Noise Traders Incarnate:
Describing a Realistic Noise Trading Process

*Joel Peress & Daniel Schmidt**

November 6, 2017

ABSTRACT

We estimate a realistic process for noise trading to help theorists calibrate noisy rational expectations models. For this purpose, we characterize the trades initiated by individual investors, who are natural candidates for the role of noise traders because their trades are, on average, cross-correlated and loss making. We use transactions data from a retail brokerage house, small TAQ trades, and flows to retail mutual funds, obtaining consistent results. We find that noise trading can be treated as approximately i.i.d. at monthly and lower frequencies but that weekly and daily trades are serially correlated; the distribution of noise trading is less heavy-tailed at lower frequency but conforms to a normal only for quarterly data. We provide a complete description of these processes, including estimates of their standard deviation. In line with theory, the estimates are higher for more liquid and volatile stocks; they also suggest that the prevalence of noise trading has declined over time.

---

* Joel Peress is at INSEAD, Boulevard de Constance, 77300 Fontainebleau, France; e-mail: joel.peress@insead.edu. Daniel Schmidt is at HEC Paris, 1 rue de la Libération, 78350 Jouy-en-Josas, France; e-mail: schmidt@hec.fr. For helpful comments, we thank Giovanni Cespa, Bernard Dumas, Thierry Foucault, Tim Johnson, and Bart Yueshen as well as seminar participants at INSEAD and at the 2016 AFA meeting (San Francisco). We are grateful to Terry Odean for sharing the large discount brokerage data. Joel Peress thanks the AXA Research Fund and the Institut Europlace de Finance for their financial support; he also thanks the Marshall School of Business at the University of Southern California for its hospitality while some of this research was developed.

Since its inception three decades ago, the noisy rational expectations equilibrium (NREE) paradigm has led to myriad models of trading under asymmetric information.[1] "Noise" or "liquidity" trading is an essential ingredient of these models. Without it, asset prices would perfectly reveal traders' private information, thereby undermining the incentive to collect costly information in the first place (the Grossman–Stiglitz paradox). To avoid this paradox, NREE models typically hypothesize an exogenous noise process for the residual stock supply available to speculators. Although several insights do not depend on how this process is specified, many others—both quantitative and qualitative—crucially do (see the examples to follow). Yet little is known about the *empirical properties of a realistic noise process*, so theorists are mostly in the dark regarding its broad features and how best to calibrate their models.

In this paper, we document the properties of a realistic noise trading process. Noise trading comes in many guises and is invoked in several literatures; our focus is specifically on NREE models. In line with this literature, we define a noise trade as any trade that is unrelated (i.e., orthogonal) to fundamental information. We then estimate a process for noise trading from retail trading data under the identifying assumption that retail trades are representative of noise trades. The literature provides ample evidence in support of this assumption (e.g., the poor performance of retail trades) and has examined some of its implications (e.g., the literature on measuring investor sentiment).

To appreciate the importance of our task, consider three aspects of noise trading.[2] The first is its *persistence*. There are at least three reasons why such persistence plays a central part in NREE models. First, it determines the degree to which arbitrageurs are willing to correct any mispricing, which in turn determines the informativeness of asset prices (e.g., Stein 1987, Grundy and McNichols 1989, He and Wang 1995, Cespa and Vives 2012). Second, persistence controls the serial correlations of stock returns and of trading volume (e.g., Wang 1993, Makarov and Rytchkov 2012). Finally, it affects the measurement of liquidity in financial markets.

[1] Grossman (1976), Grossman and Stiglitz (1980), and Hellwig (1980) laid the foundations for noisy rational expectations models in competitive markets; Kyle (1985) offered the seminal analysis of strategic markets. According to Google Scholar, Grossman and Stiglitz (1980) and Kyle (1985) have together been cited more than 17,000 times.

[2] We elaborate on all these aspects in Section 1.

Indeed, standard measures of stock price liquidity fail to detect the presence of adverse selection when noise trading is predictable, such as when it is persistent (Collin-Dufresne and Fos 2014a, Kacperczyk and Pagnotta 2016).

The second set of features underlying the importance of noise trading are its *distribution* and its *correlation with fundamentals*. Standard models assume that noise trades are both normally distributed and uncorrelated with the asset's fundamental value. Yet, recent theoretical work suggests that neither assumption is innocuous. In fact, both are required to rule out strategic complementarities in information acquisition and hence the possibility of multiple equilibria; see Barlevy and Veronesi (2000, 2008) and Breon-Drish (2010, 2014).

Finally, the third crucial aspect of noise trading is its *intensity* (i.e., its standard deviation). Although the NREE paradigm was originally developed to deliver qualitative insights, many authors have used such models to make quantitative predictions and to estimate orders of magnitude of the phenomena being studied. Yet, because noise trading is not directly observable, these scholars usually either pick an arbitrary value for its variance (e.g., Campbell and Kyle 1993, Watanabe 2008, Biais et al. 2010) or choose a value such that the model's predicted moments match sample moments estimated from market data—thereby forgoing a testable restriction (e.g., Campbell et al. 1993, Peress 2003, Banerjee 2011).

By estimating a realistic noise process, we enable researchers to evaluate the plausibility of their assumptions and to test additional restrictions on the data. To accomplish this, we analyze purchases and sales of equity by retail investors. These comprise trades directly executed by retail investors, as well as flows in and out of equity mutual funds, which subsequently trigger trades. Retail investors are natural candidates for the role of noise traders because previous research has documented that they perform poorly on average, even before transactions costs,[3] and that they trade "in concert" (Kumar and Lee 2006, Barber et al. 2009b). Thus their

[3] We do not argue that all retail investors lose from trading, only that they do so on average. Some retail investors may be skilled. For evidence on stock trades, see Odean (1999), Barber and Odean (2000), and Grinblatt and Keloharju (2000); for evidence on investments in mutual funds, see Frazzini and Lamont (2008), Ben-Rephael et al. (2012), and Akbas et al. (2015). We review this evidence in detail in Section 2. Also, note that our analysis of mutual fund flows is not inconsistent with fund managers picking stocks, possibly with skill: flows force them to

trades contain a common systematic component that is unrelated to the future value of the asset and that, far from washing out in the aggregate, can actually "blur" the price signal. As a matter of fact, a large body of empirical research views retail investors as the archetypal noise traders (e.g., the literature on measuring investor sentiment, and Stambaugh 2014).

Our analyses are based on three complementary data sources. The first source is retail trading data from a large discount brokerage house. The second is a set of small trades from the Trade and Quote (TAQ) database, which prior to decimalization in 2001 were most likely to have been made by retail investors (Hvidkjaer 2008, Barber et al. 2009a). The third source is data on net flows to mutual funds invested in US domestic equity and sold to retail investors. Each of these data sets has its pros and cons: the retail brokerage data allow us to follow individual traders but represent only a fraction of all retail trading; TAQ small trades are more comprehensive but do not reveal traders' identities. These two data sets cover trades directly executed by retail investors, the importance of which has been declining over time. Mutual funds currently account for the bulk of retail activity, but that data cannot be directly related to trades in individual stocks. The data sets also complement each other by covering different periods: 1991–1996, 1991–2000, and 1999–2013 for (respectively) the retail brokerage, TAQ, and mutual fund data. We confirm that the trades and flows in our samples conform to the NREE literature's definition of what constitute noise trades; in particular, they are unprofitable on average and cross-correlated (as reported in previous studies).

We acknowledge that not all retail trades are noise trades; that is, some are actually informed trades. Conversely, not all noise trades are retail trades because some institutions also trade on noise. Yet the aforementioned evidence suggests that retail investors, notwithstanding their heterogeneity, do behave *on average* as noise traders. Our results are valid to the extent that the error in our measure of noise trades—which is equal to institutional noise trades minus informed retail trades—either resembles retail noise trades (e.g., is equally persistent) or makes up only a small fraction of total noise trades.[4] We acknowledge that the

---

trade in and out of the stock market as a whole, but they have discretion regarding which stocks to trade. For this reason, we restrict our use of mutual fund data to the estimation of a noise trading process for the entire market.

[4] In Appendix B, we derive formal conditions for this approach to be valid.

validity of our identifying assumption cannot be established directly, short of asking investors why they traded. Even so, we view our work as a reasonable first attempt to measure and calibrate a realistic noise trading process for NREE models.

We serve theorists by providing an accurate description of the noise trading process within the canonical framework they employ. For tractability, most models assume that investors are risk neutral or exhibit constant absolute risk aversion (CARA), so that their demand—given as a number or turnover of shares (after dividing by the number of shares outstanding)—is linear in random variables, including prices. Hence these models assume that aggregate noise trader demand is also measured in number or turnover of shares, either because each noise trader trades a random number of shares or because noise traders randomly participate in the market. We accordingly analyze the turnover of shares traded by retail investors, which we define as the aggregate value of their trades divided by the total value of the market in the brokerage and TAQ data sets or, in the case of mutual funds, as the aggregate value of fund flows divided by funds' total assets. In addition, we use the brokerage data to study the fraction of households that trade. All variables are net in the sense that they measure the difference between buys and sells: the buy turnover minus the sell turnover for shares traded in the brokerage and TAQ data sets; the purchase turnover minus the redemption turnover for mutual funds flows; and the fraction of households buying minus the fraction of households selling in the brokerage data set.[5]

NREE models typically assume that the distribution of noise trades is normal, and either independent and identically distributed (i.i.d.) or with an autoregressive component. We therefore seek to fit a parsimonious autoregressive process to households' aggregate trades. For all three data sets we find that trades can be considered i.i.d. at monthly and lower frequencies (quarterly). In contrast, daily and weekly trades are serially

[5] These data exhibit seasonal patterns. In line with prior studies, we find that net buys are lower in December, which is consistent with households realizing losses for tax purposes, and also over the summer (when households are on vacation); see Badrinath and Lewellen (1991) and Hong and Yu (2008). Our own analysis is performed after purging the data of such calendar effects.

correlated: the latter require one to six lags and the former at least five lags.[6] Focusing on the first-order autoregressive, or AR(1), processes commonly postulated by theorists, we find—across these data sets—that the first-order autocorrelation coefficient declines as the duration of time periods increases (as conjectured by He and Wang 1995, Cespa and Vives 2012 in their theoretical work). More specifically, our results indicate that the coefficient drops by 0.5%–1% for each additional trading day.

Turning now to the parametric form of noise trades, we find that they cannot, in general, be treated as normally distributed—contra to what most theorists assume. The distributions of noise trade are less "heavy tailed" at lower frequencies, but conform to a normal distribution only with quarterly data. In short, retail aggregate trades at the quarterly frequency are the closest to matching standard model assumptions: they are both i.i.d. and normally distributed. Monthly trades are i.i.d. but their distribution may not be normal. Weekly and daily trades are serially correlated, and their residuals are not normally distributed.

Next we attempt to quantify the intensity of noise trading, by no means an easy task. Even assuming (as we do) that the trades in our samples are noise trades, we cannot say what fraction of total noise trading they represent. Do the traders in our brokerage sample represent 1%, or only 1/10,000, of the noise trading in a stock? We answer this question by comparing retail investors' trading volume with the total trading volume in the market. We demonstrate that a regression of total trading volume (observed in CRSP) on retail investors' trading volume provides bounds on the fraction of noise trading volume accounted for by our retail trades, which in turn enables us to derive bounds on the standard deviation of noise trading in the market. Our methodology is fairly general in that these bounds are valid in the two canonical NREE frameworks—namely, the Grossman and Stiglitz (1980) competitive model and the Kyle (1985) strategic model—and also under various information structures (e.g., dispersed and hierarchical information sets).

[6] This result confirms the intuition in Banerjee (2011). When bringing his model to the data, Banerjee argues that "[f]rom an empirical perspective, while we may expect to find persistence in supply shocks at short horizons (e.g., over days or weeks), the independence assumption is not likely to be restrictive over the monthly horizon at which the predictions are tested" (p. 3032).

We find that the households in our brokerage sample account for at least 0.039%, 0.025%, and 0.024% of all noise trades at (respectively) the daily, weekly, and monthly frequency. The implication is that the standard deviation of noise trading represents no less than 38%, 44%, and 37% of (respectively) the standard deviation of total daily, weekly, and monthly trading volume in the market. Our estimates using small TAQ trades are remarkably similar; mutual fund data (which cover a more recent period) yield larger estimates of the standard deviation of noise trading but lower estimates of its ratio to the standard deviation of total trades. A subperiod analysis reconciles these findings and indicates that, although the standard deviation of noise trading has increased over time, it has not increased as rapidly as the standard deviation of total trades—a finding that suggests informed trading has become relatively more important. This result contributes to the debate over the economic value of a growing financial sector (e.g., Greenwood and Scharfstein 2013, Philippon 2015). It is consistent with recent work by French (2008) and Bai et al. (2016), who document, respectively, an increase in spending on price discovery and an improvement in stock price informativeness over recent decades.

Using the brokerage and TAQ data sets, we also measure noise trading intensity for groups of stocks.[7] This analysis serves a double purpose. First, it confirms that our approach to estimating the variance of noise trading is reasonable. Indeed, in accordance with extant theory, we find that the variance of noise trading is greater among more liquid stocks (Kyle 1985) and among stocks exhibiting greater return volatility (Hellwig 1980, He and Wang 1995). Second, the cross-sectional estimates reported here are of interest in their own right because they can help calibrate multi-stock NREE models.

Finally, we document that stock trades in the brokerage and TAQ data sets are but weakly correlated with fundamentals captured by firms' earnings news; fund flows' correlation with news about aggregate earnings is similarly low. Together these results suggest that strategic complementarities in information acquisition and multiple equilibria are unlikely to arise through the channel outlined by Barlevy and Veronesi (2000, 2008).

[7] Recall that mutual fund data do not allow us to identify which stocks are traded by fund managers in response to flows.

Our paper speaks to the large stream of theoretical research that specifies an exogenous noise trading process. This stream comprises models building on the seminal works of Grossman and Stiglitz (1980) and Kyle (1985), which describe investors' trading behavior and price formation in the presence of asymmetric information. Our contribution is to suggest a plausible process for noise trading that enables theorists (i) to make qualitatively realistic assumptions and (ii) to calibrate and simulate their models without having to choose parameters arbitrarily or match moments, thus freeing up testable restrictions. Our analysis also contributes to the broader debate on the efficiency of the stock market by tracking the relative importance of noise trading over time.

We remark that a theoretically appealing alternative to our approach would be to endogenize noise trading. Indeed, several papers (e.g., Dow and Gorton 1994, Wang 1994, Dow and Gorton 1997) follow this approach. These models offer qualitatively interesting predictions, but they are too stylized to capture a realistic noise-trading process. For example, Wang (1994) assumes that rational agents have access to a private investment opportunity whose return is random but correlated with stock returns; shocks to the private investment returns cause random shifts in investors' demand for stocks. These noise trades inherit all the time-series and cross-sectional properties assumed for private investment returns. But barring data on those returns, the noise trading process remains largely arbitrary.

The rest of our paper proceeds as follows. Section 1 discusses the motivation for our work in more detail. Section 2 presents evidence from the literature regarding our identifying assumption that retail trading proxies for noise trading, and Section 3 describes the data. In Section 4, we explore the time-series properties of noise trading; in Section 5, we estimate its intensity and examine how it changes over time and across types of stocks. Section 6 examines the correlation between noise trades and fundamentals. We conclude in Section 7 with a summary of our approach and results.

1) Motivation: Why estimating a realistic process for noise trading is valuable

There are three main reasons why it is critical to know what noise trading actually looks like. The first two reasons involve qualitative features of the noise trading process while the third is of a quantitative nature.

a. The persistence of noise trades

As mentioned previously, there are at least three reasons why the persistence of noise trades plays a central part in NREE models. First, it determines the extent to which arbitrageurs are willing to correct any mispricing and thereby determines the informativeness of asset prices—for example, whether they are better or worse predictors of fundamentals than is the consensus opinion. Indeed, in their analysis of short-termism in financial markets, Cespa and Vives (2015, p. 2101) acknowledge that "a crucial hypothesis of our model is that liquidity trading displays persistence." Other examples include Stein (1987), Grundy and McNichols (1989), Campbell and Kyle (1993), He and Wang (1995), and Cespa and Vives (2012).

Second, the persistence of noise trading controls the serial correlations of stock returns and of trading volume (e.g., Wang 1993). Makarov and Rytchkov (2012, p. 949) show that "the autocorrelation structure of returns in rational expectation[s] equilibria is determined to a large extent by the assumed process for the stochastic supply of equity." Thus their Theorem 2 establishes that, when noise trading follows an AR(1) process, the sign of the autocorrelation of returns depends entirely on the magnitude of the first-order autocorrelation coefficient. This is a crucial result because it implies that asymmetric information alone cannot generate return momentum when the demand from noise traders follows an AR(1) process. The authors note also that this property no longer holds if noise trading follows an AR(2) process, which underscores the sensitivity of stock return behavior to the noise trading process.

A similar dependence on the persistence of noise trading holds for the serial correlation of trading volume. As noted by Banerjee and Kremer (2010, pp. 1271–72), "one can generate serial correlation in volume by assuming serial correlation in the aggregate supply shocks [i.e., in noise trading], or [one] can generate trade without price changes by forcing aggregate supply shocks to perfectly offset aggregate information shocks. However, this is unappealing in terms of providing insight into what generates these patterns, since the noise process is assumed to be unexplained and exogenous."

Third, the persistence of noise trades is central to the debate concerning how the liquidity of financial markets should be measured. In a recent empirical study, Collin-Dufresne and Fos (2014a) document that standard measures of stock price liquidity—and, in particular, of the adverse selection component (e.g., estimates of Kyle's (1985) lambda)—fail to capture the presence of informed trading (see also Kacperczyk and Pagnotta (2016)). These authors inspect trades executed by informed investors and uncover a strong positive relation between liquidity and the likelihood of informed trades; thus, contrary to traditional models, informed trades are associated with high liquidity and not with low liquidity. The leading explanation, developed further in Collin-Dufresne and Fos (2014b), is that informed investors choose when to trade and participate only when they expect the market and/or the target stock to be liquid. Because liquidity is usually associated with the presence of noise traders, this explanation presumes that noise trading is predictable, such as when it is persistent.

b. Distribution of noise trading and its correlation with fundamentals

Standard models assume that noise trades are both normally distributed and uncorrelated with the asset's fundamental value. However, recent theoretical work suggests that neither assumption is innocuous. In fact, both are required to preclude strategic complementarities in information acquisition and thus to rule out multiple equilibria. In Breon-Drish (2010, 2014), complementarities arise because of departures from the normal distribution. The intuition is that a price signal's informativeness varies with the price level, which can lead to a backward-bending demand curve for uninformed traders (meaning that the demand for the asset can increase with its price). This, in turn, clouds the price signal and may render the value of information non-monotonic in the number of informed traders. In Barlevy and Veronesi (2000, 2008), complementarities arise because there is a positive correlation between the asset's fundamental value and its supply. Indeed, a high price is associated not only with a high fundamental value (as in standard models with a zero correlation) but also with a low fundamental value when supply is low. In such a setting, prices tend to be less informative as more traders become informed, thus spurring further information acquisition. It follows that equilibrium uniqueness is fragile outside the normal and independently distributed framework. We shall assess the plausibility of these assumptions in Sections 4.b and 6.

c. *Noise trading intensity*

Finally, another fundamental aspect of noise trading is its intensity, or standard deviation. Though most authors focus on the qualitative predictions of NREE models, many seek to quantify the size of the effects they model and to assess their economic relevance. Then noise trading intensity becomes a key input for calibrating or simulating models. But because noise trading cannot be observed directly, most theorists simply pick an arbitrary level of variance or choose a value so as to match the model's predicted moments with sample moments that are estimated from market data.

As an example of the former strategy, Watanabe (2008, p. 246) argues as follows: "Since no estimate is available for the variance of individual endowment noises, it is set somewhat arbitrarily at $\Sigma_{\zeta}^{1/2} \equiv 4\Sigma_{\eta}^{1/2}$ throughout the rest of the calibration." Biais et al. (2010) likewise arbitrarily assume, in their numerical analysis, that the variance of noise (in their setup, of "endowment shocks") is 1% and that its serial correlation is zero. Campbell and Kyle (1993) similarly estimate their model under various yet arbitrary assumptions about its correlation with fundamentals (see also Brennan and Cao 1996, Bernardo and Judd 2000).

An example of the latter strategy is given by Campbell et al. (1993, p. 931) who state that "[t]he trickiest part of the calibration is to specify the dynamics of the Z_t process [here Z_t is the marginal investor's risk aversion, which is subject to shocks and thus generates noise trading]. We would like to pick a process that generates realistic stock price behavior" (see also Peress 2003, Banerjee 2011). Although matching moments is a sensible approach, it offers no way to gauge the plausibility of the chosen noise trading parameters. More importantly, once stock market moments are matched, the empirical validity of a model's predictions about those moments can no longer be evaluated. By pinning down a realistic noise process, we enable researchers to test additional restrictions on the data.

2) Empirical strategy: Are retail trades a good proxy for noise trades?

Our analysis builds on the premise that trades due to retail investors (both stock trades and flows to mutual funds) are noise trades. In this section, we discuss the evidence presented in the literature regarding this premise.

a. Retail stock trades

i. Retail investors perform poorly as a group, even before transactions costs

Noise trades are difficult to detect in practice. They are defined in the NREE literature as trades that are not motivated by traders' rational beliefs about assets' fundamentals. Instead, they might be driven by liquidity needs, preference shifts, random stock endowments, private risky investment opportunities, or behavioral traits, among other possibilities. A defining characteristic—one that is implied by this definition—is that they lead to monetary losses. As observed by Black (1986, p. 531), "most of the time, the noise traders as a group will lose money by trading, while the information traders as a group will make money." Individual investors meet this criterion because they have been consistently found to lose money (for a review of the evidence, see Barber and Odean 2013).

Although transactions costs (e.g., commissions and bid-ask spreads) contribute significantly to their poor performance, individuals also lose money on their trades before costs. For example, Odean (1999), using the same brokerage data as in our paper, estimates that stocks bought by retail investors underperform stocks sold by 23 basis points per month in the year after the transaction. Using small TAQ trades as a proxy for the trading of individual investors, as we do here, Hvidkjaer (2008) and Barber et al. (2009a) document that stocks heavily bought by individuals over horizons ranging from one month to one year subsequently underperform stocks heavily sold by individuals. Grinblatt and Keloharju (2000) and Barber, Lee et al. (2009) report similar results for individual investors located in Finland and in Taiwan, respectively.

This poor performance contrasts with that of institutional investors. Although the question is still not settled, several studies find that fund managers do earn superior returns, at least gross of fees. For instance, Wermers

(2000) reports that mutual funds hold stocks that (on average) outperform the market by 1.3% per year—but underperform after deducting all expenses (viz., trading costs, management fees, and costs associated with non-stock holdings). Chen et al. (2000) estimate that the stocks managers buy outperform the stocks they sell by 2% per year. Fama and French (2010) find that fund managers earn positive returns gross of management fees.[8]

With regard to the performance of retail investors, two caveats are in order. First, there is some evidence that retail trades are associated with superior returns over short horizons (i.e., of up to a week; see Kaniel et al. 2008, Kelley and Tetlock 2013).[9] These returns can be explained by individual investors supplying liquidity to institutions that require immediacy. We believe that this finding does not invalidate our empirical design because retail investors actually hold stocks for much longer than a week—for 16 months on average, according to Barber and Odean (2000). Furthermore, there is no evidence of short-term overperformance associated with mutual fund flows, which we also analyze.

Second, although the performance of retail investors is low on average, it varies considerably across individuals. Some investors appear to outperform consistently, at least before trading costs, but they are rare (Coval et al. 2005). We aggregate trades across all individuals in order to focus on the average investor's behavior.

ii. Characteristics of the stocks traded by retail investors

A small number of empirical studies attempt to identify noise traders by examining the stocks they predominantly trade. Most of these studies find noise trading to be associated with retail trading. For example, Foucault et al. (2011) exploit a decline in retail trading triggered by a reform of the French stock market that raised the cost of trading for retail investors. These authors report, following the reform, a reduction in stock return volatility, accompanied by a decrease in the magnitude both of return reversals and of the price impact of trades, for stocks targeted by the reform—and only those stocks. Almost all theories link these observations

[8] See also Grinblatt and Titman (1989) and Daniel et al. (1997).

[9] See also Hvidkjaer (2008) and Barber et al. (2009a). Evidence on these superior returns is mixed outside the United States (Andrade et al. 2008, Barber, Lee et al. 2009).

to a reduction in the intensity of noise trading. In a similar vein, Peress and Schmidt (2016) study episodes of sensational news (e.g., the O. J. Simpson trial) that is exogenous to the market and distracts investors. They find that, on “distraction days”, retail investors trade less while both liquidity and volatility decline among stocks owned predominantly by retail investors. Again, these findings are consistent with retail investors behaving as noise traders.

iii. Retail trades are cross-correlated

An important requirement for a group of trades to qualify as noise trades is that they be correlated. Otherwise, they will wash out in the aggregate without materially affecting asset prices—and thus cannot blur the price signal in NREE models. This is requirement is especially important for retail investors given their trades tend to be small. Kumar and Lee (2006) and Barber et al. (2009b) report that retail trades have a common systematic component.

iv. Why do retail investors trade?

If individuals lose money from trading, then why do they trade? Besides liquidity shocks (e.g., excess cash that needs to be invested, or a consumption need requiring divestment)—which seem insufficiently compelling to justify the vast amount of trading observed in the data—the evidence points to various psychological heuristics and biases. For example, survey evidence and trading records indicate that investors who are more confident about their skills trade more aggressively (Dorn and Huberman 2005, Glaser and Weber 2007, Graham et al. 2009, Grinblatt and Keloharju 2009). Barber and Odean (2001) show that men, who are more likely (it is argued in the psychology literature) to be overconfident than are women, trade more than women and perform worse. Individuals who enjoy sensation-seeking activities, as proxied by speeding tickets (Grinblatt and Keloharju 2009) or the availability of lotteries in their state (Dorn et al. 2014), also trade more.

In addition, individual investors appear to be strongly influenced by the media. Barber and Odean (2008) show that retail investors buy stocks that are featured in the media, and Engelberg and Parsons (2011) find that they trade in response to their local newspapers’ business coverage. Engelberg et al. (2010) report that the market reaction to trading recommendations on the *Mad Money* television show is greater when viewership is higher. Another illustration of how biases affect investors' trading behavior is the well-documented "disposition effect" (Shefrin and Statman 1985), whereby individuals would rather sell winning stocks (i.e., those that have increased in value since being purchased) than losing stocks (those that have decreased in value). Collectively, these findings indicate that psychological traits, errors, and biases unrelated to stocks' actual fundamentals are drivers of retail investors' trades.

b. *Retail flows to mutual funds*

Retail investors' poor skills at selecting stocks are also apparent in their mutual fund investments. Sirri and Tufano (1998) show that they tend to "chase performance" by directing money to mutual funds with strong recent performance yet fail to withdraw from funds with poor recent performance. Using aggregate flows, Friesen and Sapp (2007) document that mutual fund investors display poor market timing ability. Ben-Rephael et al. (2012) investigate retail flows between bond and equity funds within fund families and find that aggregate net exchanges toward equity funds are negatively correlated with market returns over the subsequent four to ten months. Also, Akbas et al. (2015) report that mutual fund flows, in contrast to hedge fund flows, exacerbate stock return anomalies (especially for growth, accrual, and momentum).[10] Analyzing the cross section of funds, Frazzini and Lamont (2008) find that fund flows are directed to funds that subsequently underperform.

Mutual fund managers must respond to positive (resp. negative) flows with stock purchases (resp. sales).[11] Since these trades are motivated by necessity rather than information (recall that mutual fund investors display no ability to pick funds or time the market), they clearly qualify as noise trades. As explained by Edelen (1999, p. 443): "Consider a fund manager who initially holds some target efficient portfolio. Suppose that the manager experiences a cash flow shock (a random number of redemptions and new sales). . . . The flow shock that the

[10] The literature also reports a positive flow–return correlation in the short run, which is attributed to price pressure. That is, flows to funds lead to short-lived excess demand—for assets held by mutual funds—that subsequently reverses (e.g., Edelen and Warner 2001, Ben-Raphael et al. 2012, Lou 2012).

[11] Scholars have examined the consequences of flow-induced trading on mutual fund performance (e.g., Edelen 1999, Christoffersen et al. 2006) and on asset pricing (e.g., Coval and Stafford 2007, Chen et al. 2010, Edmans et al. 2012).

fund experiences moves the fund away from the target portfolio. Getting back to an efficient portfolio requires trade in some or all stocks. . . . This liquidity component of the fund managers' trading plays the role of the exogenous supply-noise trading in standard rational expectations models of trade."

In light of these considerations, we use mutual fund flows to proxy for noise trading. We stress that our approach is not inconsistent with mutual fund managers picking stocks (successfully or otherwise): their need to satisfy flows forces them to trade in and out of the stock market as a whole, but they do have discretion about which stocks to trade. Indeed, such trading decisions might well be informed ones. Hence we restrict our use of mutual fund data to estimation of a noise trading process for the entire market and not for individual stocks.

Our analysis of flows makes one important assumption—namely, that the trades induced by flows have similar statistical properties (with respect to, e.g., persistence and distribution) as the flows themselves.[12] So, for example, if flows have a first-order autocorrelation coefficient of 0.2, then flow-driven trades should have the same first-order autocorrelation coefficient. This assumption is strictly correct only if mutual fund managers meet all positive (negative) net flows with stock purchases (sales). In practice, managers carry a proportion of their assets in cash so they can absorb flows without trading immediately.[13] Yet cash positions are typically quite low because holding cash is costly: it causes performance to deviate from investment benchmarks (Wermers 2000). Yan (2008) estimates that the median fund holds 3.68% of its assets in cash and that funds' cash holdings are extremely persistent, with a first-order autocorrelation coefficient of 0.96 at the monthly frequency. Edelen (1999) and Chernenko and Sunderam (2016) report consistent findings using semi-annual and monthly data, respectively; they find that a dollar of fund flows is associated with 70 and 77 cents (respectively) in trading activity. In short, funds do not constantly draw down and build up cash reserves in

[12] In Section 5 we discuss the sensitivity of our findings to this assumption.

[13] Chernenko and Sunderam (2016) report that mutual funds use alternative liquidity management tools (e.g., redemption restrictions, credit lines, interfund lending programs) much less frequently than cash for that purpose.

response to flows. It follows that the statistical properties of flow-induced trades likely resemble those of the flows themselves.

c. Summary of the evidence

Because retail trades are unprofitable, cross-correlated, implicated in stock volatility and liquidity, and motivated by behavioral biases, such trades are commonly viewed as the archetypal noise trades. For example, Stambaugh (2014), in his Presidential Address on the influence of noise trading in investment management, uses the fraction of US equity owned directly by individuals as a proxy for noise trading. He argues that the decline in that fraction over the past three decades explains several concomitant trends, including the shift by active managers toward lower fees and the rise of index-like investing. Most of the literature on investor sentiment similarly attributes sentiment to individual investors (e.g., Lee et al. 1991) and uses retail flows to mutual funds as a measure of that sentiment (Frazzini and Lamont 2008, Ben-Rephael et al. 2012, Da et al. 2014). Our work builds on this stream of applied research.

We acknowledge that not all noise trades are retail trades; indeed, some institutional trades also qualify as noise. Mutual fund flows capture such trades, but they are unlikely to comprise all institutional noise trades. We have, unfortunately, no way to detect institutional noise trades. Hence we simply assume either that institutional noise trades behave similarly to retail noise trades or that they constitute a small portion of total noise trades. In Appendix B we derive the formal conditions under which retail trading is a good proxy for noise trading.

3) Data

We use three data sets. The first two—from a brokerage house and from TAQ—contain individual stock transactions, and the last one consists of retail flows to mutual funds.

a. Households' trading data

The first data set comprises the trades made by retail investors or "households" through a large discount brokerage firm. These data are described in detail by Barber and Odean (2000) and amount to some 1.9 million common stock trades executed by 78,000 households between January 1991 and November 1996 (inclusive). Hirshleifer et al. (2008) argue that this data set is representative of individual investors as a whole; with 1.25 million clients (from which the 78,000 households were randomly drawn), the broker accounts for 4% of the population of individual shareholders. Moreover, Ivković et al. (2005) document that the patterns of stock sales recorded in this data set are similar to those reported by individuals on their income-tax returns. Because the number of households in this data set displays structural breaks in January of each year—breaks that are likely due to how the brokerage house recorded the data and not to actual changes in its client base—we follow Barber and Odean (2002) in focusing on the trades of 12,743 households with portfolio holdings throughout the 1991–1996 sample period. We obtain virtually identical results when we use instead all the households in this data set.

[[INSERT Figure 1 about Here]]

Our main measure of households' net buys is the net turnover (henceforth simply "turnover"), which is defined as the aggregate value of their buys minus the aggregate value of their sells, divided by total market capitalization. We also consider the net number of households buying shares, defined as the number of households buying minus the number of households selling. Both variables are constructed at the daily, weekly, and monthly frequency. Figure 1 displays the monthly time series of households' aggregate trades, and Table 1 presents summary statistics for the different frequencies.

[[INSERT Table 1 about Here]]

b. TAQ data

Our second data source consists of transactions involving NYSE/AMEX/NASDAQ stocks recorded in the TAQ/ISSM database since 1991.[14] Research has shown that, until decimalization was introduced in 2001 (and thereby made order splitting cost-effective), small trades were likely due to individual investors whereas large trades were typically placed by institutions (Hvidkjaer 2008). We therefore use small trades over the 1991–2000 period to identify retail trades.[15] Trades are classified as being buyer- or seller-initiated according to the Lee and Ready (1991) algorithm, and they are classified by size via a procedure described in Hvidkjaer (2006). This procedure sorts stocks into quintiles based on NYSE/AMEX firm size cutoff points, where those quintiles reflect the following small-trade (resp. large-trade) cutoff points: $3,400 (resp. $6,800) for the smallest firms; $4,800 ($9,600), $7,300 ($14,600), and $10,300 ($20,600) for the three middle quintiles; and $16,400 (resp. $32,800) for the largest firms. We then aggregate dollar buys and dollar sells over the entire data set separately for small and large trades and by day, week, and month. Next we calculate the difference between buys and sells before dividing by the total market capitalization to obtain a measure of net turnover. Thus we produce three pairs of time series for net turnover, or one pair of turnovers (representing small and large trades) for each frequency. Figure 1 displays the daily time series of net turnover estimated from small trades in TAQ, and Table 1 presents summary statistics at daily, weekly, and monthly frequencies.

c. Mutual fund data

Our final data source consists of net daily flows to mutual funds. We restrict our analysis to funds that are invested in US domestic equity and are sold to retail investors. The data, obtained from TrimTabs Investment Research, Inc., cover the period from January 1999 through August 2013. According to Kaniel and Parham (2017), the coverage of TrimTabs increases from approximately 5% percent of the universe of US-based mutual

---

[14] The ISSM (Institute for the Study of Securities Market) data set includes all transactions in all stocks listed on NYSE/AMEX/ NASDAQ in 1991 and 1992, while TAQ covers 1993 to 2000.

[15] The data for 1991 and 1992 come from the ISSM database. In analyzing various transaction databases, including the one we use here, Lee and Radhakrishna (2000) and Barber et al. (2009a) confirm that trade size is an effective proxy for identifying retail trades over the 1991–2000 period.

funds at the beginning of the sample period to approximately 20% toward the end of that period. A detailed description of TrimTabs data is given by Edelen and Warner (2001). These data are matched to the CRSP Mutual Fund files—based on fund ticker, total net assets (TNA), and monthly flows—to identify such fund characteristics as return and investment style. Note that TrimTabs does not report purchases and redemptions separately, only their difference (net flows). We aggregate net flows and TNA at the daily, weekly, and monthly frequency, and we define net turnover as the aggregated net flow divided by funds' aggregate TNA.

d. Complementarity of data sources

Our three data sources complement each other. One advantage of the brokerage data is that it covers retail investors exclusively—that is, noise traders as we define them. Furthermore, investors are identified and followed over time, thus enabling the measurement of investor-level variables such as the number of investors trading. A drawback of this data set is that it covers only a subset of the retail population and their corresponding stock trades.

In contrast, the TAQ data set covers all NYSE and AMEX stocks and offers a broad view of the market. It thus enables examination of small trades with less concern about the sample's representativeness. One shortcoming of the TAQ data is that they do not contain traders' identities, which makes it impossible to confirm that small trades are actually executed by retail investors. Some small trades are likely made by institutional investors breaking up their informed trades to "pass" as retail traders.

Mutual fund flows account for the bulk of retail activity. Indeed, households nowadays hold most of their equity indirectly through institutions; in 2007 they directly owned only 37% of the stock market, down from 57% in 1991 and 46% in 2000 (Rydqvist et al. 2014). A drawback of using mutual fund flows is that they cannot be related to noise trading in individual stocks. We have mentioned that mutual fund managers, when responding to flows, have some discretion about which stocks to trade and that they might select stocks on the basis of fundamental information. For this reason, we restrict our use of mutual fund data to estimating a noise trading process for the entire market rather than for individual stocks.

Finally, the data sets complement each other also by covering different periods: 1991–1996 for the brokerage data, 1991–2000 for the TAQ data, and 1999–2013 for the mutual fund data. Reporting results for all three data sets allows us to assess their robustness and to gain insights into the evolution of noise trading over time.

e. Seasonality

The data exhibit seasonal patterns. Regressing net turnover on calendar month dummies yields results that are consistent with prior studies (to save space, we do not report these regressions). We find that net turnover is lower in December, which is consistent with individual investors realizing losses for tax purposes (Badrinath and Lewellen 1991), and also in August and September, which coincides with summer vacation (Hong and Yu 2008). We also find some evidence for day-of-the-week effects when we regress daily data on day-of-the-week dummies, but the coefficient estimates tend to be statistically insignificant. Throughout the analysis, we purge the data—for households' trades, TAQ trades, and mutual fund flows—of calendar effects and time trends by using the residuals from regressions on indicator variables for day of the week, month of the year, and year.[16]

4) Time-series properties of aggregate trades

Here we investigate the time-series properties of aggregate households' trades, TAQ small trades, and mutual fund flows.[17]

a. Fitting an autoregressive process to the data

Models typically assume either that noise trading is i.i.d. or that it follows an autoregressive process. We evaluate these assumptions and determine the number of lags to include. We fit households' net turnover, the net number of households trading, TAQ small net turnover, and mutual fund net flows to autoregressive models with up to 30 lags. In Figure 2 we plot the *p*-value of a white-noise *Q*-test for the residuals (left axis). High *p*-values indicate that we cannot reject the null hypothesis of residuals from the fitted process being serially

[16] Results are qualitatively unchanged if we use the raw data instead.

[17] Dickey–Fuller tests (not reported here) confirm that these time series are stationary.

uncorrelated. We also show the value of Akaike's information criterion (dashed line and right axis) as a function of the number of lags.[18] Lower values of this criterion correspond to better models.

[[INSERT Figure 2 about Here]]

A comparison of the four panels in this figure reveals that fewer lags are required to fit the data at lower frequencies. At the daily frequency, multiple lags are needed to eliminate serial dependence in the residuals. The number of lags ranges from 5 for small TAQ net buys to 15 for the number of households trading (at the 10% significance level). At the weekly frequency, one lag or less is sufficient to produce uncorrelated residuals for household and small TAQ trades; for fund flows, 6 lags are needed (down from 14 at the daily frequency). For monthly data, an AR(0) model offers a reasonable approximation in all three data sets, since we cannot reject the hypothesis that trades are serially uncorrelated. This is good news for theorists because, when there are fewer lags, the models are less complex. The information criterion at times selects at least one lag, so an AR(1) model may prove to fit the data best.[19]

[[INSERT Figure 3 about Here]]

We now examine the performance of AR(1) processes in more detail. Indeed, several theoretical papers model noise trading as an AR(1) process and argue that the magnitude of the first-order autocorrelation coefficient decreases with the duration of a period (see e.g. He and Wang 1995, Cespa and Vives 2012). This conjecture is consistent with our previous analysis of the lag order. It is also consistent with Figure 3, which displays the first-order autocorrelation coefficient as a function of the time period's duration (in days). A downward trend is visible in all four panels, as hypothesized by theorists. For households' turnover (upper left panel), the fitted line has a slope of −0.0066, which means that extending the period by one day reduces the coefficient by

[18] Akaike's information criterion is used to discriminate among nested econometric models. It trades off goodness of fit against model complexity (in our case, the number of lags).

[19] When we use households' trades, the first-order autocorrelation coefficients for net turnover are 0.157, 0.119, and 0.108 at (respectively) the daily, weekly, and monthly frequency. The corresponding values when we use small TAQ trades are 0.496, 0.480, and 0.305; the corresponding values when using mutual fund flows are −0.086, 0.133, and 0.038.

0.0066. The slopes for the number of households trading (upper right panel), small-trade turnover in TAQ (lower left panel), and mutual fund flows (lower right panel) are in the same neighborhood: respectively −0.0102, −0.0086, and −0.0050. The solid circles in Figure 3 mark coefficients that are statistically significant at the 10% level. The plot becomes noisier as duration increases (rightward movement in the graph) because the number of periods decreases, magnifying variations in the coefficient and reducing the number of statistically significant coefficients.

Summary: Daily trades require multiple lags in all three data sets. For weekly trades our findings are mixed: they can be accurately described with no lag or a single lag in the brokerage and TAQ data sets, yet require six lags in the mutual fund data set. Monthly trades can be treated as i.i.d. in all three data sets.

b. *Parametric form*

Here we examine the parametric shape of noise trades. Figure 4 plots their histograms. The curves are hump-shaped like a normal distribution, yet fat tails are clearly visible. Figure 5 displays quantile-to-quantile (Q-Q) plots; that is, this figure plots quantiles of trades against quantiles of a normal distribution. Points along the 45° line conform to a normal distribution. The daily and weekly data deviate from the 45° line in the tails across all trading measures, behavior that reflects the presence of extreme values (first two columns of graphs). In contrast, the monthly data are better aligned with the 45° line for all households' variables; this outcome suggests that households' aggregate trades are approximately normally distributed at that frequency (last column, top two rows). However, small TAQ trades and fund flows continue to display deviations from the 45° line even at the monthly frequency (last column, bottom two rows).

[[INSERT Figure 4 about Here]]

[[INSERT Figure 5 about Here]]

We use the Shapiro–Wilk test to formally test the hypothesis that trades and their residuals from the fitted AR(1) process are normally distributed. Table 2 presents the results. Consistently with a visual inspection of Figure 4, we find that the null hypothesis of normality is rejected across all measures of noise trading at the daily and weekly frequencies. For monthly trades, the results are mixed: normality is rejected for small TAQ trades and mutual fund flows but not for household trades. For all noise measures, test statistics decrease with frequency; this finding indicates that the data become closer to normally distributed at lower frequencies. A natural question is whether the distribution of small TAQ trades and mutual fund flows is normal at a frequency lower than monthly. Hence Table 2 also reports results of the Shapiro–Wilk test conducted on quarterly data. Quarterly trades, regardless of the data set from which they are drawn, all conform to the normal distribution.[20]

[[INSERT Table 2 about Here]]

Recall that our use of retail fund flows as a proxy for noise trades rests on the assumption that flow-induced stock trades inherit the statistical properties of those flows. We assess the sensitivity of our findings to this assumption by repeating the analysis after excluding mutual funds that are more prone to buffer flows with cash. Flow-induced trading harms performance more when it involves illiquid assets. Therefore, funds that hold such assets (e.g., small-cap funds) are likely to rely on a larger cash buffer (see Chen et al. 2010, Chernenko and Sunderam 2016). We find that excluding mid-cap, small-cap or micro-cap funds has no qualitative or quantitative effect on our findings (results available upon request), which leads us to conclude that cash holdings do not significantly alter the properties of fund flows.

Summary: Across all three data sets, retail aggregate trades conform well to standard model assumptions at the quarterly frequency in that they can be considered i.i.d. normal. Monthly trades are i.i.d. but their distribution may not be normal. Weekly and daily trades are serially correlated, and their residuals are definitely not normal. One lag suffices to describe weekly trades in the brokerage and TAQ data sets; more are needed for mutual fund flows and for daily trades.

[20] With respect to persistence, we confirm that quarterly trades can be treated as i.i.d. (results available upon request).

5) Noise trading intensity

An essential aspect of noise trading is its intensity, parameterized in NREE models as the variance of the stock's net supply. Measuring this variance is a challenge and a long-standing question in finance, as reflected by the vast literature on stock market efficiency. Although we assume that households' trades, small TAQ trades, and fund flows are noise trades, we do not know how much of total noise trading they constitute. Do they represent a small percentage or perhaps the majority of noise trading in a stock?[21] We describe here how to address this question, starting with an overview of our strategy. We then formalize that strategy within the two canonical NREE frameworks: the Grossman and Stiglitz (1980) competitive model and the Kyle (1985) strategic model. For concreteness, we show how our procedure works in the case of households; it applies equally well to small TAQ trades and fund flows.

a. Overview

Our key identifying assumption is that households' trades account for a fraction 1/*b* of all noise trading in the economy. That is, *b* represents the ratio of total noise trades in the market to our households' trades. Knowing the coefficient *b* will allow us to scale up the standard deviation of household trading volume in our sample to obtain an estimate of the standard deviation of noise trading for the market as a whole. To estimate *b*, we relate total trading volume in the market to the trading volume of our households:

$$\text{Total trading volume}_t = \frac{1}{2}(\text{Noise trading volume}_t + \text{Rational trading volume}_t)$$

$$= \frac{1}{2}(b \times \text{Households' trading volume}_t + \text{Rational trading volume}_t)$$

In these equations, the second term $\text{Rational trading volume}_t$ represents the aggregate trades of agents who are *not* noise traders, such as informed traders and market makers. Here the factor $\frac{1}{2}$ avoids the double counting of trades; see He and Wang (1995, p. 942, eq. 28) or Admati and Pfleiderer (1988, p. 14, eq. 7).

[21] The aspects of noise trading discussed previously (e.g., lag order, autocorrelation coefficients, shape of the distribution) are independent of scale, so this question does not arise for them.

A (time-series) regression of total trading volume in the market (readily available from CRSP) on households' trading volume yields an unbiased estimate of $b/2$, provided that rational trading volume is uncorrelated with retail trading volume. That would be the case if rational agents traded solely on signals that were uncorrelated with noise trades, such as private information about the asset's fundamental value (informational trades). However, this condition is unlikely to hold: rational investors—both market makers and rational speculators—trade also in reaction to noise trades, accommodating their excess stock demand (resp. supply) with their own sales (resp. purchases). Because of these market-making activities (or non-informational trades), the least-squares estimate of $b/2$, denoted $\hat{b}$, is biased. The bias is upward; that is, $\hat{b}$ is an overestimate of $b/2$ because the volume of rational trading is positively correlated with the volume of noise trading. Indeed, rational investors intensify their market-making trades when noise trading strengthens.

Thus a simple regression—one that does not assume any particular market structure—yields an upper bound on the standard deviation of noise trading for the market as a whole. This upper bound is calculated as twice the estimated regression coefficient multiplied by the standard deviation of households' trading volume. Finding a lower bound requires that we impose some structure on the market. Yet that requirement is mild in that the same lower bound obtains under the two canonical models of trading and under various information structures, as we show next.

b. Bounds in a competitive market

We first derive bounds on the intensity of noise trading within Grossman and Stiglitz's (1980) framework of competitive trading. Toward that end, we employ He and Wang's (1995) intertemporal extension that characterizes the dynamics of trading volume. For ease of reference, we adopt their notation.

In the He and Wang (1995) model, trading is performed by two groups of investors: noise traders and rational traders. A representative noise trader has an inelastic (exogenous) demand for stocks that induces supply shocks. The residual supply of shares, θ_t, that is available to rational agents follows an AR(1) process:

$$\theta_t = a_\theta \theta_{t-1} + \varepsilon_{\theta,t}, \quad \text{where } -1 < a_\theta < 1 \text{ and } \varepsilon_{\theta,t} \sim N(0, \sigma_\theta^2).$$

The change in that supply, $\Delta\theta_t$, is equal to the net number of shares sold by noise traders in aggregate or, equivalently, to the negative of their net buys. By market clearing, $\Delta\theta_t$ also equals the aggregate net buys of rational traders. Normalizing the supply of shares to 1, $\Delta\theta_t$ can be interpreted as the net share turnover.

Rational investors maximize expected (CARA) utility from consuming their wealth at the terminal date. There is a continuum of such agents, who are indexed by *i* and have unit mass. They receive both private and public information about a stock's fundamental value. Errors in private signals are i.i.d. across investors, and public information includes the market price. Rational investors trade either to accommodate supply shocks (a.k.a. market making or non-informational trading) or to speculate on future price changes based on their information (informational trading). Formally, He and Wang (1995, p. 942) establish that the trades of a rational agent *i* can be expressed as the sum of two *uncorrelated* components, $\Delta\theta_t$ and Δx_t^i, which represent (respectively) non-informational trading and informational trading.

As before, total trading volume in the market includes both noise trades and rational trades:

$$\text{Total volume}_t = \frac{1}{2}\left(|-\Delta\theta_t| + \int_i \ |\Delta\theta_t + \Delta x_t^i|\right);$$

once again, the factor $\frac{1}{2}$ prevents trades from being double counted.

Let $\Delta\hat{\theta}_t$ denote the net number of shares purchased by households in our brokerage data set. As explained previously, we assume that these trades account for a fraction *b* of all noise trading in the economy: $\Delta\hat{\theta}_t = -\frac{1}{b}\Delta\theta_t$.[22] It follows that

$$\text{Total volume}_t = \frac{1}{2}\left(\left|b\Delta\hat{\theta}_t\right| + \int_i \ \left|-b\Delta\hat{\theta}_t + \Delta x_t^i\right|\right).$$

[22] The negative sign in this expression accounts for our using $\Delta\theta_t$ to denote the change in the supply of shares available to rational traders (which equals the number of shares *sold* by noise traders) while using $\Delta\hat{\theta}_t$ to denote the number of shares *purchased* by households (which equals the number of shares bought by all noise traders divided by *b*).

We establish in Appendix B that the coefficient $\hat{b}$, which is obtained by regressing total trading volume on households' trading volume $|\Delta\hat{\theta}_t|$, is given by $\hat{b} = b(1+\sqrt{1+r}-\sqrt{r})/2$, where $r \equiv \mathrm{var}(\Delta x_t^i)/\mathrm{var}(\Delta\theta_t)$. The parameter r reflects rational traders' behavior, depends on unobservable investor parameters (e.g., their risk aversion or signal precision), and in principle can take any positive value. Nonetheless, observing that $\sqrt{1+d} - \sqrt{d}$ lies between 0 and 1 for any positive *d* allows us to bound *b* as follows (and confirms the upper bound just derived):

$$\hat{b} \leq b \leq 2\hat{b}.$$

Lower and upper bounds on the standard deviation of noise trading now follow as $\hat{b}\sqrt{\mathrm{var}(|\Delta\hat{\theta}_t|)}$ and $2\hat{b}\sqrt{\mathrm{var}(|\Delta\hat{\theta}_t|)}$, respectively, where $\mathrm{var}(|\Delta\hat{\theta}_t|)$ is the (observed) time-series variance of households' trades.

Our analysis, which is based on He and Wang's (1995) model of disperse information, applies to more general information structures—including those with hierarchical information sets. For example, if some rational investors are informed while others are not (as in Grossman and Stiglitz 1980), then the demand of informed and uninformed investors can be expressed as $\Delta\theta_t + \Delta x_t^I$ and $\Delta\theta_t + \Delta x_t^U$, respectively, where neither Δx_t^I nor Δx_t^U is correlated with $\Delta\theta_t$. The only change in our derivations is that Δx_t^i is now replaced by either Δx_t^I or Δx_t^U. Thus the same bounds obtain.

c. Bounds in a strategic market

We now derive bounds on the intensity of noise trading in a strategic market à la Kyle (1985). The economy is populated by a representative noise trader, a rational market maker and a rational (strategic) trader. The analysis extends straightforwardly to multiple rational traders. The rational trader is assumed to be risk-neutral. In contrast to the competitive case, investors submit orders to a market maker who sets prices. As a result, total trading volume in the market includes not only noise trades and rational trades but also the market marker's trades. To see this, suppose that the noise trader sells 100 shares and that the rational trader buys 150 shares; hence 100 shares will be crossed between traders, and the residual 50 shares will be met by the market maker. Then the total amount of trading is $150 = \frac{1}{2}(|-100| + |150| + |-50|)$. We can express trading volume formally as follows, where the factor $\frac{1}{2}$ again compensates for the double counting of trades (see Admati and Pfleiderer 1988, p. 14):

$$\text{Total volume}_t = \frac{1}{2}(|-\Delta\theta_t| + |\Delta\theta_t + \Delta x_t| + |-\Delta x_t|).$$

The three terms on the right-hand side of this equation represent the volume traded by (respectively) the noise trader, the rational trader, and the market maker. As in the competitive case, Δx_t is a function of rational agents' signals about fundamentals and is uncorrelated with noise trading $\Delta\theta_t$. Plugging in our proxy for noise trading, $-b\Delta\hat{\theta}_t$, yields

$$\text{Total volume}_t = \frac{1}{2}\left(\left|b\Delta\hat{\theta}_t\right| + \left|-b\Delta\hat{\theta}_t + \Delta x_t\right| + |-\Delta x_t|\right).$$

Appendix B shows that the coefficient $\hat{b}$ derived from regressing total trading volume on $\left|\Delta\hat{\theta}_t\right|$ is equal to $b(1 + \sqrt{1+r} - \sqrt{r})/2$, just as in the competitive case. It follows that identical bounds obtain.

Summary: Under both competitive and strategic trading, the standard deviation of noise trading is bounded (a) from below by the standard deviation of our households' aggregate trades multiplied by the regression coefficient of CRSP trading volume on households' aggregate trades and (b) from above by twice that product.

We remark that our estimates of noise trading could be biased upward or downward. On the one hand, to the extent that some noise trades are unrelated to households' trades, small TAQ trades, or mutual fund flows, our approach underestimates the variance of noise trading because those trades are ignored in our calculations. On the other hand, if some households' trades, small TAQ trades, or mutual fund flows reflect information rather than noise, then our approach overestimates the variance of noise trading by treating them as noise trades. The more these biases balance each other, the more accurate are our estimates of noise trading intensity.

d. Noise trading intensity in the overall market

Table 3 gives the results of our estimation procedure for the market at large.[23] The share turnover for the overall market is defined, analogously to that for households, as the value of shares traded in the market (obtained from CRSP) divided by the value of the market. The 12,743 households in our sample (i.e., those with with 71 consecutive months of common stock positions) account for 0.039%–0.078%, 0.025%–0.049%, and 0.024%–0.049% of all noise trades at (respectively) the daily, weekly, and monthly frequency; here the upper (resp., lower) percentages in these ranges are calculated as 1 divided by the corresponding estimate of $\hat{b}$ (resp., half of that quotient). Because these traders represent about 1% of the broker's clients, our figures are consistent with Hirshleifer et al.'s (2008) back-of-the-envelope estimate that those clients account for approximately 4% of all US retail traders.

[[INSERT Table 3 about Here]]

The standard deviation of noise trading at the daily, weekly, and monthly frequency is in the respective ranges 0.029%–0.057%, 0.150%–0.302%, and 0.459%–0.918% when we use households' trades, which constitute anywhere from one third to three quarters of the standard deviation of total trades in the market. These estimates are close to those obtained using small TAQ trades. At the daily frequency, for example, the bounds on the standard deviation of noise trades are 0.030% and 0.060%, or 19% and 38% of the standard deviation of total trades in the market. The standard deviation of noise trading is about twice as large when we use mutual fund flows (0.042%–0.084%, 0.324%–0.647%, and 1.021%–2.042% at the daily, weekly, and monthly frequency), although it accounts for a smaller fraction (than do households' and small TAQ trades) of the standard deviation of total trades in the market.

[[INSERT Figure 6 about Here]]

[23] We perform this analysis only for turnover. The reason is that we have no data on the number of traders in the stock market as a whole.

Differences in estimates across data sets may be attributable to their covering different periods. To check this, we split the TAQ sample into halves. Over the first five years (1991–1995), TAQ estimates are extremely close to those derived from the brokerage data that cover almost the same period (1991–1996). In the second TAQ data subperiod (1996–2000), those estimates are considerably larger than their brokerage data counterparts—even though the former's ratio to the standard deviation of total trades is lower. Figure 6 confirms these observations;[24] it does so by plotting the noise trading intensity from 1991 through 2013.[25] Panel A of the figure shows that our estimates are close to one another across data sets. Panel B reveals a clear downward trend in the ratio of noise trading intensity to standard deviation of total trades; in other words, the standard deviation of noise trading has not kept pace with the standard deviation of total trades. This pattern is not an artefact of the declining share of direct retail trades because it is observed also with respect to mutual fund flows. It suggests that informed trades have become relatively more prevalent in the stock market. This finding contributes to the ongoing debate concerning the economic value of an expanding financial sector (see e.g. Greenwood and Scharfstein 2013, Philippon 2015). The greater prevalence of informed trades that we uncover is consistent with results reported in French (2008) and Bai et al. (2016), using different approaches. Estimates from the former indicate that spending on price discovery has risen from 0.3% to 1% of GDP from 1980 to 2006; according to the latter, US stock prices became more informative over the period 1960–2014.

e. Noise trading intensity by groups of stocks

We now estimate the noise trading intensity for groups of stocks. We can perform this estimation only for households' trades and small TAQ trades because mutual fund flows are not specific to any particular stock (i.e., we do not know which stocks are traded by fund managers in response to flows).

---

[24] Figure 6 uses weekly data to mitigate any problem due to the concurrence of CRSP trades and mutual fund flows. Indeed, if mutual fund managers trade not on the same day they receive flows but instead with a few days' lag, then CRSP trading volume and mutual fund flows will be out of sync in the daily regression that yields the coefficient estimate *b.* However, this issue does not invalidate the rest of our daily analysis.

[25] We obtain a similar picture when we do not purge the data of time trends.

We continue to assume that households' trades and small TAQ trades are scaled-down versions of noise trades, but we allow the scaling factor to vary over groups of stocks. Formally, for stock group *k* and day *t*, we write

$$\Delta\hat{\theta}_t^k = -\frac{1}{b^k}\Delta\theta_t^k;$$

here $\Delta\hat{\theta}_t^k$ and $\Delta\theta_t^k$ denote (respectively) households' trades and noise trades in stock group *k* on day *t,* and b^k is a group-specific constant. A larger scaling factor b^k indicates that a stock group is more under-represented, relative to its noise trading intensity, in our sample of households' trades.[26]

Our procedure for measuring the marketwide scaling factor can be readily applied on a stock-by-stock basis. For each stock *k*, we regress total trading volume on households' trading volume and use $\hat{b}^k$ to denote the resulting regression coefficient. The standard deviation of noise trading in stock *k* is bounded from below by the time-series standard deviation of households' aggregate trades in stock *k*, or $\sqrt{\text{var}(|\Delta\hat{\theta}_t^k|)}$, multiplied by $\hat{b}^k$; it is bounded from above by twice that product.

Our estimation of the noise trading intensity for groups of stocks proceeds in four steps. First, for each month we sort stocks into deciles based on their capitalization, share price, turnover, Amihud illiquidity ratio (a measure of price impact of trades), the (closing) bid–ask spread, return volatility, and return autocovariance. All seven of these variables are estimated every month from daily observations. Capitalization, price, turnover, and bid–ask spread are monthly averages. The Amihud illiquidity ratio is the monthly average of the daily ratio

[26] That scaling factors can vary across stocks has no bearing on our previous analysis of the noise trading process's scale-independent aspects (e.g., lag order, autocorrelation coefficients, shape of the distribution). Consider, as an illustration, the persistence of noise trading, and suppose that noise trades follow an AR(1) process whose coefficient of autocorrelation *r* is identical across stocks:

$$\Delta\theta_{t+1}^k = r\Delta\theta_t^k + \epsilon_{t+1}^k, \quad \text{where } E[\epsilon_{t+1}^i \mid \Delta\theta_t^k, \Delta\theta_t^j] = 0.$$

It follows that $-b^k\Delta\hat{\theta}_{t+1}^k = -rb^k\Delta\hat{\theta}_t^k + \epsilon_{t+1}^k$ or, equivalently, that $\Delta\hat{\theta}_{t+1}^k = r\Delta\hat{\theta}_t^k - \epsilon_{t+1}^k/b^k$. Summing over all stocks (*k* = 1, …, *N*), we obtain

$$\textstyle\sum_k^N \Delta\hat{\theta}_{t+1}^k \equiv \Delta\hat{\theta}_{t+1} = r\Delta\hat{\theta}_t - \sum_k^N \epsilon_{t+1}^k/b^k, \quad \text{where } E[\sum_k^N \epsilon_{t+1}^k/b^k \mid \Delta\hat{\theta}_t] = 0.$$

Thus the autocorrelation coefficient can be estimated equivalently at the market level or the stock level.

of a stock's absolute return to its dollar trading volume. Return volatility and return autocovariance are the monthly standard deviation and autocovariance of the stock's daily raw returns.

Second, within each decile, we aggregate trading volume in our samples of households' trades and of small TAQ trades and in CRSP—over daily, weekly, and monthly frequencies—to generate a 7 × 10 × 3 × 2 time series (one for each sorting variable, decile, frequency, and trading measure). The third step is to obtain the coefficients $\hat{b}^k$ (k = 1, …, 10) by regressing, decile by decile, CRSP trading volume on households' trading volume or TAQ small trade volume. Finally, to derive bounds on the standard deviation of noise trading within each decile, we multiply the regression coefficient b^k by the standard deviation of households' trades for that decile.

[[INSERT Figure 7 about Here]]

The results of this procedure are graphed in Figure 7 and detailed in Table 4.[27] For almost all sorting variables, the variance of households' trades and of small TAQ trades—as well as our estimate of the scaling factor b^k—varies across deciles. This finding suggests that stocks differ not only in the intensity with which they are traded within our data sets but also in the fraction of noise trading for which they account. Panel E of the table, where sorting is in terms of the Amihud illiquidity ratio, illustrates the importance of scaling the variance of households' trades and small TAQ trades decile by decile. The standard deviation of households' and small TAQ trades is greater for stocks that are less liquid, which reflects the prevalence of retail investors among small stocks. However, this does not imply that noise trading is higher among more illiquid stocks. Indeed, the regression coefficient $\hat{b}^k$ is considerably lower for these stocks, too, which implies that the high standard deviation of households' and small TAQ trades is scaled by a smaller factor. The bounds on noise trading depend on the product of the regression coefficient and the standard deviation, so the overall effect of illiquidity on the standard deviation of noise trading is unclear a priori. According to the values reported in Table 4, the scaling factor's effect dominates: noise trading is less volatile for stocks that are more illiquid. Sorts by other measures of liquidity (viz., CRSP turnover in Panel B of the table, bid–ask spreads in Panel C) confirm the positive

[27] Negative estimates of the noise trading intensity correspond to estimates of the slope coefficient that are statistically insignificant; hence they can safely be ignored.

association between noise trading and liquidity. This is precisely what adverse selection models in the spirit of Kyle (1985) predict.

[[INSERT Table 4 about Here]]

Greater volatility in returns is associated with greater volatility in noise trading, as shown in Panel F of Table 4 and as implied by most NREE models (e.g., Hellwig 1980, He and Wang 1995). In contrast, the standard deviation of noise trading does not vary much with stock size (Panel A), most likely because size is related to many stock characteristics and sometimes in opposite ways (as with, e.g., liquidity and volatility). The standard deviation of noise trading appears to be somewhat higher for high-priced stocks (Panel C), but this tendency is also weak. Indeed, despite the standard deviation of households' and small TAQ trades being an order of magnitude higher in the lowest price decile than in the highest price decile, our scale adjustment mitigates the stock price's effect on the standard deviation of noise trading. With respect to daily households' trades, for example, the lower bounds are $271 = 2.39 \times 113.32$ in the lowest decile versus $277 = 0.23 \times 1219.39$ in the highest decile. Once again, these results highlight that our estimates do not simply mirror households' preferences for some stocks but capture noise trading more broadly—that is, by agents not in our sample (e.g., other retail investors and institutions).

The autocovariance results in Panel G of the table are less clear-cut. Whereas noise trading intensifies with increasing autocovariance of daily stock returns when TAQ data are used, the pattern is U-shaped when household data are used. Measured as a fraction of total trades, noise trading is more volatile in the upper deciles than in the lower deciles in both data sets. This tendency is consistent with most theoretical models (e.g., Grossman and Stiglitz 1980, Kyle 1985), where noise trades induce temporary price shifts that encourage investors or market makers to accommodate those trades—for instance, a price increase after noise buying to encourage the sale of those shares. Such price shifts subsequently revert (here, resulting in a price reduction) because they are unrelated to fundamentals; hence they generate a negative autocorrelation in returns.

Summary: The bounds on noise trading are remarkably consistent across data sets. Over time, they have grown less rapidly than the standard deviation of total trades—an indication that the stock market has become more informationally efficient. In the cross section of stocks, noise trading bounds vary in ways that are consistent with extant theory. In particular, greater liquidity and return volatility are associated with greater noise trading volatility, as predicted by virtually all NREE models.

6) Correlation between noise trades and fundamentals

In this section, we assess the extent to which noise trading and fundamentals are correlated. As argued in Section 1, that correlation is important for determining whether decisions to acquire information exhibit strategic substitutability—as in the standard Grossman–Stiglitz (1980) framework—or rather complementarity (as in Barlevy and Veronesi 2000, 2008).

For this determination, we examine how closely retail investments in stocks and mutual funds are associated with news about firms' earnings. We start with stock trades from the brokerage and TAQ data sets. For each stock and quarter we measure the earnings surprise as the difference between actual and expected earnings, where the latter are derived from a seasonal random walk with drift (cf. Bernard and Thomas 1990). To normalize earnings surprises, we divide them by their standard deviation and label the resulting variable *standardized unexpected earnings* (SUE); this variable is defined formally as follows:

$$\mathrm{SUE}_{i,q} = \frac{E_{i,q} - (E_{i,q-4} + \mathrm{drift}_{i,q})}{\sigma_{i,q}}, \quad \text{where } \mathrm{drift}_{i,q} = \frac{1}{8}\sum_{n=1}^{8}(E_{i,q-n} - E_{i,q-n-4}).$$

Here $E_{i,q}$ denotes the actual earnings of firm *i* in quarter *q* (Compustat's earnings per share, excluding extraordinary items) and $\sigma_{i,q}$ is the standard deviation of earnings surprises estimated over the preceding eight quarters. To mitigate the effect of outliers, we sort SUE into deciles and use the decile number as the dependent variable. Then, for each firm and quarter, we aggregate households' and small TAQ net buys over a window ending on the day a firm announces its earnings. We report results for windows of various durations (1, 5, 10, 20, and 40 trading days) because it is not obvious how best to evaluate the (static) Barlevy and Veronesi model in terms of these data. We restrict the analysis of households' trades to stocks that were traded at least 100 times over the period 1991–1996, and we restrict the analysis of TAQ trades to stocks with at least $100,000 worth of small trades over the period 1991–2000.[28] Finally, we estimate a panel regression model of net buys on (contemporaneous) earnings surprise deciles. The regression includes firm, quarter, and month-of-the-year fixed effects; standard errors are clustered by firm.

[[INSERT Table 5 about Here]]

Results for households are reported in Table 5. The estimated coefficients for net buys vary both in sign and in statistical significance. These coefficients are negative in the brokerage data set (Panel A) but positive in the TAQ data set (Panel B). We must emphasize that, throughout, the coefficient estimates are small in terms of economic magnitude. For example, the coefficient of −0.553 in the first row and column of the table indicates that a decrease in earnings surprises from the top decile to the bottom decile is associated with a 5×10^{-6} (= $0.553 \times [10 - 1]/(1$ million)) decrease in net turnover over the 40-day pre-announcement window, or one fifth (= $[5 \times 10^{-6}]/[25.28 \times 10^{-6}]$) of a standard deviation. Even more striking is that the coefficient of 0.076 for small TAQ trades—in the bottom row of in Panel B (1-day pre-announcement window)—implies that a similar decrease in earnings surprises is associated with a 0.7×10^{-6} (= $0.076 \times [10 - 1]/(1$ million)) increase in net turnover on the announcement day, or about one hundredth (= $[0.7 \times 10^{-6}]/[61 \times 10^{-6}]$) of a standard deviation. For turnover and the number of households in the brokerage data, the corresponding values are (respectively) 1% and 6% of a standard deviation. This weak economic significance is also reflected in the low R^2 values (less than 0.2%).

We conduct a similar analysis of mutual fund flows to equity funds. Since we do not know which stocks managers trade in response to these flows, we relate flows to surprises about *aggregate* earnings. Toward this end, we aggregate earnings across all stocks and define the standardized aggregate earnings surprise using the same method as in the case of individual firms.[29] That is, we estimate the difference between actual and

[28] Our findings are not sensitive to the choice of these filters.
[29] We omit from the aggregation all firms whose fiscal year does not end on December 31. These firms account for 30% of the firm universe.

expected aggregate earnings, where the latter are derived from a seasonal random walk with drift, and then divide this difference by its standard deviation:

$$\mathrm{SUAE}_q = \frac{AE_q - (AE_{q-4} + \mathrm{drift}_q)}{\sigma_q}, \quad \text{where } \mathrm{drift}_q = \frac{1}{8}\sum_{n=1}^{8}(AE_{q-n} - AE_{q-n-4}).$$

Here AE_q represents the aggregate earnings in quarter *q* while σ_q denotes the standard deviation of aggregate earnings surprises estimated over the previous eight quarters. For a given quarter, we use the median announcement date across all firms. We run a quarterly time-series regression of fund flows (net turnover) on SUAE, where flows are aggregated every quarter over windows of 40, 20, and 10 trading days ending on the day before the median announcement date. Findings for shorter windows are not meaningful and so are not reported.[30] The regressions include quarter fixed effects.

Our results, displayed in Panel C of Table 5, are consistent with those based on stock trades; more specifically, they reveal a weak association between flows and earnings surprises. The estimated coefficients are statistically insignificant—possibly owing to the relatively small number of quarters—and their economic significance is again modest. For example, the coefficient of 520 in the first row of the table indicates that a decline in aggregate earnings surprises from the top decile to the bottom decile is associated with a 0.5% (= 520 × [10 − 1]/(1 million)) decrease in net flows over the 40-day pre-announcement window, or 37% (= [0.5%]/[12,630 × 10^{-6}]) of a standard deviation.

Summary: The weak correlation between noise trades (households' trades, small TAQ trades and fund flows) and fundamentals suggests that the scope for complementarities in information acquisition through the channel outlined by Barlevy and Veronesi (2000, 2008) is limited.

7) Conclusion

In this paper we breathe life into noise trading, an essential but mostly impalpable component of trading models. We characterize the trades executed by investors who are natural candidates for the role of noise

[30] We obtain similar results when we aggregate fund flows at a quarterly frequency and then estimate a time-series regression of fund flows in a quarter on aggregate earnings surprises for the subsequent quarter.

traders: individual (retail) investors. Using three different data sources (a brokerage house, the TAQ database, and mutual fund flows data), we estimate a realistic process for noise trading that can help theorists make qualitatively plausible assumptions about noise trading, perform comparative statics that account for any potential effect on noise trading, and—no less importantly—calibrate their models.

Our data sources yield remarkably consistent findings in spite of their dissimilarity. We first document that noise trading can be treated as i.i.d. normal at the quarterly frequency, in conformance with theorists' assumptions. Monthly trades are i.i.d. but are generally not normally distributed. Daily and weekly trades require multiple lags and are not normal.

Next, we take up the challenge of measuring the intensity of noise trading. We develop a fairly general methodology that is valid in the two canonical NREE frameworks—the Grossman and Stiglitz (1980) competitive model and the Kyle (1985) strategic model—and under various information structures (e.g., under disperse and hierarchical information sets). Although exact numbers vary with the data used and its frequency, we find overall consistent estimates. An important source of variation in our estimates is due to their evolution over time; in particular, we find that the standard deviation of noise trading has grown over the past two decades. Yet it has not kept pace with the standard deviation of total trades, which suggests that the stock market has become more informationally efficient. This result is relevant to the ongoing debate on the economic value of a growing financial sector and it is line with research documenting an increase in stock price informativeness.

In addition, we quantify the noise trading intensity over groups of stocks in order to validate our estimation strategy and to help calibrate multi-stock NREE models. We find that our estimates vary across stocks in ways that are largely consistent with the predictions of NREE models. In particular, our results confirm that noise trader risk is higher for stocks that are more liquid and/or exhibit greater return volatility.

Finally, we document that stock trades in the brokerage and TAQ data sets are weakly correlated with fundamentals as proxied by firms' earnings news; fund flows' correlation with news about aggregate earnings is similarly low. Hence multiple equilibria and strategic complementarities in information acquisition—through the channel outlined by Barlevy and Veronesi (2000, 2008)—are unlikely to arise.

References

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

Akbas, F., Armstrong, W. J., Sorescu, S., and Subrahmanyam, A. ,2015, Smart money, dumb money, and capital market anomalies. *Journal of Financial Economics*, 118(2), 355-382.

Andrade, S. C., Chang, C., & Seasholes, M. S. (2008). Trading imbalances, predictable reversals, and cross-stock price pressure. *Journal of Financial Economics*, *88*(2), 406-423.

Badrinath, S. G., and W. G. Lewellen, 1991, Evidence on Tax-Motivated Securities Trading Behavior, *Journal of Finance* 46, 369–382.

Bai, J., Philippon, T., & Savov, A., 2016, Have financial markets become more informative? *Journal of Financial Economics*, 122(3), 625-654.

Banerjee, S., 2011, Learning from Prices and the Dispersion in Beliefs, *Review of Financial Studies* 24, 3025-3068.

Banerjee, S., and I. Kremer., 2010, Disagreement and Learning: Dynamic Patterns of Trade, *Journal of Finance* 65, 1269–302.

Barber, B.M., Y. Lee, Y. Liu, and T. Odean, 2009, Just How Much Do Individual Investors Lose by Trading? *Review of Financial Studies* 22:609-632.

Barber, B. M., and T. Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *Journal of Finance* 55, 773–806.

Barber, Brad, and Terrance Odean, 2001, Boys will be boys: Gender, overconfidence, and common stock investment, *Quarterly Journal of Economics* 116, 261–92.

Barber, B. M., and T. Odean, 2002, Online investors: Do the slow die first? *Review of Financial Studies* 15, 455–89.

Barber, Brad, and Terrance Odean, 2008, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *Review of Financial Studies* 21, 785–818.

Barber, B. M., and T. Odean, 2013, The Behavior of Individual Investors, *Handbook of the Economics of Finance*, Elsevier, edited by Constantnides, Harris, and Stulz, 1533-69.

Barber, B. M., T. Odean, and N. Zhu, 2009a, Do retail trades move markets? *Review of Financial Studies* 22, 151–186.

Barber, Brad, Terrance Odean, and Ning Zhu, 2009b, Systematic noise, *Journal of Financial Markets* 22, 547–569.

Barlevy, G., and P. Veronesi, 2000, Information Acquisition in Financial Markets, *Review of Economic Studies* 67, 79–90.

Barlevy, G. and P. Veronesi, 2008, Information Acquisition in Financial Markets: a Correction, Working Paper.

Ben-Rephael, A., Kandel, S., Wohl, A., 2012. Measuring investor sentiment with mutual fund flows. *Journal of Financial Economics* 104, 363-382.

Bernard, V. L., and J. Thomas, 1990, Evidence that stock prices do not fully reflect the implications of current earnings for future earnings, *Journal of Accounting and Economics* 13, 305–340.

Bernardo, A. E., & Judd, K. L., 2000. Asset market equilibrium with general tastes, returns, and informational asymmetries. *Journal of Financial Markets*, 3(1), 17-43.

Biais, B., Bossaerts, P., & Spatt, C., 2010, Equilibrium asset pricing and portfolio choice under asymmetric information. *Review of Financial Studies*, 23(4), 1503-1543.

Black, F., 1986, Noise. *Journal of Finance* 41, 529–43.

Brennan, M. J., & Cao, H. H., 1996. Information, trade, and derivative securities. *Review of Financial Studies,* 9(1), 163-208.

Breon-Drish, B., 2010, Asymmetric Information in Financial Markets: Anything Goes, Working Paper.

Breon-Drish, B., 2014, On Existence and Uniqueness of Equilibrium in a Class of Noisy Rational Expectations Models, Working Paper.

Campbell, J. Y., S. J. Grossman, and J. Wang, 1993, Trading volume and serial correlation in stock returns, *Quarterly Journal of Economics* 108, 905–939.

Campbell, John Y., and Albert S. Kyle, 1993, Smartmoney, noise trading and stock price behaviour, *Review of Economic Studies* 60.

Cespa, G., and X. Vives, 2012, Dynamic Trading and Asset Prices: Keynes vs. Hayek, *Review of Economic Studies* 79, 539–580.

Cespa, G., & Vives, X., 2015, The Beauty Contest and Short-Term Trading, *Journal of Finance*, 70(5), 2099-2154.

Chen, Q., Goldstein, I., & Jiang, W., 2010. Payoff complementarities and financial fragility: Evidence from mutual fund outflows. *Journal of Financial Economics*, *97*(2), 239-262.

Chen, Hsiu-Lang, Narasimhan Jegadeesh, and Russ Wermers, 2000, The value of active mutual fund management: An examination of the stockholdings and trades of fund managers. *Journal of Financial and quantitative Analysis* 35.03, 343-368.

Chernenko, S., & Sunderam, A. , 2016. *Liquidity transformation in asset management: Evidence from the cash holdings of mutual funds* (No. w22391). National Bureau of Economic Research.

Christoffersen, S., Keim, D., Musto, D., 2006.Valuable information and costly liquidity: evidence from individual mutual fund trades. Working Paper, McGill University, University of Pennsylvania.

Collin-Dufresne, P., and V. Fos, 2014a, Do prices reveal the presence of informed trading? *Journal of Finance*, forthcoming.

Collin-Dufresne, P., and V. Fos, 2014b, Insider Trading, Stochastic Liquidity, and Equilibrium Prices, NBER Working Paper No.18451.

Coval, J. D., D. A. Hirshleifer, and T. Shumway, 2005, Can Individual Investors Beat the Market? HBS Finance Working Paper No. 04-025. Available at SSRN: http://dx.doi.org/10.2139/ssrn.364000.

Coval, J., Stafford, E., 2006. Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics* 86, 479–512.

Da, Z., Engelberg, J., and Gao, P., 2014. The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies*, 28(1), 1-32.

Daniel, K., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of finance*, *52*(3), 1035-1058.

Dorn, A.J., D. Dorn, and P. Sengmueller, 2014, Trading as Gambling., *Management Science*, 61(10), 2376-2393.

Dorn, D. and G. Huberman, 2005, Talk and Action: What Individual Investors Say and What They Do, *Review of Finance* 9:437-481.

Dow, J., and G. Gorton, 1994, Arbitrage Chains, *Journal of Finance* 49, 819–849.

Dow, J., and G. Gorton, 1997, Noise Trading, Delegated Portfolio Management, and Economic Welfare, *Journal of Political Economy* 105, 1024–1050.

Edelen, R., 1999. Investor flows and the assessed performance of open-end mutual funds. *Journal of Financial Economics* 53, 439–466

Edelen, R. M., and Warner, J. B., 2001, Aggregate price effects of institutional trading: a study of mutual fund flow and market returns. *Journal of Financial Economics*, 59(2), 195-220.

Edmans, A., Goldstein, I., and Jiang, W., 2012. The real effects of financial markets: The impact of prices on takeovers. *The Journal of Finance*, 67(3), 933-971.

Engelberg, J. E., & Parsons, C. A., 2011, The causal impact of media in financial markets. *Journal of Finance*, 66(1), 67-97.

Engelberg, J., C. Sasseville, and J. Williams, 2010, Market Madness: The Case of Mad Money, *Management Science*, *58*(2), 351-364.

Fama, E. F., and K. R. French, 2010, Luck versus skill in the cross-section of mutual fund returns, *Journal of Finance* 65, 1915-1947.

Foucault, T., D. Sraer, and D. Thesmar, 2011, Individual Investors and Volatility. *Journal of Finance* 66, 1369–1406.

Frazzini, A., & Lamont, O. A., 2008. Dumb money: Mutual fund flows and the cross-section of stock returns. *Journal of Financial Economics*, *88*(2), 299-322.

French, K. R., 2008. Presidential address: The cost of active investing. *The Journal of Finance*, *63*(4), 1537-1573.

Friesen, G.C., Sapp, T., 2007. Mutual fund flows and investor returns: an empirical examination of fund investor timing ability. *Journal of Banking and Finance* 31, 2796–2816.

Glaser, M., & Weber, M. (2007). Overconfidence and trading volume. *The Geneva Risk and Insurance Review*, *32*(1), 1-36.

Graham, J.R., C.R. Harvey, and H. Huang, 2009, Investor Competence, Trading Frequency, and Home Bias, *Management Science* 55:1094-1106.

Greenwood, R., and Scharfstein, D., 2013, The growth of finance. *The Journal of Economic Perspectives*, *27*(2), 3-28.

Grinblatt, M. and M. Keloharju, 2000, The Investment Behavior and Performance of Various Investor Types: A Study of Finland's Unique Data Set, *Journal of Financial Economics*, 55:43-67.

Grinblatt, M. and M. Keloharju, 2009, Sensation Seeking, Overconfidence, and Trading Activity, *Journal of Finance*, 64:549-578.

Grinblatt, M., & Titman, S. ,1989. Mutual fund performance: An analysis of quarterly portfolio holdings. *Journal of business*, 393-416.

Grossman, S. J., 1976, On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information*,* *Journal of Finance* 31, 573–585.

Grossman, S. J. and J. E. Stiglitz, 1980, On the Impossibility of Informationally Efficient Markets, *American Economic Review* 70, 393–408.

Grundy, B., and M. McNichols, 1989, Trade and Revelation of Information through Prices and Direct Disclosure, *Review of Financial Studies* 2, 495–526.

He, H. and J. Wang, 1995, Differential Information and Dynamic Behavior of Stock Trading Volume, *Review of Financial Studies* 8, 919–972.

Hellwig, M., 1980, On the Aggregation of Information in Competitive Markets, *Journal of Economic Theory* 22, 477–98.

Hirshleifer, D. A., J. N. Myers, L. A. Myers, and S. H. Teoh, 2008, Do Individual Investors Cause Post-Earnings Announcement Drift? Direct Evidence from Personal Trades, *Accounting Review* 83, 1521–1550.

Hong, H.G. and J. Yu, 2008, Gone Fishin': Seasonality in Trading Activity and Asset Prices, *Journal of Financial Markets* 12, 672–702.

Hvidkjaer, S., 2006, A Trade-based Analysis of Momentum, *Review of Financial Studies* 19, 457–491.

Hvidkjaer, S., 2008, Small Trades and the Cross-section of Stock Returns, *Review of Financial Studies* 21, 1123–1151.

Ivković, Z., J. Poterba, and S. Weisbenner, 2005, Tax-Motivated Trading by Individual Investors, *American Economic Review* 95, 1605–1630.

Kacperczyk, Marcin, and Emiliano Pagnotta, 2016, Chasing prive information, Working paper.

Kaniel, R., and Parham, R., 2017, WSJ Category Kings–The impact of media attention on consumer and mutual fund investment decisions. *Journal of Financial Economics*, 123(2), 337-356.

Kaniel, R., Saar, G. and Titman, S. 2008, Individual Investor Trading and Stock Returns, *Journal of Finance* 63, 273–310.

Kelley, E., and P. Tetlock, 2013, How Wise are Crowds? Insights from Retail Orders and Stock Returns, *Journal of Finance* 68, 1229–1265.

Kumar, A., and C. Lee, 2006, Retail Investor Sentiment and Return Comovements, *Journal of Finance* 61, 2451–2486.

Kyle, A., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1335.

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

Lee, Charles and Mark J. Ready, 1991, Inferring Trade Direction from Intraday Data, *Journal of Finance* 46, 733-46.

Lee, C., Shleifer, A., & Thaler, R. H., 1991. Investor sentiment and the closed-end fund puzzle. *The Journal of Finance*, *46*(1), 75-109.

Lou D. 2012. A Flow-Based Explanation for Return Predictability. *Review of Financial Studies* 25: 3457-3489.

Makarov, I., & Rytchkov, O., 2012, Forecasting the forecasts of others: Implications for asset pricing. *Journal of Economic Theory*, 147(3), 941-966.

Odean, T., 1999, Do Investors Trade Too Much? *American Economic Review* 89, 1279–98.

Peress, J. (2003). Wealth, information acquisition, and portfolio choice. *Review of Financial Studies*, 17(3), 879-914.

Peress, Joel and Schmidt, Daniel, 2016, Glued to the TV: Distracted Retail Investors and Stock Market Liquidity, INSEAD Working Paper.

Philippon, T., 2015, Has the US finance industry become less efficient? On the theory and measurement of financial intermediation. *The American Economic Review*, *105*(4), 1408-1438.

Rydqvist, Kristian, Spizman, Joshua, and Strebulaev, Ilya, 2014, Government policy and ownership of equity securities, . *Journal of Financial Economics,* 111(1), 70-85.

Shefrin, H.M. and M.S. Statman, 1985, The Disposition to Sell Winners too Early and Ride Losers too Long: Theory and Evidence, *Journal of Finance* 40:777-790.

Sirri, E. R., and P. Tufano. 1998. Costly search and mutual fund flows. *Journal of Finance* 53:1589.622.

Stambaugh, R. F., 2014, Investment Noise and Trends, *Journal of Finance* 69, 1415–1453.

Stein, J. C., 1987, Informational externalities and welfare-reducing speculation. *Journal of political economy*, *95*(6), 1123-1145.

Wang, J., 1993, A Model of Intertemporal Asset Prices Under Asymmetric Information, *Review of Economic Studies* 60, 249–282.

Wang, J., 1994, A model of competitive stock trading volume, *Journal of Political Economy* 102, 127–168.

Watanabe, M., 2008, Price Volatility and Investor Behavior in an Overlapping Generations Model with Information Asymmetry, *Journal of Finance* 63, 229–272.

Wermers, R., 2000, Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses, *Journal of Finance*, 55(4),1655-1703.

Yan, X. (2008) Liquidity, investment style, and the relation between fund size and fund performance, *Journal of Financial and Quantitative Analysis* 43, 741–768.

Appendix A: Are the sample trades a reasonable proxy for noise trades?

In this appendix, we check that our data display the noise trade characteristics that have been reported in the literature. Specifically, we examine whether (on average) households' trades, small TAQ trades, and mutual fund flows (a) are cross-correlated and (b) perform poorly.

a. Correlation among trades

We first check that households' and small TAQ net buys contain a common component that does not wash out in the aggregate and hence might blur the price signal (Kumar and Lee 2006, Barber et al. 2009b). We start by looking at the household data. Following Kumar and Lee, we document two related findings. First, in a stock–month panel setting, a given stock is more likely to be bought by households at times when they are buying other stocks. Second, in a household–month panel setting, a given household tends to buy stocks at times when other households are buying stocks. To establish the first result, we regress a stock's net buys (measured as turnover, the number of trades, and the number of households trading) in a given month on the average net buys across all other stocks—where this average excludes, to prevent inducing "automatic" correlation, the stock's own net buy. As in Kumar and Lee (2006), we include the contemporaneous market return as a control variable to remove the common component in investor net demand that is due to overall market movements. We proceed in a similar fashion for the second result. Namely, we run a household–month panel regression of a household's net buys of all stocks in a given month (in addition to the previous measure, we now include the number of distinct stocks bought by a household) on the average net buys across all other households (where this average again excludes the household's own net buy) and the market return. The estimation results given in Panels A and B of Table A.1 show positive and statistically significant coefficients for average net buys in both regressions and across all trading measures. These coefficients range from 0.5 to 1, which means that a one-unit increase in average net buys increases a given stock's or household's net buys by as much as one unit and by no less than half a unit.

[[INSERT Table A.1 about Here]]

We conduct a similar stock–month panel analysis using the TAQ data. As before, we regress a stock's small-trade net turnover in a given month on the average of that turnover across all other stocks (here, too, the average excludes the stock's own net turnover). Panel C of Table A.1 reports a positive and statistically significant coefficient estimate for the average small-trade net turnover. For comparison purposes, we run the same regression for large-trade net turnover. The estimated coefficient for the average large-trade net turnover is also positive and statistically significant, but its magnitude is only half of that for small trades.

Finally, we perform a related analysis using mutual fund flows. Our data allow us only to check that flows are correlated across funds and hence can be no more than indicative of a correlation between the stocks that managers trade (since we do not know which stocks are traded in response to flows). Panel D shows the results from a fund–month panel regression of net flows to an equity fund on the average aggregate net flow to all other equity funds and on the contemporaneous market return. The coefficient, which is close to 1 and statistically significant, suggests that a given fund is more likely to receive inflows at times when other funds receive inflows.

Overall, these findings confirm the existence of a strong common directional component in the trades of households, in small TAQ trades, and in mutual fund flows.

b. Performance of trades

Here we investigate the performance of households' trades, small TAQ trades, and fund flows. Following Odean (1999), we measure the post-trade–return difference between buy and sell transactions. Thus we calculate the equal-weighted average return of all buy and sell transactions over horizons of four months (84 trading days) and one year (252 trading days) subsequent to the transaction date and then take the difference. Because noise traders lose when trading against informed investors, we expect this return difference to be negative; Table A.2 confirms that expectation. Panel A shows that households' average post-trade–return difference (based on raw returns) is a marginally significant −0.5% (*t*-statistic of 1.7) after four months and a highly significant −2.6% (t = 3.9) after one year. In other words, the stocks that individuals buy earn lower returns than the stocks they sell. Results are similar when the post-trade–return difference is measured using market-adjusted returns. The values derived here are strongly similar to those reported in Odean (1999).[31]

[[INSERT Table A.2 about Here]]

Panel B of Table A.2 reports our findings from the corresponding analysis based on TAQ trades. Small trades underperform significantly at both horizons irrespective of the return adjustment. As a comparison, we report the performance of large TAQ trades, which we find to be indistinguishable from zero. Thus small trades perform poorly but large trades do not. Panel C of Table A.2 reports the findings based on retail flows to mutual funds. Here we use fund purchases/redemptions and fund returns (net of fees) in lieu of, respectively, stock buys/sells and stock returns. We find that the post-trade–return difference is not distinguishable from zero. That is, the funds that individuals purchase do not generate higher returns than the funds they redeem.

Summary: Retail investments are not profitable. Households' trades and small TAQ trades lead to losses, on average, and their mutual fund choices do not yield superior returns. Overall, these investors have not earned superior returns by selling out of their current positions to enter new ones.

[31] We reach a similar conclusion when looking at the average portfolio returns of our sample households (as in Barber and Odean 2000). Although households earn positive raw returns (thanks to the equity risk premium), they significantly underperform their own benchmark—that is, the return they would have earned had they simply held their beginning-of-the-year portfolio for the entire year.

Appendix B: Relation between the first-order autocorrelation coefficients of retail trades and noise trades

We wish to estimate the first-order autocorrelation coefficient for the noise trading process z_t. We observe the time-series process for a sample x_t of retail trades. Retail trades are an imperfect proxy for noise trades: on the one hand, some retail trades are not noise trades (they are motivated by information); on the other hand, not all noise trades are retail trades (some are institutional trades). We shall use y_{1t} to denote retail informed trades and y_{2t} to denote institutional noise trades. Given this classification, noise trades can be broken up into retail trades minus retail informed trades plus institutional noise trades:

$$z_t = x_t - y_{1t} + y_{2t} = x_t + y_t;$$

here $y_t \equiv -y_{1t} + y_{2t}$ captures the difference between noise trade and retail trades.

Suppose that x_t and y_t are both AR(1) processes governed by the respective first-order autocorrelation coefficients ρ_X and ρ_Y. Let σ_X, σ_Y, and σ_{XY} denote (respectively) the variance of x_t, the variance of y_t, and the covariance between x_t and y_t. It is then straightforward to show that z_t is also an AR(1) process and that its first-order autocorrelation coefficient ρ_Z is given by

$$\rho_Z \equiv \frac{cov(z_t, z_{t-1})}{var(z_{t-1})} = \rho_X \frac{1 + \frac{\rho_Y}{\rho_X}\frac{\sigma_Y}{\sigma_X} + \left(1 + \frac{\rho_Y}{\rho_X}\right)\frac{\sigma_{XY}}{\sigma_X}}{1 + \frac{\sigma_Y}{\sigma_X} + 2\frac{\sigma_{XY}}{\sigma_X}}.$$

We now list sufficient conditions for the first-order autocorrelation coefficient of the retail trading process (which is what we observe) to be a good proxy for the first-order autocorrelation coefficient of the noise trading process (what we wish to infer)—that is, sufficient conditions for $\rho_X \approx \rho_Z$ to hold.

1. $\sigma_X \gg \sigma_Y$ and $\sigma_X \gg \sigma_{XY}$. In words: The difference between the component of noise trades that we are missing (institutional noise trades) and the component of retail trades that are not noise (informed retail trades) is small relative to retail noise trades. That is, "what we observe swamps what we don't observe."

2. $\rho_X \approx \rho_Y$. In words: The difference between the component of noise trades that we are missing (institutional noise trades) and the component of retail trades that are not noise (informed retail trades) has the same first-order autocorrelation coefficient as retail noise trades. That is, "what we don't observe behaves similarly to what we do observe."

3. $\sigma_Y \approx -\sigma_{XY}$. In words: The difference between the component of noise trades that we are missing (institutional noise trades) and the component of retail trades that are not noise (informed retail trades) affects the autocovariance and the variance of noise trades (respectively, $\text{cov}(z_t, z_{t-1})$ and $\text{var}(z_{t-1})$) in the same way and so "washes out".

Appendix C: Coefficient estimate derived from regressing total trading volume on retail trading volume

Here we prove Section 5's claim that the coefficient $\hat{b}$—obtained by regressing total trading volume on households' trading volume $|\Delta\hat{\theta}_t|$—is in fact given by $\hat{b} = b(1 + \sqrt{1+r} - \sqrt{r})/2$, where b represents the ratio of total noise trades in the market to households' trades and where r is a positive scalar.

The regression coefficient is given by $\hat{b} \equiv \text{cov}(\text{Total volume}_t,\ |\Delta\hat{\theta}_t|)/\text{var}(\ |\Delta\hat{\theta}_t|)$ Here the expression for total trading volume depends on the market structure, as described next.

In a *competitive* market, $\text{Total volume}_t = \frac{1}{2}\left(|b\Delta\hat{\theta}_t| + \int_i\ |-b\Delta\hat{\theta}_t + \Delta x_t^i|\right)$. In computing $\hat{b}$ we note that, for two jointly normal random variables z and ε and a scalar *a*,

$$\text{cov}(|z|, |az+\varepsilon|) = \left(1 - \tfrac{2}{\pi}\right)\left(1 - \sqrt{1 - \text{corr}^2(z, az+\varepsilon)}\right)\sqrt{\text{var}(z)}\sqrt{\text{var}(az+\varepsilon)};$$

see Wang (1994, Apx. B). If z and ε are uncorrelated then $\text{var}(az+\varepsilon) = a^2\,\text{var}(z) + \text{var}(\varepsilon)$ and $\text{corr}(z, az+\varepsilon) = \frac{a\,\text{var}(z)}{\sqrt{\text{var}(z)\,\text{var}(az+\varepsilon)}} = \frac{a\sqrt{\text{var}(z)}}{\sqrt{a^2\,\text{var}(z)+\text{var}(\varepsilon)}}$, from which it follows that

$$\text{cov}(|z|, |az+\varepsilon|) = \left(1 - \tfrac{2}{\pi}\right)\left(1 - \sqrt{\tfrac{\text{var}(\varepsilon)}{a^2\,\text{var}(z)+\text{var}(\varepsilon)}}\right)\sqrt{\text{var}(z)}\sqrt{a^2\,\text{var}(z) + \text{var}(\varepsilon)}. \qquad (*)$$

The denominator of $\hat{b}$ can be computed by setting $z = \Delta\hat{\theta}_t$ and $a = 1$ in equation (*):

$$\text{var}(\ |\Delta\hat{\theta}_t|) = \left(1 - \tfrac{2}{\pi}\right)\text{var}(\Delta\hat{\theta}_t).$$

Turning now to the numerator of $\hat{b}$, we can substitute $z = \Delta\hat{\theta}_t$, $a = -b$, and $\varepsilon = \Delta x_t^i$ into (*), re-arrange the expression, and then sum over all rational agents *i*. The result is

$$\text{cov}(\text{Total volume}_t, |\Delta\hat{\theta}_t|) = \tfrac{1}{2}b\left(1 - \tfrac{2}{\pi}\right)\text{var}(\Delta\hat{\theta}_t)(1 + \sqrt{1+r} - \sqrt{r}),$$

where $r \equiv \frac{\text{var}(\Delta x_t^i)}{\text{var}(\Delta\theta_t)} = \frac{\text{var}(\Delta x_t^i)}{b^2\,\text{var}(\Delta\hat{\theta}_t)}$ is positive. Therefore, the coefficient from regressing total trading volume on $|\Delta\hat{\theta}_t|$ is $\hat{b} = b(1 + \sqrt{1+r} - \sqrt{r})/2$.

In a *strategic* market, $\text{Total volume}_t = \frac{1}{2}\left(|b\Delta\hat{\theta}_t| + |-b\Delta\hat{\theta}_t + \Delta x_t| + |-\Delta x_t|\right)$. Proceeding as in the competitive case and noting that Δx_t is uncorrelated with $\Delta\hat{\theta}_t$, we obtain

$$\text{cov}(\text{Total volume}_t, |\Delta\hat{\theta}_t|) = \tfrac{1}{2}b\left(1 - \tfrac{2}{\pi}\right)\text{var}(\Delta\hat{\theta}_t)(1 + \sqrt{1+r} - \sqrt{r}),$$

where $r \equiv \frac{\text{var}(\Delta x_t)}{\text{var}(\Delta\theta_t)} = \frac{\text{var}(\Delta x_t)}{b^2\,\text{var}(\Delta\hat{\theta}_t)}$. As a result, $\hat{b} = b(1 + \sqrt{1+r} - \sqrt{r})/2$.

Figure 1: Time series of households' monthly aggregate trades

The graphs in this figure display time series of households' monthly aggregate trades at a large discount broker (from January 1991 through November 1996), of small TAQ trades (from January 1991 through December 2000) and of monthly flows to equity mutual funds (from January 1999 through August 2013). The household data are for those holding common stock positions for 71 consecutive months. The TAQ data include all small trades in NYSE/AMEX stocks, where trades are classified as small or large based on the procedure described in Hvidkjaer (2006); see Section 1.b. Mutual funds data include all retail equity mutual funds reported in the TrimTabs data set. The top panel considers households' net turnover: the aggregate value of their buys, minus the aggregate value of their sells, divided by the market's total value. The second panel considers the net number of households buying shares: the number of households buying minus the number of households selling. The third panel considers the net turnover for small trades in the TAQ data set: the aggregate value of small buys, minus the aggregate value of small sells, divided by the market's total value. The bottom panel considers the net turnover for flows to equity mutual funds in the TrimTabs data set: the aggregate value of purchases, minus redemptions, of retail equity mutual funds divided by these funds' aggregate total net assets. We adjust all variables for seasonality and time trends by regressing them on dummy variables for month of the year, and year, and then taking the residuals.

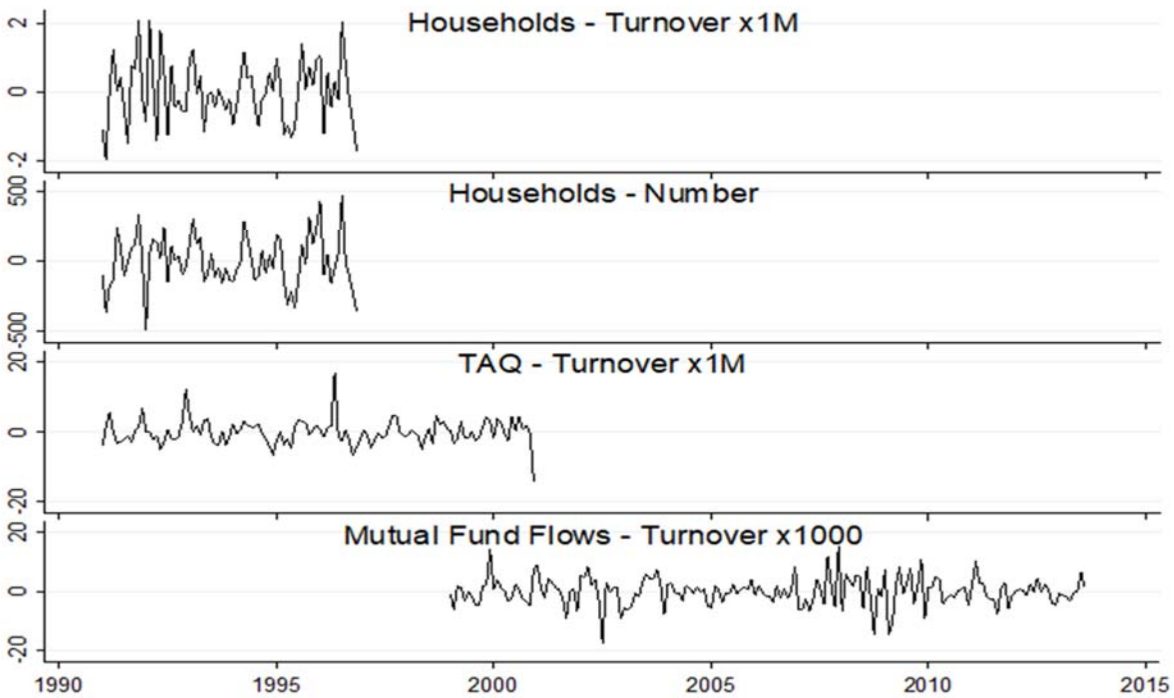
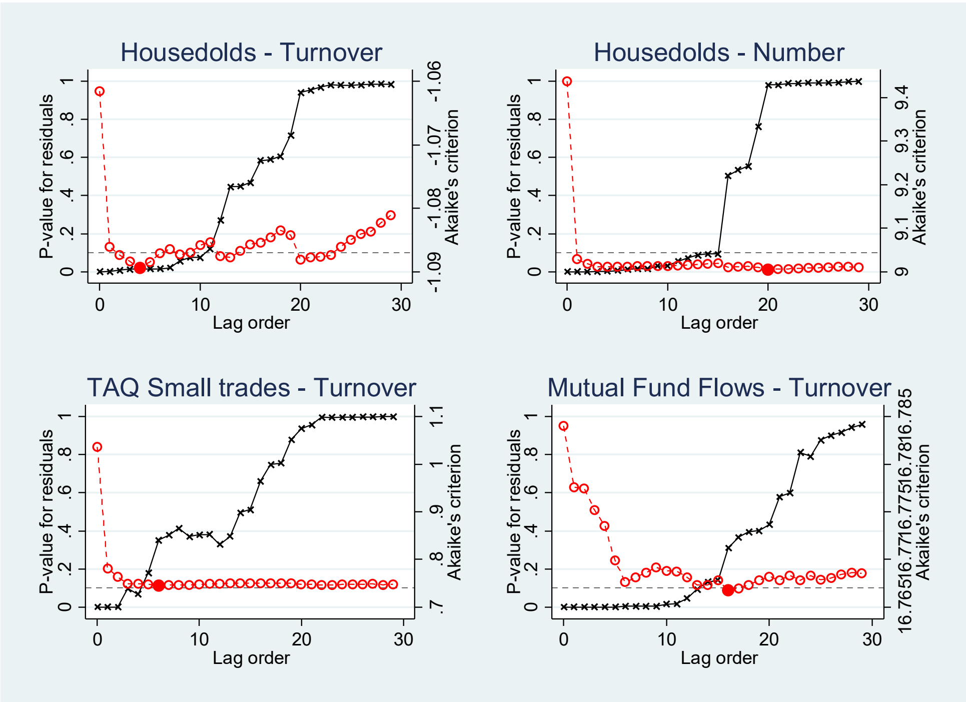


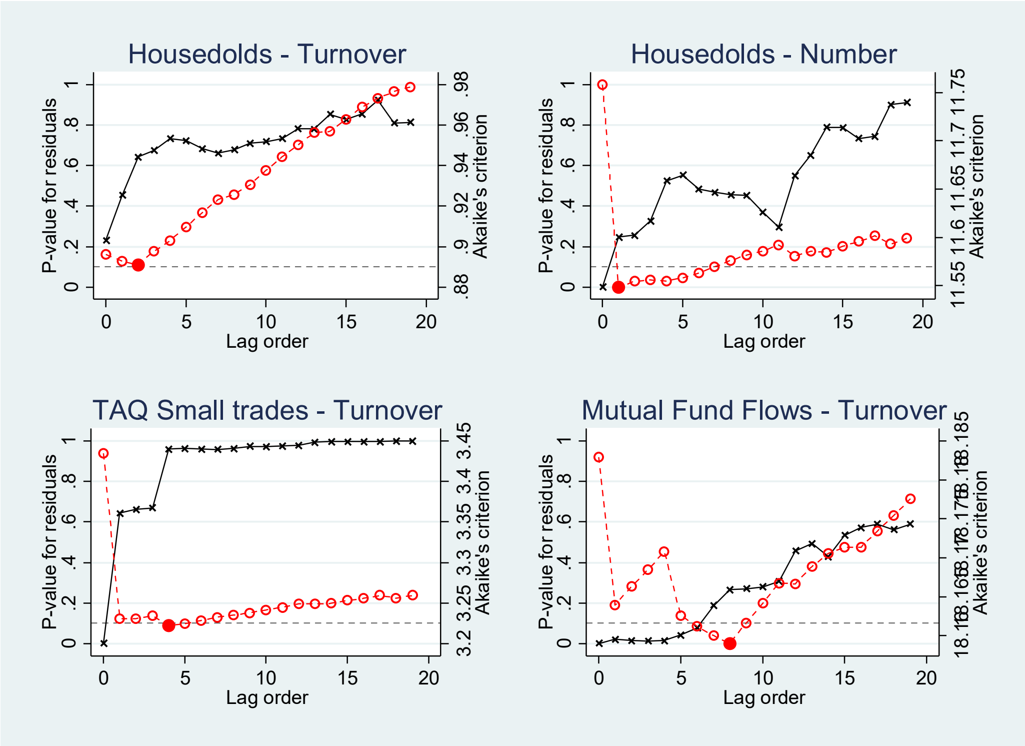
Figure 2: Lag-order selection

This figure displays the performance of autoregressive models fitted to households' aggregate trades (top left and right panels), to TAQ small trades (bottom left panel), and to mutual fund flows (bottom right panel) as a function of the number of lags. The number of lags ranges from 0 to 30, 0 to 20, and 0 to 10 at (respectively) daily (Panel A), weekly (Panel B), and monthly (Panel C) frequencies. The graphs' crosses and left axes mark *p*-values of a white-noise *Q*-test for residuals of the fitted data. High *p*-values indicate that we cannot reject the null hypothesis of the residuals being serially uncorrelated. The horizontal dashed line marks the 10% significance level. The circles and right axes mark the value of Akaike's information criterion, where lower values correspond to better models; a solid circle marks the lag order that this criterion deems optimal. We adjust all variables for seasonality and time trends by regressing them on day-of-the-week (as applies), month-of-the-year, and year dummy variables and then taking the residuals.

Panel A: Daily frequency



Panel B: Weekly frequency



Panel C: Monthly frequency

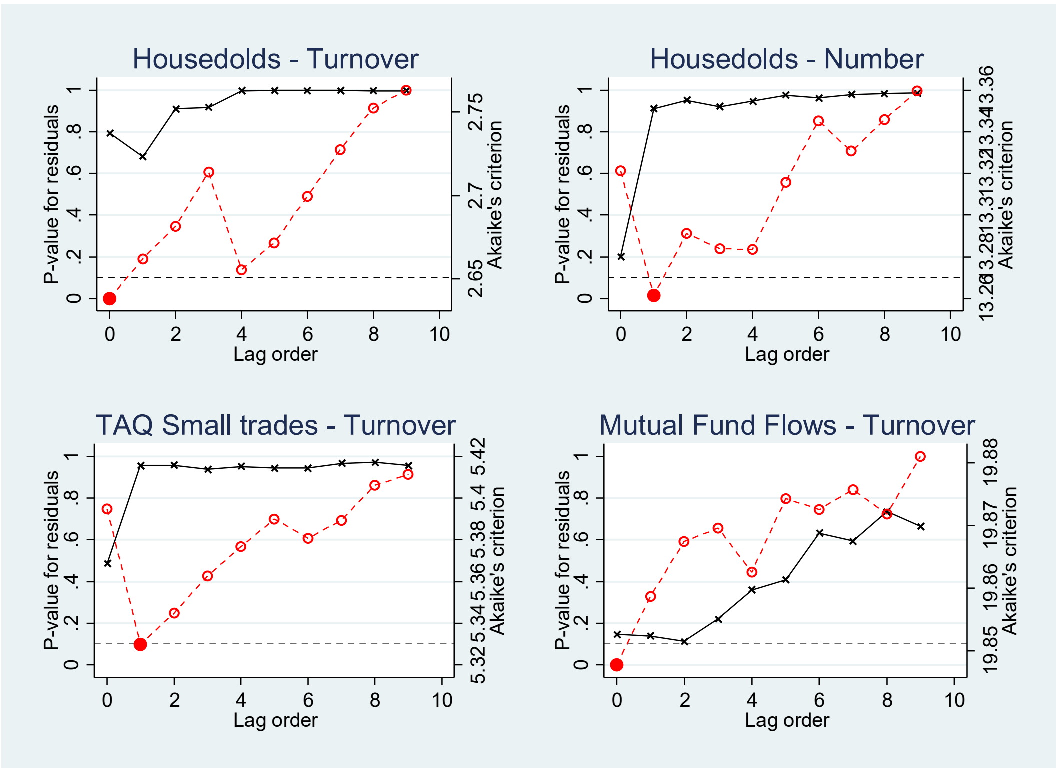


Figure 3: Fitting an AR(1) process to households' aggregate trades

The graphs in this figure plot the first-order autocorrelation coefficient of aggregate trades as a function of the duration of a time period in days. Solid circles mark coefficients that are statistically significant at the 10% level. The upper left panel considers households' net turnover: the aggregate value of their buys, minus the aggregate value of their sells, divided by the market's total value. The upper right panel considers the net number of households buying shares: the number of households buying minus the number of households selling. The lower left panel considers the net turnover for small trades in the TAQ data set: the aggregate value of small buys, minus the aggregate value of small sells, divided by the market's total value. The TAQ data include all small trades in NYSE/AMEX stocks, where trades are classified as small or large based on the procedure described in Hvidkjaer (2006); see Section 1.b. The lower right panel considers the net turnover for flows to retail equity mutual funds in the TrimTabs data set: the aggregate value of purchases, minus redemptions, of equity mutual funds divided by these funds' aggregate TNA. We adjust all variables for seasonality and time trends by regressing them on dummies for month of the year and year, and then taking the residuals.

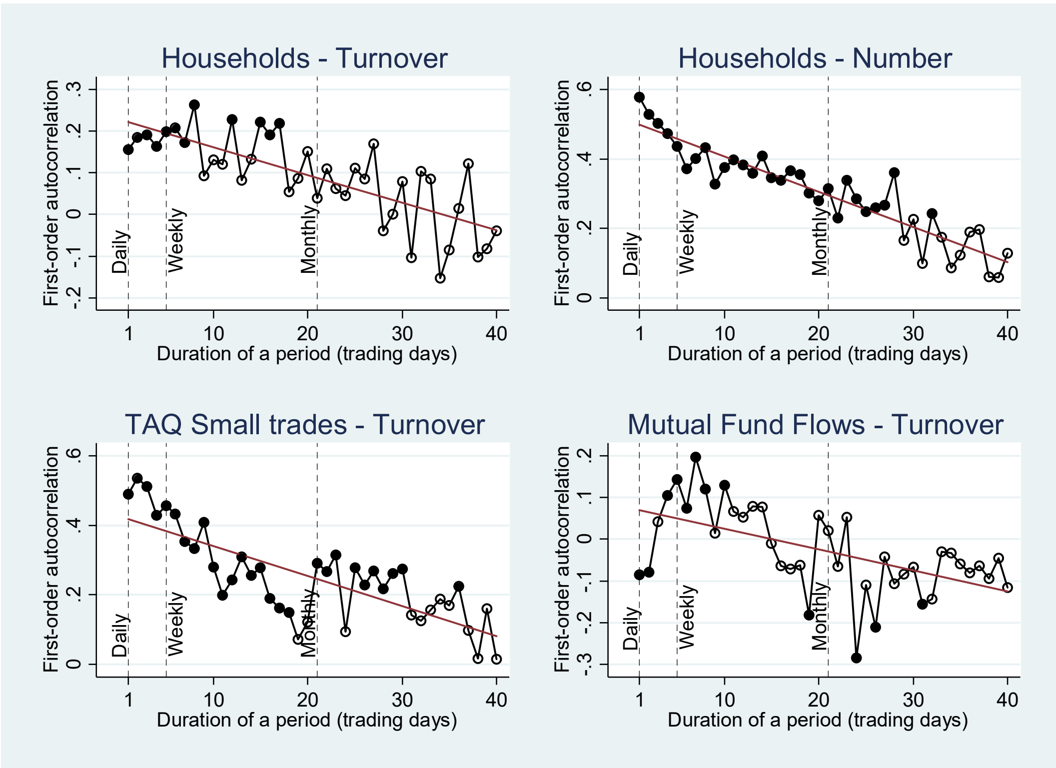


Figure 4: Histograms of households' daily aggregate trades

The graphs in this figure are histograms of aggregate trades. The left, middle, and right columns consider (respectively) daily, weekly, and monthly data. The top row considers households' net turnover: the aggregate value of their buys, minus the aggregate value of their sells, divided by the total value of the market. The second row considers the net number of households buying shares: the number of households buying minus the number of households selling. The third row considers the net turnover for small trades in the TAQ data set: the aggregate value of small buys, minus the aggregate value of small sells, divided by the market's total value. The TAQ data include all small trades in NYSE/AMEX stocks, where trades are classified as small or large based on the procedure described in Hvidkjaer (2006); see Section 1.b. The bottom row considers the net turnover for flows to retail equity mutual funds in the TrimTabs data set: the aggregate value of purchases, minus redemptions, of equity mutual funds divided by these funds' aggregate TNA. We adjust all variables for seasonality and time trends by regressing them on dummy variables for day of the week, month of the year, and year, and then taking the residuals.

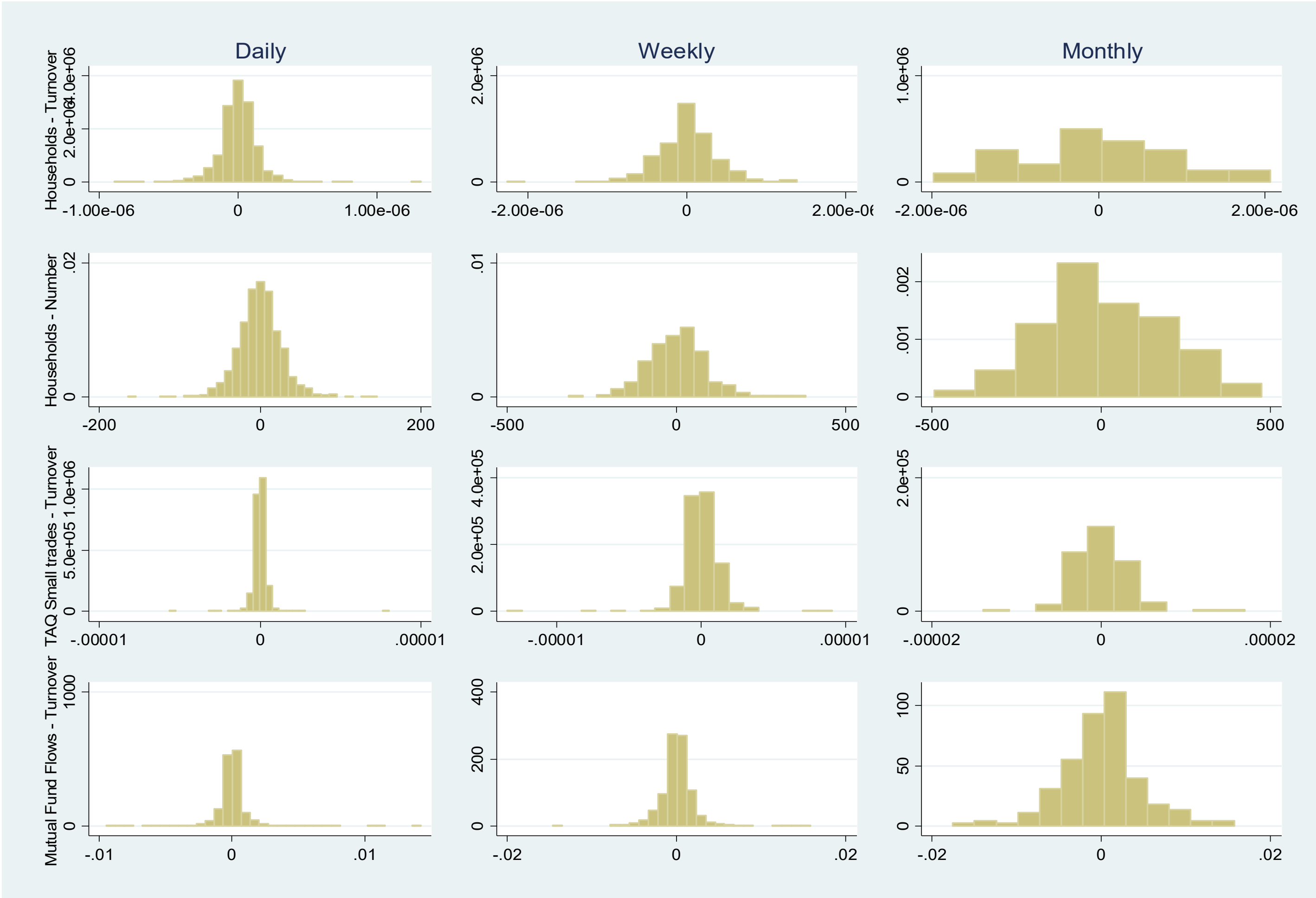


Figure 5: Probability (Q-Q) plots of households' aggregate trades

The graphs in this figure plot quantiles of households' aggregate trades against quantiles of a normal distribution at various frequencies. The left, middle, and right columns consider (respectively) daily, weekly, and monthly data. The top row considers households' net turnover: the aggregate value of their buys, minus the aggregate value of their sells, divided by the market's total value. The second row considers the net number of households buying shares: the number of households buying minus the number of households selling. The third row considers the net turnover for small trades in the TAQ data set: the aggregate value of small buys, minus the aggregate value of small sells, divided by the total value of the market. The TAQ data include all small trades in NYSE/AMEX stocks, where trades are classified as small or large based on the procedure described in Hvidkjaer (2006); see Section 1.b. The bottom row considers the net turnover for flows to retail equity mutual funds in the TrimTabs data set: the aggregate value of purchases, minus redemptions, of equity mutual funds divided by these funds' aggregate TNA. We adjust all variables for seasonality and time trends by regressing them on day-of-the-week, month-of-the-year, and year dummy variables, and then taking the residuals.

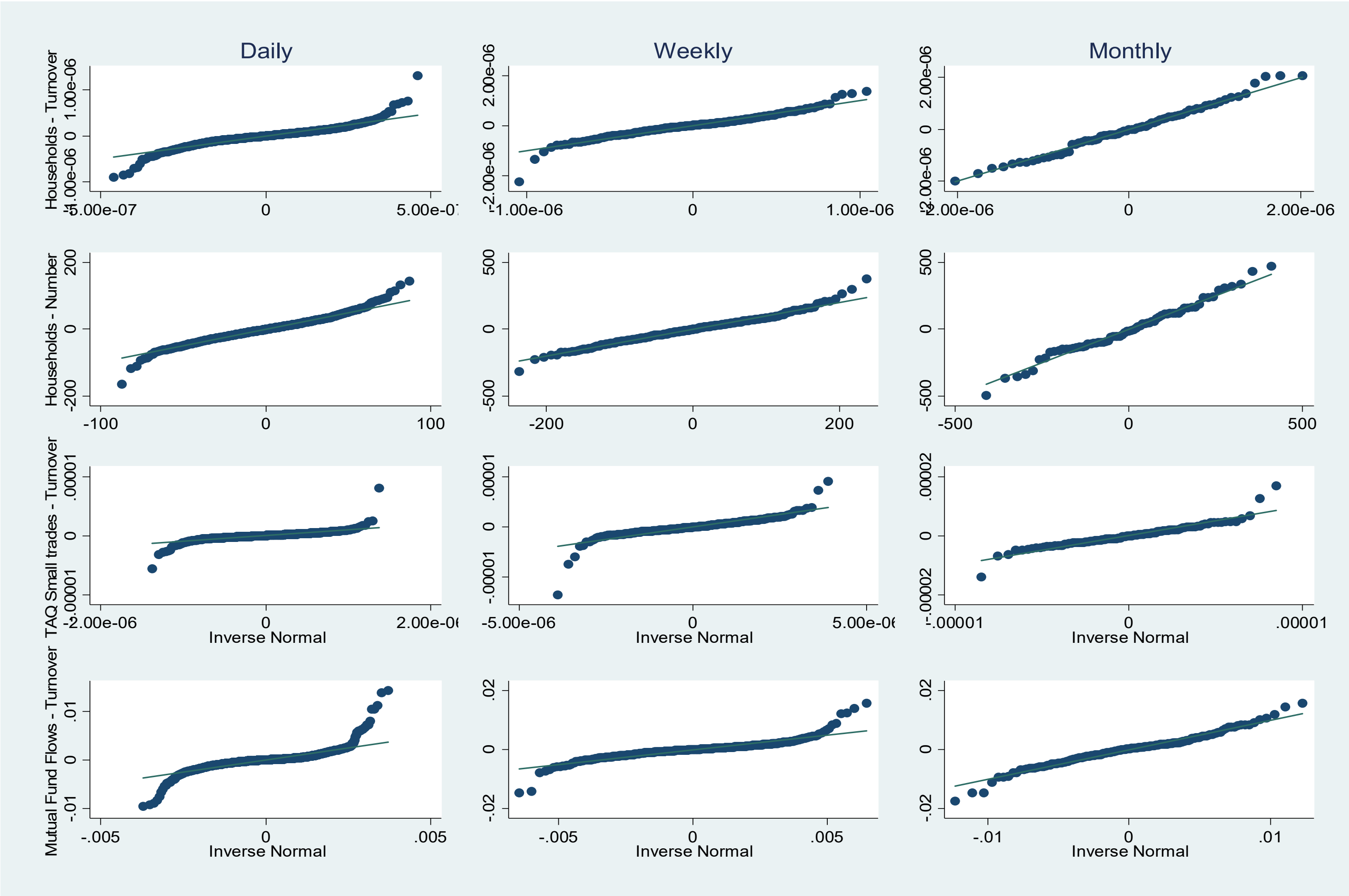
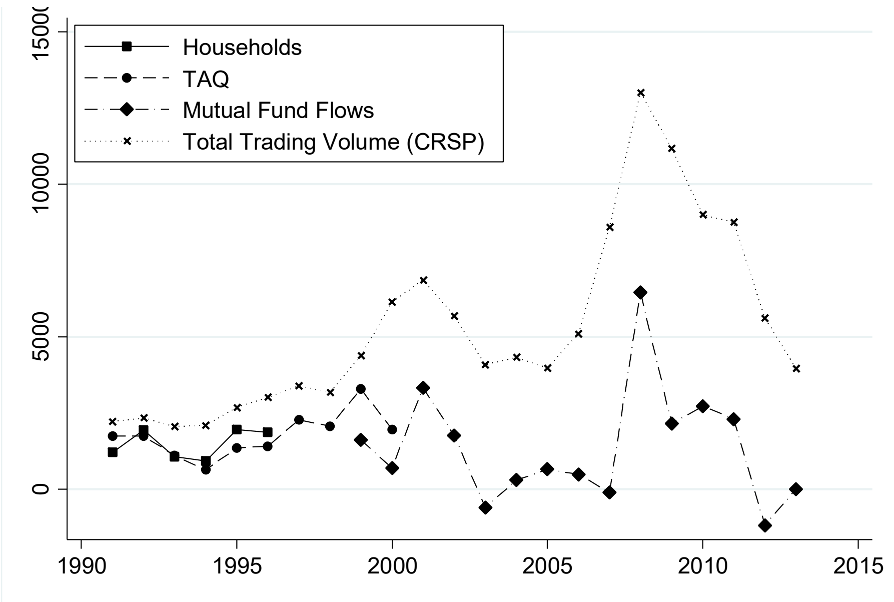


Figure 6: Estimating the intensity of noise trading for stock groups

This figure plots the lower bound on the standard deviation of noise trades measured over time. The bound is estimated using households' trades (solid lines, square markers), small TAQ trades (dashed lines, circles) and mutual fund flows and small TAQ trades (dot-dashed lines, diamonds). Each year, we regress weekly total turnover (CRSP trading volume divided by the market's total value) on the weekly retail turnover (the sum of buys and sells divided by the market's total value) as measured using households' trades, small TAQ trades, and mutual fund flows. The lower bound on the standard deviation of noise trading in any year is given by the time-series standard deviation of turnover in that decile multiplied by the regression coefficient; the upper bound (not shown) is equal to twice the lower bound. The bound is displayed in terms of levels in Panel A and also, in Panel B, as a fraction of the standard deviation of total turnover in the market.

Panel A: Standard deviation of noise trading



Panel B: Ratio of the standard deviation of noise trading to the standard deviation the total turnover in the market

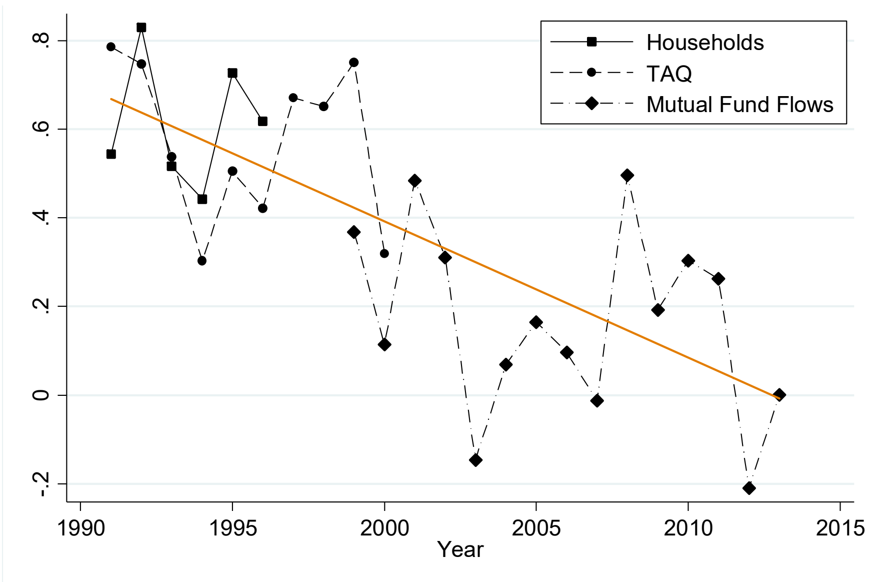


Figure 7: Estimating the intensity of noise trading for stock groups

The graphs in this figure plot the lower bound on the standard deviation of noise trades, across stock characteristic deciles, as measured using households' trades (solid lines, square markers) and small TAQ trades (dashed lines, circles). For each month, we sort stocks into deciles according to their capitalization, share price, turnover, closing bid–ask spread, Amihud illiquidity ratio, return volatility, and return autocovariance. All variables are estimated every month from daily observations. For a stock's capitalization, price, bid–ask spread, and turnover we use its respective monthly average. The Amihud illiquidity ratio is the monthly average of the daily ratio of the stock's absolute return to its dollar trading volume. The return volatility is the standard deviation of the stock's daily raw return over a month, and the return autocovariance is the autocovariance of the stock's daily returns over a month. Then, decile by decile, we regress daily total turnover (CRSP trading volume divided by the market's total value) on the daily retail turnover (the sum of buys and sells divided by the market's total value) as measured using households' trades and small TAQ trades. The lower bound on the standard deviation of noise trading in any decile is given by the time-series standard deviation of turnover in that decile multiplied by the regression coefficient; the upper bound (not shown) is equal to twice the lower bound. All variables are adjusted for seasonality and time trends before running the regressions.

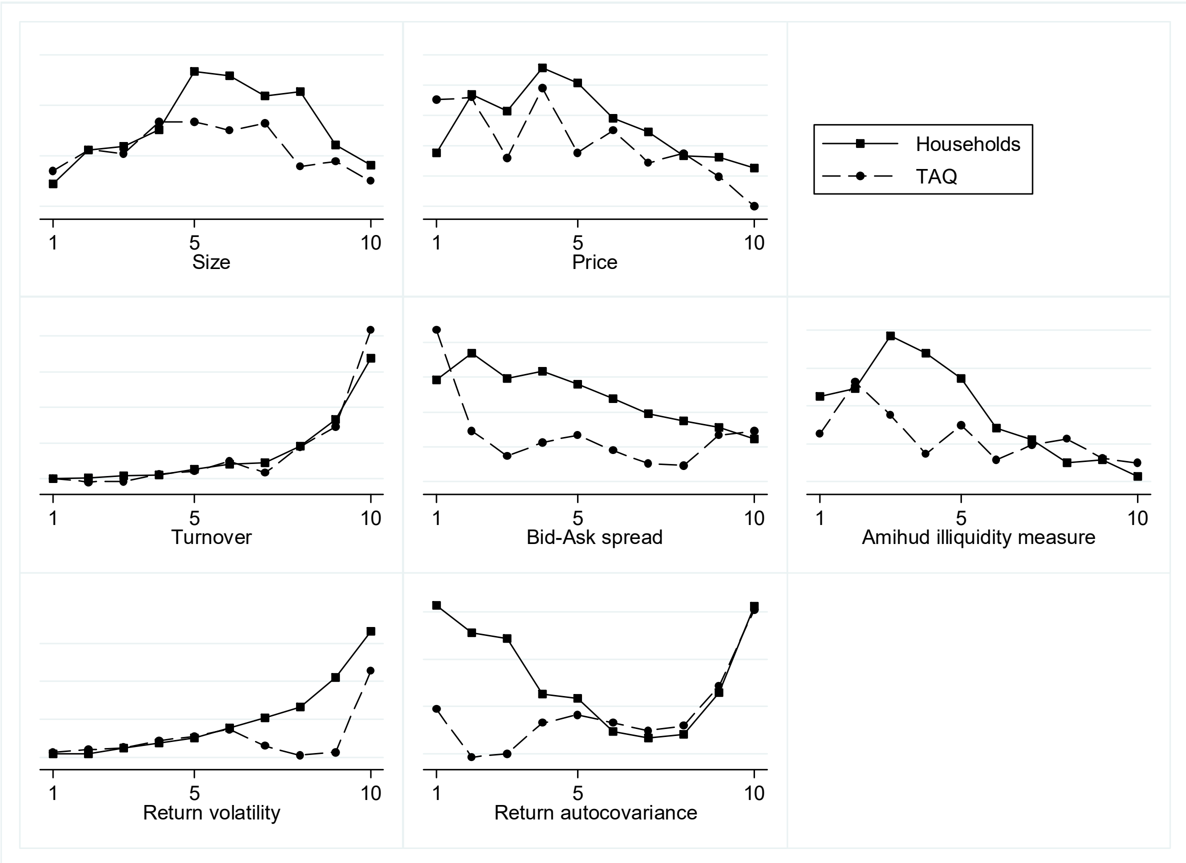


Table 1: Descriptive statistics for aggregate trades

This table presents summary statistics for the time series of households' daily aggregate trades at a large discount broker (from January 1991 through November 1996), for the time series of TAQ trades (from January 1991 through December 2000), and for daily flows to retail equity mutual funds (from January 1999 through August 2013). The household data are for those holding common stock positions for 71 consecutive months. The TAQ data include all small and large trades in NYSE/AMEX stocks, where trades are classified in terms of size based on the procedure described in Hvidkjaer (2006); see Section 1.b. Mutual funds data include all retail equity mutual funds reported in the TrimTabs data set. Panel A considers households' net turnover: the aggregate value of their buys, minus the aggregate value of their sells, divided by the market's total value (in millions). Panel B considers the net number of households buying shares: the number of households buying minus the number of households selling. Panel C considers the net turnover for small trades in the TAQ data set: the aggregate value of small buys, minus the aggregate value of small sells, divided by the market's total value (in millions of dollars); Panel D does likewise for large TAQ trades. Panel E considers the net turnover for flows to retail equity mutual funds in the TrimTabs data set: the aggregate value of purchases, minus redemptions, of equity mutual funds divided by these funds' aggregate total net assets. We adjust all variables for seasonality and time trends by regressing them on dummy variables for day of the week, month of the year, and year, and then taking the residuals.

Frequency	Obs.	min.	mean	median	max.	std. dev.	skewness	kurtosis
	Panel A: Households - Turnover x 1M							
Daily	1497	-0.890	0.000	0.003	1.319	0.143	0.111	12.850
Weekly	309	-2.260	0.000	0.007	1.382	0.382	-0.421	7.750
Monthly	71	-1.979	0.000	0.005	2.069	0.919	0.183	2.693
Number of firms : 9,158								
	Panel B: Households - Number							
Daily	1497	-164.540	0.000	-0.247	145.291	27.130	0.135	5.700
Weekly	309	-318.232	0.000	-0.216	380.303	87.002	0.281	4.677
Monthly	71	-493.814	0.000	-14.064	473.152	186.214	0.100	3.223
Number of firms : 9,158								
	Panel C: TAQ Small trades - Turnover x 1M							
Daily	2526	-5.636	0.000	0.007	8.006	0.409	1.361	80.450
Weekly	522	-13.445	0.000	-0.032	9.052	1.346	-1.402	28.643
Monthly	120	-13.988	0.000	-0.093	16.983	3.539	0.762	8.769
Number of firms : 11,828								
	Panel D: TAQ Large trades - Turnover x 1M							
Daily	2526	-22.136	0.000	0.155	31.857	4.176	-0.078	6.229
Weekly	522	-43.300	0.000	0.448	38.708	12.232	-0.226	3.241
Monthly	120	-76.881	0.000	-0.068	70.290	30.073	-0.145	2.774
Number of firms : 11,790								
	Panel E: Mutual Fund Flows - Turnover x 1,000							
Daily	3662	-9.468	0.000	0.010	14.221	1.072	1.799	38.515
Weekly	764	-14.609	0.000	0.021	15.820	2.147	0.581	17.405
Monthly	176	-17.537	0.000	0.232	15.675	4.857	-0.131	4.697
Number of funds : 2,453								

Table 2: Shapiro–Wilk test for normality

This table reports results of a Shapiro–Wilk test that households' aggregate trades, small TAQ trades, mutual fund flows, and their residuals from a fitted AR(1) process are normally distributed at daily, weekly, and monthly frequencies. The null hypothesis is that these series are normal, and the alternative is that they are not normal. The top panel considers households' net turnover: the aggregate value of their buys, minus the aggregate value of their sells, divided by the market's total value. The second panel considers the net number of households buying shares: the number of households buying minus the number of households selling. The third panel considers the net turnover for small trades in the TAQ data set: the aggregate value of small buys, minus the aggregate value of small sells, divided by the market's total value. The TAQ data include all small trades in NYSE/AMEX stocks, where trades are classified as small or large based on the procedure described in Hvidkjaer (2006); see Section 1.b. The bottom panel considers the net turnover for flows to equity mutual funds in the TrimTabs data set: the aggregate value of purchases, minus redemptions, of equity mutual funds divided by these funds' aggregate TNA. We adjust all variables for seasonality and time trends by regressing them on day-of-the-week, month-of-the-year, and year dummy variables and then taking the residuals.

	Variables		Residuals from fitted AR(1)	
	Test Statistic	p-value	Test Statistic	p-value
	Households - Turnover		Households - Turnover	
Daily	11.213	0.000	11.077	0.000
Weekly	5.706	0.000	5.914	0.000
Monthly	-0.363	0.642	0.044	0.482
Quarterly	0.528	0.299	0.106	0.458
	Households - Number		Households - Number	
Daily	7.735	0.000	7.124	0.000
Weekly	3.105	0.001	1.618	0.053
Monthly	-0.440	0.670	-2.045	0.980
Quarterly	0.220	0.413	-0.080	0.532
	TAQ Small trades - Turnover		TAQ Small trades - Turnover	
Daily	14.899	0.000	14.799	0.000
Weekly	10.148	0.000	9.781	0.000
Monthly	4.762	0.000	4.937	0.000
Quarterly	-0.699	0.758	-0.649	0.742
	Mutual Fund Flows - Turnover		Mutual Fund Flows - Turnover	
Daily	16.393	0.000	16.354	0.000
Weekly	10.799	0.000	10.726	0.000
Monthly	2.908	0.002	2.875	0.002
Quarterly	-0.521	0.699	0.233	0.408

Table 3: Estimating the intensity of noise trading

We regress total turnover (CRSP trading volume divided by the market's total value) on retail turnover (sum of buys and sells divided by the market's total value) as measured using households, small TAQ trades, and mutual fund flows over different frequencies. All variables are first adjusted for seasonality and time trends by regressing them on dummy variables for day of the week, month of the year, and year and then taking residuals. The regression coefficient $\hat{b}$ determines bounds on the fraction of noise turnover in our sample that is due to retail traders. The standard deviation of noise trading is bounded from below by the standard deviation of retail trades multiplied by the regression coefficient $\hat{b}$; it is bounded from above by twice that product. These bounds are displayed in terms of levels and also as a fraction of the standard deviation of total turnover in the market. We use * and *** to indicate statistical significance at (respectively) the 10%, and 1% level.

Frequency	Std. dev. of retail turnover (x1million)	Regression coefficient $\hat{b}$		Std. dev. of noise trading: Lower bound (x1,000)	Lower bound % of std. dev. of total turnover	Upper bound (x1,000)	Upper bound % of std. dev. of total turnover
Households, 1991-1996							
Day	0.222	1,293	***	0.287	38%	0.574	76%
Week	0.741	2,040	***	1.512	44%	3.024	88%
Month	2.232	2,057	***	4.591	37%	9.182	74%
Quarter	5.602	1,535		8.598	25%	17.195	50%
TAQ small trades							
1991-2000							
Day	1.601	187.0	***	0.299	19%	0.599	38%
Week	7.305	228.7	***	1.671	23%	3.342	45%
Month	25.630	155.9		3.995	13%	7.989	26%
Quarter	56.471	173.7		9.806	11%	19.613	22%
1991-1995							
Day	0.949	255.3	***	0.242	36%	0.485	72%
Week	4.445	296.0	***	1.316	44%	2.631	87%
Month	14.997	149.5	*	2.242	22%	4.484	43%
Quarter	35.139	29.1		1.023	4%	2.046	7%
1996-2000							
Day	2.055	172.4	***	0.354	23%	0.709	46%
Week	9.337	213.5	***	1.993	27%	3.987	55%
Month	33.167	157.2		5.212	18%	10.425	37%
Quarter	72.876	207.3		15.104	18%	30.208	37%
Mutual fund flows, 1999-2013							
Day	900.357	0.465	***	0.419	18%	0.838	36%
Week	1687.236	1.917	***	3.235	29%	6.469	58%
Month	3314.858	3.079	***	10.208	24%	20.415	48%
Quarter	5225.371	2.252		11.765	10%	23.530	20%

Table 4: Estimating the intensity of noise trading for stock groups

The seven panels in this table report bounds on the standard deviation of noise trades across groups of stocks sorted on various characteristics. For each month, we sort stocks into deciles according to their capitalization (Panel A), turnover (Panel B), share price (Panel C), closing bid–ask spread (Panel D), Amihud illiquidity ratio (Panel E), return volatility (Panel F), and return autocovariance (Panel G). All variables are estimated every month from daily observations. For a stock's capitalization, turnover, price, and bid–ask spread we use its respective monthly average. The Amihud illiquidity ratio is the monthly average of the daily ratio of the stock's absolute return to its dollar trading volume. The return volatility is the standard deviation of the stock's daily raw returns over a month, and the return autocovariance is the autocovariance of the stock's daily returns over a month. Then, decile by decile, we regress total turnover (CRSP trading volume divided by the market's total value) on the turnover (sum of buys and sells divided by the market's total value) as measured using households' trades (columns 2–6) or small TAQ trades (columns 7–11). We use $\hat{b}^k$ to denote the regression coefficient. All variables are first adjusted for seasonality and time trends by regressing them on day-of-the-week, month-of-the-year, and year dummy variables and then taking residuals. The standard deviation of noise trading in decile k is bounded from below by the time-series standard deviation of trades in that decile multiplied by $\hat{b}^k$—and from above by twice that product. The bounds on the standard deviation of noise trades are reported in terms of levels and also as a fraction of the standard deviation of total trades in a decile. Negative estimates of the noise trading intensity correspond to statistically insignificant estimates of the slope coefficient.

Panel A: By stock size

	Households - Turnover x 1M					TAQ Small trades - Turnover x 1M				
				Lower bound on SD of noise trading					Lower bound on SD of noise trading	
Decile	Median stock size ($M)	SD of trades (x1M)	Regression coef. $\hat{b}$	(x1M)	(% of SD of total trades)	Median stock size ($M)	SD of trades (x1M)	Regression coef. $\hat{b}$	(x1M)	(% of SD of total trades)
	Daily frequency									
1	5.2	4.91	20.01	98.26	8%	5.6	6.74	120.51	812.79	36%
2	12.0	1.41	130.34	184.15	15%	14.3	2.38	345.06	820.74	45%
3	20.9	1.63	77.83	127.12	11%	25.9	1.28	433.04	554.06	38%
4	34.5	1.28	208.52	266.81	22%	43.2	0.94	732.30	686.24	38%
5	54.6	1.56	241.86	377.01	29%	70.0	0.59	1095.63	649.12	39%
6	87.9	1.04	398.08	414.89	31%	114.0	0.28	2144.96	593.98	37%
7	148.0	0.85	485.63	413.96	31%	189.0	0.29	1283.10	366.68	25%
8	281.0	0.73	644.09	473.02	38%	351.0	0.08	5032.34	408.81	32%
9	710.0	0.43	802.86	341.44	33%	823.0	0.04	12232.32	439.25	42%
10	6310.0	0.22	1180.89	259.00	34%	3380.0	0.01	43194.94	354.21	49%
	Weekly frequency									
1	5.1	9.70	70.28	681.79	12%	5.5	27.56	105.20	2898.89	34%
2	12.1	3.74	359.65	1343.88	24%	14.2	9.40	292.90	2753.29	40%
3	21.0	3.97	259.75	1032.03	18%	26.0	5.24	424.23	2224.24	37%
4	34.3	3.56	547.73	1951.30	35%	43.2	4.17	632.55	2638.24	36%
5	54.9	4.36	706.40	3080.50	51%	69.9	2.64	1046.57	2760.24	39%
6	88.2	3.14	974.82	3059.59	49%	114.0	1.20	2166.85	2602.18	37%
7	148.0	2.57	1157.73	2972.00	50%	189.0	1.10	1779.45	1955.32	30%
8	281.0	2.38	1299.69	3086.94	54%	351.0	0.38	4691.16	1768.56	31%
9	707.0	1.24	1636.46	2035.93	44%	823.0	0.17	12206.56	2092.62	44%
10	6260.0	0.68	1985.80	1344.30	40%	3380.0	0.01	43194.94	354.21	49%
	Monthly frequency									
1	5.2	19.28	114.41	2206.13	9%	5.5	89.87	39.01	3505.54	15%
2	12.4	10.52	529.40	5571.40	24%	14.2	31.33	178.36	5587.48	26%
3	21.7	9.50	626.91	5953.41	27%	26.0	18.74	279.60	5240.15	25%
4	34.6	10.10	749.45	7570.97	35%	43.3	15.52	537.38	8340.77	31%
5	54.9	12.15	1099.31	13357.78	57%	70.0	9.45	889.60	8405.68	33%
6	89.2	9.52	1355.68	12908.08	56%	114.0	4.32	1740.61	7521.31	29%
7	147.0	7.42	1472.93	10923.42	50%	189.0	3.27	2522.10	8254.33	36%
8	281.0	7.42	1532.77	11367.09	56%	351.0	1.38	2874.97	3963.37	20%
9	712.0	3.62	1680.30	6077.67	39%	825.0	0.64	7048.16	4480.06	33%
10	6310.0	1.86	2197.11	4092.48	34%	3390.0	0.14	17798.83	2495.94	30%

Panel B: By stock turnover

Decile	Households - Turnover x 1M					TAQ Small trades - Turnover x 1M				
	Median turnover	SD of trades (x1M)	Regression coef. b	Lower bound on SD of noise trading (x1M)	Lower bound on SD of noise trading (% of SD of total trades)	Median turnover	SD of trades (x1M)	Regression coef. b	Lower bound on SD of noise trading (x1M)	Lower bound on SD of noise trading (% of SD of total trades)
Daily frequency										
1	0.02%	0.15	1.52	0.23	0%	0.03%	0.01	738.27	7.17	12%
2	0.06%	0.20	34.73	7.10	4%	0.07%	0.02	-583.56	-10.72	-6%
3	0.10%	0.16	185.19	29.70	9%	0.11%	0.02	-474.32	-7.39	-3%
4	0.14%	0.16	379.84	61.96	15%	0.16%	0.01	15148.20	107.13	31%
5	0.19%	0.17	562.36	96.32	18%	0.22%	0.01	33810.98	199.68	46%
6	0.26%	0.25	651.11	164.38	24%	0.30%	0.01	32678.92	303.31	51%
7	0.34%	0.39	499.83	194.43	21%	0.40%	0.02	11161.75	205.93	26%
8	0.48%	0.47	809.68	377.89	28%	0.55%	0.03	17906.00	523.89	41%
9	0.74%	1.00	649.57	647.47	30%	0.83%	0.06	14232.04	811.25	39%
10	1.71%	2.09	896.96	1871.49	39%	1.65%	0.27	6405.78	1708.07	38%
Weekly frequency										
1	0.02%	0.29	33.30	9.80	4%	0.03%	0.04	952.70	36.44	16%
2	0.06%	0.42	112.98	47.59	5%	0.07%	0.08	-1246.19	-102.40	-15%
3	0.10%	0.40	435.19	173.32	13%	0.11%	0.07	-724.25	-52.79	-5%
4	0.14%	0.47	649.46	305.81	17%	0.16%	0.03	16947.44	578.47	39%
5	0.19%	0.45	1327.54	592.10	25%	0.22%	0.03	33634.29	1014.04	53%
6	0.26%	0.71	1439.66	1025.10	33%	0.30%	0.05	35054.59	1627.33	60%
7	0.34%	1.11	1111.16	1235.76	30%	0.40%	0.09	13453.05	1196.35	33%
8	0.47%	1.37	1675.52	2289.12	38%	0.55%	0.14	19040.10	2731.05	49%
9	0.74%	3.08	1210.04	3723.79	39%	0.83%	0.27	15710.60	4257.12	46%
10	1.71%	6.96	1455.47	10124.86	47%	1.65%	1.15	8539.49	9813.21	50%
Monthly frequency										
1	0.02%	0.63	174.57	110.20	10%	0.03%	0.13	85.30	11.32	2%
2	0.06%	0.85	376.58	318.73	9%	0.07%	0.33	-2505.35	-814.36	-36%
3	0.10%	1.02	956.36	977.45	19%	0.11%	0.32	-2109.73	-665.53	-22%
4	0.14%	1.32	849.21	1120.99	16%	0.16%	0.13	9624.83	1291.86	31%
5	0.19%	1.20	2216.93	2656.40	29%	0.22%	0.10	23079.16	2195.54	40%
6	0.26%	1.93	2104.06	4061.78	34%	0.30%	0.16	31698.87	4980.00	61%
7	0.34%	3.53	1307.67	4614.55	29%	0.40%	0.31	5367.30	1689.34	15%
8	0.47%	3.80	2419.67	9186.96	40%	0.55%	0.53	17238.86	9071.81	48%
9	0.74%	8.45	1959.13	16554.54	46%	0.83%	1.03	14210.11	14616.22	44%
10	1.71%	21.50	1575.82	33886.86	44%	1.64%	4.14	10063.06	41658.80	67%

Panel C: By stock price

Decile	Households - Turnover x 1M: Median stock price	SD of trades (x1M)	Regression coef. *b*	Lower bound on SD of noise trading (x1M)	Lower bound on SD of noise trading (% of SD of total trades)	TAQ Small trades - Turnover x 1M: Median stock price	SD of trades (x1M)	Regression coef. *b*	Lower bound on SD of noise trading (x1M)	Lower bound on SD of noise trading (% of SD of total trades)
Daily frequency										
1	0.99	2.39	113.32	271.20	18%	1.09	9.38	92.68	869.45	52%
2	2.67	1.67	198.78	332.88	27%	2.76	2.20	364.95	801.58	47%
3	4.39	1.46	197.99	289.22	25%	4.61	0.78	777.28	603.48	45%
4	6.56	1.29	358.21	462.41	37%	6.82	0.40	2205.93	878.38	58%
5	9.23	1.05	344.93	361.74	31%	9.52	0.22	2411.93	523.04	44%
6	12.65	0.81	476.02	385.79	34%	12.90	0.12	5045.47	581.30	53%
7	16.44	0.73	429.41	312.81	29%	17.04	0.06	8689.63	489.55	52%
8	21.54	0.45	635.05	285.96	34%	22.52	0.03	16091.56	464.39	55%
9	28.76	0.35	686.17	240.03	29%	30.19	0.02	24653.95	386.27	54%
10	49.95	0.23	1219.39	277.70	34%	47.29	0.00	90540.63	410.03	50%
Weekly frequency										
1	1.02	5.82	320.41	1863.47	27%	1.09	39.64	89.41	3544.24	51%
2	2.70	4.89	457.59	2237.92	38%	2.76	9.67	327.22	3163.91	45%
3	4.40	4.09	512.06	2092.22	38%	4.63	3.45	664.74	2293.40	39%
4	6.56	3.99	720.02	2874.15	50%	6.83	1.74	1810.60	3159.06	53%
5	9.21	3.14	829.48	2607.96	49%	9.57	1.00	2230.27	2233.64	43%
6	12.61	2.50	926.63	2319.78	46%	12.90	0.54	4708.05	2549.86	54%
7	16.51	1.96	1068.49	2099.08	45%	17.04	0.27	8282.97	2196.37	53%
8	21.50	1.32	1257.63	1659.85	46%	22.52	0.14	15376.38	2114.38	57%
9	28.76	1.03	1386.36	1423.19	39%	30.19	0.08	24220.19	1856.08	58%
10	50.11	0.72	2022.22	1446.22	40%	47.27	0.02	78724.11	1715.22	48%
Monthly frequency										
1	1.02	12.79	432.73	5533.63	21%	1.09	128.52	70.36	9042.24	40%
2	2.71	14.29	655.38	9365.37	40%	2.76	32.11	286.15	9187.45	38%
3	4.37	11.53	718.99	8292.79	39%	4.61	12.27	422.23	5180.96	26%
4	6.58	12.10	921.93	11151.29	52%	6.83	6.46	1520.16	9818.10	49%
5	9.26	9.90	1025.24	10145.18	52%	9.57	3.75	1476.29	5543.26	34%
6	12.60	7.95	983.65	7815.47	43%	12.90	2.04	3445.81	7031.46	49%
7	16.46	5.04	1375.48	6931.24	43%	17.04	0.97	5028.39	4884.62	44%
8	21.47	3.84	1387.82	5325.67	46%	22.52	0.49	11218.97	5504.92	51%
9	28.80	2.90	1802.38	5230.28	39%	30.27	0.28	14349.41	3980.70	48%
10	49.29	2.00	2263.47	4518.69	34%	47.27	0.07	27582.39	2011.41	20%

Panel D: By stock bid–ask spread

Decile	Households - Turnover x 1M					TAQ Small trades - Turnover x 1M				
	Median bid-ask spread (basis points)	SD of trades (x1M)	Regression coef. $\hat{b}$	Lower bound on SD of noise trading (x1M)	Lower bound on SD of noise trading (% of SD of total trades)	Median bid-ask spread (basis points)	SD of trades (x1M)	Regression coef. $\hat{b}$	Lower bound on SD of noise trading (x1M)	Lower bound on SD of noise trading (% of SD of total trades)
Daily frequency										
1	74	0.98	448.19	440.13	17%	84	0.03	30781.63	998.65	57%
2	143	0.63	807.38	506.84	32%	134	0.04	3499.10	153.78	20%
3	201	0.72	481.32	346.68	29%	184	0.14	-180.99	-25.07	-3%
4	268	0.80	324.87	260.16	26%	239	0.13	497.05	63.40	8%
5	348	0.99	223.64	222.10	22%	299	0.09	1858.06	166.19	21%
6	438	0.83	177.77	146.82	16%	374	0.16	330.07	52.03	6%
7	556	1.15	155.35	178.80	20%	468	0.24	-1.38	-0.32	0%
8	723	1.19	86.09	102.24	14%	608	0.39	-71.90	-28.03	-3%
9	1001	0.97	86.97	84.40	12%	859	2.19	116.13	254.44	31%
10	1502	0.84	88.33	74.23	15%	1442	1.72	117.40	202.32	29%
Weekly frequency										
1	73	2.79	855.04	2383.96	20%	84	0.15	30779.70	4604.64	58%
2	143	1.99	1718.76	3423.73	48%	134	0.19	3072.38	590.86	17%
3	201	2.04	1152.88	2352.44	45%	184	0.56	-498.79	-280.83	-8%
4	268	1.97	1201.56	2361.46	52%	239	0.61	575.41	351.97	10%
5	349	2.57	772.42	1982.99	45%	299	0.42	1958.50	812.93	24%
6	439	2.23	616.31	1375.56	32%	376	0.70	383.21	269.67	6%
7	561	3.10	406.65	1260.89	31%	468	1.06	-43.14	-45.82	-1%
8	734	3.08	277.86	855.58	25%	608	1.76	-90.81	-160.08	-4%
9	1036	2.42	210.63	509.94	17%	859	8.64	112.83	974.29	29%
10	1538	2.05	195.38	400.66	20%	1442	6.96	126.95	883.55	33%
Monthly frequency										
1	72	7.71	1241.52	9573.28	19%	84	0.59	28154.31	16733.05	57%
2	143	5.95	2261.80	13449.41	48%	135	0.57	4048.68	2296.18	20%
3	200	6.20	1585.81	9828.27	53%	185	1.73	-752.91	-1300.15	-13%
4	268	5.08	2144.60	10895.33	70%	241	1.73	344.99	595.91	6%
5	347	7.31	1237.84	9043.59	57%	300	1.65	1036.54	1707.76	16%
6	440	6.56	1052.48	6903.36	43%	377	2.64	-212.40	-560.12	-4%
7	558	8.48	558.79	4738.15	29%	468	4.48	-524.20	-2350.27	-17%
8	738	8.42	439.59	3701.84	27%	610	7.34	-367.16	-2693.57	-18%
9	1016	5.00	562.43	2812.96	23%	861	28.01	61.08	1710.72	14%
10	1533	4.36	268.71	1170.93	14%	1442	23.63	95.15	2248.28	24%

Panel E: By stock Amihud illiquidity measure

Decile	Households - Turnover x 1M					TAQ Small trades - Turnover x 1M				
	Median Amihud illiquidity measure (x1M)	SD of trades (x1M)	Regression coef. $\hat{b}$	Lower bound on SD of noise trading (x1M)	Lower bound on SD of noise trading (% of SD of total trades)	Median Amihud illiquidity measure (x1M)	SD of trades (x1M)	Regression coef. $\hat{b}$	Lower bound on SD of noise trading (x1M)	Lower bound on SD of noise trading (% of SD of total trades)
Daily frequency										
1	0.00	0.25	1157.69	285.52	33%	0.00	0.01	44742.69	376.25	49%
2	0.01	0.33	741.40	248.08	29%	0.01	0.04	11445.15	420.95	47%
3	0.03	0.49	626.51	307.07	35%	0.02	0.08	4122.48	315.09	33%
4	0.08	0.60	366.15	220.81	30%	0.06	0.14	1803.64	248.68	29%
5	0.21	0.54	297.79	159.90	23%	0.15	0.27	1187.27	314.65	42%
6	0.53	0.52	192.68	100.11	18%	0.34	0.32	589.04	187.22	34%
7	1.26	0.51	125.14	63.59	15%	0.83	0.81	195.17	159.03	34%
8	3.02	0.48	103.92	49.94	16%	1.98	0.66	392.16	259.76	44%
9	8.04	0.54	65.21	35.52	13%	5.34	0.91	197.72	179.23	41%
10	38.20	3.35	9.53	31.89	11%	28.90	1.82	74.11	135.09	24%
Weekly frequency										
1	0.00	0.76	1962.03	1496.38	39%	0.00	0.04	42576.58	1763.75	52%
2	0.01	1.03	1578.29	1625.02	41%	0.01	0.17	11829.20	2065.03	51%
3	0.03	1.52	1366.56	2081.80	52%	0.02	0.35	4240.02	1462.88	34%
4	0.08	1.66	1049.83	1744.27	52%	0.06	0.61	1609.18	978.19	26%
5	0.21	1.55	783.73	1215.79	37%	0.15	1.11	1092.54	1210.60	39%
6	0.54	1.36	588.26	798.51	31%	0.34	1.36	549.01	748.57	31%
7	1.28	1.28	439.58	564.12	31%	0.83	3.15	220.47	694.91	37%
8	3.03	1.28	243.41	312.13	25%	1.98	2.52	369.67	931.13	40%
9	8.33	1.34	191.18	255.38	24%	5.35	3.79	159.17	603.28	36%
10	39.20	6.44	27.10	174.55	16%	28.90	7.33	68.99	505.34	28%
Monthly frequency										
1	0.00	2.09	2151.54	4500.26	33%	0.00	0.14	17673.60	2533.73	29%
2	0.01	3.10	1589.54	4931.83	34%	0.01	0.66	7998.34	5253.44	40%
3	0.03	4.82	1597.33	7705.66	53%	0.02	1.33	2645.36	3515.69	22%
4	0.08	4.78	1422.63	6793.22	56%	0.06	2.26	644.82	1455.19	11%
5	0.22	4.31	1264.52	5454.56	44%	0.15	3.82	784.24	2999.11	28%
6	0.53	3.75	756.65	2841.20	30%	0.34	4.79	243.85	1168.83	14%
7	1.31	3.59	624.04	2237.62	33%	0.82	9.54	204.27	1948.46	32%
8	3.03	3.20	314.91	1008.87	23%	1.97	8.54	264.89	2262.73	30%
9	8.27	2.85	408.55	1163.12	30%	5.21	13.48	91.11	1228.50	23%
10	36.70	12.21	23.73	289.64	7%	28.90	25.94	38.26	992.46	23%

Panel F: By stock return volatility

Decile	Households - Turnover x 1M					TAQ Small trades - Turnover x 1M				
	Median SD of stock returns	SD of trades (x1M)	Regression coef. $\hat{b}$	Lower bound on SD of noise trading (x1M)	Lower bound on SD of noise trading (% of SD of total trades)	Median SD of stock returns	SD of trades (x1M)	Regression coef. $\hat{b}$	Lower bound on SD of noise trading (x1M)	Lower bound on SD of noise trading (% of SD of total trades)
	Daily frequency									
1	1.02%	0.19	288.21	54.64	14%	1.03%	0.01	24117.12	177.07	45%
2	1.47%	0.20	552.30	113.12	21%	1.51%	0.01	52911.33	284.09	58%
3	1.89%	0.34	697.49	234.39	31%	1.94%	0.01	36409.21	355.44	54%
4	2.34%	0.72	428.84	308.16	22%	2.37%	0.02	32577.72	577.65	56%
5	2.81%	0.98	472.50	462.58	26%	2.84%	0.03	24322.01	736.30	46%
6	3.34%	1.15	717.97	826.36	33%	3.41%	0.06	15889.72	952.97	43%
7	3.99%	1.63	691.67	1128.69	35%	4.05%	0.12	5123.88	590.36	20%
8	4.89%	2.05	628.85	1291.12	30%	4.95%	0.22	2358.78	517.42	12%
9	6.32%	3.50	572.00	2004.77	39%	6.41%	0.59	984.25	581.74	11%
10	10.13%	7.03	557.45	3916.48	43%	10.21%	3.26	802.21	2614.83	22%
	Weekly frequency									
1	1.04%	0.47	874.42	413.11	25%	1.03%	0.04	24053.00	843.62	52%
2	1.47%	0.57	1088.34	619.96	27%	1.51%	0.03	52850.57	1366.88	64%
3	1.89%	1.00	1197.61	1199.17	37%	1.94%	0.05	34558.29	1627.42	57%
4	2.34%	1.78	1084.88	1927.47	32%	2.37%	0.09	32571.40	2820.15	62%
5	2.81%	2.95	822.45	2426.81	32%	2.85%	0.14	24325.60	3485.57	50%
6	3.34%	3.54	1211.60	4285.95	40%	3.41%	0.28	16199.82	4568.54	48%
7	3.99%	4.86	1362.73	6621.94	48%	4.07%	0.55	5013.72	2768.04	23%
8	4.88%	6.27	1258.21	7886.46	42%	4.95%	1.00	2770.38	2774.01	16%
9	6.29%	10.60	1072.47	11371.11	52%	6.41%	2.72	981.51	2664.83	12%
10	10.03%	24.33	803.15	19538.46	52%	10.22%	14.18	844.97	11978.77	25%
	Monthly frequency									
1	1.03%	1.10	1589.85	1748.82	31%	1.03%	0.14	17097.52	2451.10	55%
2	1.47%	1.46	1266.59	1847.99	23%	1.51%	0.10	41449.22	3995.07	61%
3	1.89%	2.80	1685.55	4711.51	41%	1.94%	0.18	27543.91	5014.07	54%
4	2.34%	4.58	1649.30	7552.35	34%	2.37%	0.33	26146.08	8667.63	59%
5	2.80%	7.40	1343.66	9941.16	35%	2.84%	0.55	20041.36	10933.99	45%
6	3.34%	10.87	1402.67	15246.15	38%	3.41%	1.05	14138.62	14776.11	44%
7	3.98%	14.64	1409.66	20633.93	41%	4.06%	2.22	2673.18	5932.66	15%
8	4.89%	18.27	1443.71	26373.04	37%	4.95%	3.85	272.52	1049.01	2%
9	6.30%	31.33	1339.36	41958.78	52%	6.41%	10.61	248.57	2637.65	4%
10	10.13%	75.39	881.85	66482.60	48%	10.21%	49.18	920.84	45282.39	26%

Panel G: By stock return autocovariance

Decile	Households - Turnover x 1M					TAQ Small trades - Turnover x 1M				
	Median return autocovariance (basis points)	SD of trades (x1M)	Regression coef. $\hat{b}$	Lower bound on SD of noise trading (x1M)	Lower bound on SD of noise trading (% of SD of total trades)	Median return autocovariance (basis points)	SD of trades (x1M)	Regression coef. $\hat{b}$	Lower bound on SD of noise trading (x1M)	Lower bound on SD of noise trading (% of SD of total trades)
Daily frequency										
1	-3.06	3.18	463.37	1474.34	32%	-3.16	2.47	338.47	834.39	11%
2	-1.07	1.95	521.11	1018.51	29%	-0.99	0.54	409.12	219.82	6%
3	-0.51	1.40	815.03	1139.30	36%	-0.48	0.29	626.14	184.06	7%
4	-0.28	0.93	580.66	539.78	25%	-0.26	0.10	4049.12	411.36	25%
5	-0.15	0.64	773.62	497.36	33%	-0.14	0.03	16811.02	505.90	45%
6	-0.07	0.42	431.85	181.49	20%	-0.06	0.02	25556.35	394.85	53%
7	-0.03	0.22	655.54	147.01	25%	-0.02	0.01	43515.93	339.39	51%
8	0.00	0.24	710.39	172.27	28%	0.01	0.01	41983.20	381.33	55%
9	0.05	0.39	1454.13	568.20	54%	0.07	0.03	32352.17	811.50	60%
10	0.29	1.66	873.22	1448.38	46%	0.32	0.36	3936.36	1400.60	38%
Weekly frequency										
1	-3.07	9.19	905.47	8325.60	45%	-3.16	10.70	273.65	2929.25	10%
2	-1.06	5.45	1168.22	6364.75	41%	-1.01	2.36	423.44	998.16	6%
3	-0.52	4.40	1510.74	6651.84	48%	-0.49	1.35	740.87	1001.36	9%
4	-0.28	2.52	1275.34	3215.67	34%	-0.26	0.47	4509.41	2125.58	30%
5	-0.15	1.95	1394.01	2714.63	40%	-0.14	0.14	16482.44	2365.20	49%
6	-0.07	1.06	1007.68	1066.54	27%	-0.06	0.07	25765.57	1905.79	58%
7	-0.03	0.62	1510.54	939.05	36%	-0.02	0.04	42198.73	1602.91	54%
8	0.00	0.75	1366.47	1031.41	38%	0.01	0.04	43372.15	1866.26	61%
9	0.05	1.37	2303.63	3166.10	67%	0.07	0.12	32644.76	3818.87	63%
10	0.28	6.22	1266.15	7880.95	57%	0.33	1.48	4838.60	7167.73	45%
Monthly frequency										
1	-2.93	30.07	1043.06	31369.16	43%	-3.16	40.54	230.60	9348.16	9%
2	-1.07	15.19	1686.68	25617.28	42%	-1.01	8.93	-85.61	-764.83	-1%
3	-0.52	12.49	1950.53	24357.56	47%	-0.48	5.90	-12.35	-72.90	0%
4	-0.28	7.98	1577.95	12598.54	34%	-0.26	2.01	3204.98	6436.37	26%
5	-0.15	5.82	2006.83	11682.28	44%	-0.14	0.59	13913.84	8219.63	49%
6	-0.07	2.63	1804.05	4740.02	31%	-0.06	0.31	21289.28	6508.01	59%
7	-0.03	1.72	1940.95	3330.87	37%	-0.02	0.14	34323.48	4855.30	48%
8	0.00	2.21	1823.95	4025.66	42%	0.01	0.17	34892.50	5926.02	56%
9	0.05	4.97	2609.27	12964.26	72%	0.07	0.47	30223.67	14288.46	64%
10	0.28	21.76	1432.46	31164.14	57%	0.33	5.60	5430.81	30399.20	53%

Table 5: Trades and economic fundamentals

We estimate regression models of trades and fund flows on surprises about firms' earnings. Panels A and B (resp., Panel C) report results for stock trades (resp., fund flows). In Panels A and B, we estimate a stock–quarter panel regression model of stock trades on firms' earnings surprises. The independent variable is a firm's quarterly standardized unexpected earnings (SUE) decile. Such earnings surprises are defined—for each stock and quarter—as the difference between actual and expected earnings, where expected earnings are derived from a seasonal random walk with drift, and are divided by their standard deviation (see text for the mathematical details). The dependent variables in Panel A are households' aggregate trades (net turnover among households and number of households trading), where the analysis is restricted to stocks that have at least 100 trades over the 1991–1996 sample period; the dependent variables in Panel B are TAQ trades (net turnover among small and large trades), where the analysis is restricted to stocks that have at least $100,000 worth of small trades over the 1991–2000 sample period. In all regressions, turnover is scaled by one million. In panels A and B, the dependent variables are trades, in a particular stock and quarter, aggregated over windows of 40, 20, 10, and 5 trading days ending on the day before the firm announces its earnings and on the announcement day (day 0). The regressions include firm, quarter, and month-of-year fixed effects; standard errors (in parentheses) are clustered by firm.

Panel C reports results for fund flows. We estimate a quarterly time-series regression of fund flows on aggregate earnings surprises. Every quarter, we aggregate earnings across all stocks whose fiscal year ends on December 31 and define the standardized aggregate earnings surprise as the difference between actual and expected aggregate earnings. Expected aggregate earnings are derived from a seasonal random walk with drift, and divided by their standard deviation (see text for details). The independent variable is the quarterly standardized unexpected aggregate earnings (SUAE) decile; the dependent variable is the net turnover, scaled by one million, for flows to retail equity mutual funds in the TrimTabs data set: the aggregate value of purchases, minus redemptions, of equity mutual funds divided by these funds' aggregate TNA. This variable is aggregated every quarter over windows of 40, 20 and 10 trading days ending on the day before the median announcement date across firms. The regressions include quarter fixed effects. We use *, **, and *** to indicate statistical significance at (respectively) the 10%, 5%, and 1% level.

Panel A: Households' trades and firms' earnings

	Households - Turnover x 1M	Households - Number
	40-day window	
SUE	-0.553***	-0.260***
	(-5.063)	(-6.940)
R-square	0.9%	1.9%
	20-day window	
SUE	-0.332***	-0.128***
	(-3.280)	(-6.840)
R-square	0.7%	1.5%
	10-day window	
SUE	-0.184**	-0.067***
	(-2.515)	(-6.059)
R-square	0.5%	1.2%
	5-day window	
SUE	-0.073**	-0.026***
	(-2.078)	(-5.565)
R-square	0.4%	0.9%
	Day 0	
SUE	-0.012	-0.012***
	(-0.805)	(-3.565)
R-square	0.1%	0.3%
Obs.	12,841	12,841
Firms	670	670

Panel B: TAQ small trades and firms' earnings

	TAQ Small trades - Turnover x 1M	TAQ Large trades - Turnover x 1M	Small minus Large trades - Turnover
		40-day window	
SUE	0.084	0.482**	-0.398*
	(1.047)	(2.216)	(-1.763)
R-square	0.3%	0.3%	0.2%
		20-day window	
SUE	0.049	0.275**	-0.226*
	(0.794)	(1.966)	(-1.710)
R-square	0.3%	0.2%	0.2%
		10-day window	
SUE	0.048	0.240**	-0.191**
	(1.112)	(2.355)	(-2.012)
R-square	0.2%	0.2%	0.2%
		5-day window	
SUE	-0.022	0.06	-0.083
	(-0.506)	(0.909)	(-1.059)
R-square	0.2%	0.1%	0.2%
		Day 0	
SUE	0.076***	0.104***	-0.028
	(4.326)	(4.376)	(-0.939)
R-square	0.2%	0.1%	0.1%
Obs.	54,495	54,495	54,495
Firms	2,750	2,750	2,750

Panel C: Mutual fund flows and firms' earnings

	Fund flows (x 1M)
	40-day window
SUAE	519.961
	(0.795)
R-square	9%
	20-day window
SUAE	339.131
	(0.899)
R-square	8%
	10-day window
SUAE	226.183
	(1.131)
R-square	14%
Obs.	58

Table A1: Correlation among trades

This table examines whether households' trades (Panels A and B), small TAQ trades (Panel C) and mutual fund flows (Panel D) have a systematic component. In Panel A we regress, in a stock–month panel setting, a stock's aggregate trades in a given month on the average aggregate trades across all other stocks (where this average excludes the stock's own trades)—labeled "Mean Dep. Var." in the table—and on the contemporaneous market return. The results indicate that a given stock is more likely to be bought by households at times when they are buying other stocks. In Panel B we estimate a household–month panel regression; here the dependent variable is a household's trades in all stocks in a given month, and the independent variables are the average trades across all other households (where this average excludes the household's own trades) and the contemporaneous market return. Panels A and B consider households' net turnover (defined as the aggregate value of their buys, minus the aggregate value of their sells, divided by the total value of the market) as well as the net number of households buying shares (defined as the number of households buying minus the number of households selling). Panel B includes, as a dependent variable, the number of distinct stocks bought by a given household. The results indicate that a given household tends to buy stocks at times when other households are buying stocks. Panel C is similar to Panel A and reports results from a stock–month panel regression of a stock's aggregate trades—measured in the TAQ data set for a given month—on the average aggregate trades across all other stocks (where this average excludes the stock's own trades) and on the contemporaneous market return. The table reports estimates for small and large trades and also for their difference. The results indicate that a given stock is more likely to be bought at times when other stocks are bought and that this effect is stronger (by a factor of 2) for small trades than for large trades. Panel D is similar to Panel A and reports results from a fund–month panel regression of net flows to an retail equity fund —measured in the TrimTabs data set for a given month—on the average aggregate net flow across all other retail equity funds (where this average excludes the stock's own trades) and on the contemporaneous market return. The results indicate that a given fund is more likely to receive inflows at times when other funds receive inflows. Standard errors are double clustered by month and by either firm (Panels A and C) or household (Panel B). We use ** and *** to indicate statistical significance at (respectively) the 5% and 1% level.

Panel A: Correlation among stocks traded by households

	Number of households	Turnover
Mean Dep. Var.	1.017***	0.498***
	(13.099)	(3.919)
Mkt return	-0.074	-0.000***
	(-0.233)	(-2.740)
Firms	9,158	9,158
Obs.	123,133	123,133
R-square	9.8%	2.4%

Panel B: Correlation among households

	Number of households	Turnover	Number of stocks
Mean Dep. Var.	1.033***	0.589***	0.941***
	(31.595)	(5.705)	(13.843)
Mkt return	0.197	-0.000**	0.075
	(1.536)	(-2.064)	(0.277)
Households	11,268	11,268	11,268
Obs.	159,305	159,305	159,305
R-square	16.5%	3.3%	32.7%

Panel C: Correlation among stocks traded in TAQ

	Small trades - Turnover	Large trades - Turnover	Small minus Large trades - Turnover
Mean Dep. Var.	0.922***	0.514***	0.500***
	(14.229)	(9.170)	(8.133)
Mkt return	-0.000	0.000***	-0.000***
	(-0.111)	(4.727)	(-4.537)
Firms	11,850	11,850	11,850
Obs.	454,467	454,467	454,467
R-square	0.2%	0.4%	0.3%

Panel D: Correlation among fund flows

	Mutual Fund Flows - Turnover
Mean Dep. Var.	0.906***
	(25.432)
Mkt return	0.000
	(0.063)
Funds	2,468
Obs.	125,578
R-square	0.019

Table A2: Performance of trades

We estimate the post-trade, buy–sell return difference as in Odean (1999). That is: for each day, we first calculate the average return across all buy and sell transactions executed on that day over the subsequent four months (84 trading days) and then take the difference between those transaction types; we also perform and report this calculation for a twelve-month holding period (252 trading days). For fund flows, we use fund purchases/redemptions and fund returns in lieu of stock buys/sells and stock returns. Average post-trade–return differences are estimated using both raw and market-adjusted returns. Households' trade returns and fund flow returns are equal-weighted; TAQ trade returns are weighted according to the value of the trade. The *t*-statistics (in parentheses) test the extent to which the mean return differs from zero. To account for overlap in the return window, standard errors are adjusted for autocorrelation (for up to 252 lags) via the Newey–West correction. We use * and *** to indicate statistical significance at (respectively) the 10% and 1% level.

Panel A: Performance of households' trades

	4-month holding period	12-month holding period
Raw Returns	-0.0054*	-0.0260***
	(-1.71)	(-3.92)
Market-adjusted Return	-0.0052*	-0.0221***
	(-1.75)	(-3.74)

Panel B: Performance of TAQ trades

	4-month holding period			12-month holding period		
	TAQ Small trades	TAQ Large trades	Small minus Large trades	TAQ Small trades	TAQ Large trades	Small minus Large trades
Raw Returns	-0.0242***	-0.0028	-0.0213***	-0.0965***	-0.0012	-0.0953***
	(-5.75)	(-1.37)	(-4.16)	(-9.67)	(-0.14)	(-8.27)
Market-adjusted Return	-0.0226***	-0.0020	-0.0206***	-0.0856***	-0.0017	-0.0839***
	(-5.51)	(-0.94)	(-4.11)	(-8.11)	(-0.21)	(-8.27)

Panel C: Performance of fund flows

	4-month holding period	12-month holding period
Raw Returns	0.0009	-0.0021
	-0.54	(-0.42)
Market-adjusted Return	0.001	-0.0015
	-0.6	(-0.29)