Not Just Noise

An Empirical Study of Irrational Noise Trading and its Role in Financial Markets

Ehrström, Valter* Sigurdson, Linus†

May 15, 2022

Masters Thesis in Finance

Stockholm School of Economics

Abstract

This paper explores the role of irrational 'noise' traders in financial markets. Theory suggests that a lower share of irrational or uninformed trading in the market should lead to higher adverse selection costs, and that irrational trading should be more susceptible to exogenous, non-economic events that capture traders' time and attention. We examine this empirically by testing for lower than expected trading volumes and higher than expected bid-ask spreads on days when major sporting events occur, using a predictive model by Bender et al. (2013). By mapping certain events to relevant geographies, we find with significance that unexplained volumes increase negatively on such days. Moreover, our findings indicate significantly higher adverse selection costs on such dates at the aggregate level, however we fail to consistently confirm this relationship across subsamples. We recognize the need for further empirical study into this phenomenon using account-level broker information and intra-day financial data.

Key Words: Microstructure theory, Noise trading, Retail trading, Sports, Transaction costs
Thesis Advisor: Alvin Chen, Assistant Professor, Department of Finance, Stockholm School of Economics

*24074@student.hhs.se
†24125@student.hhs.se

The topic of noise trading and its role in financial markets has been debated for several decades. Some scholars have taken the view that noise trading is a wholly irrational activity. Black (1986) expressed it as follows: “people sometimes trade on noise as if it were information. If they expect to make profits from noise trading, they are incorrect.” Later studies have provided some nuance to this view, identifying different types of noise trading, and deeming some of them rational. For example, noise trading might occur for liquidity reasons, like a private individual selling stocks to make a down payment on a house. Similarly, insurance companies or pension funds forced to make fixed payment obligations will have to trade on what appears to be noise, rather than information. This stands in contrast to the completely irrational form of noise trading, whose role is the main topic of this paper. The fundamental question we pose is: does irrational trading play an important role in providing liquidity and efficiency to equity markets?

Retail traders have been identified as a likely source of true, irrational, noise trading. Tending to lose money on average, the link between retail traders and noise traders, and how their presence or absence influence the behavior of informed traders, has been previously explored by Ozik et al. (2021). Our study empirically explores this relationship further by identifying days when the market activity of retail traders (and other irrational market participants) could be expected to be extraordinarily low. Our chosen proxy for such days is those with major sporting events. Sporting events suit this purpose well as they grab a large portion of irrational traders’ attention, while being of no economic importance in the way that e.g. a presidential election would be. For the purposes of this paper, we have singled out soccer, Grand Slam tennis, and golf as the disciplines of choice. These have been chosen on several parameters, such as popularity (soccer is arguably the most watched sport in the world), calendar (we exclude e.g. the Super Bowl and Formula 1 as they occur exclusively on Sundays), and demographics (golf enthusiasts are overrepresented among active US traders). In addition, we have chosen a wide set of countries’ most traded stocks in order to map certain games to its participating nations, assuming for example that

French traders are more likely to abstain from trading when France's national soccer team is playing.

We base our predictive model and general approach on the paper by Bender et al. (2013). They test for unusually high trading volumes following certain technical analysis indicators (which they show to be unprofitable, i.e. noise), as well as the resulting effect on the market's pricing efficiency, reflected in narrower bid-ask spreads. Essentially, we perform the inverse of these tests, checking for low volumes and wide bid-ask spreads, in order to assess the role of noise trading in the market. Our study aims to contribute to the limited existing empirical literature on the role of irrational noise traders by gathering evidence on their presence in the market, and how general market behavior changes in their absence.

After narrowing down our sample to specific sporting events that have the highest likelihood of distracting traders in each particular country, we find that unexplained trading volume in the US and UK is significantly more negative on days with large sporting events. The observed volume effect in the US is larger in the later sample periods, a trend with similar development to the rise of retail traders' share of market volume. Moreover, we record a significant increase in excess bid-ask spreads in conjunction with the selected events, especially in well-known stocks and more recent time periods.

Our findings suggest that the share of irrational noise trading in the market declines during large sporting events. This has a palpable impact on market participants' bottom lines, as it increases the transaction costs arising from adverse selection. Informed traders focused on maximizing profits, such as brokers acting on behalf of corporate insiders or firms buying back shares, would want to ensure a sufficient level of irrational activity in the market before executing their trades. Lastly, we suggest future research to use more detailed data such as account-level information or intra-day bid-ask spreads to account for this effect with higher confidence.

This paper proceeds as follows. Section 1 defines key concepts and explores the theoretical and empirical background from the literature. Section 2 describes our equity data and the rationale behind our chosen sporting events. Section 3 describes our volume and bid-ask spread predictive models, and results are presented in Section 4. Section 5 concludes.

I. Literature Review

Market Dynamics and Efficiency

Although the importance of irrational traders has been studied and discussed extensively on a theoretical level, few have considered their practical implication for the functioning of financial markets. Milgrom and Stokey (1982) highlight the theoretical dysfunction of any financial market solely inhabited by informed, rational traders, which results in little voluntary trading taking place and no ways for investors to profit from private information. Consider a scenario with financial markets in a Pareto-optimal state, and where only fully rational market participants are present. The only motive to trade is if you believe that it would make you better off. However, as all traders are aware of each other's rationality, no one is going to accept your trade. Why? Because they know that the only reason you would want to trade at the current price is to profit at their expense. Given rational expectations, both traders know with certainty that one of them is making a mistake entering the trade. Resultingly, Milgrom and Stokey conclude that, in such situations, traders can never agree to any non-null trade, nor can any trader beneficially trade on private information.

The above conclusions pose a larger question: if investors cannot profit from becoming informed, and acquiring information is expensive, why bother? Grossman and Stiglitz (1980) also reflect on the impossibility of perfectly efficient markets, and illustrate the failure to reach equilibrium in efficient markets where information is costly to obtain. Hence, they argue that markets by nature must consist of both informed and uninformed traders. The uninformed market participants solve the no-trade scenario illustrated above, allowing informed market participants to know that they will not, with certainty, enter a trade in which they are the losing party. Further, Grossman and Stiglitz derive that although the process of arbitrageurs bidding up prices serves as a way of mediating their private information to the uninformed, market prices will never fully reflect all available information.

Defining Noise Trading

Several noise rational expectations models on microstructure theory have reached the same conclusion, that financial markets cannot function in the absence of uninformed traders and equilibrium prices can consequently never reflect all available information. In these models, the term 'noise' is often used interchangeably with non-informational trading, although the proposed source of the noise varies. Many scholars relate noise trading to constraints on liquidity or hedging (Admati, 1985, and Glosten and Milgrom, 1985). As such, all agents will trade with the goal of maximizing a lifetime utility function, i.e. all are behaviorally rational. However, influenced by liquidity constraints or being forced to place hedges, some agents must transact at suboptimal prices and points in time. For the purposes of this study, however, noise trading due to liquidity reasons (so called liquidity traders), are of lesser interest. Intuitively, exogenous distractions will likely not impact a liquidity trader's decision, forced to trade at a specific time due to liquidity constraints, of whether to transact in the market or not.

Other scholars relate noise to fully irrational behavior in the market, i.e. agents whose behavior diverges from the maximization of their utility function for reasons that cannot be rationally explained, as opposed to liquidity constraints. Revisiting Black's (1986) quote on noise: "...people sometimes trade on noise as if it were information. If they expect to make profits from noise trading, they are incorrect", he uses the term 'noise' as a general contrast to information. Nonetheless, Black acknowledges noise's essence to the survival of well-functioning and liquid markets, as previously discussed. As opposed to the view of noise trading being equal to suboptimal trading due to constraints placed on the trader, Black takes a broader view on the origin of noise trading, arguing that some market participants might trade on noise incorrectly assuming it is information. He goes even further, implying that some may even trade simply for the love of the game.

The emergence of online brokerage firms at the offset of the 2000s considerably lowered the threshold for inexperienced investors to access the stock market. Although online traders are presented with almost unlimited information at their fingertips, the illusion of knowledge and control has contrarily led to greater dispersion of beliefs and overconfidence (Barber and Odean, 2001). In line with this added ease of trading and overconfidence among investors, Linnainmaa (2010) finds that retail traders often exhibit suboptimal (and unprofitable) trading behavior. Foucault et al. (2011) make this connection empirically by showing that retail trading activity impacts return volatility positively. Consequently, retail traders are most often perceived as noise traders.

Unlike liquidity traders, the market presence of these irrationally driven market participants can be thought to vary to a much greater extent when influenced by external non-economic factors. Thus, it is mainly these irrational noise traders' (including but not limited to retail traders) market presence that is studied in this paper.

Noise Traders' Impact on Volume and Bid-Ask Spreads

This paper borrows the construction of its empirical model from the paper by Bender et al. (2013), which examines the impact of noise trading on trading volume and bid-ask spreads. They consider a share price pattern typically used by technical analysts to identify a buying opportunity, and test for excess volume and (negative) excess bid-ask spreads in conjunction with these non-economic signals. They succeed in identifying said trading pattern as an unprofitable 'illusory correlation', and use the increase in volume and decrease in spreads to link the pattern to imperfectly rational noise trading. In our paper, we mirror this approach by using their predictive model to compute excess volume and spreads, but applying it to test for the absence, rather than presence, of irrational noise traders. The technical aspects of this predictive model are discussed in Section 3.

In his seminal paper on the topic, Kyle (1985) theorizes that noise trading activity should be inversely related to the price impact of trading, but with no impact on efficiency. Glosten and Milgrom (1985), as well as Admati and Pfleiderer (1988), propose similar models but with the implication that noise trading and efficiency are positively related. This is consistent with Greene and Smart (1999) who show empirically that efficiency increases with a higher prevalence of noise trading.

In this paper, we use bid-ask spreads to study the potential effects on market efficiency in the absence of uninformed noise trading. This builds on the assumption that, all else equal, bid-ask spreads narrow in the presence of uninformed noise trading, as proposed in microstructure theory by Glosten and Milgrom (1985). Lei and Wu (2005) show this empirically by identifying the time-varying probability of informed trading as a good predictor of contemporaneous bid-ask spreads. Consider a case where the probability of informed trading increases. From a market maker's perspective, the risk of entering a trade where the counterparty will profit at your expense is now higher. The market maker will then quote a higher bid-ask spread for taking on this increased adverse selection risk, giving rise to this phenomenon.

The Importance of Noise Traders

So what, then, is the practical importance of noise traders in financial markets? As stated, noise trading is a fundamental and vital part of functioning markets, enabling all market participants to trade with each other. Studying account-level trading data, Collin-Dufresne and Fos (2015, 2016) find that informed traders exhibit higher activity and trade more aggressively at times when market liquidity is high. Their sample of informed traders, corporate insiders, are found to time their trades with points in time when the level of uninformed trading is high. Similar empirical evidence is found by Cookson et al. (2021), who conclude that informed traders respond meaningfully to disagreement in the market.

Gaining a better practical understanding of the role of uninformed noise trading in the markets, as well as the connection to retail trading, would be of great value to practitioners and scholars alike. This paper aims to add to the existing literature by empirically examining how market efficiency is impacted in times where the share of uninformed noise trading decreases. We perform this by studying if the level of irrational trading changes during popular sporting events (chosen, as they convey no economic information), and how they impact market functionality through bid-ask spreads.

II. Data

A Sports Data - Selected Sporting Events

The distinction between informed and uninformed traders is opaque by nature, which is effectively what creates the adverse selection problem faced by traders. Thus, it is difficult to definitively assign specific trades to either category. We base our empirical study on the assumption that the aforementioned irrationally driven noise traders are the only (or at least most likely) market participants to allow themselves to be distracted by non-economic events. Using a broad time window and a large sample of events, we are able to exclude any other external factors that may impact market activity on a given day. This allows us to implicitly study how noise traders' market activity varies on said days.

We select major sporting competitions as our events of choice, having several desirable characteristics for the purpose of this study. First, sports is an entertainment form that is both widely and closely followed across the globe. Furthermore, these are broadly televised events, which increases their reach and consequently the probability of being followed by market participants.

Second, sporting events bear no connection to nor influence on the greater economy, as they convey no economic information to the market. This is highly important, as events like political rallies or elections, policy changes, or other significant political or economic announcements, pose the risk of carrying fundamental information that could cause a rational change in traders' behavior.

Third, unlike other large entertainment events, such as music festivals, which often occur during weekends or eventide, it is easier to compile a sample of sporting events that occur during market hours. This is vital to our study, as we aim to examine if the events result in traders being absent or preoccupied during that day's trading.

Our sports data set comprises major sports events from 2006 to 2022. We elect to focus on the disciplines of soccer, tennis, and golf. Below, we explain these choices on a more granular level. We choose the mentioned time frame, which is skewed towards the present day, as it represents a period of increased penetration of stock market trading among retail consumers. In total, 100 tournaments over the period are sampled, comprising 671 individual games played over 323 unique days.

The particular sports are chosen on, among others, three important criteria. Firstly, popularity. The chosen events need to be popular enough and attract significant enough viewership to have an impact on trading volumes. This rules out any obscure sports. Secondly, the event has to take place on days when markets are open for trading. Consequently, high-viewership events such as the Super Bowl or the Formula 1 Monaco Grand Prix that occur on Sundays do not fit the study. This also becomes a limitation of the events included in our sample, as expanded on for the individual sports below. Lastly, we account for the demographics of traders. We believe that golf allows us to capture a large portion of the US trading community, as the sport skews toward the more affluent (i.e. those who tend to be active investors). Our inclusion of tennis follows a similar logic, and soccer, being the most popular sport in the world, meets this criterion by default.

Some events that do fulfill all three criteria are not included, most notably the Olympic Games. We choose to omit both the Summer and Winter Games since they, despite high viewerships, are difficult to distill down to single events or days. Additionally, the sheer breadth of ongoing events, and the range of the perceived importance of different disciplines in different countries, lie beyond the scope of this paper. On a related note, while we recognize its importance and prominence, we have not included the women's game for soccer and golf given its historically low viewership figures. However, given the palpable rise in viewership for women's soccer tournaments in later years, potential future studies into this effect may wish to include such events.

Our large sample over time and across companies ensures that our results are not impacted by any individual, non-recurring, company-specific or macroeconomic events. However, we look out for significant recurring macroeconomic news which may coincide with our events throughout the sample period. For example, a situation where a major yearly, scheduled Federal Reserve announcement should occur at the same date as the Wimbledon semi-finals, could significantly skew our results. In our view, there are two main types of major recurring macroeconomic news. Firstly, meetings and minutes from the Federal Reserve and European Central Bank. These occur irregularly, and thus do not pose a problem for our study. Secondly, the publication of the two most closely followed macroeconomic indicators, the US non-farm payrolls and the Purchasing Managers' Index from the Institute for Supply Management. The former of the two is published on the first Friday of every month, while the latter is published on the first (manufacturing index) and second (non-manufacturing index) day of each month. Although these indicators are both frequently and regularly recurring, they seldom diverge from analyst expectations and thus most often do not result in any notable market reactions.

Soccer

For the Beautiful game, we include the FIFA men's World Cup and UEFA men's Euro tournaments. Both reach high viewership, the Euros particularly so in our chosen European countries. We include every men's World Cup since 2006, but only include the men's Euro 2016 and Euro 2020 (played in 2021) as previous tournaments' games were all played after European trading hours. World Cup games played after European trading hours are still included as we expect some US viewership, especially in the later stages of the tournaments.

As both of these events are quadrennial, only every other year includes soccer data. Unfortunately for our analysis, the most watched games, namely the finals, occur on Sundays.

Tennis

Our set of tennis matches includes all the four Grand Slam tournaments, i.e. the Australian Open, French Open, Wimbledon, and US Open, from 2006 and onward. These events are held annually, attracting global audiences, with Wimbledon being the most popular of the four. As is the case with soccer, finals are played on the weekend (with the notable exception of the 2012 French Open men's singles final which was postponed and played on the subsequent Monday). We elect to include the women's and men's singles semifinals in our data set, typically played on Thursdays and Fridays, respectively. Worth noting is that the Australian Open generally occurs in January, extending our scope beyond the (Northern Hemisphere) summer months. Additionally, the US Open further motivates the inclusion of US equity data. This is important as the US market has the highest trading volumes in the world, making it apt for this analysis.

Golf

We include the Masters and British Open, two of the most prestigious tournaments in the golfing world, from 2006 to 2021. The Masters is ranked as the sixth most watched non-American football event in the US and is an event that is often closely followed by the trading community. Both of these tournaments take place over the course of four days, from Thursday to Sunday, with one round being played each day. As with many of the tournaments within the other chosen disciplines, the final two rounds of both the Masters and British Open are played during the weekend. Hence, we only include rounds 1 and 2 in our data set.

Summary of Selected Events

For a given event in our sports data set, we record information about the date and time, which tournament it belongs to, and, in the case of soccer and tennis, who is playing. We record the local start time and the location's offset to Coordinated Universal Time (UTC) of each event, enabling us to account for time zone differences and determine if each event occurs within an exchange's trading hours. We call such days *EventDays*. If the relevant country's national team is playing (or certain other criteria, described in detail in Appendix

B, such a day also qualifies as a *GameDay*. Table 1 contains summary statistics for our sports data.

Table 1: Summary statistics - selected sporting events

Discipline	No. of tournaments	No. of games/events	Unique no. of days
Soccer	6	358	145
Tennis	64	253	123
Golf	30	60	60
Total	**100**	**671**	**323**

B Trading Data

Our financial data set consists of daily trading data from 2005 to 2022 for the constituents of the main indices for 9 European stock exchanges, as well as the UK and the US. Hence, the chosen companies are among the largest and most traded companies on each country's exchange. The latter condition is especially important as the presence of noise traders in absolute terms can be perceived to be higher in these more popular equities, making it both more relevant and easier to study changes in noise. Worth noting, however, is that our sample of companies is based on the chosen index constituents as of February 2022. Hence, some of the companies have not been included in the indices during the entirety of the studied time period, with some companies even having been listed after the start of the sample. Also notable is that, in a few instances, a company may be listed on multiple exchanges, and thus duplicated in the data. However, as the aim of this study is to examine the activity of traders in the market, i.e. how actively traders are conducting transactions in financial instruments, having companies with securities listed on several exchanges does not pose a problem, in our view.

In total, the sample comprises 11 indices totalling 446 equities. Daily trading data for all companies, including price, bid-ask quotes, trading volumes and outstanding shares, are gathered through Thomson Reuters' Eikon database. Table 2 provides an overview of our equity data set.

Summary of Equity Data

Collectively, our sampled companies have a market capitalization of nearly €30 trillion, representing close to a third of the value of all publicly traded equities worldwide. As is evident from Table 2, the combined market capitalization differs significantly between countries. For example, our US sample comprises approximately 20 percent of the total number of sampled equities, yet accounts for almost two thirds of the total market value.

Also evident is that, despite all chosen companies belonging to the most traded companies on their respective exchanges, trading activity differs significantly between the different exchanges. Perhaps unsurprisingly, activity (measured as the daily turnover of shares outstanding) is to a large extent skewed towards the US Nasdaq exchange where, on average, the daily turnover of shares across all index constituents is almost 6 percent of the total shares outstanding. Compared to the least traded exchange in our sample, Portugal, daily turnover of US shares is higher by a factor of nearly 30. This higher activity could have multiple possible explanations, including a larger, more established stock exchange and financial system, a higher retail presence, or the included companies themselves being more popular among investors. In addition, the Nasdaq contains many of the tech stocks most widely subjected to media coverage and social media attention. With market activity being exponentially higher in the US than the rest of our sample, this domain is of particular interest for the purpose of this study.

Interestingly, Sweden exhibits the highest daily share turnover among the European exchanges, although still significantly lagging behind our US sample with a turnover of 0.8 percent. The rest of our European and UK exchanges largely show similar trading activity, with Germany slightly above the rest.

Despite our large sample of events and broad time window, the number of *GameDays*, i.e. days when a specific country could be assumed to be watching, is rather low for the majority of our exchanges. This is mainly due to many of the sporting events in question being outside that market's trading hours. In the US and UK, however, our definition of *GameDay* is broader, allowing for a higher sample of *GameDays* in these markets. Definitions of the *EventDay* and *GameDay* variables can be found in the section below.

Table 2: Overview of equity data

Domicile of exchange	**No. of equities**	**Combined market cap. (€bn)**	**Avg. daily trading volume (% of total shares)**	**Avg. no. of *EventDays***	**Avg. no. of *GameDays***
Belgium	18	260	0.3%	120	2
Finland	25	293	0.5%	113	3
France	38	2,239	0.4%	123	2
Germany	40	1,738	0.6%	106	3
Ireland	20	153	0.3%	93	1
Netherlands	29	1,233	0.4%	101	2
Norway	25	272	0.4%	98	2
Portugal	19	70	0.2%	109	3
Sweden	30	894	0.8%	112	2
United Kingdom	99	2,572	0.4%	117	46
United States	101	18,334	5.8%	93	28
Total	**446**	**28,058**	**1.6%**	**106**	**n.a.**

III. Methodology

A Constructing our Testing Variables

In order to perform our statistical tests we construct a number of dummy variables. Firstly, our *EventDay* variable, which assumes a value of one when an event (as outlined in the previous section) occurs within the trading hours of a particular firm's stock exchange. *GameDay* is a more narrowly defined version of *EventDay*, which only includes days when a particular country could be assumed to be watching. This includes when the country's national team is playing a soccer game, as well as the widely popular Djokovic, Federer, and Nadal clashing in Grand Slam tennis. Moreover, days when Wimbledon and the British Open are played have been assigned the label of *GameDay* for the UK market, along with the Masters for the US market. We do not label the entire US Open *GameDay* for the US market, opting instead for all Grand Slam tennis matches played by at least one of the Williams sisters.

This approach allows us to study the effect of sporting events on trading on a more general level, as well as studying country-specific behavior. For a comprehensive overview of *EventDay* and *GameDay*, see Appendix B.

B Volume Regression

We use a predictive model presented by Bender et al. (2013) to estimate the expected trading volume on any given day. Our prediction of each day's trading volume is then compared to the observed volume on that day, giving us a measure of unexplained trading volume (henceforth also called excess, or residual volume). To calculate predicted volume on a given trading day, we run the following regression on every firm in our sample. The regression residual serves as our measure of excess volume at any given date.

$$\ln(TradVol_t) = \alpha + \sum_{i=1}^{50} \beta_i \ln(TradVol_{t-i}) + \sum_{i=1}^{10} \chi_i \ln(High_{t-i}/Low_{t-i}) + \delta t + \Phi \ln(p_{t+10}) + \epsilon_t \tag{1}$$

1. *Trading volume, TradVol.* Our independent variable is the natural logarithm of daily trading volume. We also use fifty lags of the daily trading volume in our regression fitting.

2. *Volatility, High/Low.* Our measure of volatility is calculated by dividing the daily *High* price by the daily *Low* price. This is is consistent with the method suggested by Rogers and Satchell (1991), and accounts for the positively correlated relationship between volume and volatility. We use ten lags of this measure.

3. *Linear trend, t.* We use the date to account for a linear volume trend. Moreover, this helps iron out the difference in trading volume levels across time, as for example firms attract more investors or overall retail stock market participation increases.

4. *Price, p.* Lastly, we use a 10-day advancement of the firm's (logged) closing share price. This measure complements the linear trend by factoring in trends in volume, and is typically only significant when the linear trend is not. Advancing the price ensures that the residual volume is not influenced by price movements ahead of our studied event days. Although this study is not concerned with price movements in the same way as the paper by Bender et al. (2013), we nonetheless mimic this approach in the name of consistency.

We fit this model to our data and record the excess volume, a variable we call *residualVolume*. A change in the excess volume on our sampled event days would thus indicate a change in, according to our predictive model, the unexplained share of trading on these days. In line with our hypothesis of decreased noise due to traders being distracted elsewhere on the selected days, we expect the unexplained (excess) trading volume to be significantly negative for said days. It is important that the chosen model does not already perfectly factor in the effect of non-economic events. If the model picks up the effect in the lagged values of volume, suggesting that trading activity consistently exhibits the same patterns on days leading up to the event, this should cause higher predicted volume on the days studied, and by extension less negative excess volume. Effectively, this means that should we in fact observe negative excess volume on *EventDays* or *GameDays*, we can be confident in our results insofar that the model does not already capture this effect.

C Bid-Ask Spread Regression

In the second part of our study, we perform a similar exercise to examine how market efficiency is impacted on the same days. Similar to the predictive model on trading volume, we run a regression to estimate the firms' bid-ask spreads on a given day (Bender et al., 2013). In the same fashion as excess volumes, we use the residuals from the regression on spreads to study how these behave on our sampled event days. According to microstructure theory, less noise should translate into decreased market efficiency as the share of informed traders increases. Hence, should the share of irrational noise trading drop on our event days, we expect excess spreads to react in the opposite direction, i.e. widen.

$$\ln(Spread_t) = \phi + \sum_{j=1}^{50} \lambda_j \ln(Spread_{t-j}) + \sum_{j=1}^{10} \mu_j \ln(High_{t-j}/Low_{t-j}) + \theta t + \rho \ln(p_{t+10}) + \eta_t \quad (2)$$

The only newly introduced variable in this regression is *Spread*.

1. *Spread*. For each date and firm, we calculate the closing bid-ask spread by taking the difference between closing ask and bid prices relative to the mid price. We use fifty lags of the bid-ask spread as dependent variables. Our measure of the spread is limited by a number of factors, including minimum tick sizes. This creates a bias where firms

with lower prices get overstated measures of the bid-ask spread, further discussed in the Limitations section.

We subsequently end up with a measure of the excess bid-ask spread, *residualSpread* (also 'excess spread'), which we use in the tests described below.

D Statistical Tests

Our main contribution to the literature consists of testing whether excess volumes and bid-ask spreads stand out on the dates when our selected sporting events occur. This aims to provide empirical evidence on current microstructure theory and examines market behavior in the absence of irrational noise trading. Our principal method for performing such tests is a fixed effects panel regression, with equities (firms) as the group variable. This mitigates the risk of our aggregate results being driven by a handful of firms that have more shares outstanding, and thus higher excess volume in absolute terms. We run regressions on our two independent variables, *residualVolume* and *residualSpread*, using both our broader *EventDay* and narrower *GameDay* variables.

Initially, we perform these tests on the entirety of our data set, followed by groupings by country of listing. We do this in part to observe effects on the country level that differ because of variations in average trading volume, market capitalization and the like, but also to single out markets that have higher retail participation in the stock market. As previously mentioned, this is especially interesting in the cases where the total market activity is considerably higher, such as the US market.

We perform a number of additional tests to affirm the robustness of our results. These are expanded on in the following section.

IV. Results

A Impact on Event Days

Below, we present the results from regressing our *EventDay* variable on the residual volume and bid-ask spread obtained from the predictive models outlined in Section 3. This provides a general indication of whether major sporting events have an impact on unexplained trading volume and, if so, whether we can observe any adverse selection costs in the form of higher bid-ask spreads. We assume that only irrational noise traders account for such a drop in unexplained volume, as no rational market participant would trade abnormally in reaction to non-economic events.

Volume

Table 3 shows that using our broad *EventDay* variable, and including all our selected geographies, we observe a negative effect on residual trading volume, significant at the ten percent level. While the effect is rather slight, it does support our hypothesis that trading volume is lower than expected on days when popular sporting events coincide with exchanges' trading hours.

Given our assumption that only irrational noise traders are likely to be distracted by such non-economic factors, the result therefore implies that the level of irrational noise trading decreases on days with major sporting events. To empirically test the resulting impact on adverse selection costs, we run the same test on bid-ask spreads.

Table 3: Regression of *EventDay* on excess trading volume

residualVolume	Coefficient	Std. err	t	p	95% confidence interval	
EventDay	-3.93*10^{-3}	2.05*10^{-3}	-1.91	0.056	-7.95*10^{-3}	9.63*10^{-5}
Constant	1.16*10^{-4}	3.52*10^{-4}	0.33	0.743	-5.74*10^{-4}	8.06*10^{-4}

Fixed effects (within) regression	Number of observations = 1,604,318
Grouped by firm	Number of groups = 446

Bid-Ask Spreads

In Table 4 we break down the results from running the *EventDay* regression on residual bid-ask spreads. This yields a positive coefficient, in line with our hypothesis, significant at the five percent level. This is consistent with current microstructure theory and the general observation that decreases in the presence of noise trading lead to lower market efficiency. The results strengthen the notion that prices are less informed on days when popular sporting events are held. Put in other words, the presence of higher bid-ask spreads suggests higher adverse selection costs driven by the absence of uninformed traders.

All in all, the results using the *EventDay* variable indicate that unexplained volume is indeed more negative on days with major sporting events, leading to higher adverse selection costs. However, due to the modest size of the observed effects, we see the need for increased granularity in the test, as *EventDay* does not take into detailed account which events may entice traders on each stock exchange.

Table 4: Regression of *EventDay* on excess bid-ask spreads

residualSpread	Coefficient	Std. err	t	p	95% confidence interval	
EventDay	$9.59*10^{-5}$	$4.79*10^{-5}$	2.00	0.045	$2.00*10^{-6}$	$1.90*10^{-4}$
Constant	$-2.81*10^{-6}$	$8.20*10^{-6}$	-0.34	0.732	$-1.89*10^{-5}$	$1.33*10^{-5}$
Fixed effects (within) regression			Number of observations = 1,564,413			
Grouped by firm			Number of groups = 446			

B Impact on Game Days

To improve the quality of our results, as well as better adhering to the underlying logic of our analysis, we perform the same tests with our *GameDay* variable. This only includes events that investors in a particular jurisdiction would be likely to watch. For example, one would expect trading activity in France to be more sensitive to the French national soccer team's fixtures than, say, a game between Iceland and Croatia. Similarly, we expect American audiences to be more captivated by the Masters, famously played in Augusta, Georgia, than they would be by the UEFA Euros. With this more realistic measure, we would expect clearer results.

Volume

As seen in Table 5, the effect on volume using the *GameDay* variable is both economically and statistically more significant than when using *EventDay*. The observed (logged) residual volume effect on *GameDays* is about eight times that of *EventDays*. Aside from further supporting our hypothesis of higher negative excess volume on such days, the bigger, more significant effect and narrower definition of *GameDay* add credibility to the notion that the sporting events are in fact what drive the results. We can thus, with some confidence, confirm empirically that the share of irrational noise trading is lower on days when major sporting events occur.

Table 5: Regression of *GameDay* on excess trading volume

residualVolume	Coefficient	Std. err	t	p	95% confidence interval	
GameDay	-3.36*10^{-2}	4.96*10^{-3}	-6.77	0.000	-4.33*10^{-2}	-2.38*10^{-2}
Constant	1.66*10^{-4}	3.48*10^{-4}	0.48	0.634	-5.16*10^{-4}	8.47*10^{-4}

Fixed effects (within) regression
Grouped by firm

Number of observations = 1,604,318
Number of groups = 446

Bid-Ask Spreads

Table 6 summarizes how *GameDay* is correlated with excess bid-ask spreads. We observe a positive effect, which once again suggests that the absence of irrational noise traders leads to higher adverse selection costs. However, we fail to show this with any meaningful degree of significance. To see if more conclusive results can be obtained by factoring in differences in volume, market capitalization, and retail trading activity, we elect to break down the above tests by country of exchange.

Table 6: Regression of *GameDay* on excess bid-ask spreads

residualSpread	Coefficient	Std. err	t	p	95% confidence interval	
GameDay	1.03*10^{-4}	1.15*10^{-4}	0.90	0.370	-1.22*10^{-4}	3.28*10^{-4}
Constant	-5.16*10^{-7}	8.10*10^{-6}	-0.06	0.949	-1.64*10^{-5}	1.54*10^{-5}

Fixed effects (within) regression
Grouped by firm

Number of observations = 1,564,413
Number of groups = 446

C Country-Level Results

To further investigate the relationship between trading volume, bid-ask spreads, and popular sporting events, we run the same *EventDay* and *GameDay* regressions for each of our eleven stock exchanges. This sheds some light on which countries are driving the aggregate results, i.e. where the effect is most prevalent. We expect countries with higher activity, market capitalization, and number of *GameDays* to show the most significant effects. As outlined in the Section 2, the US leads the pack by a wide margin when it comes to market capitalization and average trading activity, and together with the UK has a significantly higher number of *GameDays*. See Table 2 for reference.

Event Days

Looking at the results shown in Table 7, we note that results vary highly between *EventDay* and *GameDay*. The effect of *EventDay* on excess volume differs decidedly by country, from Finland and Norway showing statistically significant negative effects to the Netherlands showing a significantly positive effect. We believe these results run the risk of being somewhat illusory, as the wide set of events do not accurately represent the country-specific viewership, hence why we look at *GameDay* in an effort to eliminate some randomness from the sample. Additionally, given the large US share of the overall sample and its highly significant negative excess volume, we find it fair to say that the Nasdaq firms are driving the aggregate negative effect on excess volume by *EventDay*. We also note that the US has the only significantly positive effect on bid-ask spreads. This relationship is in line with the theory that transaction costs increase when the share of irrational noise traders decreases in the market.

Game Days

Unsurprisingly, as shown in Table 7, the countries with the largest sample sizes and highest trading volumes, namely the UK and the US, give off the most significant outcomes. The highly significant negative excess trading volume exhibited in the US gives us confidence in the conclusion that excess volume does in fact increase negatively on days with major sporting events, seeing as it has the highest average daily market turnover by far. Given that no rational market participant would refrain from trading because of sporting events, the corresponding positive effect on excess bid-ask spreads indicates the adverse selection costs caused by irrational noise traders leaving the market, however we do not observe this with significance when using the *GameDay* variable.

To translate the observed effect on volume and spreads into more digestible terms: The excess bid-ask spread for US firms is roughly 3 basis points on any given *GameDay*. Given the average daily trading volume on our sampled Nasdaq companies, a back-of-the-envelope calculation would suggest roughly \$10 million in increased adverse selection costs on such days. This would mean that over the course of our sample period, large sporting events have led to increased transaction costs in excess of \$300 million - not an inconsequential amount of money.

Furthermore, narrowing down the sample to *GameDays* also eliminates some of the significant results from the *EventDay* regression. We view this as an indication that the *EventDay* results for certain exchanges may not be driven by the effect we aim to isolate in this paper.

We observe a less intuitive effect in Sweden, which notably has a high share of retail participation in the market. The excess volume effect on *GameDays* is significantly positive, with excess bid-ask spreads seemingly unaffected. While we would expect bid-ask spreads to narrow if this increase in volume was driven by an influx of uninformed market participants, we hesitate to draw any conclusions from this result.

Seeing as the effect on bid-ask spreads shows limited significance even when examining results on the country level, we see the need to hone in on specific stocks that attract a higher share of retail traders in an attempt to further isolate the effect in question.

Table 7: Regression of *EventDay* and *GameDay* on excess trading volume and bid-ask spreads, by country

	EventDay		*GameDay*	
Country	Volume	Spread	Volume	Spread
Belgium	-1.20*10^{-2}	3.79*10^{-5}	-9.64*10^{-2}	-4.27*10^{-4}
(t-score)	(-1.28)	(0.49)	(-1.33)	(-0.73)
Finland	-3.27*10^{-2}***	1.61*10^{-5}	3.79*10^{-2}	-1.51*10^{-5}
	(-4.42)	(0.54)	(0.84)	(-0.08)
France	1.36*10^{-2}***	-1.09*10^{-6}	1.13*10^{-2}	9.66*10^{-7}
	(2.89)	(-0.07)	(0.31)	(0.01)
Germany	8.22*10^{-3}	7.87*10^{-6}	5.20*10^{-2}	5.60*10^{-6}
	(1.62)	(0.42)	(1.67)	(0.05)
Ireland	2.54*10^{-2}	-7.32*10^{-5}	1.54*10^{-1}	-2.57*10^{-3}
	(1.13)	(-0.14)	(0.90)	(-0.65)
Netherlands	1.57*10^{-2}**	2.66*10^{-5}	-5.03*10^{-2}	-6.08*10^{-4}
	(2.19)	(0.41)	(-1.08)	(-1.53)
Norway	-2.90*10^{-2}***	7.86*10^{-5}	1.37*10^{-1}*	-1.58*10^{-3}
	(-2.72)	(0.76)	(1.69)	(-2.02)
Portugal	-2.77*10^{-2}*	1.85*10^{-4}	3.61*10^{-2}	1.05*10^{-3}
	(-1.75)	(1.29)	(0.36)	(1.20)
Sweden	1.15*10^{-4}	1.64*10^{-5}	9.93*10^{-2}**	-6.55*10^{-5}
	(0.02)	(0.86)	(2.23)	(-0.46)
United Kingdom	6.44*10^{-3}	2.97*10^{-5}	-3.92*10^{-2}***	2.34*10^{-5}
	(1.50)	(0.77)	(-5.79)	(0.38)
United States	-2.20*10^{-2}***	3.65*10^{-4}*	-3.84*10^{-2}***	3.09*10^{-4}
	(-6.59)	(1.83)	(-6.36)	(0.85)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, values show the coefficient of the dependent variable

D Impact on Well-Known Stocks

As especially highlighted by the recent retail trading frenzy in stocks such as AMC Entertainment and GameStop, some stocks arguably have a higher following of irrational traders. Barber and Odean (2008) find that individual (i.e. retail) investors have a higher likelihood of trading in attention-grabbing stocks, such as stocks recently mentioned in the media. In other words, retail traders have a high probability of buying shares in companies that have caught their attention in various ways, as opposed to trading based on e.g. fundamental analysis. Similarly, Ozik et al. (2021) emphasize that retail traders' presence is often largely skewed towards high-attention stocks.

To examine whether our studied effect is larger in this certain type of stocks, we develop a new classification of our sample companies. We develop three classifications: *WidelyKnown*, *BestKnown*, and *TechStocks*. We then conduct an evaluation of each of our 446 sample companies individually, deciding whether or not they fit into each respective category. Evaluation is based solely on the companies' operational characteristics and public perception, as opposed to financial or trading characteristics. Each classification is defined more thoroughly below and in Table 8.

WidelyKnown. In our broadest category, we include all recognizable brands and company names. This mainly pertains to consumer-facing, B2C firms, such as consumer goods and services, banks, domestically prominent industrial companies, well-known technology firms, as well as companies that are often discussed in business media. In total, 101 companies are included in the *WidelyKnown* category, i.e. roughly one fourth of our entire company sample. Notably, the selected companies account for about 60 percent of the total market value of our sample.

BestKnown. We narrow down the previously defined *WidelyKnown*, only including the most known consumer goods and services firms and technology companies from the previous category. Companies such as banks and telecommunications operators, previously included in the *WidelyKnown* category, are disregarded in this subsample. The new category includes, in the authors' view, the top 10 percent of our company sample in terms of conspicuousness. As is evident from Table 8, our measure of daily trading rises significantly for the new category, with average turnover reaching almost 13 percent across the *BestKnown* companies. This can be related to approximately 6 percent for the US as a whole and 2 percent for our entire sample, and supports our view that these well-known companies are actively traded to a much larger extent.

TechStocks. We also include a category for the most recognized technology firms in our sample, as high-profile tech stocks such as Apple, Google, and Tesla have been getting large attention from both traders and financial news outlets in recent years. In total, we include 39 technology companies in this category. One should note that the majority of these companies are US based, and hence the category is largely skewed towards our US sample. Like the *BestKnown* category and in line with our expectations, the daily trading in our technology firms far outweighs the total average, as can be seen in Table 8.

Table 8: Summary of stock classifications based on salience

	No. of equities	**Combined market cap. (€bn)**	**Avg. daily trading volume (% of total shares)**
WidelyKnown	101	16,870	5.0%
BestKnown	40	13,664	12.8%
TechStocks	39	12,810	14.1%

Consistent with the results from the larger UK and US markets, both our *WidelyKnown* and *BestKnown* categories exhibit negative excess volumes on *GameDays*, significant at the 10 percent level. When narrowing down our sample to arguably the most known stocks, the observed, decreased activity of irrational noise also translates into decreased market efficiency, with a positive effect on excess spreads significant at the ten percent level. We can, however, not identify a similar relationship for the *WidelyKnown* and *TechStocks* categories. Interestingly, our most actively traded subsample, *TechStocks*, does not exhibit a significant effect on excess volumes on *GameDays*, although the coefficient remains negative. First, this could point to brand recognition being more important to irrational noise traders than sector belonging. Second, it could indicate that the general level of irrational noise in technology companies is consistently lower than for our sample as a whole. Third, the inability of finding a relationship between decreasing noise and market inefficiency makes our previous findings on the empirical relationship less conclusive. We further expand on possible reasons for this in the Limitations section.

Table 9: Regression of *GameDay* on excess trading volume and bid-ask spreads, by company category

	Excess volume	Excess spread
WidelyKnown	-1.87*10^{-2}*	1.20*10^{-4}
(t-score)	(-1.94)	(0.52)
BestKnown	-2.06*10^{-2}*	1.66*10^{-4}*
	(-1.72)	(1.82)
TechStocks	-2.49*10^{-2}	4.22*10^{-5}
	(-1.61)	(0.25)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, values show the coefficient of the dependent variable *GameDay*

E Robustness Checks

We aim to ensure that our conclusion regarding lower excess trading volume on days with major sporting events is in fact driven by these events. Thus, we run two complementary robustness checks.

Placebo Test

We perform a series of placebo tests, where we assign random dates the status of event days to track whether our observed effects can be unambiguously attributed to the occurrence of major sporting events. Five hundred iterations are performed, where each iteration randomly assigns *GameDay* status to approximately ten days out of the total sample period. We then proceed to perform regressions on residual volume and residual bid-ask spreads using these *Placebo* days. A summary of the regression results can be found in Appendix D.

As one would expect from a random trial on residuals, we observe a more or less equal split between positive and negative coefficients on both residual volume and bid-ask spreads. However, the volume regression results are significant far more often than should be the case for randomly assigned dates. Running smaller-scale simulations on the country level, the same effect still seems prevalent, however to a lesser degree than on the aggregate level. We see two reasons for this anomaly. First, it could be that our construction of the placebo test is flawed in some capacity. Second, we cannot ignore the possibility that the data from our *residualVolume* and *residualSpread* predictive regressions are in some way impaired.

While this does not rule out our effects being driven by something other than sporting events, it also does not disprove our hypothesis. We thus reiterate the need for further research into trading activity on *GameDays*.

Time Periods

To further solidify our results, we perform our *EventDay* regression on residual volume using different subsamples of time. We divide our sample into four roughly equal time periods: 2005-2009, 2010-2013, 2014-2017, and 2018-2022. As can be seen in Table 10, we observe that the two latter periods show larger economic and statistical significance, the most recent period in particular. This is in line with our expectations, and underpins the argument that the negative excess volumes are driven by the absence of (irrational) retail traders, as the proportion of retail traders in the market has increased in these periods. Figure 1 shows the development of retail market participation in the US from 2010 to 2021. While certainly not the only source of irrational noise trading, a further rise in the share of retail participation should, if our results are any indication, exacerbate the adverse selection costs discussed in this paper on days with large non-economic events. Consequently, this phenomenon should become an increasingly important aspect for market participants to consider going forward. Note that we only include US data in this test, as it is the subsample in which we see the largest and most significant effect on excess trading volume, and that we use *EventDay* due to lack of sufficient *GameDays* in certain timeframes.

Figure 1: Retail market share of overall US equity trading volumes

Source: *Financial Times (Martin and Wigglesworth, 2021)*

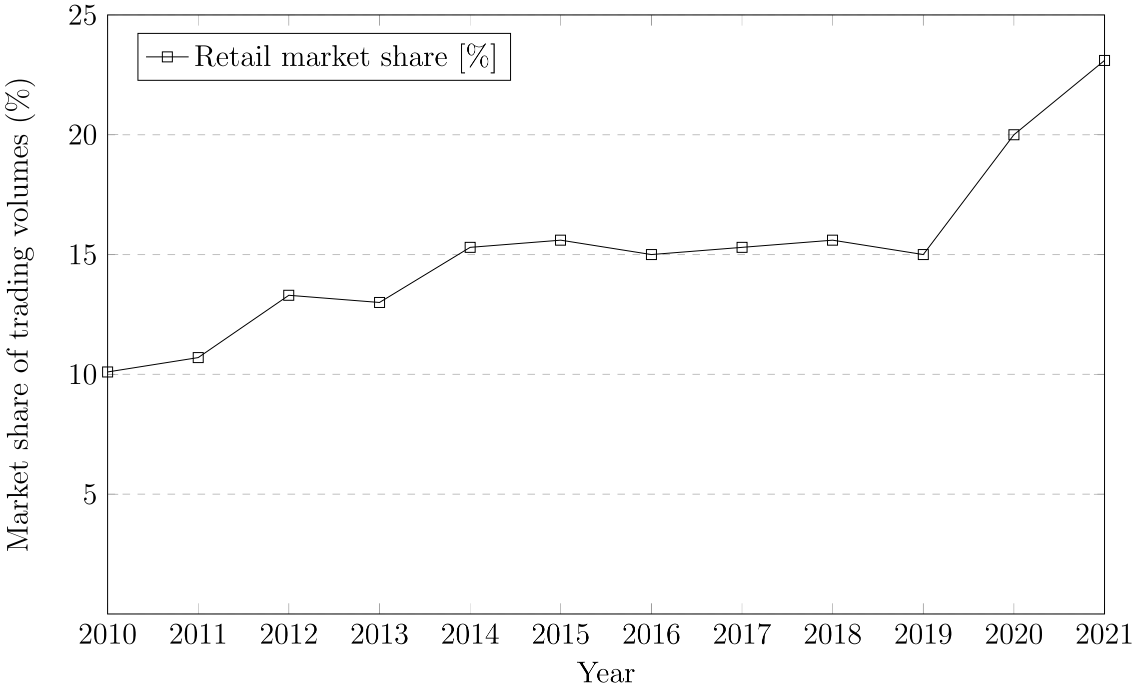

Table 10: Regression of *EventDay* on excess trading volume and bid-ask spreads, by time period (US data)

	Excess volume	Excess spread
2005-2009	$-5.08*10^{-3}$	$1.24*10^{-3}$
(t-score)	(-0.57)	(1.16)
2010-2013	$1.13*10^{-2}$	$-1.42*10^{-4}$
	(1.24)	(-0.99)
2014-2017	$-1.14*10^{-2}$*	$4.91*10^{-4}$***
	(-1.99)	(3.58)
2018-2022	$-5.32*10^{-2}$***	$6.34*10^{-5}$**
	(-10.45)	(1.52)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, values show the coefficient of the dependent variable *EventDay*

F Limitations

A central difficulty in studying market efficiency is its two-dimensional nature. Whereas our study focuses on bid-ask spreads, market efficiency is in practice also influenced by a second, less palpable component, namely depth. This component refers to the volume offered at each bid and ask quote. Unlike spreads, which are recorded and quoted continuously by exchanges, depth may only exist in the minds of market makers and is thus inherently challenging to study. In theory, a drop in market efficiency could be caused entirely by a shallower market (with spreads remaining at the same level), a scenario which we are unable to capture in this study.

As previously stated, bid-ask spreads are limited by minimum tick sizes, which to an extent restricts market participants' ability to adjust their bid-ask quotes and could, at times, propel participants towards reacting through depth. Likewise, our ability to study the impact on bid-ask spreads is limited by only having access to closing spread quotes, i.e. the bid and ask prices at the end of each trading session. While trading volume is cumulative, meaning the absence of some market participants during a part of the session will be implicitly recorded in the session's total trading volume, spreads only reflect a specific point in time. Hence, in the case that an event has ended earlier in the day and market participants have had time to return to trading, their absence earlier in the trading session might not be observable in our data set.

As the sampled events used in this study on average span the course of two hours, ideally one would limit the event window to only reflect this period. This, however, requires the use of intra-day trading data on volumes and bid-ask quotes at the hourly or even more granular level, which falls outside the scope of this study. We suggest further research to leverage intra-day data, or perhaps even account-level trading data from online brokers, as in Cookson et al. (2021) and Berkman & Koch (2008).

Another limitation lies inherent in our chosen non-economic events, i.e. the major sporting competitions. Being a popular form of entertainment, the bulk of high-profile games and events occur during prime time, on evenings and weekends. This limits the number of days that can reasonably be classified as *GameDays*. While we manage to, in some capacity, circumvent this issue through our choice of events, one might be able to identify an even more appropriate non-economic event on which to perform a similar study. We also refrain from expanding our data set of sporting events, as less known and followed events run a higher risk of not being watched by participants in financial markets. Such inclusion would in fact risk diluting the effect we are trying to study.

Lastly, we acknowledge the possibility of error stemming from our choice of predictive model. It is inevitably so that another model may yield different results, to say nothing of its degree of accuracy.

V. Conclusion

We attempt to find and analyze irrational noise trading in equity markets. Using a predictive model of volume and bid-ask spreads constructed by Bender et al. (2013), we empirically test for changes in unexplained volume and spreads on days with major sporting events. Given that no fundamental economic change affects markets on these days, we assume that any change in unexplained volume is attributable to irrational noise traders.

Our results show that the observed trading volume in the US and UK is significantly lower than expected on days with major sporting events. We hypothesize this to be a result of irrational traders' attention being drawn away from the market. In line with microstructure theory, we observe an indication that bid-ask spreads widen in conjunction with the decreased share of irrational trading, implying increased adverse selection costs. In the US, where we see the strongest, most significant effect, we show that this effect has increased over time. This would also suggest that the larger (negative) unexplained volume is partially driven by retail traders, as their share of market participation has increased greatly over the past years.

Given our results, we suggest that a comprehensive predictive model should account for the expected share of uninformed trading on a given day. The adverse selection costs that arise when uninformed traders leave the market have real economic impact and could, over time, hurt market participants' bottom lines. An informed trader would consequently want to time their trades with periods that exhibit sufficient irrational activity, and shy away from trading during events like those considered in this paper. This echoes the observations made by Collin-Dufresne and Fos (2015, 2016), that corporate insiders use liquidity provided by the presence of uninformed retail trading to their advantage. The same principle would be applicable when brokers execute share buybacks on behalf of firms.

Additionally, given the indications from our placebo test, we cannot rule out our results being driven by other, unknown factors. To investigate this further, we suggest future research to widen the scope to more granular trading data, and look beyond the closing bid-ask spread to e.g. intra-day spreads or market depth. Using account-level data from online brokers as in Cookson et al. (2021) could also yield more conclusive results as it allows for the isolation of retail traders and their specific impact on markets.

In conclusion, our findings contribute to the existing work on noise trading and microstructure theory by showing empirically that excess market volume, and consequently irrational noise trading, decreases on days with exogenous non-economic distractions. We can also, to some degree of significance, confirm the positive impact on adverse selection costs suggested by the literature. However, given our limitations on data, we suggest further empirical research into irrational noise trading. Understanding its role and impact on markets is in the interest of any market participant seeking to make a profit. Irrational market activity could result in real economic consequences, and not just noise.

References

Admati, A. R. (1985), 'A noisy rational expectations equilibrium for multi-asset securities markets', *Econometrica, 53(3)* pp. 629–657.

Admati, A. R. & Pfleiderer, P. (1988), 'A theory of intraday patterns: Volume and price variability', *The Review of Financial Studies, 1(1)* pp. 3–40.

Barber, B. M. & Odean, T. (2001), 'The internet and the investor', *The Journal of Economic Perspectives, 15(1)* pp. 41–54.

Barber, B. M. & Odean, T. (2008), 'All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors', *The Review of Financial Studies, 21(2)* pp. 785–818.

Bender, J. C., Osler, C. L. & Simon, D. (2013), 'Noise trading and illusory correlations in US equity markets', *Review of Finance, 17(2)* pp. 625–652.

Berkman, H. & Koch, P. D. (2008), 'Noise trading and the price formation process', *Journal of Empirical Finance, 15(2)* pp. 232–250.

Black, F. (1986), 'Noise', *The Journal of Finance, 41(3)* pp. 529–543.

Collin-Dufresne, P. & Fos, V. (2015), 'Do prices reveal the presence of informed trading?', *The Journal of Finance, 70(4)* pp. 1555–1582.

Collin-Dufresne, P. & Fos, V. (2016), 'Insider trading, stochastic liquidity, and equilibrium prices', *Econometrica, 84(4)* pp. 1441–1475.

Cookson, J. A., Fos, V. & Niessner, M. (2021), 'Does disagreement facilitate informed trading? evidence from activist investors', *Working Paper, University of Colorado* .

Foucault, T., Sraer, D. & Thesmar, D. J. (2011), 'Individual investors and volatility', *The Journal of Finance, 66(4)* pp. 1369–1406.

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)* pp. 71–100.

Grossman, S. J. & Stiglitz, J. E. (1980), 'On the impossibility of informationally efficient markets', *The American Economic Review, 70(3)* pp. 393–408.

Kyle, A. S. (1985), 'Continuous auctions and insider trading', *Econometrica, 53(6)* pp. 1315–1335.

Lei, Q. & Wu, G. (2005), 'Time-varying informed and uninformed trading activities', *Journal of Financial Markets, 8(2)* pp. 153–181.

Linnainmaa, J. T. (2010), 'Do limit orders alter inferences about investor performance and behavior?', *The Journal of Finance, 65(4)* pp. 1473–1506.

Martin, K. & Wigglesworth, R. (2021), 'Rise of the retail army: The amateur traders transforming markets', *Financial Times* . Accessed on 08/05/2022.
URL: *https://www.ft.com/content/7a91e3ea-b9ec-4611-9a03-a8dd3b8bddb5*

Milgrom, P. & Stokey, N. (1982), 'Information, trade and common knowledge', *Journal of Economic Theory, 26(1)* pp. 17–27.

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)* pp. 2356–2388.

Thomson Reuters Eikon Database (2022), 'Daily trading data'. Data retrieved on 14/02/2022.

Appendices

A Data Construction

Table 11: Variable definition and construction. Input goes into our predictive models.

Variable	**Description**	**Source**	**Construction**
Bid	Closing bid quote	Eikon	
Ask	Closing bid quote	Eikon	
High	Highest transaction price	Eikon	
Low	Lowest transaction price	Eikon	
Price	Closing transaction price	Eikon	
Spread	Relative bid-ask spread	-	(*Ask* - *Bid*) / *Price*
TradVol	Total trading volume	Eikon	
Volatility	Measure of volatility	-	*High* / *Low*

All data are recorded at the daily level

B Description of Variables

Table 12: Description of variables

Variable	**Description**
EventDay	Dummy variable with value 1 on dates when a sports game is played and the country's stock exchange is open at the time of playing, and 0 otherwise.
GameDay	Dummy variable with value 1 on dates when the country's stock exchange is open at the time of playing and; a) the country's national team is playing (For the UK, this includes England, Northern Ireland, Scotland, and Wales), b) Novak Djokovic, Roger Federer, or Rafael Nadal meet each other in a tennis match, c) For the UK, Wimbledon or the British Open is being played, or d) For the US, the Masters are being played, or Serena or Venus Williams is playing, and 0 otherwise.
High	Highest transacted price for each date and firm.
Low	Lowest transacted price for each date and firm.
Placebo	Dummy variable with value 1 for a set of randomly selected dates (roughly 10 days), and 0 otherwise.
Price	Closing stock price for each date and firm.
residualSpread	Observed bid-ask spread less value obtained in the regression fitting.
residualVolume	Observed trading volume less value obtained in the regression fitting.
Spread	Relative bid-ask spread for each date and firm. The relative bid-ask spread is defined as the difference between (closing) ask and bid quotes, divided by the closing price (*Price*).
TradVol	Trading volume for each date and firm.
Volatility	Measure of volatility, defined as the highest price (*High*) divided by the lowest price (*Low*) for each date and firm.

C Summary Statistics

Table 13: Summary of input data

Variable	No. of obs.	Mean	Min	Max
Ask	1,631,077	371.17	0	301,000
Bid	1,631,074	370.24	0	67,290
DATE			02/08/2005	12/02/2022
Price	1,631,998	370.35	0	17,135
TradVol	1,632,126	6,466,791	1	$3.37*10^9$
Volatility (log)	1,631,986	0.02	-0.06	2.76

All data are recorded at the daily level

D Placebo Test Results

Table 14: Summary of regressions of *Placebo* on excess trading volume and bid-ask spreads

	Volume	Spread
No. of iterations	500	500
No. of positive coefficients	263	190
Of which significant*	211	32
No. of negative coefficients	237	310
Of which significant*	190	3

* at a 10% significance level ($p < 0.10$)