structural changes" (Schmeling, 2007), and, as such, their impacts should be frequently researched and updated to reflect the latest trends and technologies.

Additionally, the current literature focuses primarily on US markets, with a few focusing on special cases in other markets.[1] The impacts of US investors are more widely publicised by global media outlets, and hence, their impacts and role are more commonly discussed. The UK is an important financial market both in Europe and globally, with the London Stock Exchange having an aggregate market capitalisation of £3.8 trillion, the second largest in Europe. As such, it is important to understand factors, such as

[1] For example, Foucault et al. (2011) focus on French markets, Cheng et al. (2021) focus on Chinese markets, and Barber, Lee, et al. (2009) analyse Taiwanese investors.

retail investor trading, that have an impact on UK markets. Whilst they share many similarities, there are some notable differences between the UK and US retail investment markets that necessitate research into the impact on UK markets in more depth. Firstly, the UK is a smaller market than the US, with the US leading globally in terms of market capitalisation. The size of the US market means that there are structural differences compared to other markets. Furthermore, there are key differences in the behaviour of US and UK retail investors, caused not only by regulatory differences, but also cultural ones. This naturally has an effect on the trading of these investors.

Another consideration is the availability and type of trading platforms that are available to investors in each country. Naturally, there are marked differences in the options available. One example is a very popular US platform, Robinhood, which attracted a large number of retail investors during the COVID-19 pandemic and caught global attention when many users used the platform to purchase infamous stocks such as AMC and GameStop. Barber et al. (2022) find that Robinhood attracts inexperienced investors due to its "unique features" and game-like nature, and that these investors are "unusually active" and are considerably more likely to trade speculatively. Some of the current literature is limited to a sample of US Robinhood traders (e.g., Ozik et al. (2021), Pagano et al. (2021), and van der Beck and Jaunin (2021)). Whilst the data from Robinhood used in these papers is extensive, one should take care when drawing conclusions about a general population of retail investors from such a limited sample. Inference on retail investors drawn from this data may suffer from endogeneity as there could exist certain characteristics that attract investors to this platform. The UK does not currently have access to the Robinhood platform, and hence, the nature of the retail investing landscape is likely different.

In this paper, I use a sample of stocks from the UK FTSE 100 index, combined with two retail investor proxies, website traffic analysis to the major trading platforms and Google Search Volume Index data, to conduct the analysis of the impact on liquidity and volatility. I test for heterogeneity in the impacts across levels of stock market capitalisation and by industry sector, as well as testing for structural breaks during the COVID-19 pandemic. Overall, I find that there is a positive relationship between retail trading and stock volatility, with some evidence of stronger effects of these traders for stocks with a lower market capitalisation. Retail investors have a lower than average impact on financial stocks, and a limited impact on stocks that are less well-known to them, such as materials and energy stocks. During the pandemic, retail investors had a greater impact on stock volatility, with some evidence to suggest that this effect remain, to some extent, after the end of lockdown period. However, the effect of these investors on liquidity is less clear, although there is some evidence that attention on a stock leads to increased liquidity the following week, implying that this effect occurs with some delay.

The remainder of the paper will be set out as follows: Section 2 provides a review of the literature on retail investor impacts in other markets; Section 3 explains the data sets used and provide summary statistics; Section 4 discusses the methodology used; Section 5 reports the empirical results; Section 6 discusses the results and their implications; Section 7 concludes.

2 Related Literature

Older studies on retail investors showed that they make systematic, not random, mistakes when trading, including being prone to the disposition effect,[2] over-extrapolation of past returns, and overconfidence (Barber and Odean, 2013; Grinblatt and Keloharju, 2009; Shleifer and Summers, 1990). Traditionally, retail investors have been thought of as noise traders.[3] However, there is some debate over the extent to which retail investors can be classified as noise traders. Whilst some papers (e.g., Ozik et al. (2021), Barber, Odean, and Zhu (2009), Schmeling (2007), and Peress and Schmidt (2020)) find that retail traders are predominantly noise traders, other papers suggest that only a portion of these investors act as noise traders, with the remaining portion trading on valuable information. Foucault et al. (2011)'s findings suggest that "some retail investors play the role of noise traders but they do not imply that all retail investors are noise traders", and, therefore, one should refrain from classifying all retail traders as noise traders and

[2] The disposition effect is the tendency "to sell winning investments while holding on to their losing investments" (Barber and Odean, 2013).

[3] As suggested by Black (1986), noise trading is directly contrasted with trading on information. A noise trader may trade based on a variety of factors that they believe to be helpful in predicting returns, but these factors do not actually contain new information about an asset.

"take care when using retail trading as an empirical proxy for noise trading" (Kelley and Tetlock, 2013). This is important when we consider volatility in markets, given that, "if noise traders affect prices, ...the risk they cause is volatility" (Brown, 1999). Similarly, Black (1986) states that "anything that changes the amount or character of noise trading will change the volatility". If we find that retail investors are predominantly noise traders, we would expect their presence to be positively correlated with additional volatility. Additionally, this would imply that these investors increase market liquidity, if they are uninformed noise traders. The Glosten-Milgrom Model (1985) suggest that the bid-ask spread is decreasing in the ratio of uninformed to informed traders. Hence, if there are more uninformed traders, we would expect higher market liquidity.

However, the retail investing landscape has changed dramatically over the past decade, so it follows that the impacts of these investors may also change. During the pandemic, the major brokerages reported record numbers of new accounts; one example is Trading 212, which saw one million new accounts opened during the pandemic, a 250% growth, and saw its number of daily active users grow from 28 thousand to 600 thousand (Trading 212, 2021).[4]

Given that retail investing on a scale this large is a fairly new phenomenon, much of the literature focuses on the impact during the COVID-19 pandemic. Before considering the impact of retail investors on liquidity during periods of market stress, it is essential to consider their impact under regular market conditions. Several studies find that individual traders may provide liquidity to institutional investors using their personal wealth (Barber, Odean, and Zhu, 2009; Kelley and Tetlock, 2013). According to Kelley and Tetlock (2013), it is the limit orders of retail investors that provide liquidity to markets, whereas market orders are trading on information about cash flows, and, hence, these trades do not provide liquidity. Kaniel et al. (2008) find that "risk-averse individuals provide liquidity to meet institutional demand for immediacy". Using data between 2011 and 2017, Cheng et al. (2021) discover a positive relationship between retail investor attention on stocks and stock liquidity in Chinese markets.

The impact of these investors on liquidity is particularly profound during periods

[4]This growth is from February 2020 to February 2021.

of high market stress, given that, unlike institutional investors, retail traders are not subject to "agency problems, career concerns, or liquidity constraints" (Kelley and Tetlock, 2013). Barrot et al. (2016) show that, during the 2008 financial crisis, French retail traders "stepped up to the plate" by providing liquidity at a time when "conventional liquidity providers were constrained".

Studies show that, during the pandemic, retail investors replaced much of the liquidity lost in US markets following the withdrawal of many institutional investors. The effective spread during the lockdown was almost 200% larger than during the period preceding the pandemic, however, "retail trading attenuated the rise in illiquidity by roughly 40%" (Ozik et al., 2021). Other papers (e.g., van der Beck and Jaunin (2021), and Pagano et al. (2021)) find similar results, showing that retail investors provided liquidity to US markets in Q1 and Q2 of 2020, aiding the recovery. Using a counterfactual model, van der Beck and Jaunin find that, in the absence of Robinhood traders, "aggregate market capitalisation of the smallest quintile of US stocks would have been 25% lower in Q2" of 2020. According to Ozik et al., the effect of retail investors on liquidity provision ended when the lockdown restrictions ended.

Let us now consider the literature that focuses on the impact of these investors on volatility. There is a significant volume of literature that suggests increased retail participation causes higher volatility in markets. Brandt et al. (2010) find stocks with higher retail ownership exhibit greater levels of idiosyncratic volatility. Foucault et al. (2011) use a French policy reform in 2000 that raised the relative cost for speculative retail investors to show that there exists a positive correlation between retail investors and volatility in financial markets.

The pandemic caused significant shocks to markets, more so than any previous disease outbreak (Baker et al., 2020), which, consequently, increased market volatility. These shocks are, at least in part, due to the uncertainty of the lockdown restrictions and the subsequent impact on the real economy (Albulescu, 2021). However, another cause may be due to the trading conducted by retail investors. Baig et al. (2022) suggest that, during the pandemic, retail trading had a "negative, persistent impact" on the price stability of US equities, showing that the effect of this impact on stability is greater than the period preceding the pandemic by around 30%. Eaton et al. (2022) find that

"Robinhood traders contribute to volatility, in line with noise trading models", but other retail investors have a less significant impact, suggesting that effects of retail investors on volatility may vary by brokerage.

As part of my analysis, I use the Google Search Volume Index (SVI) as a measure of retail investor attention. It is worth considering how this has been used in prior studies and how this measure can contribute to this paper. The Google SVI has been identified as an accurate indicator of retail investor attention (Da et al., 2011; Ding and Hou, 2015; Smales, 2021). Whereas institutional investors use services such as Bloomberg, the vast majority of retail investors do not. As such, they use the internet, namely Google, as their source of information. Da et al. (2011) argue that the Google SVI is a direct and accurate measure of retail investor attention for several reasons. Firstly, Google remains the most popular search engine; moreover, "search is a revealed attention measure"; that is, if someone is searching for something, they are paying attention to it.

One may wonder: *to what extent does attention translate into trading?* Whilst the literature has not looked at the transmission of the Google SVI to trading, there are many studies that have looked at the buying behaviour of retail investors on stocks that have been identified as "attention-grabbing". Barber and Odean (2008) find that retail investors are net-buyers of attention-grabbing stocks; Gavish et al. (2021) find that less sophisticated investors are more prone to buying attention-grabbing stocks. Seasholes and Wu show that attention-grabbing events "induce individual investors to buy stocks they have not previously owned". Overall, there seems that an increase in the level of attention on a stock likely translates into increased purchasing of the stock.

3 Data

The data used in this analysis is split into two categories: UK financial market data (Section 3.1) and proxies of retail trading (Section 3.2).

3.1 Refinitiv FTSE 100 Data

The FTSE 100 pricing data was obtained from the Refinitiv Data platform, and ranges between 01 January 2017 and 01 July 2022. The data contains opening and closing prices, daily high and low values, daily market capitalisation of each stock, daily trading volume, daily turnover (summation of the value of trades during the market day), and closing bid and ask prices. This data is used to create the measures of liquidity and volatility, which are, in turn, used to form the analysis.

There were 5 stocks for which the data is truncated (see Table A.1 in the Appendix); this is due to different listing dates on the London Stock Exchange. The stock symbols for these listings are: *AAF, AVST, EDV, MNG,* and *PSH.* As such, these observations have been removed from the analysis and the remaining analysis has been conducted using a sample of 95 stocks from the FTSE 100 index. For a full list of the constituents in the data set, see Table A.1.

Table 1 shows the summary statistics for the data retrieved from the Refinitiv platform. Table 2 shows the summary statistics for the spreads and returns generated from the original data.

Table 1: Summary Statistics for FTSE 100 Constituents

	Mean	**Std. Dev.**	**Min.**	**Max.**
Open Price	50.65	286.02	0.24	4267.65
Close Price	50.65	285.93	0.24	4263.64
High Price	51.29	289.91	0.25	4299.74
Low Price	50.00	282.00	0.24	4195.46
Closing Bid	50.63	285.83	0.24	4259.63
Closing Ask	50.67	286.04	0.24	4263.64
Daily Volume	8.30M	24.70M	11099	1.41B
Market Cap.	20.30B	29.00B	104.00M	233.00B
Turnover	63.40M	247.0M	63882.70	27.70B

All values rounded to 2 decimal places.
Observations: 131,941 (95 stocks).
Note: M refers to million - i.e., 1M $= 1 \times 10^6$. B refers to billion - i.e., 1B $= 1 \times 10^9$.

3.2 Retail Investor Proxies

Given the limited nature of transaction level data on UK financial transactions without significant cost, I am unable to obtain a measure of trading that can be linked to retail traders with complete accuracy. As such, my analysis is based upon two proxies of retail trading that should closely follow the true investing behaviour of retail investors.

Table 2: Summary Statistics for FTSE 100 Returns & Spreads

	Mean	**Std. Dev.**	**Min.**	**Max.**
Returns				
1-Day Return	0.0003	0.0206	-0.5747	0.5668
5-Day Return	0.0014	0.0461	-0.8134	0.9648
Spreads				
Absolute Spread	0.0360	0.2975	0*	24.0657
Relative Spread	0.0006	0.0011	0*	0.2026

All values rounded to 4 decimal places.
*There are only three instances in the sample for which the closing spread is 0. Excluding these three instances, we get that the minimum absolute and relative spreads are 0.00005 and 0.0001, respectively.

3.2.1 Website Traffic Analysis

Firstly, I use website traffic analysis data for the major trading platforms between January 2017 and July 2022. This was obtained from 'Semrush', a search engine optimisation (SEO) tool that provides the number of visitors to a given domain (or subdomain, in some cases). The data measures the monthly number of 'unique' visitors to the trading platforms.[5] This data will form the basis of my correlational analysis on the impacts. Given the aggregate nature of the data, the number of traders cannot be linked to a specific stock; as such, I will be limited to market-level correlations using this data.

The platforms used are: *CMC Markets, eToro, FreeTrade, Hargreaves Lansdown, IG*, and *Trading 212*. These platforms were selected based on of the number of users using the platforms, as well as a competitor analysis using Semrush for similar sites with a large number of users. These sites commonly listed each other as their top competitors. Whilst this not a comprehensive list of all platforms available to UK investors, it serves as a representative index of the change over time. Figure 1 shows the the number of unique visitors to each of the platforms.

[5] Where possible, the actual trading pages are used, however, when limited (i.e., for eToro and FreeTrade), the login page is used.

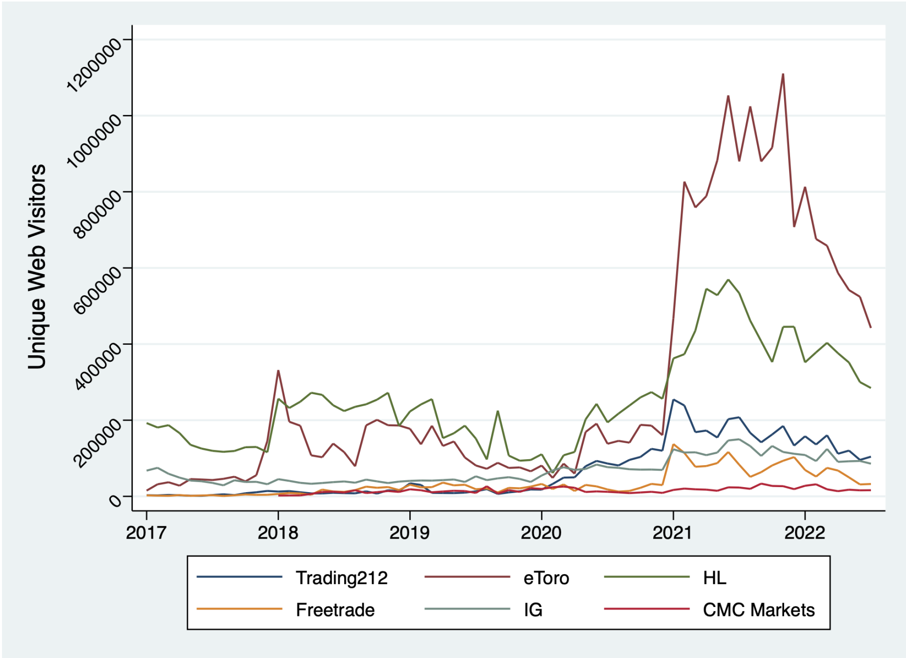

Figure 1: Number of Unique Visitors by Site

3.2.2 Google Search Volume Index

The second proxy is attention-based in nature. I use the weekly Google SVI for the stock's 'ticker' symbol, following Da et al. (2011). The SVI data provides a weekly index for searches of a given term; as it is an index, the values range between 0 and 100.[6] Da et al. choose to use the ticker symbol rather than the company name as it is "less ambiguous", stating that if an investor is searching for a particular stock symbol, they are likely doing so because they are interested in the financial information of the company. In some cases, the search term was combined with the word 'share' to avoid ambiguity with other similar searches; this judgement was made based on the 'related topics' component from Google Trends.[7] To ensure robustness, I conduct dummy variable regressions to test for statistical significance of the effects of noisy searches. I find no evidence of statistical differences when including the 'noisy' stock symbols.[8]

There are eight stocks for which the Google Trends data is excluded. This is due

[6]This data is obtained from Google Trends, which is available at https://trends.google.co.uk/.

[7]For a further explanation, I refer the reader to Appendix A.2.

[8]These results are reported in Appendix B.1.

to the fact that the ticker symbol only is too ambiguous, but there is missing data when using the ticker symbol appended with ‘share’. These stocks are: *DPH, FLTR, HIK, ICP, LAND, RS1, SDR,* and *SKG*. Thus, when combining the financial pricing data (five stocks removed) and Google SVI data (eight stocks removed), the analysis is conducted with 87 stocks. Table 3 shows the summary statistics for the data retrieved from the Refinitiv platform for the 87 stocks that are used in the Google SVI analysis. Whilst the exclusion of these stocks is not ideal, this is unlikely to significantly affect the empirical results, given that the models are run at the stock-level and the SVI data is an index, not an absolute value, meaning that the analysis considers *changes* in this index.

Table 3: Summary Statistics for FTSE 100 Constituents - 87 Stocks

	Mean	**Std. Dev.**	**Min.**	**Max.**
Open Price	50.94	298.64	0.24	4267.65
Close Price	52.94	298.54	0.24	4263.64
High Price	53.61	302.70	0.25	4299.74
Low Price	52.27	294.44	0.24	4195.46
Closing Bid	52.93	298.44	0.24	4259.63
Closing Ask	52.96	298.66	0.24	4263.64
Daily Volume	8.98M	25.70M	11288	1.41B
Market Cap.	21.60B	29.90B	104.00M	233.00B
Turnover	68.10M	257.0M	63882.70	27.70B

All values rounded to two decimal places.
Observations: 120,835 (87 stocks).
Note: M refers to million - i.e., 1M = 1×10^6. B refers to billion - i.e., 1B = 1×10^9.

4 Methodology

This section will discuss the empirical methods used in the analysis of retail investing impact on UK markets. The results of this analysis are reported in Section 5 and discussed in Section 6.

4.1 Proxies of Dependent Variables

Firstly, one must consider the methodology used in producing proxies for the dependent variables, that is, the measures of liquidity and volatility. For a given variable, X, let X_d refer to a daily time interval, X_t to a weekly time interval and X_T to a monthly time interval.

4.1.1 Volatility Measures

Let us first start with volatility. Financial volatility refers to the fluctuations in the returns of an asset, and most commonly refers to the standard deviation, $\hat{\sigma}$, or variance, $\hat{\sigma}^2$, over a set of observations (Poon and Granger, 2003). The variance is given by

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{t=1}^{N}(R_d - \bar{R})^2, \tag{1}$$

where R_d is the return on day d and $\bar{R}$ is the sample mean. I will use the standard deviation measure, $\hat{\sigma}$, in my analysis; in the regressions below, this measure will be denoted by $Volatility$.

Following Baig et al. (2022), I will also use a range-based volatility measure, *LogRange*, à la Alizadeh et al. (2002), in which the volatility is given as the difference in the intraday log high and low quoted prices,

$$LogRange_d = ln(high_d) - ln(low_d). \tag{2}$$

This variable will then be averaged at either a weekly or monthly level, depending on whether the retail investor measure is weekly (Google SVI) or monthly (website traffic analysis).

4.1.2 Liquidity Measures

In my analysis, I use two measures of liquidity (or, conversely, illiquidity); the quoted closing bid-ask spread, and Amihud (2002)'s illiquidity measure. Firstly, the quoted bid-ask spread is simple measure of the liquidity in the market; a narrower spread refers to increased liquidity and lower trading costs. Therefore, if we expect an increase in retail investors to increase market liquidity, then we would expect to see a negative coefficient when regressing on market spreads. In absolute terms, the bid-ask spread is:

$$S_d = a_d - b_d.$$

However, we can look at this in relative terms by dividing this by the midpoint of the two prices:

$$s_d \equiv \frac{S_d}{m_d} = \frac{a_d - b_d}{m_d}, \tag{3}$$

where $m_d = (a_d + b_d)/2$. In the models presented in Section 4.2, I will denote the relative spread, s_d, by *Spread*.

Unfortunately, given the limited access to microstructure (intra-day) data on bid-ask spreads, this analysis is limited to closing bid-ask spreads. As such, finer, intra-day data would be superior in the analysis. This may be a limitation to the statistical significance of some results that focus on liquidity.

Amihud (2002)'s illiquidity measure is given as:

$$Illiq_d = \frac{|R_{i,d}|}{Volume_{i,d}}, \tag{4}$$

where $|R_{i,d}|$ is the absolute daily return of stock i and $Volume_{i,d}$ is the dollar trading volume of the corresponding stock. In my analysis, I will use the logarithmic transformation of the average Amihud ratio over the week/month to make the regression coefficients more interpretable.

The measures of liquidity, as given in Equations (3) and (4), will again by averaged at either the weekly or monthly level for a given stock, depending on the regression model in question.

4.2 Regression Models

This section will present the Autoregressive Distributed Lag (ARDL) models used to measure the impact on liquidity and volatility when using both the attention proxy and the number of users.

In the models below, I account for the potentially persistent effects of volatility and liquidity over the previous months; I consider this using both the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). When there are differences between the two criteria, I will favour the use of the lags recommended by the AIC. I also consider the persistence in the number of users over the preceding period; both criteria recommend the use of one lag of $ln(Users)$. For the weekly attention data, $Attention_t$, both criteria recommend the use of two weeks of lagged attention.

4.2.1 Monthly Regression Models

Models 1, 2, 3, and 4 are monthly ARDL models using the $ln(Users)_T$ measure (website traffic analysis). For both measures of volatility, the AIC and BIC recommend the use of one month of lagged volatility. For the relative spreads, one month is suggested by both criteria. For the Amihud illiquidity ratio, one lag is suggested by both criteria at the monthly level.

Regression Model 1

$$Volatility_{i,T} = \alpha + \beta_1 ln(Users)_T + \beta_2 ln(Users)_{T-1} + \gamma Volatility_{i,T-1} + \delta \cdot \Theta,$$

where Θ is a vector of stock-specific control variables, following Baig et al. (2022). These controls are turnover ($Turnover$), closing bid-ask spreads ($ClosingSpread$), and market capitalisation ($MarketCap$).

Regression Model 2

$$LogRange_{i,T} = \alpha + \beta_1 ln(Users)_T + \beta_2 ln(Users)_{T-1} + \gamma LogRange_{i,T-1} + \delta \cdot \Theta$$

In the liquidity models below, Φ is a vector of control variables similar to those above, however, closing spreads are no longer included as a control. As such, Φ is vector of $MarketCap$ and $Turnover$.

Regression Model 3

$$Spread_{i,T} = \alpha + \beta_1 ln(Users)_T + \beta_2 ln(Users)_{T-1} + \gamma Spread_{i,T-1} + \delta \cdot \Phi$$

Regression Model 4

$$ln(Illiq)_{i,T} = \alpha + \beta_1 ln(Users)_T + \beta_2 ln(Users)_{T-1} + \gamma ln(Illiq)_{i,T-1} + \delta \cdot \Phi$$

4.2.2 Weekly Regression Models

Models 5, 6, 7, and 8 are the weekly regression models using $Attention_t$, the Google SVI data. For $Volatility_t$, the AIC recommends three lags, whilst the BIC recommends two lags; in Model 6, I follow the AIC and use three weeks of lagged $Volatility$. Both the AIC and BIC suggest a lag of one week for $LogRange_t$. For the relative spreads, lags of four weeks are suggested by both criteria. For the Amihud illiquidity ratio, the recommendations are three and two weeks for the AIC and BIC, respectively. As above, I will follow the AIC; thus, I use three weeks of lags.

Regression Model 5

$$Volatility_{i,t} = \alpha + \beta_0 Attention_{i,t} + \sum_{j=1}^{2} \left(\beta_j \cdot Attention_{i,t-j}\right) + \sum_{i=1}^{3} \left(\gamma_k \cdot Volatility_{t-k}\right) + \delta \cdot \Theta$$

Regression Model 6

$$LogRange_{i,t} = \alpha + \beta_0 Attention_{i,t} + \sum_{j=1}^{2} \left(\beta_j \cdot Attention_{i,t-j}\right) + \gamma LogRange_{i,t-1} + \delta \cdot \Theta$$

Regression Model 7

$$Spread_{i,t} = \alpha + \beta_0 Attention_{i,t} + \sum_{j=1}^{2} \left(\beta_j \cdot Attention_{i,t-j}\right) + \sum_{k=1}^{4} \left(\gamma_k \cdot Spread_{i,t-k}\right) + \delta \cdot \Phi$$

Regression Model 8

$$ln(Illiq)_{i,t} = \alpha + \beta_0 Attention_{i,t} + \sum_{j=1}^{2} \left(\beta_j \cdot Attention_{i,t-j}\right) + \sum_{k=1}^{3} \left(\gamma_k \cdot ln(Illiq)_{i,t-k}\right) + \delta \cdot \Phi$$

4.3 Regressions by Market Capitalisation

It is also of interest to consider how the effects of retail investors on liquidity and volatility vary across stocks with different levels of market capitalisation. In the models below, I segment the stocks by using the average market capitalisation over the sample period; these are split into the bottom 25%, middle 50% (interquartile range), and the top 25% of stocks by market capitalisation. As there are 95 stocks used for the monthly level data and 87 for the weekly level, these quartiles vary across the two samples. The weekly and monthly quartile values are given in Table 4.

Table 4: Weekly and Monthly Market Capitalisation Quartiles

	Bottom 25%	Middle 50%	Top 25%
Weekly	6.10×10^{9}	8.64×10^{9}	2.13×10^{10}
Monthly	5.99×10^{9}	8.32×10^{9}	2.04×10^{10}

Weekly: $N = 87$. Monthly: $N = 95$.

I run the models in Section 4.2 conditional on subsets of the data set, split into the three groups. I then test for statistically significant differences across the regression coefficients using:

$$Z = \frac{\hat{\beta} - \tilde{\beta}}{\sqrt{\hat{se}^2 + \tilde{se}^2}}, \tag{5}$$

under the assumption that $Z \sim N(0, 1)$. $\hat{\beta}$ and $\hat{se}$ refer to the regression coefficient and corresponding standard error of one subset of the data, and $\tilde{\beta}$ and $\tilde{se}$ refer to the regression coefficient and corresponding standard error of another subset.

4.4 Structural Breaks Regression Models

I will also consider how the impact of these investors varies over time by considering the effects before the pandemic, during the lockdown, and following the easing of lockdown restrictions. I will define the COVID-19 'lockdown' period as 23 March 2020 to 26 January 2022, as suggested by the timeline given by the British Foreign Policy Group (2022). The period preceding this will be referred to as 'pre-COVID' and the period following as 'post-COVID'.

To conduct this analysis, I use three dummy variables that separate the time series into their respective periods. These variables are $PreCOVID, Lockdown$, and $PostCOVID$. They take the value 1 if they satisfy the time frames below, or 0 otherwise. Given the nature of the data sets, these dummy variables must be computed separately for the weekly and monthly data sets. When some dates are mid-week/mid-month, I round it to the nearest full week/month to ensure that it to matches the other data sets used; I extend the lockdown period (shortening the pre-COVID and post-COVID periods) accordingly.

Let us take Model 1 as an example. I will change this model to include the dummy

	Weekly			Monthly		
PreCOVID	02 Jan 2017	-	22 Mar 2020	Jan 2017	-	Feb 2020
Lockdown	23 Mar 2020	-	31 Jan 2022	Mar 2020	-	Jan 2022
PostCOVID	01 Feb 2022	-	01 Jul 2022	Feb 2022	-	Jul 2022

variables as above:

$$\begin{aligned} Volatility_{i,T} &= \alpha + \beta_1 ln(Users)_T + \beta_2 ln(Users)_{T-1} \\ &+ \sum_{i=0}^{1} \phi_i Lockdown \times ln(Users)_{T-i} \\ &+ \sum_{i=0}^{1} \psi_i PostCOVID \times ln(Users)_{T-i} \\ &+ \gamma Volatility_{i,T-1} + \delta \cdot \Theta \end{aligned}$$

I omit the $PreCOVID$ variable in these models to avoid collinearity. The β_1 and β_2 coefficients account for the effect of retail investors during the pre-COVID period; ϕ_1 and ϕ_2 account for the effects during lockdown; ψ_1 and ψ_2 are the post-COVID effects. The other structural break models will follow the setup in the model above. These results are reported in Section 5.4.

To test for structural breaks in the models, I test for statistical significance of the coefficients on the interaction terms using the following F-test:

$$H_0 : \phi_0 = ... = \phi_{N-1} = \psi_0 = ... = \psi_{N-1} = 0,$$

where N is the number of lags of the retail investing proxy that is used, i.e., in the example above, $N = 2$ and so the null hypothesis would be:

$$H_0 : \phi_0 = \phi_1 = \psi_0 = \psi_1 = 0.$$

4.5 Sectoral Heterogeneity Models

I also consider how the effect of retail investors may vary across industry sectors, with potentially heterogeneous effects on volatility and liquidity. This is done only at the weekly level using the Google SVI data, and so, this analysis is limited to a sample of 87 stocks. Each of these stocks is classified into one of the eight sectors explained in Table 5. The classification of each stock is listed in Table A.4 in the Appendix.

Table 5: Sector Classification

Sector	Description	No. of Stocks
Financial	*Financial services companies, banks, insurance providers, investment trusts, etc.*	20
Communication	*Telecommunication services and media companies.*	10
Materials	*Firms involved in the production of manufacturing goods, including chemicals and construction materials, as well as mining stocks.*	8
Consumer	*Consumer discretionary and consumer staples - including retail companies, food, beverages, tobacco, travel, automobiles, clothing, etc.*	20
Real Estate	*Property development firms, construction, property management, and real estate trusts.*	9
Energy	*Includes both energy producing firms (i.e., oil or natural gas) and utility companies involved in energy transmission and distribution.*	7
Industrial	*Companies involved in the use of heavy machinery (including airlines and transportation) and aerospace, defence and engineering companies.*	8
Other	*Any stocks that do not classify into the sectors above.*	14

I run the models in Section 4.2 conditional on each of the sectors and test the hypothesis that the effects are greater than average in some sectors. I also use the formula in Equation (5) to test whether two coefficients are statistically different from one another.

One consideration is the small sample size of some of the sectors, i.e., there are only eight stocks classified into the energy sector. This may, to some extent, limit the interpretation of some empirical results presented in Section 5.5. An analysis studying sectoral impacts of these investors using a larger sample of stocks would be an interesting starting point for future research.

5 Results

This section presents the empirical results from the models given in Section 4. Each of these models are split into three submodels in the output below: (i) gives the most basic model, with a limited use of lags; (ii) gives the full version (using hetroskedasticity-robust standard errors) but without accounting for panel fixed effects; (iii) is the full model that accounts for fixed effects and robust standard errors.[9]

5.1 Volatility Models Results

Tables 6 and 7 present the monthly regressions for $Volatility_T$ (Model 1) and $LogRange_T$ (Model 2), respectively. The weekly regressions for $Volatility_t$ (Model 5) and $LogRange_t$ (Model 6) can be found in Tables 8 and 9, respectively.

At the monthly level, we observe a positive and statistically significant (p-values < 0.01) coefficient on $ln(Users)_T$ for both measures of volatility. When considering $Volatility$ (Table 6), we see that a 10% increase in number of retail investors correlates with a 0.00062% increase in volatility. For $LogRange$ (Table 7), a 10% increase in the number of users increases the ratio of the daily high and low prices by approximately 0.09%. Based on the average high and low values, this corresponds to an increase in the average range between the daily high and low prices of approximately 0.09 pence (£0.0009).[10] We observe negative coefficients on the $ln(Users)_{T-1}$ for both $Volatility$ and $LogRange$ at the 1% level. These results are similar, in absolute value, to the coefficients of $ln(Users)_T$, suggesting that these effects may cancel out after one month.

At the weekly level, an increase in attention, as measured by the Google SVI, is associated with an increase in the volatility of a stock, both in terms of $Volatility$ and $LogRange$. However, the coefficients are small in magnitude. For the current week, week t, the coefficient of $Attention_t$ is 0.00006 in Table 8 and 0.00004 in Table 9. A

[9]Hetroskedasticity-robust standard errors have been chosen following testing using the modified Wald test for group-wise heteroskedasticity. This test suggested a strong likelihood against the null hypothesis of homoskedasticity at the 1% level.

[10]This is based on an average high price of £51.29 and an average low of £50.00. See Table 1 for more detail.

Table 6: Monthly $Volatility$ Regression Models

	$Volatility_T$		
Model 1	(i)	(ii)	(iii)
$ln(Users)_T$	0.0014***	0.0061***	0.0062***
	(0.0002)	(0.0005)	(0.0005)
$ln(Users)_{T-1}$		-0.0058***	-0.0057***
		(0.0005)	(0.0005)
$Volatility_{T-1}$		0.5299***	0.4327***
		(0.0227)	(0.0208)
$Turnover_T$	1.43×10^{-11}***	4.75×10^{-12}**	1.34×10^{-11}***
	(1.33×10^{-12})	(2.20×10^{-12})	(4.35×10^{-12})
$ClosingSpread_T$	-0.0029**	-0.0017**	-0.0008
	(0.0013)	(0.0007)	(.0006)
$MarketCap_T$	-1.32×10^{-13}***	-2.64×10^{-14}***	-1.27×10^{-13}***
	(1.13×10^{-14})	(6.42×10^{-15})	(2.65×10^{-14})
Constant	0.0020	0.0050***	0.0064***
	(0.0024)	(0.0011)	(0.0018)
Fixed Effects	-	-	Yes
Robust s.e.	-	Yes	Yes
N	6,270	6,175	6,175
R^2	0.0298	0.3208	0.2483

Note: $Turnover$, $ClosingSpread$, $MarketCap$, and the lags of $Volatility$ are monthly averages by stock. *: $p-value < 0.1$, **: $p-value < 0.05$, ***: $p-value < 0.01$.

10-point increase in the SVI coincides with a £0.0006 increase in the fluctuations of price. In terms of $LogRange$, we again see that there is a positive correlation between attention and volatility during the current week, as indicated by $Attention_t$. A 10-point increase in the SVI correlates with a 0.04% increase in $LogRange$ at the weekly level, equivalent to a 0.04 pence (£0.0004) increase in the daily high-low range. We see that the impact of lagged attention on volatility reverses signs after the current week from positive to negative and it stays negative for $Attention_{t-2}$. This is significant at the 1% and 5% levels in Tables 8 and 9, respectively.

Table 7: Monthly *LogRange* Regression Models

	$LogRange_T$		
Model 2	(i)	(ii)	(iii)
$ln(Users)_T$	0.0018***	0.0088***	0.0090***
	(0.0002)	(0.0004)	(0.0005)
$ln(Users)_{T-1}$		-0.0087***	-0.0088***
		(0.0005)	(0.0005)
$LogRange_{T-1}$		0.6474***	0.5603***
		(0.0184)	(0.0174)
$Turnover_T$	5.77×10^{-12}***	1.95×10^{-12}	6.24×10^{-12}
	(1.54×10^{-12})	(2.20×10^{-12})	(5.21×10^{-12})
$ClosingSpread_T$	0.0002	-0.0004	0.0007
	(0.0015)	(0.0007)	(0.0004)
$MarketCap_T$	-1.63×10^{-13}***	-2.73×10^{-14}***	-1.08×10^{-13}***
	(1.36×10^{-14})	(6.00×10^{-15})	(2.59×10^{-14})
Constant	0.0036	0.0079***	0.0091***
	(0.0027)	(0.0011)	(0.0017)
Fixed Effects	-	-	Yes
Robust s.e.	-	Yes	Yes
N	6,270	6,175	6,175
R^2	0.0365	0.4653	0.4261

Note: *Turnover*, *ClosingSpread*, *MarketCap*, and the lags of *LogRange* are monthly averages by stock. *: $p-value < 0.1$, **: $p-value < 0.05$, ***: $p-value < 0.01$.

5.2 Liquidity Models Results

Tables 10 and 11 present the monthly regressions for $Spread_T$ (Model 3) and $ln(Illiq)_T$ (Model 4), respectively. The weekly regressions for $Spread_t$ (Model 7) and $ln(Illiq)_t$ (Model 8) can be found in Tables 12 and 13, respectively. The regression output in the tables below will follow a similar style to those for volatility, including increasingly richer specifications and controls for fixed effects.

The results for the impact of retail investors on market liquidity are less clear than those for volatility. In terms of *Spread*, we do not see statistically significant coefficients at the monthly or weekly levels when considering panel fixed effects; when not using

Table 8: Weekly $Volatility$ Regression Models

	$Volatility_t$		
Model 5	(i)	(ii)	(iii)
$Attention_t$	0.00005***	0.00007***	0.00006***
	(3.06×10^{-6})	(9.53×10^{-6})	(9.65×10^{-6})
$Attention_{t-1}$		-0.00003***	-0.00003***
		(5.57×10^{-6})	(5.60×10^{-6})
$Attention_{t-2}$		-0.00002***	-0.00002***
		(4.79×10^{-6})	(4.34×10^{-6})
$Volatility_{t-1}$	0.4811***	0.3279***	0.3035***
	(0.0055)	(0.0108)	(0.0129)
$Volatility_{t-2}$		0.2059***	0.1872***
		(0.0089)	(0.0095)
$Volatility_{t-3}$		0.1343***	0.1143***
		(0.0108)	(0.0084)
$Turnover_t$	5.46×10^{-12}***	5.29×10^{-12}**	1.07×10^{-11}***
	(4.22×10^{-13})	(2.04×10^{-12})	(3.69×10^{-12})
$ClosingSpread_t$	-0.0017***	-0.0019***	-0.0007
	(0.0003)	(0.0005)	(0.0006)
$MarketCap_t$	-3.30×10^{-14}***	-2.22×10^{-14}***	-9.50×10^{-14}***
	(2.62×10^{-15})	(6.16×10^{-15})	(2.03×10^{-14})
Constant	0.0084***	0.0057***	0.0081***
	(0.0002)	(0.0003)	(0.0006)
Fixed Effects	-	-	Yes
Robust s.e.	-	Yes	Yes
N	24,882	24,708	24,708
R^2	0.2654	0.3292	0.3000

Note: $Turnover$, $ClosingSpread$, $MarketCap$, and the lags of $Volatility$ are all weekly averages by stock. *: $p-value < 0.1$, **: $p-value < 0.05$, ***: $p-value < 0.01$.

fixed effects, the results are only significant at the 10% level. In Table 10, we see that, in column (ii), there is a coefficient on $ln(Users)_T$ of -0.0062, suggesting, to some extent, that an increase in retail investors is associated with greater liquidity through lower spreads. However, this result has a p-value of 0.074 and does not account for the potential for stock-level effects by controlling for fixed effects. In the weekly regression models

Table 9: Weekly *LogRange* Regression Models

	$LogRange_t$		
Model 6	(i)	(ii)	(iii)
$Attention_t$	0.00008***	0.00005***	0.00004***
	(4.26×10^{-6})	(7.31×10^{-6})	(7.55×10^{-6})
$Attention_{t-1}$		-0.00002***	-0.00002***
		(5.73×10^{-6})	(5.61×10^{-6})
$Attention_{t-2}$		-4.30×10^{-6}	-6.04×10^{-6}**
		(3.50×10^{-6})	(2.67×10^{-6})
$LogRange_{t-1}$		0.7287***	0.6812***
		(0.0082)	(0.0184)
$Turnover_t$	6.71×10^{-12}***	2.35×10^{-12}	4.90×10^{-12}
	(6.68×10^{-13})	(2.04×10^{-12})	(3.07×10^{-12})
$ClosingSpread_t$	0.0007	-0.0006	4.57×10^{-6}
	(0.0007)	(0.0004)	(0.0004)
$MarketCap_t$	-2.12×10^{-13}***	-2.34×10^{-14}***	-8.36×10^{-14}***
	(9.33×10^{-15})	(6.25×10^{-15})	(2.04×10^{-14})
Constant	0.0259***	0.0064***	0.0087***
	(0.0006)	(0.0003)	(0.0008)
Fixed Effects	-	-	Yes
Robust s.e.	-	Yes	Yes
N	24,969	24,795	24,795
R^2	0.0358	0.5517	0.5352

Note: *Turnover*, *ClosingSpread*, *MarketCap*, and the lags of *LogRange* are all weekly averages by stock. *: $p-value < 0.1$, **: $p-value < 0.05$, ***: $p-value < 0.01$.

for relative spreads (Table 12), there are no significant coefficients on the *Attention* variable. A cause of the lack of significance when using $Spread_t$ may be the nature of the variable itself; the spread data obtained is the closing bid-ask data, rather than the intra-day quotes. This may, therefore, affect the statistical power of the models.

On the other hand, there are statistically significant results when using the Amihud illiquidity ratio. At the monthly level, a 10% increase in the number of retail users correlates with a decline in illiquidity (an increase in liquidity) of 0.73%. There is a reversal in the signs of the coefficients when considering the number of users during the previous month; this has a coefficient of 0.1069. At the weekly level, only the effect of the

Table 10: Monthly *Spread* Regression Models

		$Spread_T$	
Model 3	(i)	(ii)	(iii)
$ln(Users)_T$	-0.0002	-0.0062*	-0.0051
	(0.0014)	(0.0034)	(0.0032)
$ln(Users)_{T-1}$		0.0058*	0.0048
		(0.0005)	(0.0032)
$Spread_{T-1}$		0.9483***	0.6518***
		(0.0079)	(0.0120)
$Turnover_T$	8.46×10^{-11}***	3.78×10^{-11}	8.81×10^{-12}
	(1.15×10^{-11})	(2.78×10^{-11})	(1.27×10^{-11})
$MarketCap_T$	-1.24×10^{-13}	-5.83×10^{-14}	6.84×10^{-14}
	(1.25×10^{-13})	(4.34×10^{-14})	(4.96×10^{-14})
Constant	0.0307	0.0059*	0.0145***
	(0.0246)	(0.0034)	(0.0027)
Fixed Effects	-	-	Yes
Robust s.e.	-	Yes	Yes
N	6,270	6,175	6,175
R^2	0.2617	0.9241	0.9236

Note: *Turnover*, *MarketCap*, and the lags of *Spread* are monthly averages by stock. *: $p - value < 0.1$, **: $p - value < 0.05$, ***: $p - value < 0.01$.

previous week's attention has a significant impact on illiquidity, whilst the current week has no impact. A 10-point increase in the index during the previous week is associated with a 0.05% decrease in illiquidity. This will be discussed in Section 6.

5.3 Market Capitalisation Regression Results

The models in this section use Models 1 through 8 with fixed effects and robust standard errors, but condition on subsets of the data, namely, quartiles of market capitalisation. The tables below report only the coefficients and standard errors for the retail investor proxies, however, the regression is run using the full model, including lags of the dependent variables and the controls.

At the monthly level, $ln(Users)_T$ is positively correlated with both measures of volatility across all three levels of market capitalisation. I find negative and statisti-

Table 11: Monthly $ln(Illiq)$ Regression Models

		$ln(Illiq)_T$	
Model 4	(i)	(ii)	(iii)
$ln(Users)_T$	0.1683***	-0.1223***	-0.0733***
	(0.0080)	(0.0167)	(0.0179)
$ln(Users)_{T-1}$		0.1288***	0.1069***
		(0.0161)	(0.0163)
$ln(Illiq)_{T-1}$		0.8862***	0.7344***
		(0.0108)	(0.0196)
$Turnover_T$	-9.94×10^{-10}***	-4.13×10^{-10}***	-4.20×10^{-10}***
	(6.56×10^{-11})	(3.57×10^{-11})	(5.11×10^{-11})
$MarketCap_T$	-2.01×10^{-11}***	-2.35×10^{-12}***	-5.58×10^{-12}***
	(6.88×10^{-13})	(4.94×10^{-13})	(1.16×10^{-12})
Constant	-23.3300***	-2.4568***	-6.0243***
	(0.1188)	(0.2741)	(0.4695)
Fixed Effects	-	-	Yes
Robust s.e.	-	Yes	Yes
N	6,270	6,175	6,175
R^2	0.5122	0.9233	0.9171

Note: $Turnover$, $MarketCap$, and the lags of $ln(Illiq)$ are monthly averages by stock. *: $p-value < 0.1$, **: $p-value < 0.05$, ***: $p-value < 0.01$.

cally significant coefficients of $ln(Users)_{T-1}$ across all levels of market capitalisation, suggesting a similar pattern of mean-reversion. The correlation between $Volatility$ and $ln(Users)_T$ is larger for the bottom 25% of stocks than for the middle 50% and the top 25% (one sided p-values are 0.0720 and 0.0653, respectively). A 10% increase in the number of users coincides with a 0.0008% increase in volatility for the bottom 25%, but only 0.0005% for the top 25%. The coefficients for the middle 50% and the top 25% are statistically indifferent. For $LogRange$, I find a similar effect; the coefficient on $ln(Users)_T$ is stronger for the bottom 25% of stocks when compared to both the middle 50% and top 25% (one sided p-values are 0.0899 and 0.0106, respectively).

At the weekly level, we see that the coefficients on $Attention_t$ are positive across all three levels of market capitalisation at the 1% level. The first lags of attention are negative across groups for both $Volatility$ and $LogRange$; the second lags are negative and significant for $Volatility$, but are only significant for $LogRange$ for the top 25% (at

Table 12: Weekly *Spread* Regression Models

		$Spread_t$	
Model 7	(i)	(ii)	(iii)
$Attention_t$	0.00004	0.00005	0.00006
	(0.00004)	(0.00005)	(0.00007)
$Attention_{t-1}$		-0.00003	-9.65×10^{-6}
		(0.00003)	(7.03×10^{-6})
$Attention_{t-2}$		-0.00006	-0.00004
		(0.00006)	(0.00004)
$Spread_{t-1}$		0.2731***	0.1877***
		(0.0035)	(0.0014)
$Spread_{t-2}$		0.3416***	0.2645***
		(0.0015)	(0.0049)
$Spread_{t-3}$		0.2088***	0.1334***
		(0.0023)	(0.0011)
$Spread_{t-4}$		0.1175***	0.0403***
		(0.0042)	(0.0019)
$Turnover_t$	1.90×10^{-11}***	2.96×10^{-11}	-6.13×10^{-12}
	(6.05×10^{-12})	(2.64×10^{-11})	(6.35×10^{-12})
$MarketCap_t$	2.16×10^{-13}**	-5.08×10^{-14}	9.73×10^{-14}
	(9.23×10^{-14})	(4.13×10^{-14})	(6.19×10^{-14})
Constant	0.0303*	0.0025	0.0118***
	(0.0182)	(0.0018)	(0.0019)
Fixed Effects	-	-	Yes
Robust s.e.	-	Yes	Yes
N	24,969	24,621	24,621
R^2	0.0295	0.8159	0.8137

Note: $Turnover$, $MarketCap$, and the lags of $Spread$ are all weekly averages by stock. *: $p-value < 0.1$, **: $p-value < 0.05$, ***: $p-value < 0.01$.

the 5% level). For both $Volatility$ and $LogRange$, I find that the magnitudes of the coefficients on the current week of the retail investor proxy are statistically indifferent across levels of market capitalisation at the 10% level.

When considering liquidity, the coefficients for the relative spreads are largely insignificant, except for the middle 50% of stocks. This coefficient is significant at the 1%

Table 13: Weekly $ln(Illiq)$ Regression Models

		$ln(Illiq)t$	
Model 8	(i)	(ii)	(iii)
$Attention_t$	0.0003**	0.0002	0.0002
	(0.0002)	(0.0002)	(0.0002)
$Attention_{t-1}$		-0.0005**	-0.0005**
		(0.0002)	(0.0002)
$Attention_{t-2}$		0.0002	0.0001
		(0.0002)	(0.0002)
$ln(Illiq)_{t-1}$		0.3677***	0.3192***
		(0.0096)	(0.0119)
$ln(Illiq)_{t-2}$		0.2605***	0.2157***
		(0.0083)	(0.0082)
$ln(Illiq)_{t-3}$		0.2511***	0.2024***
		(0.0069)	(0.0078)
$Turnover_t$	-5.68×10^{-10}***	-3.67×10^{-10}***	-2.74×10^{-10}***
	(3.00×10^{-11})	(6.64×10^{-11})	(5.61×10^{-11})
$MarketCap_t$	-1.98×10^{-11}***	-2.36×10^{-12}***	-5.31×10^{-12}***
	(4.49×10^{-13})	(5.28×10^{-13})	(1.08×10^{-12})
Constant	-21.2766***	-2.5403***	-5.5683***
	(0.0535)	(0.3322)	(0.4535)
Fixed Effects	-	-	Yes
Robust s.e.	-	Yes	Yes
N	24,969	24,708	24,708
R^2	0.3988	0.8367	0.8306

Note: $Turnover$, $MarketCap$, and the lags of $Spread$ are all weekly averages by stock. *: $p - value < 0.1$, **: $p - value < 0.05$, ***: $p - value < 0.01$.

level and suggests that a 10% increase in the number of retail traders correlates with a 0.014% decrease in the relative spread. For $ln(Illiq)_T$, the only significant coefficient for the $ln(Users)_T$ is for the bottom 25% of stocks, with the other two coefficients being statistically insignificant at the 10% level. For the bottom 25% of stocks, a 10% increase in the number of users approximately correlates with a 1% increase in liquidity.

At the weekly level, there are no significant coefficients (at or above the 5% level) for the relative spreads across all three levels of market capitalisation. For the Amihud

Table 14: Monthly Volatility Models - Market Capitalisation

	$Volatility_T$		
Model 1	(< 25%)	(25% − 75%)	(> 75%)
$ln(Users)_T$	0.0075***	0.0056***	0.0054***
	(0.0012)	(0.0005)	(0.0007)
$ln(Users)_{T-1}$	-0.0060***	-0.0035***	-0.0048***
	(0.0012)	(0.0005)	(0.0008)
N	1,560	3,055	1,560
R^2	0.2246	0.3887	0.3254
	$LogRange_T$		
Model 2	(< 25%)	(25% − 75%)	(> 75%)
$ln(Users)_T$	0.0106***	0.0088***	0.0074***
	(0.0012)	(0.0006)	(0.0007)
$ln(Users)_{T-1}$	-0.0095***	-0.0068***	-0.0068***
	(0.0012)	(0.0006)	(0.0008)
N	1,560	3,055	1,560
R^2	0.4512	0.5030	0.4271

The models above use fixed effects and heteroskedasticity-robust standard errors. *: $p - value < 0.1$, **: $p - value < 0.05$, ***: $p - value < 0.01$. ( ) = standard errors.

illiquidity ratio, we see, contrasting the monthly correlations, that the coefficients are only significant for the middle 50% and top 25%; both of these coefficients are positive at the 5% level. These suggest that a 10-point increase correlates with a 0.9% decrease in liquidity for the top 25%. Whilst the coefficient may seem larger for the top 25% than the middle 50% (0.00094 compared to 0.00055), these coefficients are statistically indifferent (one-sided p-value = 0.1837).

5.4 Structural Break Model Results

In this section, I split the data into three parts using dummy variables for three time periods: pre-COVID, lockdown, and post-COVID. The tables in this section report only the coefficients and standard errors for the retail investor proxies, as well as the F-statistic, however, the model is run using the lags of the dependent variables and the controls, using fixed effects and heteroskedasticity-robust standard errors.

The monthly models (Table 18) suggest that there are structural breaks in the cor-

Table 15: Weekly Volatility Models - Market Capitalisation

	$Volatility_t$		
Model 5	$(< 25\%)$	$(25\% - 75\%)$	$(> 75\%)$
$Attention_t$	0.00006***	0.00005***	0.00005***
	(0.00002)	(0.00001)	(0.00002)
$Attention_{t-1}$	-0.00003**	-0.00002**	-0.00005***
	(0.00001)	(0.00001)	(0.00001)
$Attention_{t-2}$	-0.00001*	-0.00002*	-0.00003***
	(0.00001)	(0.00001)	(0.00001)
N	5,985	12,540	6,270
R^2	0.5186	0.5916	0.5862
	$LogRange_t$		
Model 6	$(< 25\%)$	$(25\% - 75\%)$	$(> 75\%)$
$Attention_t$	0.00004***	0.00003***	0.00003**
	(0.00001)	(0.00001)	(0.00001)
$Attention_{t-1}$	-0.00002*	-0.00002**	-0.00004***
	(0.00001)	(0.00001)	(0.00001)
$Attention_{t-2}$	-0.00001	-0.00000	-0.00001**
	(0.00000)	(0.00001)	(0.00001)
N	5,985	12,540	6,270
R^2	0.1957	0.4086	0.4633

The models above use fixed effects and heteroskedasticity-robust standard errors. *: $p - value < 0.1$, **: $p - value < 0.05$, ***: $p - value < 0.01$. ( ) = standard errors.

relations between $ln(Users)_T$ and $Volatility$, $LogRange$, and $ln(Illiq)$, but not $Spread$. The coefficients on $ln(Users)_T$ are negative during the pre-COVID period for $Volatility$ and $LogRange$, however, the coefficients on the interaction terms suggest a positive and significant increase both during the lockdown and post-COVID periods. This is the same for $ln(Illiq)$.

The correlations for the volatility measures after the pre-COVID period become positive; the effects during the lockdown periods are 0.0011 and 0.0038 for $Volatility$ and $LogRange$, respectively. The effects of the COVID structural break more than than cancel out the negative pre-COVID coefficient on $ln(Users)_T$ for both measures (p-value = 0.0044 for $Volatility$ and p-value < 0.0001 for $LogRange$). These lockdown effects are statistically indifferent in the post-COVID period, suggesting that the correlations

Table 16: Monthly Liquidity Models - Market Capitalisation

	$Spread_T$		
Model 3	$(< 25\%)$	$(25\% - 75\%)$	$(> 75\%)$
$ln(Users)_T$	-0.0202	-0.0014***	0.0021
	(0.0136)	(0.0004)	(0.0016)
$ln(Users)_{T-1}$	0.0139	0.0008**	-0.0025
	(0.0095)	(0.0004)	(0.0018)
N	1,560	3,055	1,560
R^2	0.9200	0.8373	0.3746
	$ln(Illiq)_T$		
Model 4	$(< 25\%)$	$(25\% - 75\%)$	$(> 75\%)$
$ln(Users)_T$	-0.1029**	-0.0197	0.0089
	(0.0495)	(0.018)	(0.0241)
$ln(Users)_{T-1}$	0.1752***	0.0974***	0.0244
	(0.0425)	(0.0185)	(0.0252)
N	1,560	3,055	1,560
R^2	0.9407	0.7922	0.8314

The models above use fixed effects and heteroskedasticity-robust standard errors. *: $p - value < 0.1$, **: $p - value < 0.05$, ***: $p - value < 0.01$. ( ) = standard errors.

do not return to the pre-COVID levels.

For Amihud illiquidity measure, we see a negative coefficient on $ln(Users)_T$ at the 1%, and a positive coefficient on $Lockdown \times ln(Users)_T$. However, unlike the volatility models, this coefficient is not sufficient to swap the sign. The effect of the post-COVID interaction term at month T is also positive and statistically indifferent from the lockdown interaction term (p-value = 0.2962), suggesting no change between the lockdown and post-COVID periods.

When considering the structural break models at the weekly level (Table 19), there may be significant structural breaks for across the volatility and liquidity measures. The lockdown interaction terms are positive at the 1% level for both volatility measures. Whilst a 10-point increase in the SVI correlates with a 0.0004 point increase in $Volatility$ (0.04% for $LogRange$) in the pre-COVID period, it correlates with a 0.0013 point increase (0.07% for $LogRange$) during the lockdown period. The coefficient on $Lockdown \times Attention_t$ for the $Volatility$ model is statistically greater than

Table 17: Weekly Liquidity Models - Market Capitalisation

	$Spread_t$		
Model 7	(< 25%)	(25% – 75%)	(> 75%)
$Attention_t$	-2.43×10^{-7}	-1.29×10^{-7}	1.14×10^{-7}
	(2.65×10^{-7})	(1.79×10^{-7})	(2.7×10^{-7})
$Attention_{t-1}$	2.96×10^{-7}*	-3.36×10^{-7}*	3.45×10^{-7}
	(1.61×10^{-7})	(1.75×10^{-7})	(2.17×10^{-7})
$Attention_{t-2}$	-1.32×10^{-7}	7.22×10^{-8}	-3.11×10^{-8}
	(2.14×10^{-7})	(1.53×10^{-7})	(2.54×10^{-7})
N	5,943	12,452	6,226
R^2	0.5313	0.1676	0.0083

	$ln(Illiq)_t$		
Model 8	(< 25%)	(25% – 75%)	(> 75%)
$Attention_t$	0.00005	0.00055**	0.00094**
	(0.00025)	(0.00024)	(0.00036)
$Attention_{t-1}$	-0.00094***	0.00010	-0.00091*
	(0.00032)	(0.00029)	(0.00050)
$Attention_{t-2}$	0.00017	0.00032	-0.00043
	(0.00036)	(0.00027)	(0.00032)
N	5,964	12,496	6,248
R^2	0.8902	0.6420	0.6482

The models above use fixed effects and heteroskedasticity-robust standard errors. *: $p-value < 0.1$, **: $p-value < 0.05$, ***: $p-value < 0.01$. ( ) = standard errors.

the $PostCOVID \times Attention_t$ coefficient (one-sided p-value = 0.0002), suggesting that the effect is smaller post-COVID than during the lockdown period. For $LogRange$, the $PostCOVID \times Attention_t$ term is insignificant for $LogRange$, suggesting the effect is indifferent from the pre-COVID level when using this measure.

The structural break models for $Spread_t$ show primarily insignificant results in the pre-COVID period, but there is some evidence of an effect during the lockdown period. The negative coefficients on the lockdown interaction terms on $Attention_{t-1}$ and $Attention_{t-2}$ show that increased retail investing during the pandemic is correlated with increased market liquidity, with some delay, through lower relative spreads. The coefficient suggests that a 10-point increase in the SVI correlates with a 0.0005% decrease in the relative spread during the following week, whereas this has no effect in the pre-

Table 18: Structural Break Tests - Monthly Models

	$Volatility$	$LogRange$	$Spread$	$ln(Illiq)$
$ln(Users)_T$	-0.0023***	-0.0008***	-0.0035	-0.2619***
	(0.0004)	(0.0003)	(0.0022)	(0.0172)
$ln(Users)_{T-1}$	0.0007	0.0007**	0.0048	0.2269***
	(0.0004)	(0.0003)	(0.003)	(0.0169)
$Lockdown \times ln(Users)_T$	0.0034***	0.0046***	-0.0004	0.0675***
	(0.0002)	(0.0002)	(0.0006)	(0.0029)
$Lockdown \times ln(Users)_{T-1}$	-0.0031***	-0.0044***	0.0002	-0.0567***
	(0.0002)	(0.0002)	(0.0004)	(0.0028)
$PostCOVID \times ln(Users)_T$	0.0037***	0.0048***	-0.0005	0.0695***
	(0.0002)	(0.0002)	(0.0008)	(0.0032)
$PostCOVID \times ln(Users)_{T-1}$	-0.0032***	-0.0045***	0.0002	-0.0566***
	(0.0002)	(0.0002)	(0.0006)	(0.0032)
N	6,175	6,175	6,175	6,175
F-stat	0.0000	0.0000	0.1789	0.0000

The models above use fixed effects and heteroskedasticity-robust standard errors. *: $p-value < 0.1$, **: $p-value < 0.05$, ***: $p-value < 0.01$. ( ) = standard errors.

pandemic period. The correlation between $Attention_t$ and $ln(Illiq)_t$ is significant and negative at the 5% level during the pre-COVID period. The interaction terms suggest that the sign of these correlation changes following the start of the lockdown, with the relationship remaining unchanged in the post-COVID period (p-value = 0.7621).

5.5 Sectoral Heterogeneity Results

Table 20 shows positive and statistically significant volatility coefficients on $Attention_t$ for financial, communication, consumer goods, and real estate stocks. These sectors are statistically significant, at least at the 10%, when considering both $Volatility$ and $LogRange$. Additionally, when using the $Volatility$ measure, we see that industrial stocks are statistically significant at the 10% level. The results indicate that there are no significant effects when considering materials, energy or 'other' stocks. With stocks classified as 'other', it is likely that there are offsetting effects from the mixed nature of these stocks, resulting in an overall significant result.

Table 19: Structural Break Tests - Weekly Models

	$Volatility$	$LogRange$	$Spread$	$ln(Illiq)$
$Attention_t$	0.00004***	0.00004***	-2.54×10^{-8}	-0.00050**
	(0.00001)	(0.00001)	(1.70×10^{-7})	(0.00019)
$Attention_{t-1}$	-0.00001**	-0.00002***	2.23×10^{-7}	-0.00053**
	(0.00001)	(0.00001)	(1.43×10^{-7})	(0.00023)
$Attention_{t-2}$	-0.00002***	-0.00002***	3.22×10^{-7}**	-0.00015
	(0.00001)	(0.00001)	(1.70×10^{-7})	(0.00019)
$Lockdown \times Attention_t$	0.00009***	0.00003***	6.01×10^{-8}	0.00183***
	(0.00001)	(0.00001)	(1.70×10^{-7})	(0.00019)
$Lockdown \times Attention_{t-1}$	-0.00005***	-0.00002**	-5.30×10^{-7}**	0.00003
	(0.00001)	(0.00001)	(1.70×10^{-7})	(0.00019)
$Lockdown \times Attention_{t-2}$	-0.00001	0.00002***	-8.16×10^{-7}***	0.00052
	(0.00001)	(0.00001)	(1.70×10^{-7})	(0.00019)
$PostCOVID \times Attention_t$	0.00004**	-0.00001	1.73×10^{-7}	0.00200***
	(0.00001)	(0.00001)	(1.70×10^{-7})	(0.00019)
$PostCOVID \times Attention_{t-1}$	-0.00002	0.00001	-4.19×10^{-7}	-0.00059
	(0.00001)	(0.00001)	(1.70×10^{-7})	(0.00019)
$PostCOVID \times Attention_{t-2}$	0.00004***	0.00005***	-1.09×10^{-6}***	0.00130**
	(0.00001)	(0.00001)	(1.70×10^{-7})	(0.00019)
N	24,708	24,795	24,621	24,708
F-test	0.0000	0.0000	0.0000	0.0000

The models above use fixed effects and heteroskedasticity-robust standard errors. *: $p - value < 0.1$, **: $p - value < 0.05$, ***: $p - value < 0.01$.
( ) = standard errors.

Table 20: Sector Regressions

	Financial	Communication	Materials	Consumer	Real Estate	Energy	Industrial	Other
				$Volatility_t$				
$Attention_t$	0.000043***	0.000082*	0.000008	0.000082***	0.000046**	0.000039	0.000098*	0.00002
	(0.000013)	(0.000042)	(0.000023)	(0.000024)	(0.000017)	(0.000027)	(0.000044)	(0.000012)
N	4,828	2,556	1,998	5,396	2,272	1,988	1,704	3,976
				$LogRange_t$				
$Attention_t$	0.000018*	0.000062*	-0.000004	0.000062***	0.000037*	0.000015	0.000073	0.000012
	(0.000009)	(0.000028)	(0.000018)	(0.000019)	(0.000017)	(0.000024)	(0.000042)	(0.000009)
N	4,845	2,565	1,995	5,415	2,280	1,995	1,710	3,990
				$Spread_t$				
$Attention_t$	-0.0000003	-0.0000002	0.0000001	0.0000002	-0.0000001	-0.0000009	0.0000004	-0.0000001
	(0.0000004)	(0.0000003)	(0.0000004)	(0.0000003)	(0.0000003)	(0.0000009)	(0.0000006)	(0.0000002)
N	4,811	2,547	1,981	5,377	2,264	1,981	1,698	3,962
				$ln(Illiq)_t$				
$Attention_t$	0.000522	0.000384	-0.000108	0.000613*	0.000481	0.000593	0.000248	-0.000136
	(0.000349)	(0.000666)	(0.000887)	(0.000292)	(0.000431)	(0.000687)	(0.000414)	(0.000427)
N	4,828	2,556	1,988	5,396	2,272	1,988	1,704	3,976

The models above use fixed effects and heteroskedasticity-robust standard errors. *: $p - value < 0.1$, **: $p - value < 0.05$, ***: $p - value < 0.01$.
( ) = standard errors.

I find that the volatilities of financial and materials stocks are statistically less impacted by retail investor attention than the average stock; the average one-sided p-values across $Volatility$ and $LogRange$ are the are 0.0517 for financial and 0.0081 for materials.[11] A 10-point increase in the SVI is correlated with a 0.0006 point increase in $Volatility$, on average, but only a 0.00043 point increase when considering financial stocks. Whilst consumer good stocks are not statistically more affected than the average stock, they are more affected by attention than financial stocks (average one-sided p-value = 0.0455). These findings are discussed in Section 6.

When considering both measures of liquidity, the results are statistically insignificant across all sectors. This suggests that these models may be more affected by data limitations and limited group samples sizes than the volatility measures.[12]

6 Discussion

Before discussing the implications of the empirical results, one should consider the extent to which the results can be identified as causal. The relationship between retail investors and liquidity and volatility may not be entirely casual; as Welch (2022) suggests, there may exist a bi-directional relationship between the two variables. Whilst retail investors may be a cause of increased volatility, they may also be attracted by it. Thus, this may amplify any periods of higher volatility, and add to the complexity of identifying the direction of the effect. Given the aggregate nature of the Semrush website traffic data, it would be difficult to draw a causal relationship between the number of retail investors and liquidity/volatility at the market level. As such, I will discuss these results from a correlational relationship standpoint to consider how these variables comove. With the Google SVI data, it may be more plausible to establish a causal relationship between the two, given that this analysis is conducted at the ticker level, and is separated into groups by market capitalisation and industry sector. By controlling for the lags of both attention and the measures of volatility and liquidity, one can, to some extent,

[11] The effects of current attention on $Volatility$ and $LogRange$ without conditioning on the sector (the average across all stocks) are 0.00006 and 0.00004, respectively (see Tables 8 and 9).

[12] The results are primarily insignificant when considering both current attention, $Attention_t$, as well as its lags, $Attention_{t-1}$ and $Attention_{t-1}$.

remove the effects of reverse causality. If the lags in the models are sufficient to control for prior effects of attention and stock volatility and liquidity, then this should break the feedback loops between retail investors and liquidity/volatility, meaning that the resulting coefficients should have a causal interpretation.

At the monthly levels, we see similar patterns for both measures of volatility. There is a positive correlation between volatility and the number of users in the current month. An increase in the number of traders may be associated with increased noise trading, which increases price fluctuations. However, given the coarse nature of the monthly data, it would be difficult to asset that retail investors *cause* higher volatility at the monthly level. The reversal the following month suggests that, following a shock in the number of users, there may be evidence of mean-reversion and stabilisation, with the relationship fading and returning to the pre-shock levels. Seasholes and Wu (2007) find that prices mean-revert to pre-event levels after an attention-grabbing event induces net-buying. Whilst this is not directly related to volatility, it does suggest that the effects of retail investors may be somewhat short-lived, with markets mean-reverting after a short period of time.

The weekly attention models confirm that there is a positive relationship between retail investor attention and volatility in the current week. If attention translates into trading, as suggested by the literature, then we may expect higher retail trading to increase volatility. The aggressive trading strategies by some investors (Barber, Lee, et al., 2009), as well as the irrationality and biases of these investors, may result in their trading causing increased volatility.

In terms of liquidity, the relative spread models are largely statistically insignificant. However, this is more likely a limit of the data than the true correlation, especially when the Amihud illiquidity measure shows statistical significance. With access to intra-day bid-ask spread data, the statistical power would be improved. In terms of the Amihud ratio, we see a negative relationship between the ratio and the number of users, that is, there is a positive relationship between market liquidity and users in the current month.

The models show statistically insignificant relationships for spreads when considering the weekly models. For $ln(Illiq)_t$, the coefficient on $Attention_t$ is insignificant, but $Attention_{t-1}$ is significant and negatively correlated with the illiquidity ratio; the first lag is positively related to increased stock liquidity. A possible explanation is that liquidity-providing retail investors are slower to react, meaning there is a delay between their attention towards a stock and a trade being placed. The type of retail investor and their trading behaviour would have an impact on market liquidity. Those that trade more aggressively and sooner may be more likely to place a liquidity-consuming market order; those that wait may be more inclined to place a liquidity-providing limit order. Whilst the effect for volatility was immediate, this may be because the investors that contribute to volatility through more aggressive trading are likely to act quicker; that is, that they wait less time to trade after a stock catches their attention. Overall, there is some evidence to suggest that retail investors provide liquidity to stocks that catch their attention.

The results from the market capitalisation conditional models show that there is a degree of heterogeneity in the effects of these investors on stock volatility, depending on the size of the company in question. We see that the (positive) correlations between the number of users and volatility, for both $Volatility$ and $LogRange$, were statistically larger for stocks with a lower market capitalisation at the monthly level. However, the weekly attention models do not show any significant differences between effects on volatility for the top and bottom 25% of stocks. The magnitude of the effect of retail investor attention on volatility is relatively small, meaning that, even if there is a statistical significant heterogeneity, it may not be clear given the magnitude. Alternatively, this may be due to the nature of the stocks used in the sample; these stocks are, by definition, large-cap stocks, given that they make up the FTSE 100 index. It may be that the range of market capitalisation is too narrow to observe significant heterogeneity.

At the monthly levels, the relationship between $ln(Users)_T$ and $Spread_T$ is statistically significant only for the middle 50% of stocks, and suggests that there is a positive relationship between users and market liquidity. One hypothesis behind this relationship is that retail investors do not have enough market power to contribute to the liquidity of the top 25% of stocks; for the bottom 25%, they may be able to affect liquidity but there could be contrasting effects of limit and market orders by these investors. However, few conclusions can be drawn from this relationship. For $ln(Illiq)$, I find that the relationship between users and liquidity provision is positive for the smallest stocks by market capitalisation.

There is a negative relationship between the current week's attention and $ln(Illiq)_t$

for largest stocks. However, it seems less plausible that these traders can significantly affect the liquidity of largest stocks in a negative manner, suggesting empirical limitations. However, for $Attention_{t-1}$, I find that retail investors lead to increased liquidity, but only for the bottom 25%. This supports the hypothesis that those investors that provide liquidity may take longer to decide to trade.

The structural break results generally suggest that the pandemic created notable breaks in the effects of retail investors, with some evidence of a permanent change in the nature of retail investors in UK markets. The investors attracted by the increased free time during the pandemic may have developed a lasting interest and continue to trade in, and have an influence on, UK markets.

Both the weekly and monthly models suggest that the volatility effect is greater during lockdown and continued during the post-pandemic period. The effect of attention on volatility triples during the pandemic; this cannot be solely attributed to retail investors, given the increased market uncertainty, but it is likely that these investors increased volatility, given that many of the newly attracted investors had little experience or trading knowledge. Hence, their trading increased volatility. Whilst the amplified effect of attention on $Volatility$ continued post-COVID, it is statistically lower than the effect during lockdown. This suggests that some of the volatility-inducing investors were only trading when they had more time to do so.

In the weekly models, the greater number of retail investors may have increased market liquidity during the lockdown, as suggested by the negative coefficient on $Lockdown \times Attention_{t-1}$ in the *Spread* model. The results suggest that this effect was significant only for the lockdown period. In terms of the Amihud measure, the effect of current attention, $Attention_t$, on liquidity becomes negative during the lockdown, implying that more attention may reduce liquidity as the aggressive trades of retail investors may have become liquidity-consuming. The results suggest that this effect carries on post-COVID. For the delayed impact using the prior week's attention, $Attention_{t-1}$, the effect on liquidity remains positive and unchanged during the lockdown and post-COVID period. Whilst the break in volatility seems evident, the implications of the pandemic on liquidity are less so, with the two measures showing slightly varying results. A reason for this is the decreased liquidity at the start of the pandemic due to uncertainty, which may have biased the positive effects of retail investors on liquidity downwards.

The sectoral regression models suggest that retail investor attention can have heterogeneous effects on volatility across sectors. Namely retail investors have less of an impact on financial and materials stocks, but potentially more of an effect on stocks relating to consumer goods. Whilst the coefficient for financial stocks is statistically positive, it is lower than the average effect of retail investors on volatility. One possible explanation is that financial stocks are more heavily traded in general and so additional retail investors have a reduced impact when combined with additional traders. I find that the average daily trading volume is statistically greater for financial stocks than the average over the whole sample, lending credit to this hypothesis.[13]

The statistically insignificant results on materials and energy may be due to the nature of these stocks, in that they may be simply less appealing to retail investors, and hence, they trade in them less. Many of the materials and energy stocks are not well-known household names, and hence, without prior exposure to these stocks, retail investors may engage with them less. In contrast, communication and consumer good stocks are likely to be more well-known, and so, attention on these stocks translates into trading and, consequently, increased volatility. Whilst I do not find that the effect for consumer goods stocks is statistically greater than the overall effect, it is statistically greater than financial and materials stocks for both volatility measures, as well as energy when considering $LogRange$.

7 Conclusion

In this paper, I analyse the impacts of retail investor trading on volatility and liquidity in UK equity markets using a sample of FTSE 100 stocks. I use the Google SVI and trading platform website traffic analysis as proxies for retail investor trading. I combine these data sets with the FTSE 100 pricing data to run ARDL models on this data. I also test for heterogeneity in the effects by running regressions conditional on subsets of the data, namely by quartiles of market capitalisation and by industry sector, as well as testing for structural breaks in the effects during the COVID-19 pandemic.

[13] I obtain p-value < 0.001, using Welch's t-test, under the null hypothesis that the two groups have the same mean.

Overall, I find that there is a positive relationship between retail investing and volatility when considering both the standard deviation and daily range measures. The results suggest that increased stock attention transmits into increased trading, which, in turn, increases the volatility of a given stock. Whilst the effects of these investors on liquidity is less clear, there is some evidence that attention on a stock leads to increased liquidity the following week, that is, this effect occurs with some delay. This could be due to the nature of retail investor behaviour in that those investors that provide liquidity through limit orders are slower to react and decide to trade in markets. There is little significant impact of attention in the current week on liquidity. This may be, in part, due to the potentially offsetting effects of different types of retail investors. There are some more aggressive traders who act quickly and place liquidity-consuming market orders, which offsets the effects by those traders placing liquidity-providing limit orders. Hence, I find some potential evidence in support of the heterogeneous nature of retail investors themselves; that is, we cannot assume that all retail investors are identical.

When conditioning on stocks by market capitalisation, there is some, albeit limited, evidence that the relationship between these traders and volatility is greater for stocks with a lower market capitalisation. This relationship is visible at the monthly level, but disappears when considering this at the weekly level. The breakdown of the relationship at the weekly level may be due to the small magnitude of the effect or due to an insufficient market capitalisation range in the data set.

There is also heterogeneity in the impact of retail investors when conditioning on industry sectors. Retail investors have a positive, but lower than average, impact on financial stocks, and a limited impact on stocks that are less well-known, such as materials and energy stocks. There is some evidence to suggest that the impact of attention on volatility is greater for consumer goods stocks. There are no significant effects across sectors when considering the impact of these investors on liquidity, although this is, in part, due to the limitations of the data.

Furthermore, it appears that the COVID-19 pandemic caused structural breaks in the impact of retail investors on volatility. We see that the impact on volatility is significantly higher during the pandemic, although there is some evidence to suggest that this effect fades slightly by the end of lockdown period. For liquidity, there is no clear interpretation of a structural break during the pandemic.

References

Albulescu, C. T. (2021). COVID-19 and the United States financial markets' volatility. *Finance Research Letters*, *38*, 101699. https://doi.org/10.1016/J.FRL.2020.101699

Alizadeh, S., Brandt, M. W., & Diebold, F. X. (2002). Range-Based Estimation of Stochastic Volatility Models. *The Journal of Finance*, *57*(3), 1047–1091. https://doi.org/10.1111/1540-6261.00454

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31–56. https://doi.org/10.1016/S1386-4181(01)00024-6

Baig, A. S., Blau, B. M., Butt, H. A., & Yasin, A. (2022). Do retail traders destabilize financial markets? An investigation surrounding the COVID-19 pandemic. *Journal of Banking & Finance*, *144*, 106627. https://doi.org/10.1016/J.JBANKFIN.2022.106627

Baker, S. R., Bloom, N., Davis, S. J., Kost, K., Sammon, M., & Viratyosin, T. (2020). The Unprecedented Stock Market Reaction to COVID-19. *The Review of Asset Pricing Studies*, *10*(4), 742–758. https://doi.org/10.1093/RAPSTU/RAAA008

Barber, B. M., Huang, X., Odean, T., & Schwarz, C. (2022). Attention-Induced Trading and Returns: Evidence from Robinhood Users. *The Journal of Finance*, *77*(6), 3141–3190. https://doi.org/10.1111/JOFI.13183

Barber, B. M., Lee, Y. T., Liu, Y. J., & Odean, T. (2009). Just How Much Do Individual Investors Lose by Trading? *The Review of Financial Studies*, *22*(2), 609–632. https://doi.org/10.1093/RFS/HHN046

Barber, B. M., & Odean, T. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies*, *21*(2), 785–818. https://doi.org/10.1093/rfs/hhm079

Barber, B. M., & Odean, T. (2013). The Behavior of Individual Investors. *Handbook of the economics of finance* (pp. 1533–1570). Elsevier. https://doi.org/10.1016/B978-0-44-459406-8.00022-6

Barber, B. M., Odean, T., & Zhu, N. (2009). Do Retail Trades Move Markets? *The Review of Financial Studies*, *22*(1), 151–186. https://doi.org/10.1093/RFS/HHN035

Barrot, J.-N., Kaniel, R., & Sraer, D. (2016). Are retail traders compensated for providing liquidity? *Journal of Financial Economics*, *120*(1), 146–168. https://doi.org/10.1016/j.jfineco.2016.01.005

Black, F. (1986). Noise. *The Journal of Finance*, *41*(3), 529–543. https://doi.org/10.1111/J.1540-6261.1986.TB04513.X

Brandt, M. W., Brav, A., Graham, J. R., & Kumar, A. (2010). The Idiosyncratic Volatility Puzzle: Time Trend or Speculative Episodes? *Review of Financial Studies*, *23*(2), 863–899. https://doi.org/10.1093/rfs/hhp087

British Foreign Policy Group. (2022). COVID-19 Timeline. https://bfpg.co.uk/2020/04/covid-19-timeline/

Brown, G. W. (1999). Volatility, Sentiment, and Noise Traders. *Financial Analysts Journal*, *55*(2), 82–89. https://doi.org/10.2469/FAJ.V55.N2.2263

Cheng, F., Chiao, C., Wang, C., Fang, Z., & Yao, S. (2021). Does retail investor attention improve stock liquidity? A dynamic perspective. *Economic Modelling*, *94*, 170–183. https://doi.org/10.1016/j.econmod.2020.10.001

Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *Journal of Finance*, *66*(5), 1461–1499. https://doi.org/10.1111/J.1540-6261.2011.01679.X

Ding, R., & Hou, W. (2015). Retail investor attention and stock liquidity. *Journal of International Financial Markets, Institutions and Money*, *37*, 12–26. https://doi.org/10.1016/j.intfin.2015.04.001

Eaton, G. W., Green, T. C., Roseman, B. S., & Wu, Y. (2022). Retail trader sophistication and stock market quality: Evidence from brokerage outages. *Journal of Financial Economics*, *146*(2), 502–528. https://doi.org/10.1016/J.JFINECO.2022.08.002

Foucault, T., Sraer, D., & Thesmar, D. J. (2011). Individual Investors and Volatility. *The Journal of Finance*, *66*(4), 1369–1406. https://doi.org/10.1111/j.1540-6261.2011.01668.x

Gavish, L. R., Qadan, M., & Yagil, J. (2021). Net Buyers of Attention-Grabbing Stocks? Who Exactly Are They? *Journal of Behavioral Finance*, *22*(1), 26–45. https://doi.org/10.1080/15427560.2020.1716360

Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, *14*(1), 71–100. https://doi.org/10.1016/0304-405X(85)90044-3

Grinblatt, M., & Keloharju, M. (2009). Sensation seeking, overconfidence, and trading activity. *Journal of Finance*, *64*(2), 549–578. https://doi.org/10.1111/J.1540-6261.2009.01443.X

Kaniel, R., Saar, G., & Titman, S. (2008). Individual investor trading and stock returns. *Journal of Finance*, *63*(1), 273–310. https://doi.org/10.1111/J.1540-6261.2008.01316.X

Kelley, E. K., & Tetlock, P. C. (2013). How Wise Are Crowds? Insights from Retail Orders and Stock Returns. *The Journal of Finance*, *68*(3), 1229–1265. https://doi.org/10.1111/jofi.12028

Li, L., Strahan, P. E., & Zhang, S. (2020). Banks as Lenders of First Resort: Evidence from the COVID-19 Crisis. *The Review of Corporate Finance Studies*, *9*(3), 472–500. https://doi.org/10.1093/rcfs/cfaa009

Ozik, G., Sadka, R., & Shen, S. (2021). Flattening the Illiquidity Curve: Retail Trading During the COVID-19 Lockdown. *Journal of Financial and Quantitative Analysis*, *56*(7), 2356–2388. https://doi.org/10.1017/S0022109021000387

Pagano, M. S., Sedunov, J., & Velthuis, R. (2021). How did retail investors respond to the COVID-19 pandemic? The effect of Robinhood brokerage customers on market quality. *Finance Research Letters*, *43*. https://doi.org/10.1016/j.frl.2021.101946

Peress, J., & Schmidt, D. (2020). Glued to the TV: Distracted Noise Traders and Stock Market Liquidity. *Journal of Finance*, *75*(2), 1083–1133. https://doi.org/10.1111/JOFI.12863

Poon, S.-H., & Granger, C. W. J. (2003). Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature*, *41*(2), 478–539. https://doi.org/10.1257/002205103765762743

Schmeling, M. (2007). Institutional and individual sentiment: Smart money and noise trader risk? *International Journal of Forecasting*, *23*(1), 127–145. https://doi.org/10.1016/J.IJFORECAST.2006.09.002

Seasholes, M. S., & Wu, G. (2007). Predictable behavior, profits, and attention. *Journal of Empirical Finance*, *14*(5), 590–610. https://doi.org/10.1016/j.jempfin.2007.03.002

Shleifer, A., & Summers, L. H. (1990). The Noise Trader Approach to Finance. *Journal of Economic Perspectives*, *4*, 19–33. https://doi.org/10.1257/jep.4.2.19

Smales, L. (2021). Investor attention and global market returns during the COVID-19 crisis. *International Review of Financial Analysis*, *73*, 101616. https://doi.org/10.1016/j.irfa.2020.101616

Trading 212. (2021). Our growth in just 12 months - Trading 212 Community. https://community.trading212.com/t/our-growth-in-just-12-months/46209

van der Beck, P., & Jaunin, C. (2021). The Equity Market Implications of the Retail Investment Boom. *Swiss Finance Institute Research Paper*, *21*(12). https://doi.org/10.2139/SSRN.3776421

Welch, I. (2022). The Wisdom of the Robinhood Crowd. *Journal of Finance*, *77*(3), 1489–1527. https://doi.org/10.1111/JOFI.13128

A Data Collection

A.1 List of FTSE 100 Companies

This section contains the list of the companies that formed the FTSE 100 at the time of the start of data collection.[14] As such, the pricing data set runs from the 01 January 2017 to 01 June 2022.

Table A.1: Stock Symbols of FTSE 100 Constituents

AAF*	BP	HLMA	MRO	SGE
AAL	BRBY	HSBA	NG	SGRO
ABDN	BT-A	HWDN	NWG	SHEL
ABF	CCH	IAG	NXT	SKG
ADM	CPG	ICP	OCDO	SMDS
AHT	CRDA	IHG	PHNX	SMIN
ANTO	CRH	III	PRU	SMT
AUTO	DCC	IMB	**PSH***	SN
AV	DGE	INF	PSN	SPX
AVST*	DPH	ITRK	PSON	SSE
AVV	**EDV***	ITV	REL	STAN
AZN	ENT	JD	RIO	STJ
BA	EXPN	KGF	RKT	SVT
BARC	FLTR	LAND	RMG	TSCO
BATS	FRES	LGEN	RMV	TW
BDEV	GLEN	LLOY	RR	ULVR
BKG	GSK	LSEG	RS1	UU
BLND	HBR	MGGT	RTO	VOD
BME	HIK	MNDI	SBRY	WPP
BNZL	HL	**MNG***	SDR	WTB

*The available data for these stocks does not start from 01 January 2017. As such, observations on these stocks have been dropped and the analysis is conducted excluding these.

[14] This was the 13th June 2022. Hence, the constituent list is based on the constituents as of this date.

A.2 Google SVI Data

In some cases, the Google Trend search term was appended with the word ‘share’ to avoid ambiguity from other related terms that were unrelated to the share in question. According to Google Trends, the phrase ‘stock price’ is search more frequently than ‘share price’ in the U.S.; the opposite is true for the U.K. The average value of the search index for both phrases between 2004 and 2022 can be seen in Figure A.1 below. Therefore, given that my analysis focuses on UK investors, I choose to use the term ‘share’ over ‘stock’ when making these changes to the search term. For example, to avoid confusion with the search term ‘ITV’ for ITV plc., this was combined with ‘share’ to identify those search about the share price. Therefore, the search term for this stock is “ITV share”.

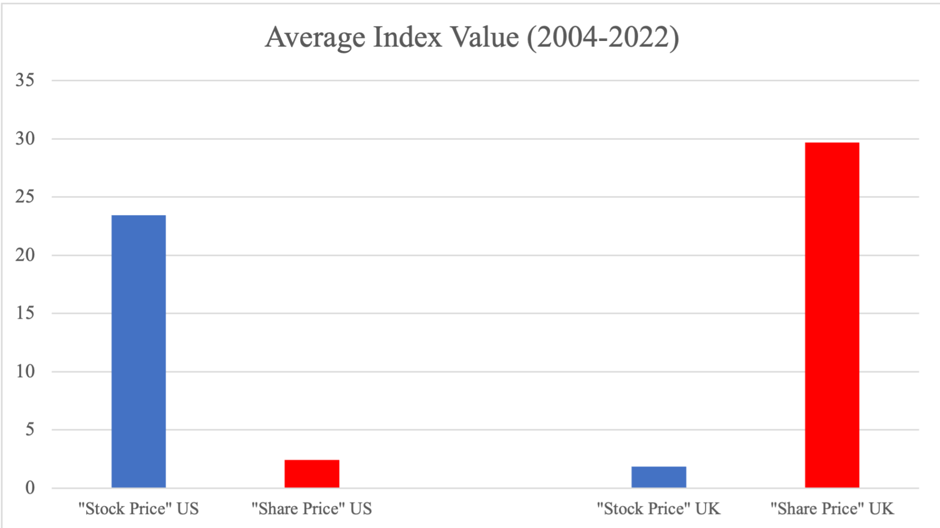

Figure A.1: Google SVI: “Stock Price” vs “Share Price”

To determine which ticker symbols were not sufficiently precise to ensure that the search was related to investment, I compared the related searches for the ticker symbol. Google Trends shows ‘related queries’ and ‘related topics’; if these sections were not related to similar searches for the stock ‘XYZ’ or to financial topics (i.e., ‘XYZ share price’ or searches for other stocks), then the search term was changed to include ‘share’.

A.2.1 List of Noisy Stock Symbols

Table A.2 contains a list of stock ticker symbols that are potentially ambiguous. In the main data set, these search terms are appended by ‘share’. For the robustness checks in the regressions that use this data, see Appendix B.1.

Table A.2: List of Noisy Google SVI Ticker Symbols

AUTO	DCC	KGF	PSN	STAN
BA	ENT	MNG	REL	SVT
BATS	GLEN	MRO	RIO	TW
BME	IHG	NG	RR	UU
BP	III	NWG	RTO	VOD
BT-A	INF	NXT	SHEL	WTB
CCH	ITV	PHNX	SN	
CPG	JD	PRU	SSE	

These search symbols are appended with the word 'share' to avoid ambiguity from using the ticker symbol alone.

There are eight stocks for which the Google SVI data is excluded. This is due to the fact that the ticker symbol only was too ambiguous, but there was missing data on the stock symbol plus 'share'. These stocks are in Table A.3.

Table A.3: List of Excluded Google SVI Searches

DPH	FLTR	HIK	ICP	LAND	RS1	SDR	SKG

A.3 Classification of Stocks by Sector

For the sectoral analysis, only those stocks that make up the Google SVI data are used. Thus, there are 87 stocks that are classified into their respective sectors. The sector of each stock is given in Table A.4.

Table A.4: Sector Classification of Stocks

AAL	Materials	BT	Communication	IMB	Consumer	REL	Communication	TSCO	Consumer
ABDN	Financial	CCH	Consumer	INF	Consumer	RIO	Materials	TW	Real Estate
ABF	Consumer	CPG	Other	ITRK	Other	RKT	Real Estate	ULVR	Consumer
ADM	Financial	CRDA	Materials	ITV	Communication	RMG	Communication	UU	Energy
AHT	Other	CRH	Real Estate	JD	Consumer	RMV	Communication	VOD	Communication
ANTO	Materials	DCC	Other	KGF	Consumer	RR	Industrial	WPP	Communication
AUTO	Communication	DGE	Consumer	LGEN	Financial	RTO	Other	WTB	Consumer
AV	Financial	ENT	Consumer	LLOY	Financial	SBRY	Consumer		
AVV	Other	EXPN	Financial	LSEG	Financial	SGE	Other		
AZN	Other	FRES	Materials	MGGT	Industrial	SGRO	Real Estate		
BA	Industrial	GLEN	Materials	MNDI	Materials	SHEL	Energy		
BARC	Financial	GSK	Other	MRO	Other	SMDS	Industrial		
BATS	Consumer	HBR	Energy	NG	Energy	SMIN	Industrial		
BDEV	Real Estate	HL	Financial	NWG	Financial	SMT	Financial		
BKG	Real Estate	HLMA	Other	NXT	Consumer	SN	Other		
BLND	Real Estate	HSBA	Financial	OCDO	Consumer	SPX	Industrial		
BME	Consumer	HWDN	Other	PHNX	Financial	SSE	Energy		
BNZL	Other	IAG	Consumer	PRU	Financial	STAN	Financial		
BP	Energy	IHG	Consumer	PSN	Real Estate	STJ	Financial		
BRBY	Consumer	III	Financial	PSON	Communication	SVT	Energy		

B Additional Results

B.1 Robustness Checks for Google Trends Data

This section contains the additional regression outputs when excluding the set of potential noisy ticker searches in the Google Trends data. The list of noisy stock symbols is given in Appendix A.2.1.

In this section, I test the null hypothesis that the regression coefficients do not change across subsets of the Google SVI data, namely, whether a stock is 'noisy' or not. I use a dummy variable, $Noisy$, which is equal to 1 if a stock is appended with the term 'share', and 0 otherwise. I then interact the variable of interest with the dummy variable, and use this to create a dummy regression model. For example, the model for $Volatility$ (Model 5) is given by:

$$Volatility_{i,t} = \alpha + \beta_0 Attention_{i,t} + \sum_{j=1}^{2} \beta_j Attention_{i,t-j} +$$
$$\sum_{x=0}^{2} \phi_x \big(Noisy \times Attention_{i,t-x}\big) + \sum_{k=1}^{3} \gamma_k Volatility_{i,t-k} + \delta \cdot \Theta_i$$

Table B.1 below shows the regression output for the interaction terms and the corresponding F-statistic, with the null hypothesis $\phi_0 = \phi_1 = \phi_2 = 0$.[15] I find that only the coefficient for $Noisy \times Attention_{t-2}$ in the $Log(Illiq)$ model is affected by the noisy measures at the 5% level. This affects the p-value of the F-test, making the interaction terms jointly statistically significant at the 5% level. When excluding this term from the test, I find that the interaction terms are jointly insignificant (p-value = 0.2463). This result is likely to be insignificant and not of meaningful interpretation. Thus, I find that, overall, the null hypothesis holds that the results are not affected by the inclusion of the 'noisy' ticker symbols.

[15] Whilst the regressions were run with the entire model (including fixed effects and hetroskedasticity-robust standard errors), only the interaction terms are reported.

Table B.1: Robustness Checks: $Noisy \times$ Variable of Interest

	$Volatility$	$LogRange$	$Spread$	$Log(Illiq)$
$Noisy \times Attention_t$	-0.000008	-0.000001	1.64×10^{-8}	0.000348
	(0.000019)	(0.000015)	(2.41×10^{-7})	(0.000305)
$Noisy \times Attention_{t-1}$	0.000015	0.000011	1.33×10^{-7}	0.000472
	(0.000012)	(0.000012)	(2.01×10^{-7})	(0.000402)
$Noisy \times Attention_{t-2}$	-0.000006	-0.000008	-8.69×10^{-7}	-0.000921**
	(0.00001)	(0.00001)	(1.70×10^{-7})	(0.00019)
N	24,708	24,795	24,621	24,708
F-test	0.2282	0.2386	0.9096	0.0369

The models above use fixed effects and heteroskedasticity-robust standard errors. *: $p - value < 0.1$, **: $p - value < 0.05$, ***: $p - value < 0.01$.
( ) = standard errors.

C Additional Materials

The data sets and stata .do files are available at:

https://www.dropbox.com/sh/eo2ai3g22qpadpm/AAAMEr-V0EeQhbceCc8PSyKFa?dl=0

- `FTSE_PricingData_Full_Clean_Py.dta` is the (cleaned) FTSE 100 pricing data set obtained from the Refinitiv Data platform.
- `Google_Trends-Stock_Symbols_Clean.dta` is the Google SVI (attention) data including the week indicators and the *Noisy* dummy variable.
- `Semrush_web_users.dta` is the website traffic analysis data for the major platforms.
- `FTSE_100_Variables.do` is the Stata commands for the formatting of the pricing variables, i.e., the liquidity and volatility measures.
- `Monthly_ARDL_Models.do` is the Stata commands for averaging at a monthly level and running the monthly ARDL models.
- `Weekly_ARDL_Models.do` is the Stata commands for averaging at a weekly level and running the weekly attention ARDL models.