Noise Traders Incarnate: Describing a Realistic Noise Trading Process

*Joel Peress & Daniel Schmidt**

October, 2020

ABSTRACT

We estimate a realistic process for noise trading to help theorists derive predictions from noisy rational expectations models. We characterize the trades of individual investors, who are natural candidates for the role of noise traders because their trades are weakly correlated with fundamentals, in line with how such models define noise trading. Data from a retail brokerage house, small and price-improved trades in TAQ, and flows to retail mutual funds yield consistent estimates. The properties of noise trading are highly sensitive to the frequency considered, with the common assumption of i.i.d.-normal noise appropriate only at monthly and lower frequencies.

---

* 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. Joel Peress thanks the AXA Research Fund and the Institut Europlace de Finance for their financial support.

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. As Fisher Black put it in his Presidential Address, "Noise makes financial markets possible" (Black (1986, p. 530)). 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 the noise trading 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*. This definition implies that portfolio rebalancing trades qualify as noise trades as long as they are motivated by shocks uncorrelated with fundamentals (e.g., liquidity shocks or exposure to a non-traded asset), whereas uninformed rational trading does not, since it involves learning about fundamentals from asset prices or order flow. We estimate a process for noise trading from retail trading data under the identifying assumption that retail trades are representative of noise trades. The empirical literature provides ample evidence in support of this assumption (e.g., the poor performance of retail trades) and has examined several of its implications (e.g., the literature on measuring investor sentiment).

To appreciate the importance of a thorough understanding of noise trading, consider the following three aspects—we elaborate on all three in Section 1 and offer references to the literature. The first is its *persistence*, which influences important properties of NREE models and financial markets: the informativeness of asset prices, the possibility of multiple equilibria, the serial correlations of stock returns and of trading volume, and finally, the measurement of liquidity. The second important aspect of noise trading concerns its *distribution* and *correlation with fundamentals*. Standard models assume

[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 19,000 times.

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. 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 since used such models to make quantitative predictions and to estimate orders of magnitude.[2] Yet, because noise trading is not directly observable, these scholars usually either pick an arbitrary value for its variance or choose a value such that the model's predicted moments match sample moments

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 retail mutual funds, which subsequently trigger trades. Retail investors are natural candidates for the role of noise traders because their trades are weakly correlated with fundamentals. Consistent with this view, previous research has documented that they amplify stock return volatility (Foucault et al. 2011, Peress and Schmidt 2020), 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). Hence, retail 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, injects "noise" into 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 rely on four 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 the spread of electronic limit order books and decimalization in 2001 were most likely to have been made by retail investors (Hvidkjaer 2008, Barber et al. 2009a). The third source is a set of trades from TAQ that received price improvements, which, as Boehmer et al. (2020)

[2] See Internet Appendix G for a list of published articles calibrating NREE models.

[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 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.

demonstrate, are initiated by retail investors. The fourth source is data on net flows to mutual funds invested in US domestic equity and sold to retail investors. Each of these datasets 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 and price-improved trades are more comprehensive but do not reveal traders' identities. These datasets 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 datasets and procedures also complement each other by covering different periods: 1991–1996, 1991–2000, 2010-2015, and 1999–2013 for (respectively) the retail brokerage data, small TAQ trades, price-improved TAQ trades, 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 weakly correlated with fundamentals captured by firms' earnings news, unprofitable on average and cross-correlated (see Internet Appendix A).

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 recognize that the 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.

The canonical NREE framework assumes 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. 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 datasets or, in the case of mutual funds, as the aggregate value of fund flows divided by funds' total

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

assets. 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 datasets, and the purchase turnover minus the redemption turnover for mutual funds flows.[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. An important result of our analysis is that *the properties of noise trading are highly sensitive to the trading frequency being considered*. Across datasets we find that trades can be considered i.i.d. at monthly and lower frequencies (quarterly). In contrast, daily and weekly trades are serially 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 datasets—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.*

These findings call for theorists to clearly state which frequency their model is intended to capture (i.e., high vs low frequency), and to make assumptions consistent with that frequency. For instance, the common CARA-normal setup with i.i.d. noise is appropriate for modelling firm managers learning

---

[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.

[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).

from prices and adjusting investment decisions as in the "feedback" literature (see, e.g., Bond et al. 2012), but not for describing a high-frequency trading environment (daily frequency and higher).

Next we attempt to quantify the intensity of noise trading. This task is difficult as it requires estimating what fraction of total noise trading is represented by the retail trades in our sample. We accomplish this by comparing retail investors' trading volume with the total trading volume in the market. Specifically, 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).

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 and price-improved TAQ trades are remarkably similar, despite covering different periods; mutual fund data yield lower estimates. A subperiod analysis suggests that noise trading has remained stable over the two decades ending in 2010. This result speaks to the debate on the evolution of price efficiency in financial markets (e.g., French 2008 and Bai et al. 2016) and supports Stein (2009)'s argument that, despite of the secular downtrend in direct retail ownership, markets are not bound to become more efficient over time.

Using the brokerage and TAQ datasets, 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 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.

[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, Llorente et al., 2002) 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. We conclude in Section 6 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 (its persistence and distribution), while the third (its intensity) is of a quantitative nature.

a. The persistence of noise trades

The NREE literature displays a wide array of assumptions concerning the persistence of noise trades: some papers assume that noise trades are i.i.d. over time (e.g., Allen, Morris, and Shin 2006, Peress 2014), while others posit that they follow a random-walk (e.g., Kyle 1985, Grundy and McNichols 1989, Brown and Jennings 1989, Vives 1995); yet others adopt a more general AR(1) structure (e.g., Campbell and Kyle 1993, Wang 1993, 1994, He and Wang 1995, Cespa and Vives 2015, Avdis 2016). As mentioned previously, there are at least four reasons why this persistence 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, it controls whether prices are better or worse predictors of fundamentals than is the consensus opinion, as demonstrated by Cespa and Vives (2015) in their analysis of short-termism in financial markets. They acknowledge (p. 2101) that "a crucial hypothesis of our model is that liquidity trading displays persistence." Likewise, persistence is central to the recent debate on how improvements in data processing—in particular, "Big Data" and Artificial Intelligence—are reshaping information production in financial markets. Begenau et al. (2018) and Farboodi and Veldkamp (2017) argue that access to order flow data (i.e., data on noise trades) improves market efficiency by enabling investors to purge price signals from their noise component and extract fundamental information. But their conclusion hinges on the assumed lack of persistence of noise trades. If noise trading is, instead, persistent, then investors might use order flow information to ride, rather than correct, the non-fundamental component of prices, thus hurting market efficiency. Other examples of the important role played by the postulated persistence of noise trading 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 trades alters the size of *complementarities in information acquisition*, and hence the *existence of multiple equilibria*. Indeed, Avdis (2016) points out that information about fundamentals can be used not only for "trading on fundamentals" but also for inferring from the price the contemporaneous level of noise trading. This level, in turn, helps predict changes in noise trading—its predictive power depends on the persistence of noise trading, and hence future price changes. It is therefore a source of short-term profits. This mechanism leads to complementarities in information acquisition: as the number of informed agents increases, prices become more informative about fundamentals and therefore less informative about the contemporaneous level of noise trading, thus making fundamental information more valuable (since agents can infer noise trading from prices only if they know the fundamental shock).

Cespa and Vives (2015) also present a model in which, noise trading persistence determines whether there can be strategic complementarities in investors' use of information. Specifically, in a dynamic setting with short-term speculation, more aggressive trading on private signals in the early period reduces uncertainty about stock prices in the later period, which in turn spurs further informed trading early on. When noise trading is persistent, this effect can dominate over the usual substitution effect in information (whereby more aggressive trading makes prices more informative and thus renders trading on private information less attractive) so that a "high information equilibrium" may arise. Our finding that noise can be treated as i.i.d. normal at monthly frequency casts doubt on the viability of this equilibrium.

Third, 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."

Finally, the persistence of noise trades is central to the debate concerning the *measurement of liquidity* in financial markets. In a recent empirical study, Collin-Dufresne and Fos (2015) 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. An explanation, developed further in Collin-Dufresne and Fos (2016), 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

Standard models assume that noise trades are normally distributed. 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. We shall assess the plausibility of these assumptions in Section 4.b.

c. Noise trading intensity

Finally, another fundamental aspect of noise trading is its intensity, or standard deviation. Though early work mostly focuses on the qualitative predictions of NREE models, more recent studies often seek to quantify predicted effects and to assess their economic relevance—Internet Appendix G presents a list of articles, many published in top journals, calibrating NREE models. 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 2004, Banerjee 2011, Manela 2014, Begenau et al. 2018, Farboodi and Veldkamp 2017). 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.

Thus, we support such calibrations with a set of parameter estimates (for, e.g., the variance of noise, the serial correlation of noise) that are both plausible (rather than arbitrary) and compatible with one another (rather than being drawn from different, possibly inconsistent, sources).

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. Direct evidence in favour of this premise comes from studies that report a drop in stock return volatility following a decline in the intensity of retail trading. Foucault et al (2011) find that a reform that discouraged retail trading on the French stock exchange permanently reduced volatility. Likewise, Peress and Schmidt (2020) report a decline following episodes of sensational news that distract (mostly) retail investors from trading.

The premise that retail trades are noise trades is further supported by evidence showing that r*etail investors perform poorly, even before transactions costs.* The NREE literature defines noise trades 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 (all of which are compatible with rationality), or behavioral traits such as overconfidence or biased beliefs. All of these motives should 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 (we review the evidence in Appendix A.1; see also Barber and Odean 2013). Their underperformance is also apparent in their mutual fund investments, which justifies our use mutual fund flows to proxy for noise trading.[8]

Another important requirement for retail trades in a stock to qualify as noise is that they be *correlated with one another*. Otherwise, they would wash out once they are aggregated, and would hence fail to inject noise into the price signal in NREE models. Kumar and Lee (2006) and Barber et al. (2009b) report that retail trades have a common systematic component.

In Appendix A.2, we verify that the trades in our datasets display all three characteristics, namely that they (i) are weakly correlated with fundamentals as captured by firms' earnings news (fund flows' correlation with news about aggregate earnings is similarly low), (ii) perform poorly, and (iii) are cross-sectionally correlated. Because of these properties, retail 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.

A caveat is in order though: 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.

[8] 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. Hence, we restrict our use of mutual fund data to the estimation of a noise trading process for the entire market and not for individual stocks. In Appendix A.1, we discuss conditions under which the trades induced by mutual fund flows have similar statistical properties (with respect to, e.g., persistence and distribution) as the flows themselves.

3) Data

We use three datasets. 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 dataset 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, through both market and limit orders, between January 1991 and November 1996 (inclusive). Hirshleifer et al. (2008) argue that this dataset 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 dataset are similar to those reported by individuals on their income-tax returns. Because the number of households in this dataset 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 dataset.

[[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. In Internet Appendix E, we show results for the net number of households buying shares, defined as the number of households buying minus the number of households selling; they are qualitatively and quantitatively similar to those obtained with the net turnover. Figure 1 displays the monthly time series of households' net turnover, and Table 1 presents summary statistics at the daily, weekly, and monthly frequency.

[[INSERT Table 1 about Here]]

b. *TAQ data*

Our second data source consists of transactions involving NYSE/AMEX/NASDAQ stocks based on market orders (limit orders are excluded) and recorded in the TAQ/ISSM database since 1991.[9] We use two distinct time periods, and correspondingly, procedures to identify retail trades. The first spans the period 1991 to 2000. 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.[10] 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.

Our second sample period spans 2010 to 2015.[11] For this period, Boehmer et al. (2020) argue, and confirm empirically, that retail-initiated trades, unlike institutional trades, receive price improvements. Hence, retail buys can be identified as those with a transaction price slightly below the round penny, whereas retail sells are those with a transaction price slightly above the round penny. The sample consists of all common stocks listed on U.S. stock exchanges with a valid share price in CRSP and a price of at least \$1.

After having identified retail trades, using either the trade's size (1991-2000) or the trade's price, we aggregate dollar buys and dollar sells over the entire dataset separately for small, large, and price-improved trades, and by day, week, and month. We then calculate the difference between buys and sells and divide by the total market capitalization to obtain a measure of net turnover. Figure 1 displays the monthly time series of net turnover estimated from small and price-improved trades in TAQ, and Table 1 presents summary statistics at daily, weekly, and monthly frequencies.

---

[9] 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. We thank Søren Hvidkjaer for sharing the aggregated ISSM and TAQ volume data, broken down by trade size, for the period 1991–2000.

[10] 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.

[11] We are grateful to Xiaoyan Zhang for sharing data on price-improved trades covering the period 2010 to 2015.

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 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 dataset is that it covers only a subset of the retail population and their corresponding stock trades.

In contrast, the TAQ dataset 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 or price-improved trades are actually executed by retail investors.

Mutual fund flows account for a significant portion 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.[12]

Finally, the datasets complement each other also by covering different periods: 1991–1996 for the brokerage data, 1991–2000 and 2010-2015 for the TAQ data, and 1999–2013 for the mutual fund data. Reporting results for all four time-series 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 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.[13]

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.[14]

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 retail net turnover as measured using households, TAQ small and price-improved trades , 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*-

[12] We find qualitatively and quantitatively similar results after excluding mutual funds that hold illiquid assets (mid-cap, small-cap or micro-cap funds; see Chen et al. 2010, Chernenko and Sunderam 2016). Since these funds are more prone to buffer flows with cash, we conclude that cash holdings do not significantly alter the properties of fund flows.

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

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

test for the residuals (solid line marked by crosses, left axis). High *p*-values indicate that we cannot reject the null hypothesis of residuals from the fitted process being serially uncorrelated. We also show the value of Akaike's information criterion (dashed line marked by circles, and right axis) as a function of the number of lags.[15] 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 14 for mutual fund flows (at the 10% significance level). At the weekly frequency, one lag or less is sufficient to produce uncorrelated residuals for household and TAQ trades; for fund flows, 6 lags are needed. For monthly data, an AR(0) model offers a reasonable approximation in all four series of data, since we cannot reject the hypothesis that trades are serially uncorrelated.[16] 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.[17]

[[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 (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 0.0066. The slopes for small TAQ trades (upper right panel), and mutual fund flows (lower right panel) are in the same neighborhood:

[15] 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).

[16] Similar results obtain for the net number of households (see Internet Appendix E).

[17] 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.

respectively −0.0086 and −0.0050.[18] For price-improved TAQ trades (lower left panel), the slope is negative too (-0.0002) but not significantly different from zero. 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.

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 the households' data; this outcome suggests that households' aggregate trades are approximately normally distributed at that frequency (last column, top two rows).[19] However, small and price-improved TAQ trades and fund flows continue to display deviations from the 45° line even at the monthly frequency (last column, bottom three 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 more nuanced: normality is rejected for all series except for households' trades. Moreover, for all 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 TAQ trades and mutual fund flows is normal at a frequency lower than monthly. Hence Table 2 also reports results of

[18] The slope for the net number of households trading is also close, at −0.0102 (see Appendix E).
[19] The distribution for the net number of households is similar to that for households' net turnover (see Internet Appendix E).

the Shapiro–Wilk test conducted on quarterly data. Quarterly trades, regardless of the dataset from which they are drawn, all conform to the normal distribution.[20,21]

[[INSERT Table 2 about Here]]

c. A falsification test

We check that the characteristics of noise trades which we report are specific to these traders and do not mechanically extend to other traders, such as speculators or liquidity providers. That is, while these trades might inherit some of the properties of noise trades, we do not expect that they inherit all those properties.

Non-noise trades might depend on noise trades through three channels. i) Informed traders camouflage their trades behind noise trades, trading strategically when noise traders are more active. ii) They learn—albeit imperfectly—about stock fundamentals from stock prices, which entails at times wrongly attributing to fundamentals the impact of noise. iii) They might be supplying liquidity to noise traders. These dependences notwithstanding, informed and liquidity-supplying trades should diverge from noise trades along some dimensions.

We repeat our estimation procedure for trades that we deliberately excluded from our TAQ and mutual fund flows samples. Specifically, we re-run our tests for large TAQ trades, TAQ trades without price improvement, and flows to non-retail mutual funds. If our procedure indeed identifies noise trades, then these series should display different properties from those we report for noise trades.

The results, reported in Internet Appendix F, confirm this prediction. The differences are the sharpest along two dimensions. First, the processes for these data are not as strongly serially correlated as for noise trading. At the daily frequency, zero or one lag suffice for residuals to be serially uncorrelated for both small trades and trades without price improvements (vs. 5 to 10 lags for small trades and price-improved trades). Second, their distributions are considerably less fat-tailed. Large trades and trades without price improvement conform to a normal at, respectively, weekly and monthly frequencies, whereas small trades and price-improved trades only do so at monthly and quarterly

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

[21] In Internet Appendix C, we investigate whether the distribution of noise trades is stable over time or whether it switches between regimes with fat and thin tails by estimating its kurtosis over time. Our results indicate that it is permanently fat-tailed.

frequencies, respectively. Flows to non-retail mutual funds display a similar tendency towards less serial correlation and fat-tailedness, but that tendency is somewhat less pronounced. Strikingly, the first-order autocorrelation coefficient of that series is an increasing function of the time period's duration, while it is decreasing for noise trading.

We also examine a sample of traders that, plausibly, are truly informed. We proxy for these using hedge fund trades contained in the institutional transactions dataset provided by Abel Noser Solutions (commonly known as ANcerno), which we aggregate as we do with noise (household) trades. We then carry out the same estimations on this time series. The results, reported in Internet Appendix F, confirm our expectation: while some of the findings obtained for noise trades extend to hedge funds' trades, many do not.

To conclude, trades other than those we identify as noise trades behave very differently from noise trades. This lends confidence in our identification strategy and suggests that our procedure specifically captures the behaviour of noise traders.

Summary: Across all four measures, 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 datasets; more are needed for mutual fund flows and for daily trades. These findings are markedly different from those for non-retail (non-noise) trades.

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?[22] We describe here how to address this question, starting with an overview of our strategy.

[22] 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.

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 TAQ trades and fund flows.

a. Overview

Our key identifying assumption is that households' (noise) 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:

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

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. 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).[23]

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

[23] For example, suppose that a noise trader buys 100 shares from a rational trader. Total trading volume (the left-hand-side of equation (1)) records one transaction of 100 shares, while the individual transactions for noise and rational traders on the right-hand-side record two transactions (a buy of 100 shares for noise traders and a sell of 100 shares for rational traders). This explains why we require the scaling factor ½ on the right-hand-side of equation (1).

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, we show next that this requirement is mild in that the same lower bound obtains under the two canonical models of trading and under various information structures.

b. Bounds under competitive trading

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:

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

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:

$$\text{Total volume}_t = \frac{1}{2}\left(|-\Delta\theta_t| + \int_i \; \left|\Delta\theta_t + \Delta x_t^i\right|\right); \tag{3}$$

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 dataset. As explained previously, we assume that these trades account for a fraction *b* of all noise trading in the economy: $\Delta\hat{\theta}_t = -\frac{1}{b}\Delta\theta_t$.[24] It follows that

$$\text{Total volume}_t = \frac{1}{2}\left(\left|b\Delta\hat{\theta}_t\right| + \int_i \; \left|-b\Delta\hat{\theta}_t + \Delta x_t^i\right|\right). \tag{4}$$

We establish in the Appendix that the coefficient $\hat{b}$, which is obtained by regressing total trading volume on households' trading volume $\left|\Delta\hat{\theta}_t\right|$, is given by $\hat{b} = b(1+\sqrt{1+r}-\sqrt{r})/2$, where $r \equiv \mathrm{var}\left(\Delta x_t^i\right)/\mathrm{var}(\Delta\theta_t)$. 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):

$$\hat{b} \leq b \leq 2\hat{b}. \tag{5}$$

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 $\mathrm{var}\left(\left|\Delta\hat{\theta}_t\right|\right)$ 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

[24] 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*).

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 under strategic trading

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):

$$\text{Total volume}_t = \frac{1}{2}(|-\Delta\theta_t| + |\Delta\theta_t + \Delta x_t| + |-\Delta x_t|). \tag{6}$$

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

$$\text{Total volume}_t = \frac{1}{2}\left(\left|b\Delta\hat{\theta}_t\right| + \left|-b\Delta\hat{\theta}_t + \Delta x_t\right| + |-\Delta x_t|\right). \tag{7}$$

The Appendix 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. 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 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 larger with price-improved TAQ trades

(0.087% –0.175%, at the daily frequency) and mutual fund flows (0.042%–0.084% at the daily frequency).[25]

[[INSERT Figure 6 about Here]]

Figure 6 plots the noise trading intensity from 1991 through 2013.[26] Panel A of the figure shows that our estimates are reasonably close to one another across datasets. Panel B shows that the ratio of the noise trading intensity to the standard deviation of total trades has remained stable over time, despite dramatic changes to the market environment and the secular downtrend in direct retail ownership. A possible explanation is that direct retail noise trades are replaced with institutional noise trades, which are executed as institutional asset managers respond to flows in and out of their funds (see, e.g., Coval and Stafford, 2007; Lou, 2012, Chinco and Fos, 2019).

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).

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

$$\Delta\hat{\theta}_t^k = -\frac{1}{b^k}\Delta\theta_t^k; \tag{8}$$

---

[25] Differences in estimates across datasets may be attributable to their covering different periods. In unreported analysis, 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.

[26] 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.

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.[27]

Our procedure for measuring the marketwide scaling factor can be readily applied on a stock-by-stock basis. For each stock group *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 group *k* is bounded from below by the time-series standard deviation of households' aggregate trades in stock group *k*, or $\sqrt{\mathrm{var}(|\Delta\hat{\theta}_t^k|)}$, 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, turnover, Amihud illiquidity ratio (a measure of price impact of trades), the (closing) bid–ask spread, return volatility, and return autocovariance. All six of these variables are estimated every month from daily observations. Capitalization, turnover, and bid–ask spread are monthly averages. The Amihud illiquidity ratio is the monthly average of the daily ratio 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 6 × 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

[27] 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.

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.[28] 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.[29] The rows labelled "1-10" in Table 4 indicate that, in almost all instances, the difference between deciles 1 and 10 in the estimate of the standard deviation of noise trading is significantly different from zero. This finding suggests that stocks differ not only in the intensity with which they are traded within our datasets but also in the fraction of noise trading for which they account. The sort in terms of the Amihud illiquidity ratio illustrates the importance of scaling the variance of retail 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., turnover, bid–ask spreads except for price-improved TAQ trades) confirm the positive 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 the fifth Panel of Table 4 and as implied by most NREE models (e.g., Hellwig 1980, He and Wang 1995). In

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

[29] The correlation between the fraction of institutional ownership (based on 13F filings) in a stock group and the standard deviation of noise trades in that group equals -21% (p-value of 0.12) on average. It equals -20% (0.13), 2% (0.91), and -29% (0.03) using households, small TAQ trades, price-improved TAQ trades, respectively. For each dataset, the correlation is based on 60 observations (10 groups x 6 sorting variables).

contrast, the standard deviation of noise trading does not vary much with stock size (first panel), most likely because size is related to many stock characteristics and sometimes in opposite ways (as with, e.g., liquidity and volatility).

The autocorrelation results in the last Panel of the table are not 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 and price-improved TAQ trades 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 datasets. 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 datasets. 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 theory. In particular, greater liquidity and return volatility are associated with greater noise trading volatility, as predicted by virtually all NREE models.

6) Conclusion

In this paper we offer a description of a realistic noise trading process. We characterize the trades executed by investors who are natural candidates for the role of noise traders: individual (retail) investors. Using four different data sources, 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. These findings have important implications for two growing streams of theories. They imply that models of the "feedback effect"—models in which financial markets not only reflect the cashflows generated by assets but also affect those cashflows by either guiding firm manager's investment decisions or by incentivising them (see Bond, Edmans, and Goldstein (2012) for a review)—in which trading rounds are thought to last months and longer can justifiably assume i.i.d. noise trading. In contrast, theories of trading at higher frequencies cannot entertain such a hypothesis. In particular, extrapolating our findings, modelling high-frequency traders (who trade at speeds of milliseconds or faster) would require allowing for severe serial correlation in noise trading as well as relaxing its normality.

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. 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 trading intensifies among stocks that are more liquid and/or exhibit greater return volatility.

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.

Avdis, E., 2016, Information tradeoffs in dynamic financial markets. *Journal of Financial Economics*, 122(3), 568-584.

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., 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.

Begenau, J., Farboodi, M., & Veldkamp, L., 2018, Big data in finance and the growth of large firms. *Journal of Monetary Economics*, 97, 71-87.

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

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.

Boehmer, Ekkehart, Charles M. Jones, Xiaoyan Zhang, and Xinran Zhang, 2020, Tracking retail investor activity. Working Paper.

Bond, P., Edmans, A., & Goldstein, I. ,2012, The real effects of financial markets. *Annual Review of Financial Economics*, *4*(1), 339-360.

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.

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.

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

Collin-Dufresne, P., and V. Fos, 2016, Insider Trading, Stochastic Liquidity, and Equilibrium Prices, *Econometrica*, 84(4), 1441-1475.

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

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. 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.

Farboodi, M. and Veldkamp, L., 2017, Long Run Growth of Financial Technology, NBER Working Paper No. w23457. Available at SSRN: https://ssrn.com/abstract=2976197

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.

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

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.

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.

Llorente, Guillermo, Roni Michaely, Gideon Saar, and Jiang Wang. 2002. “Dynamic Volume Return Relation of Individual Stocks.” *The Review of Financial Studies,* 15 (4):1005– 1047.

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.

Manela, A., 2014, The value of diffusing information. *Journal of Financial Economics,* 111 (1), 181–199.

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

Peress, J., 2004, Wealth, information acquisition, and portfolio choice, *Review of Financial Studies*, 17(3), 879-914.

Peress, J., 2014, Learning from stock prices and economic growth, *The Review of Financial Studies*, *27*(10), 2998-3059.

Peress, J. and D. Schmidt, 2020, Glued to the TV: Distracted Noise Investors and Stock Market Liquidity, *Journal of Finance,* forthcoming.

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

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.

Stein, J. C., 2009, Sophisticated Investors and Market Efficiency, *Journal of Finance* 64, 1517-1548.

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.

Appendix: 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 $\hat{b} \equiv \mathrm{cov}(\text{Total volume}_t,\ |\Delta\hat{\theta}_t|)/\mathrm{var}(\,|\Delta\hat{\theta}_t|)$ 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*,

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

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

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

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

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

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

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

where $r \equiv \frac{\mathrm{var}(\Delta x_t^i)}{\mathrm{var}(\Delta\theta_t)} = \frac{\mathrm{var}(\Delta x_t^i)}{b^2\,\mathrm{var}(\Delta\hat{\theta}_t)}$ 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

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

where $r \equiv \frac{\mathrm{var}(\Delta x_t)}{\mathrm{var}(\Delta\theta_t)} = \frac{\mathrm{var}(\Delta x_t)}{b^2\,\mathrm{var}(\Delta\hat{\theta}_t)}$. 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 trades at a large discount broker (from January 1991 through November 1996), of small TAQ trades (from January 1991 through December 2000), of price-improved TAQ trades (from January 2010 through December 2015), and of monthly 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 small-trades data include all small trades in NYSE/AMEX common stocks, where trades are classified as small or large based on the procedure described in Hvidkjaer (2006); see Section 3.b. The price-improved TAQ trades include all trades in common stocks listed on all U.S. stock exchanges with a share price of at least $1, where trades are classified as retail-initiated based on the transaction price following the procedure described in Boehmer et al. (2020); see Section 3.c. Mutual funds data include all retail equity mutual funds reported in the TrimTabs dataset. The top panel considers households' net turnover (in millions): 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 turnover (in millions) for small trades in the TAQ dataset: the aggregate value of small buys, minus the aggregate value of small sells, divided by the market's total value. The third panel considers the net turnover (in millions) for price-improved trades in the TAQ dataset: the aggregate value of price-improved buys, minus the aggregate value of price-improved sells, divided by the market's total value. The bottom panel considers the net turnover (in thousands) for flows to equity mutual funds in the TrimTabs dataset: 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.

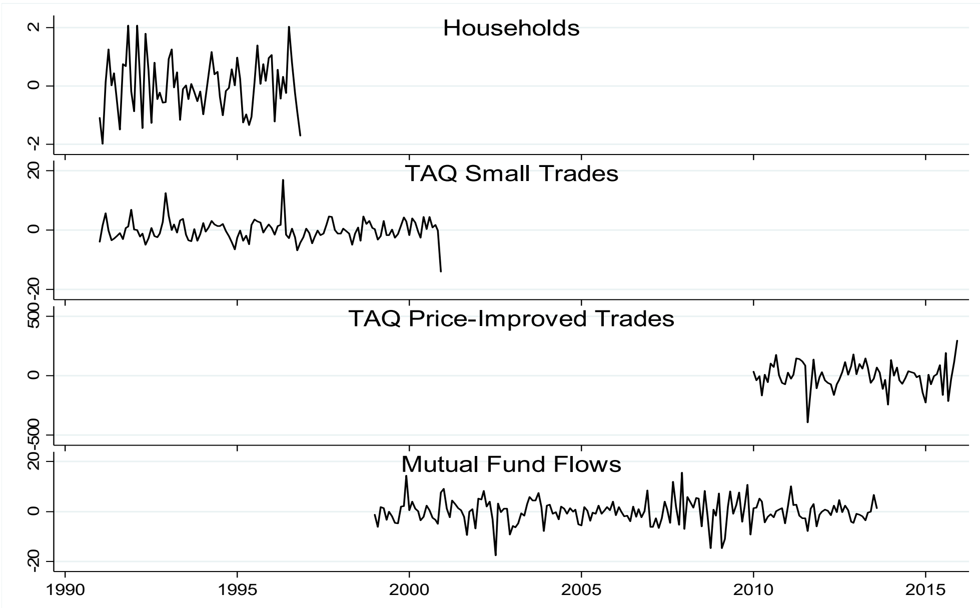
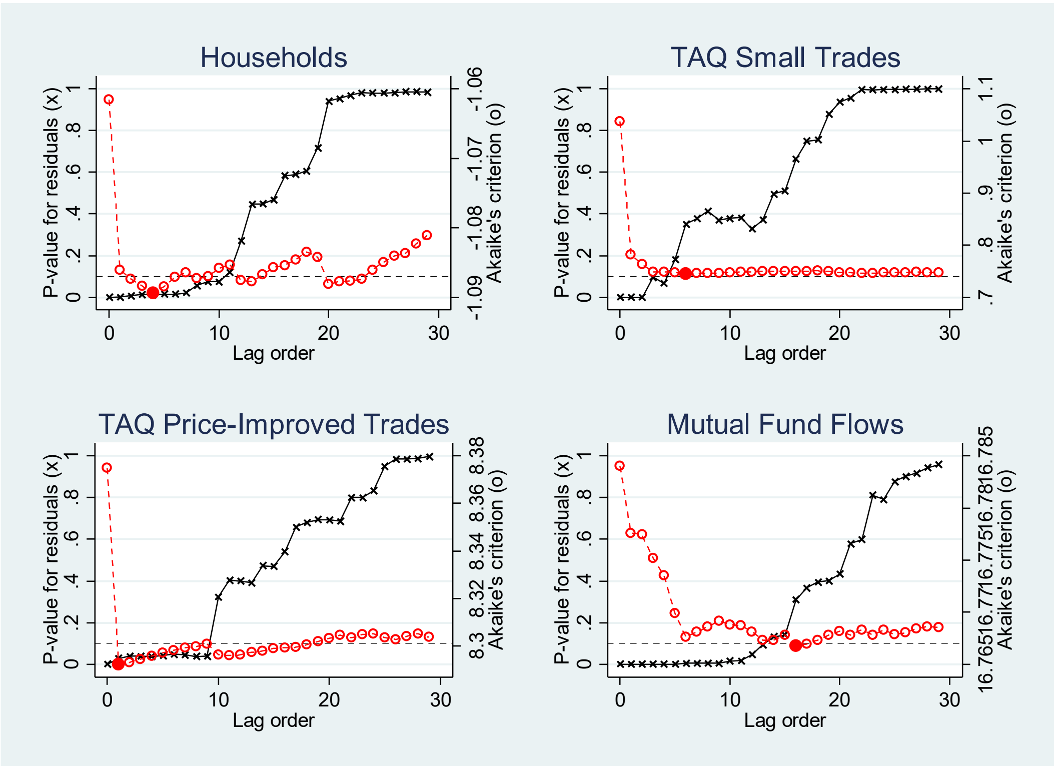


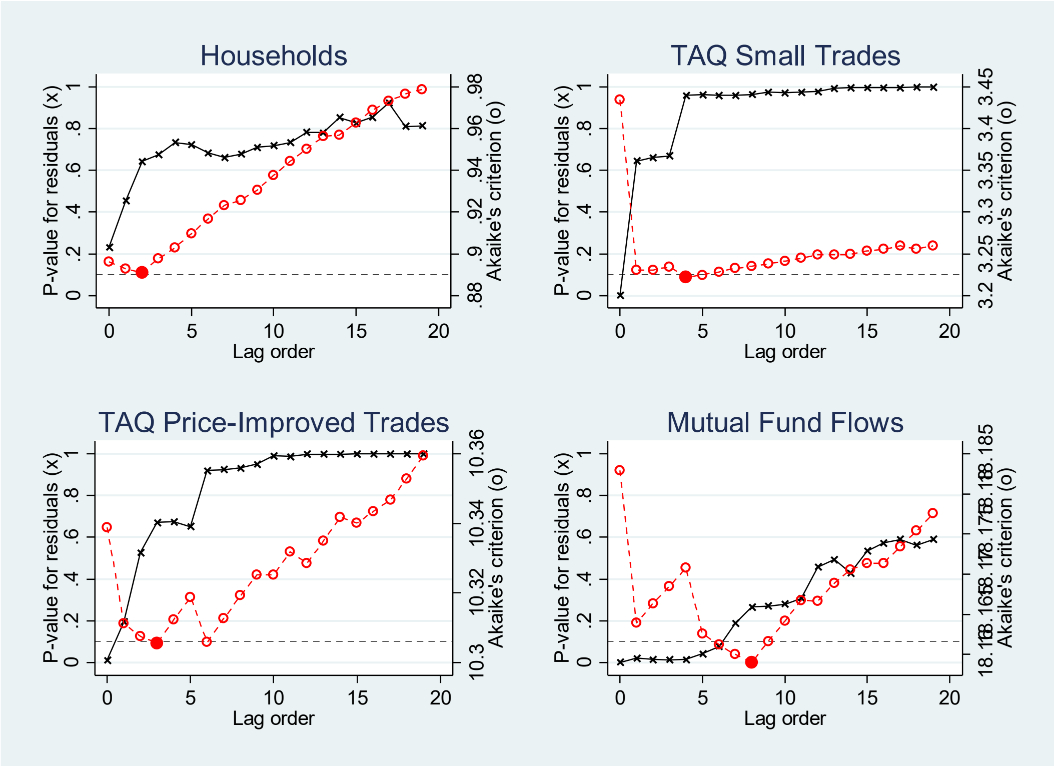
Figure 2: Lag-order selection

This figure displays the performance of autoregressive models fitted to households' aggregate trades (top left panel), to TAQ small trades (top right panel), to TAQ price-improved 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. 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 turnover for small trades in the TAQ dataset: the aggregate value of small buys, minus the aggregate value of small sells, divided by the market's total value. The lower left panel considers price-improved trades in the TAQ dataset: the aggregate value of buys with a transaction price slightly below the round penny, minus the aggregate value of sells with a transaction price slightly above the round penny, divided by the market's total value. The lower right panel considers the net turnover for flows to retail equity mutual funds in the TrimTabs dataset: 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 (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

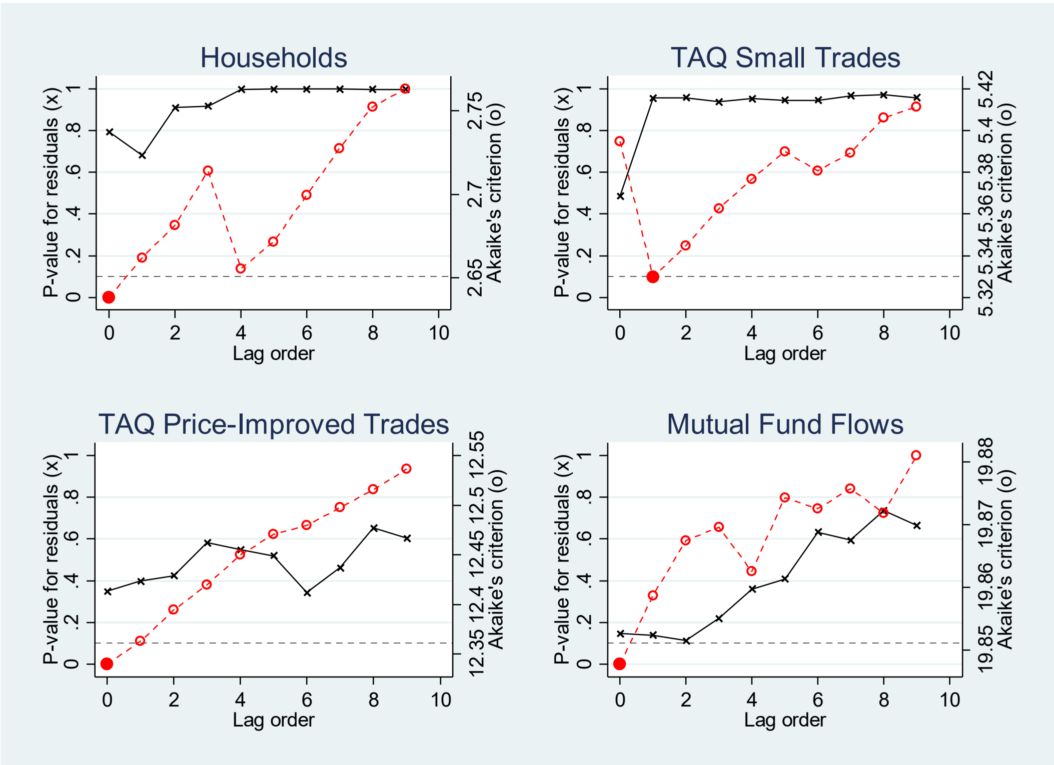


Figure 3: Fitting an AR(1) process to households' aggregate trades

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 turnover for small trades in the TAQ dataset: the aggregate value of small buys, minus the aggregate value of small sells, divided by the market's total value. The lower left panel considers price-improved trades in the TAQ dataset: the aggregate value of buys with a transaction price slightly below the round penny, minus the aggregate value of sells with a transaction price slightly above the round penny, divided by the market's total value. The lower right panel considers the net turnover for flows to retail equity mutual funds in the TrimTabs dataset: 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.

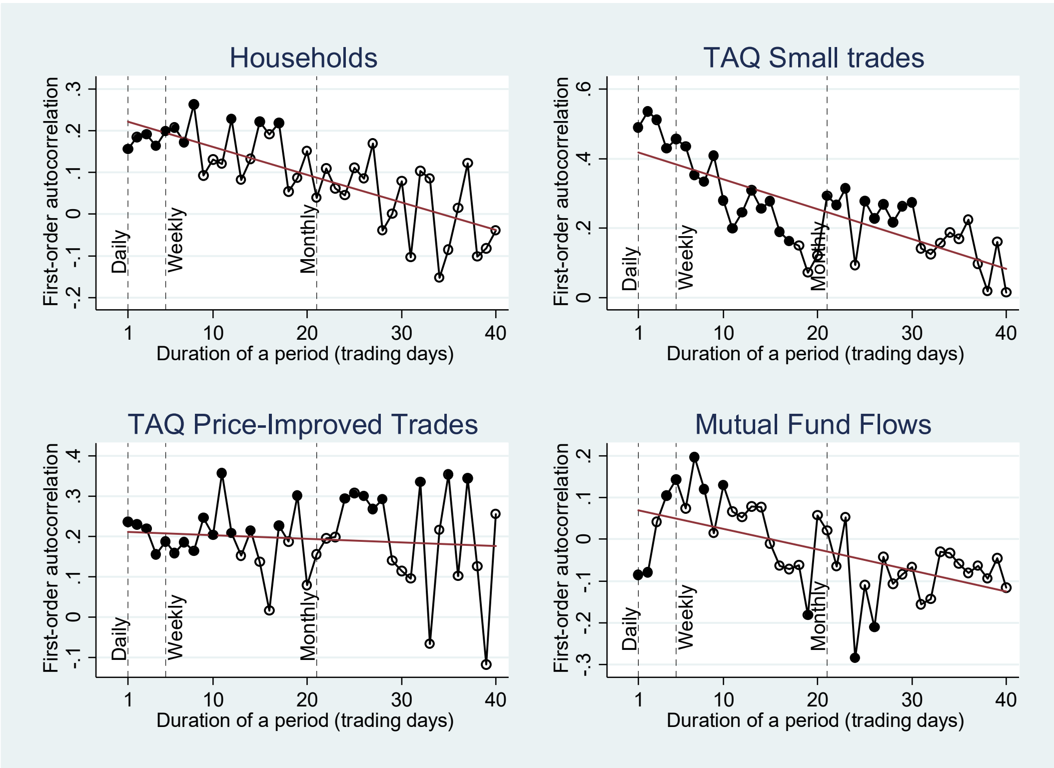


Figure 4: Histograms of households' daily aggregate trades

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 turnover for small trades in the TAQ dataset: the aggregate value of small buys, minus the aggregate value of small sells, divided by the market's total value. The third row considers price-improved trades in the TAQ dataset: the aggregate value of buys with a transaction price slightly below the round penny, minus the aggregate value of sells with a transaction price slightly above the round penny, divided by the market's total value. The bottom row considers the net turnover for flows to retail equity mutual funds in the TrimTabs dataset: 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.

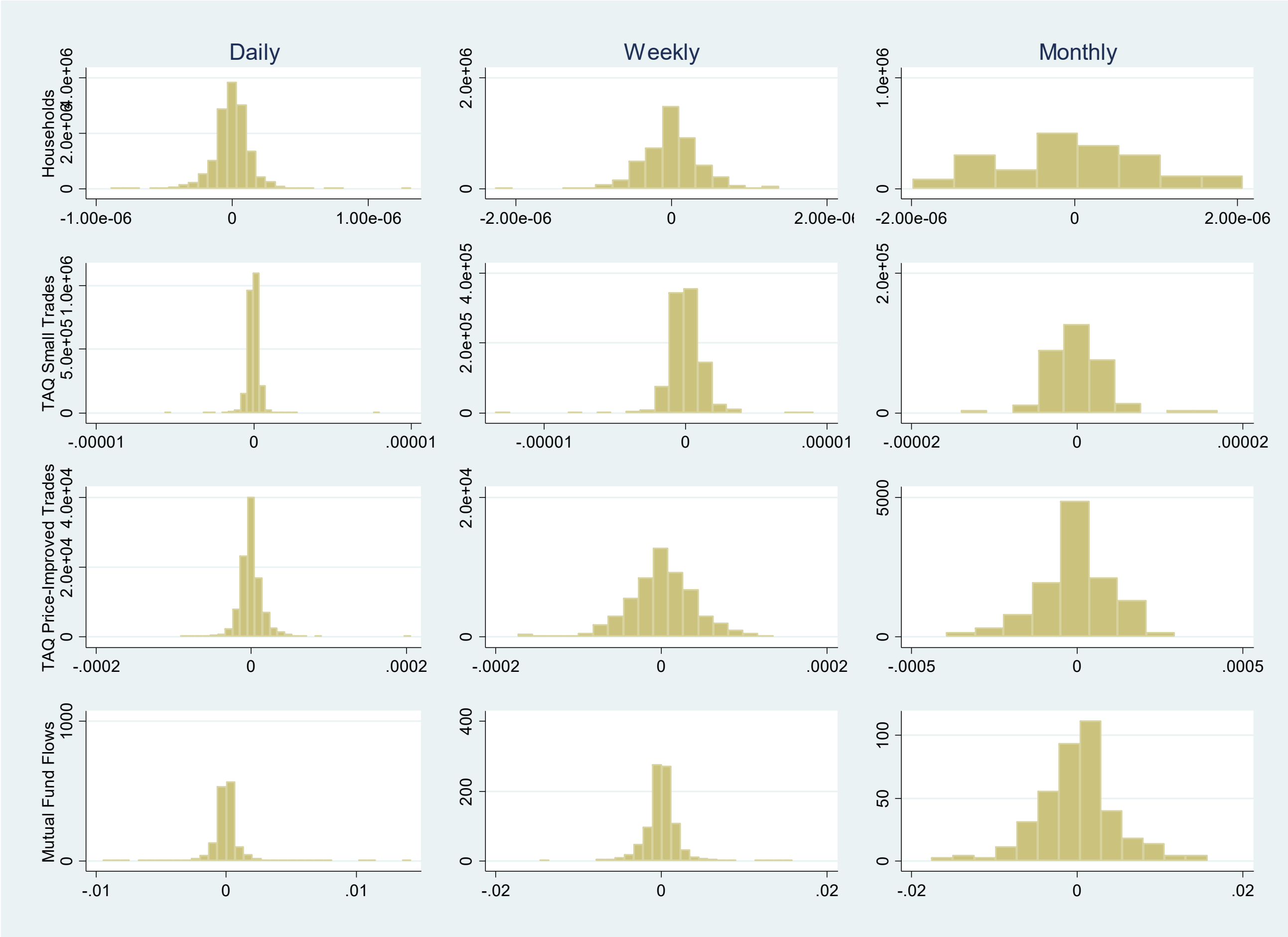


Figure 5: Probability (Q-Q) plots of households' aggregate trades

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 turnover for small trades in the TAQ dataset: the aggregate value of small buys, minus the aggregate value of small sells, divided by the total value of the market. The third row considers price-improved trades in the TAQ dataset: the aggregate value of buys with a transaction price slightly below the round penny, minus the aggregate value of sells with a transaction price slightly above the round penny, divided by the market's total value. The bottom row considers the net turnover for flows to retail equity mutual funds in the TrimTabs dataset: 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.

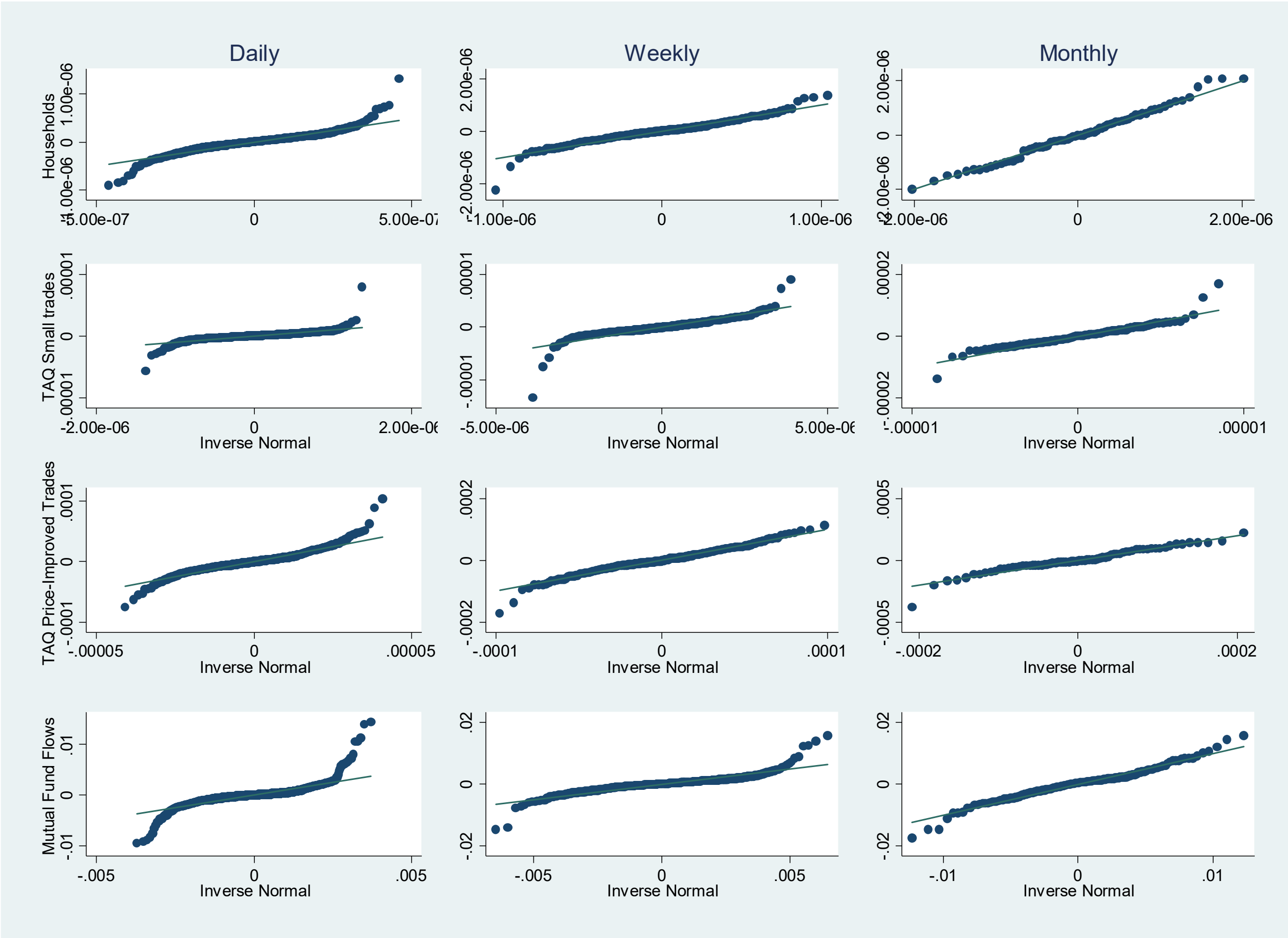
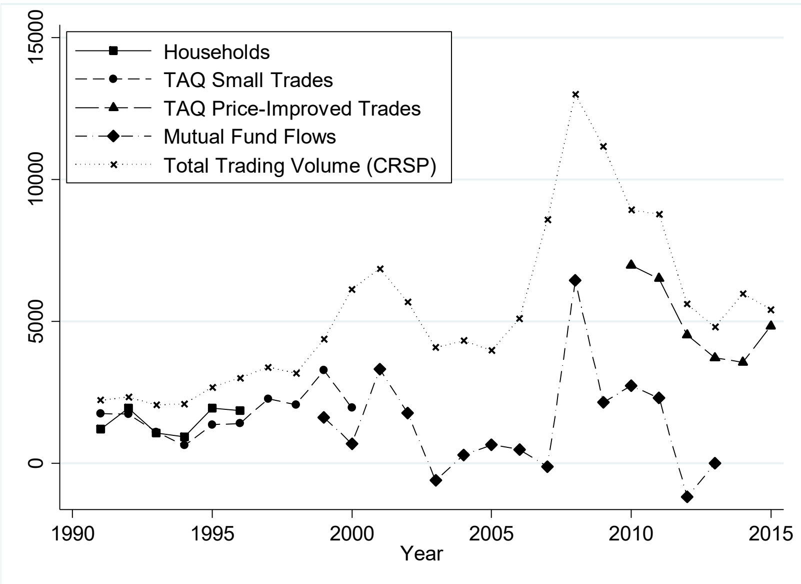


Figure 6: Estimating the intensity of noise trading for stock groups

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 (short-dashed lines, circles), price-improved TAQ trades (long-dashed lines, triangles), and mutual fund flows (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, price-improved 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 year 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. The dotted line marked with crosses in Panel A represents the standard deviation of weekly total turnover (CRSP trading volume divided by the market's total value) during the course of the year.

Panel A: Standard deviation of noise trading



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

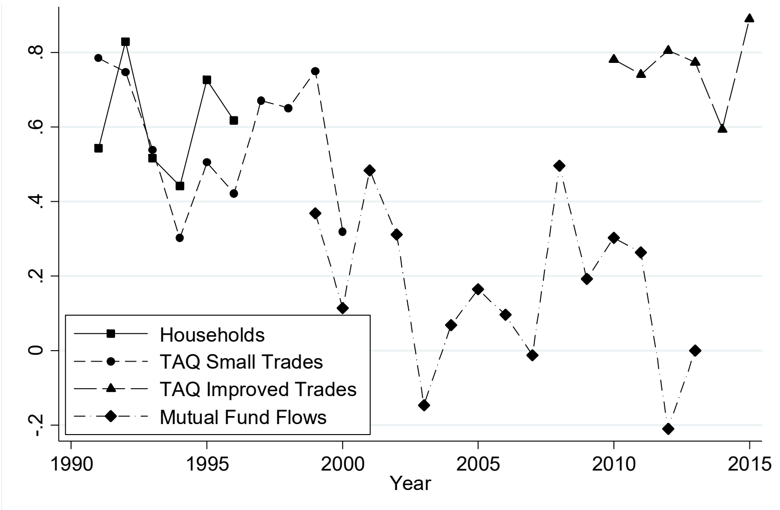


Figure 7: Estimating the intensity of noise trading for stock groups

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), small TAQ trades (short-dashed lines, circles), and price-improved TAQ trades (long-dashed lines, triangles). The x-axis represents deciles in terms of the stock characteristic variable, and ranges from 1 (e.g., smallest firms) to 10 (e.g., biggest firms). For each month, we sort stocks into deciles according to their capitalization, 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, 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, small TAQ trades and price-improved 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.

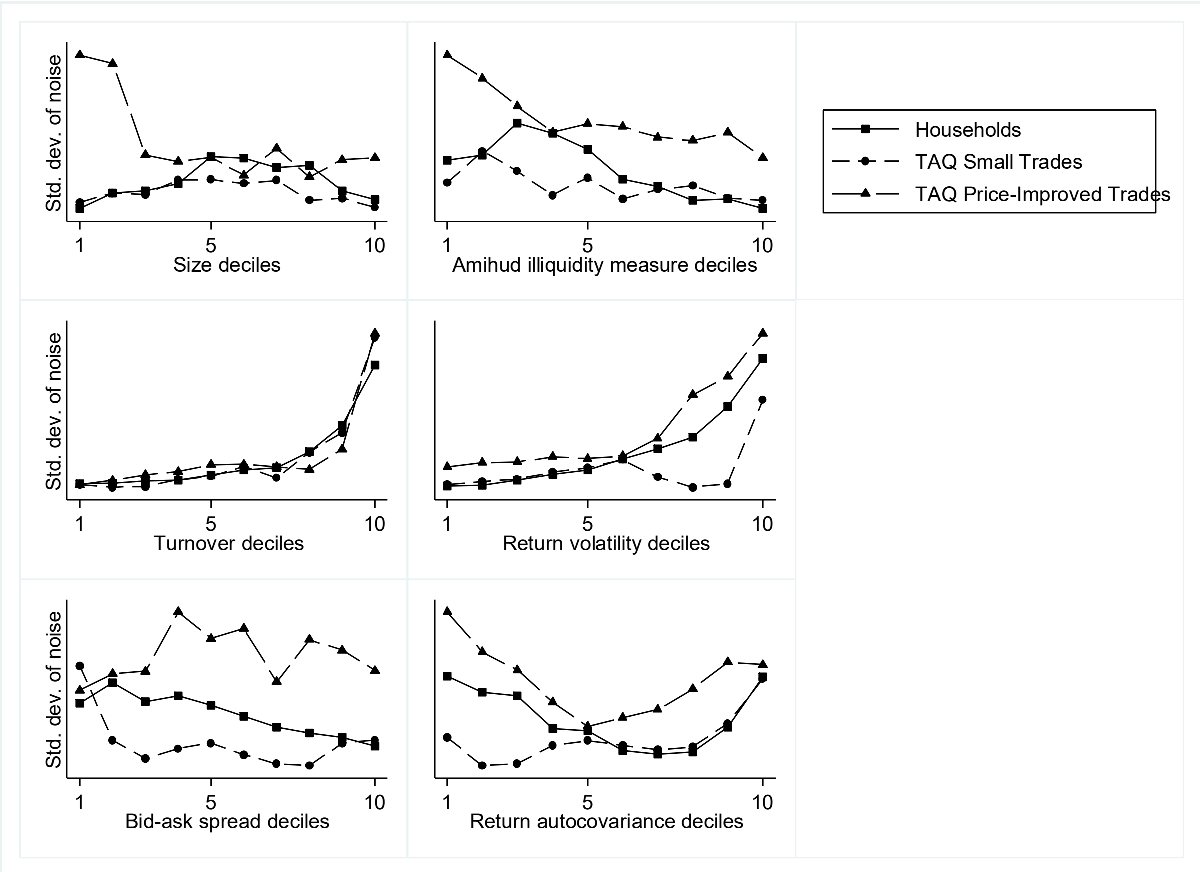


Table 1: Descriptive statistics for aggregate trades

This table presents summary statistics for the time series of households' daily trades at a large discount broker (from January 1991 through November 1996), of TAQ small and large trades (from January 1991 through December 2000), of price-improved TAQ trades (from January 2010 through December 2015), and of 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 3.b. The price-improved TAQ trades include all trades in common stocks listed on all U.S. stock exchanges with a share price of at least $1, where trades are classified as retail-initiated based on the transaction price following the procedure described in Boehmer et al. (2020); see Section 3.c. Mutual funds data include all retail equity mutual funds reported in the TrimTabs dataset. Panel A considers households' net turnover (in millions): the aggregate value of their buys, minus the aggregate value of their sells, divided by the market's total value. Panel B considers the net turnover (in millions) for small trades in the TAQ dataset: the aggregate value of small buys, minus the aggregate value of small sells, divided by the market's total value. Panel C does likewise for large TAQ trades. Panel D considers the net turnover (in millions) for price-improved trades in the TAQ dataset: the aggregate value of buys with a transaction price slightly below the round penny, minus the aggregate value of sells with a transaction price slightly above the round penny, divided by the market's total value. Panel F considers the net turnover (in thousands) for flows to retail equity mutual funds in the TrimTabs dataset: 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							
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: TAQ Small trades							
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 C: TAQ Large trades							
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 D: TAQ Improved trades							
Daily	1510	-90.795	0.000	-0.254	206.370	15.910	1.448	25.441
Weekly	313	-172.761	0.000	-0.441	135.183	41.955	-0.361	4.912
Monthly	72	-393.776	0.000	6.624	294.949	112.451	-0.528	4.421
Number of firms : 5,044								
	Panel E: Mutual Fund Flows							
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 (columns 2 and 3), and their residuals from a fitted AR(1) process (columns 4 and 5) 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 turnover for small trades in the TAQ dataset: the aggregate value of small buys, minus the aggregate value of small sells, divided by the market's total value. The third panel considers price-improved trades in the TAQ dataset: the aggregate value of buys with a transaction price slightly below the round penny, minus the aggregate value of sells with a transaction price slightly above the round penny, divided by the market's total value. The bottom panel considers the net turnover for flows to equity mutual funds in the TrimTabs dataset: 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			
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
	TAQ Small trades			
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
	TAQ Improved trades			
Daily	11.757	0.000	16.393	0.000
Weekly	3.988	0.000	10.799	0.000
Monthly	1.489	0.068	2.908	0.002
Quarterly	0.168	0.433	-0.521	0.699
	Mutual Fund Flows			
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, price-improved 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%
TAQ improved trades, 2010-2015							
Day	99.360	8.787	***	0.873	47%	1.746	94%
Week	476.121	9.093	***	4.329	50%	8.659	100%
Month	1151.786	8.492	***	9.781	32%	20.000	64%
Quarter	2697.837	8.022		22.000	26%	43.000	53%
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 six 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 (reported in $ millions), turnover, closing bid–ask spread (reported in basis points), Amihud illiquidity ratio (reported multiplied by 1 million), return volatility, and return autocovariance (reported in basis points). All variables are estimated every month from daily observations. For a stock's capitalization, turnover, 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), or price-improved TAQ trades (columns 12–16) using monthly data. 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. The row labelled ”1-10” at the bottom of each stock characteristic panel reports the difference in the estimate of the standard deviation of noise trading between deciles 1 and 10, with stars indicating the results of a standard t-test for this difference being significatively different from zero. We use *, **, and *** to indicate statistical significance at (respectively) the 10%, 5%, and 1% level.

Decile	Households					TAQ Small trades					TAQ Improved trades				
	Median value of sorting variable	SD of trades (x1M)	Regression coef. *b*	Lower bound on SD of (x1M)	Lower bound on SD of (% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD of (x1M)	Lower bound on SD of (% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD of (x1M)	Lower bound on SD of (% of SD of total trades)
By stock size															
1	5.24	19.28	114.41	2206.13	0.09	5.55	89.87	39.01	3505.54	0.15	19.96	8739.76	3.99	34857.01	0.90
2	12.37	10.52	529.40	5571.40	0.24	14.24	31.33	178.36	5587.48	0.26	53.32	7757.43	4.27	33118.75	0.97
3	21.67	9.50	626.91	5953.41	0.27	26.03	18.74	279.60	5240.15	0.25	110.65	2837.49	4.81	13635.29	0.86
4	34.57	10.10	749.45	7570.97	0.35	43.26	15.52	537.38	8340.77	0.31	204.33	2511.38	4.89	12269.04	0.84
5	54.89	12.15	1099.31	13357.78	0.57	69.99	9.45	889.60	8405.68	0.33	351.09	2226.42	5.88	13082.51	0.79
6	89.24	9.52	1355.67	12908.08	0.56	113.91	4.32	1740.61	7521.31	0.29	644.55	1963.33	4.78	9375.78	0.54
7	147.04	7.42	1472.93	10923.42	0.50	189.40	3.27	2522.10	8254.33	0.36	1129.23	2113.29	7.13	15058.57	0.65
8	280.93	7.42	1532.77	11367.09	0.56	351.25	1.38	2874.97	3963.37	0.20	2104.81	1300.57	6.89	8963.48	0.41
9	711.53	3.62	1680.30	6077.67	0.39	824.68	0.64	7048.16	4480.06	0.33	4574.06	1105.88	11.43	12641.61	0.50
10	6306.94	1.86	2197.11	4092.48	0.34	3388.98	0.14	17798.83	2495.94	0.30	18089.13	1038.27	12.62	13105.57	0.70
1-10				-1886.35					1009.6***					21751.45***	
By stock turnover															
1	0.02	0.63	174.57	110.20	0.10	0.03	0.13	85.30	11.32	0.02	0.05	120.87	-1.24	-150.04	-0.14
2	0.06	0.85	376.58	318.73	0.09	0.07	0.33	-2505.35	-814.36	-0.36	0.13	326.37	3.41	1113.31	0.39
3	0.10	1.02	956.36	977.45	0.19	0.11	0.32	-2109.73	-665.53	-0.22	0.23	434.84	6.08	2642.56	0.55
4	0.14	1.32	849.21	1120.99	0.16	0.16	0.13	9624.83	1291.86	0.31	0.34	452.95	7.79	3530.11	0.54
5	0.19	1.20	2216.93	2656.40	0.29	0.22	0.10	23079.16	2195.54	0.40	0.45	493.71	11.15	5505.00	0.61
6	0.26	1.93	2104.06	4061.78	0.34	0.30	0.16	31698.88	4980.00	0.61	0.59	740.82	7.56	5599.10	0.47
7	0.34	3.53	1307.67	4614.55	0.29	0.40	0.31	5367.30	1689.34	0.15	0.76	1024.93	4.76	4879.70	0.34
8	0.47	3.80	2419.67	9186.96	0.40	0.55	0.53	17238.86	9071.81	0.48	0.98	1417.96	2.93	4158.61	0.22
9	0.74	8.45	1959.13	16554.54	0.46	0.83	1.03	14210.12	14616.22	0.44	1.36	2449.79	4.08	9986.22	0.39
10	1.71	21.50	1575.82	33886.86	0.44	1.64	4.14	10063.06	41658.80	0.67	2.40	7329.69	5.82	42658.43	0.80
1-10				-33776.66***					-41647.47***					-42808.47***	

Decile	Households					TAQ Small trades					TAQ Improved trades				
	Median value of sorting variable	SD of trades (x1M)	Regression coef. *b*	Lower bound on SD (x1M)	Lower bound on SD (% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD (x1M)	Lower bound on SD (% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD (x1M)	Lower bound on SD (% of SD of total trades)
By stock bid-ask spread															
1	72.21	7.71	1241.52	9573.28	0.19	83.95	0.59	28154.30	16733.05	0.57	2.11	1120.40	10.72	12010.39	0.64
2	143.18	5.95	2261.80	13449.41	0.48	134.93	0.57	4048.68	2296.17	0.20	3.47	1811.12	8.44	15292.69	0.64
3	199.82	6.20	1585.81	9828.27	0.53	184.90	1.73	-752.91	-1300.15	-0.13	5.22	1987.25	7.92	15741.06	0.62
4	267.91	5.08	2144.59	10895.33	0.70	240.95	1.73	344.99	595.91	0.06	7.63	2677.99	10.20	27304.06	0.83
5	346.94	7.31	1237.84	9043.59	0.57	299.58	1.65	1036.54	1707.76	0.16	11.43	2593.24	8.53	22132.76	0.79
6	439.65	6.56	1052.48	6903.36	0.43	376.55	2.64	-212.40	-560.12	-0.04	18.73	3679.66	6.54	24070.70	0.75
7	558.15	8.48	558.79	4738.15	0.29	468.22	4.48	-524.20	-2350.27	-0.17	34.95	3002.88	4.54	13636.70	0.76
8	738.14	8.42	439.59	3701.84	0.27	609.54	7.34	-367.16	-2693.57	-0.18	76.42	3638.62	6.02	21907.62	0.89
9	1015.55	5.00	562.43	2812.96	0.23	860.98	28.01	61.08	1710.72	0.14	158.59	3100.35	6.41	19876.97	0.87
10	1533.48	4.36	268.71	1170.93	0.14	1442.46	23.63	95.15	2248.28	0.24	364.30	3705.56	4.29	15885.25	0.97
1-10				8402.36***					14484.77***					-3874.86***	
By stock Amihud illiquidity measure															
1	0.00	2.09	2151.54	4500.26	0.33	0.00	0.14	17673.60	2533.73	0.29	0.00	1081.08	12.56	13573.32	0.68
2	0.01	3.10	1589.54	4931.83	0.34	0.01	0.66	7998.34	5253.44	0.40	0.00	1000.51	11.60	11602.91	0.54
3	0.03	4.82	1597.33	7705.66	0.53	0.02	1.33	2645.36	3515.69	0.22	0.00	991.38	9.23	9149.21	0.46
4	0.08	4.78	1422.63	6793.22	0.56	0.06	2.26	644.82	1455.19	0.11	0.00	1039.69	6.66	6925.30	0.47
5	0.22	4.31	1264.51	5454.56	0.44	0.15	3.82	784.24	2999.10	0.28	0.00	1131.08	6.75	7633.71	0.60
6	0.53	3.75	756.65	2841.20	0.30	0.34	4.79	243.85	1168.83	0.14	0.01	1345.93	5.48	7377.65	0.71
7	1.31	3.59	624.04	2237.62	0.33	0.82	9.54	204.27	1948.46	0.32	0.03	1516.77	4.28	6488.95	0.79
8	3.03	3.20	314.91	1008.86	0.23	1.97	8.54	264.89	2262.73	0.30	0.09	1479.38	4.18	6179.65	0.84
9	8.27	2.85	408.55	1163.12	0.30	5.21	13.48	91.11	1228.50	0.23	0.52	1529.13	4.50	6888.53	0.80
10	36.68	12.21	23.73	289.64	0.07	28.92	25.94	38.26	992.46	0.23	5.84	1040.94	4.51	4695.64	0.84
1-10				4210.63***					1541.27***					8877.68***	

	Households					TAQ Small trades					TAQ Improved trades				
Decile	Median value of sorting variable	SD of trades (x1M)	Regression coef. *b*	Lower bound on SD		Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD		Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD	
				(x1M)	(% of SD of total trades)				(x1M)	(% of SD of total trades)				(x1M)	(% of SD of total trades)
By stock return volatility															
1	1.03	1.10	1589.84	1748.82	0.31	1.03	0.14	17097.52	2451.10	0.55	0.90	1010.20	11.13	11248.05	0.67
2	1.47	1.46	1266.59	1847.99	0.23	1.51	0.10	41449.22	3995.07	0.61	1.22	2064.44	6.55	13523.99	0.74
3	1.89	2.80	1685.55	4711.51	0.41	1.94	0.18	27543.91	5014.07	0.54	1.45	2487.54	5.60	13930.28	0.61
4	2.34	4.58	1649.30	7552.35	0.34	2.37	0.33	26146.08	8667.63	0.59	1.71	2777.35	5.92	16438.76	0.66
5	2.80	7.40	1343.66	9941.16	0.35	2.84	0.55	20041.36	10933.99	0.45	1.99	2916.38	5.35	15591.92	0.57
6	3.34	10.87	1402.67	15246.15	0.38	3.41	1.05	14138.62	14776.11	0.44	2.29	4168.17	4.05	16897.49	0.58
7	3.98	14.64	1409.66	20633.93	0.41	4.06	2.22	2673.18	5932.65	0.15	2.65	6378.79	4.08	25996.36	0.70
8	4.89	18.27	1443.71	26373.04	0.37	4.95	3.85	272.52	1049.01	0.02	3.13	9476.04	5.06	47937.86	0.83
9	6.30	31.33	1339.36	41958.78	0.52	6.41	10.61	248.57	2637.65	0.04	3.93	10928.35	5.24	57283.00	0.87
10	10.13	75.39	881.85	66482.60	0.48	10.21	49.18	920.84	45282.39	0.26	5.84	16416.21	4.82	79092.85	0.71
1-10				-64733.78***					-42831.3***					-67844.8***	
By stock return autocovariance															
1	-29.28	30.07	1043.06	31369.16	0.43	-31.61	40.54	230.60	9348.16	0.09	-5.87	13077.31	4.16	54423.12	0.69
2	-10.70	15.19	1686.68	25617.28	0.42	-10.05	8.93	-85.61	-764.83	-0.01	-2.19	8057.30	4.96	40003.95	0.83
3	-5.22	12.49	1950.53	24357.56	0.47	-4.83	5.90	-12.35	-72.90	0.00	-1.13	6044.27	5.56	33636.27	0.84
4	-2.77	7.98	1577.94	12598.54	0.34	-2.58	2.01	3204.98	6436.37	0.26	-0.62	3112.23	7.09	22052.34	0.72
5	-1.54	5.82	2006.82	11682.28	0.44	-1.36	0.59	13913.84	8219.63	0.49	-0.35	1811.29	7.36	13323.61	0.54
6	-0.75	2.63	1804.05	4740.02	0.31	-0.64	0.31	21289.28	6508.01	0.59	-0.16	2181.87	7.59	16558.07	0.73
7	-0.27	1.72	1940.95	3330.87	0.37	-0.20	0.14	34323.48	4855.30	0.48	0.00	2109.64	9.18	19375.37	0.82
8	0.04	2.21	1823.95	4025.66	0.42	0.10	0.17	34892.50	5926.02	0.56	0.21	3320.57	8.03	26656.33	0.82
9	0.55	4.97	2609.27	12964.26	0.72	0.72	0.47	30223.67	14288.46	0.64	0.65	5149.75	7.06	36379.74	0.82
10	2.84	21.76	1432.46	31164.14	0.57	3.29	5.60	5430.81	30399.20	0.53	2.44	6500.43	5.46	35504.56	0.68
1-10				205.02					-21051.05***					18918.56***	

Internet Appendix to

Noise Traders Incarnate:
Describing a Realistic Noise Trading Process

Table of Contents

Internet Appendix A: Are the trades in our sample 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 Appendix, we first discuss the evidence presented in the literature regarding this premise. Then we check that our data display similar features.

A.1: Evidence in the literature: Are retail trades a good proxy for noise trades?

A.1.a: Retail stock trades

- *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.[1]

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).[2] 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

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

[2] 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).

are rare (Coval et al. 2005). We aggregate trades across all individuals in order to focus on the average investor's behavior.

- *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 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.

- *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.

- *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.

A.1.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).[3] 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).[4] 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 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.[5] 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.[6] 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 response to

[3] 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).

[4] 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).

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

[6] 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.

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

A.1.c: Summary of the evidence in the literature

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.

A.2: Evidence based on our samples: 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 but weakly correlated with fundamentals, (b) are cross-correlated and (c) perform poorly.

A.2.a: Correlation between noise trades and fundamentals

We assess the extent to which noise trading and fundamentals are correlated. 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 datasets. 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.[7] Finally, we estimate a panel regression model of net buys on (contemporaneous)

[7] Our findings are not sensitive to the choice of these filters.

earnings surprise deciles. The regression includes firm, quarter, and month-of-the-year fixed effects; standard errors are clustered by firm.

[[INSERT Table A.1 about Here]]

Results for households are reported in Table A.1. The estimated coefficients for net buys vary both in sign and in statistical significance. These coefficients are negative in the brokerage dataset (Panel A) but positive in the TAQ dataset (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 × [10 − 1]/(1 million)) decrease in net turnover over the 40-day pre-announcement window, or one fifth (= [5 × 10−6]/[25.28 × 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 × [10 − 1]/(1 million)) increase in net turnover on the announcement day, or about one hundredth (= [0.7 × 10−6]/[61 × 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 R2 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.8 That is, we estimate the difference between actual and 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.9 The regressions include quarter fixed effects.

Our results, displayed in Panel C of Table A.1, 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.

To summarize, we find a weak correlation between retail trades (households' trades, small TAQ trades and fund flows) and fundamentals, consistent with how noise trades are defined in NREE.

---

[8] 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

[9] 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.

A.2.b: Correlation among trades

Here we 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.2 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.2 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.2 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.

A.2.c: Performance of trades

We turn to investigating 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.3 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).[10]

[[INSERT Table A.3 about Here]]

Panel B of Table A.3 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.3 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.

To summarize, 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.

[10] 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.

Table A.1: 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 dataset: 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 A.2: Correlation among trades

This table examines whether households' trades (Panels A and B), small TAQ trades (Panel C), price-improved trades (Panel D), and mutual fund flows (Panel E) 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 and D are similar to Panel A and reports results from a stock–month panel regression of a stock's aggregate trades—measured in the TAQ dataset 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. Panel C 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 reports estimates for price-improved trades. The results indicate that a given stock is more likely to be bought at times when other stocks are bought. Panel E 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 dataset 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 classified by trade size

	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 stocks traded in TAQ classified by transaction price

	Price Improved Trades
Mean Dep. Var.	0.890***
	(7.480)
Mkt return	0.000
	(1.207)
Firms	5,044
Obs.	261,681
R-square	0.003

Panel E: 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 A3: 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)

Internet 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."

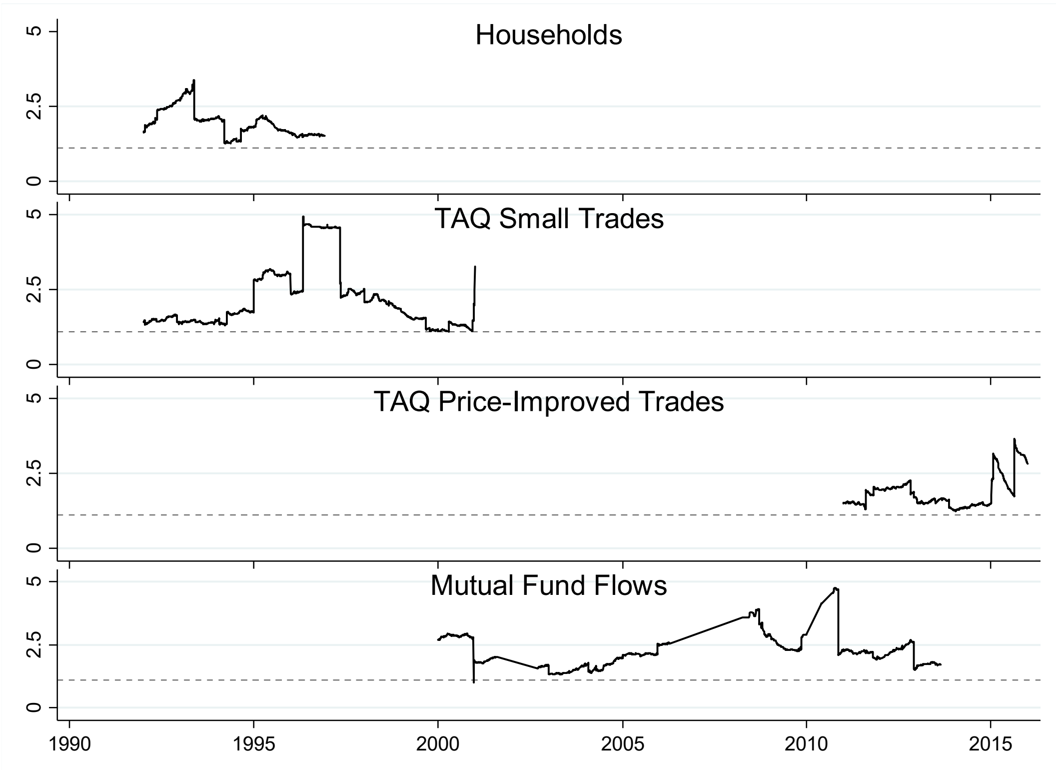
2. $\rho_X \approx \rho_Y$. In words: The difference between the component of noise trades that we are missing (institutional noise trades) and the component of retail trades that are not noise (informed retail trades) has the same first-order autocorrelation coefficient as retail noise trades. That 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, $\mathrm{cov}(z_t, z_{t-1})$ and $\mathrm{var}(z_{t-1})$) in the same way and so "washes out".

Internet Appendix C: The kurtosis of retail trades over time

Figure C.1: The kurtosis of retail trades over time

The graphs in this figure display time series of the log of the kurtosis of households' trades at a large discount broker (from January 1991 through November 1996), of small TAQ trades (from January 1991 through December 2000), of price-improved TAQ trades (from January 2010 through December 2015), and of flows to retail equity mutual funds (from January 1999 through August 2013). Each series' kurtosis is estimated over a one (calendar)-year rolling window using daily data. The horizontal dashed line marks log(3), where a kurtosis in excess of 3 is generally considered symptomatic of a leptokurtic (fat-tailed) distribution.



Internet Appendix D: Estimating the intensity of noise trading for stock groups at daily and weekly frequencies

In the paper, we report bounds on the standard deviation of noise trades estimated across groups of stocks using monthly data, i.e., based on monthly regressions. Here, we report estimates based on daily and weekly regressions.

Table D.1: Estimating the intensity of noise trading for stock groups at daily frequency

This table reports bounds on the standard deviation of noise trades estimated across groups of stocks sorted on various characteristics, using daily data. For each month, we sort stocks into deciles according to their capitalization (reported in $ millions), turnover, closing bid–ask spread (reported in basis points), Amihud illiquidity ratio (reported multiplied by 1 million), return volatility, and return autocovariance (reported in basis points). All variables are estimated every month from daily observations. For a stock's capitalization, turnover, 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), or price-improved TAQ trades (columns 12–16). This table reports estimates from daily regressions. 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. The row labelled "1-10" at the bottom of each stock characteristic panel reports the difference in the estimate of the standard deviation of noise trading between deciles 1 and 10, with stars indicating the results of a standard t-test for this difference being significatively different from zero. We use *, **, and *** to indicate statistical significance at (respectively) the 10%, 5%, and 1% level.

	Households					TAQ Small trades					TAQ Improved trades				
Decile	Median value of sorting variable	SD of trades (x1M)	Regression coef. *b*	Lower bound on SD (x1M)	Lower bound on SD (% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD (x1M)	Lower bound on SD (% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD (x1M)	Lower bound on SD (% of SD of total trades)
By stock size															
1	5.25	4.91	20.01	98.26	0.08	5.57	6.74	120.51	812.79	0.36	20.06	844.22	3.85	3249.05	0.85
2	11.98	1.41	130.34	184.15	0.15	14.26	2.38	345.06	820.74	0.45	53.30	601.50	4.29	2582.67	0.92
3	20.94	1.63	77.83	127.12	0.11	25.93	1.28	433.04	554.06	0.38	111.10	224.46	4.66	1046.00	0.46
4	34.45	1.28	208.52	266.81	0.22	43.21	0.94	732.30	686.24	0.38	206.93	164.50	5.76	947.11	0.47
5	54.64	1.56	241.86	377.01	0.29	69.99	0.59	1095.63	649.12	0.39	359.16	155.23	6.74	1045.54	0.54
6	87.93	1.04	398.08	414.89	0.31	113.58	0.28	2144.96	593.98	0.37	661.72	149.19	7.15	1066.85	0.53
7	148.05	0.85	485.63	413.96	0.31	189.34	0.29	1283.10	366.68	0.25	1157.55	140.03	8.83	1235.80	0.56
8	281.43	0.73	644.09	473.02	0.38	351.23	0.08	5032.34	408.81	0.32	2159.05	99.59	12.32	1226.54	0.57
9	710.25	0.43	802.86	341.44	0.33	823.39	0.04	12232.32	439.25	0.42	4705.68	87.44	14.98	1310.08	0.61
10	6308.62	0.22	1180.89	259.00	0.34	3381.93	0.01	43194.93	354.21	0.49	18507.94	82.74	14.25	1179.00	0.77
1-10				-160.74***					458.58***					2070.05***	
By stock turnover															
1	0.02	0.15	1.52	0.23	0.00	0.03	0.01	738.27	7.17	0.12	0.05	6.74	1.31	8.80	0.07
2	0.06	0.20	34.73	7.10	0.04	0.07	0.02	-583.56	-10.72	-0.06	0.13	19.30	5.90	113.79	0.39
3	0.10	0.16	185.19	29.70	0.09	0.11	0.02	-474.32	-7.39	-0.03	0.23	28.73	9.71	278.85	0.51
4	0.14	0.16	379.84	61.96	0.15	0.16	0.01	15148.20	107.13	0.31	0.34	36.71	13.35	490.21	0.63
5	0.19	0.17	562.36	96.32	0.18	0.22	0.01	33810.98	199.68	0.46	0.45	41.93	15.71	658.77	0.67
6	0.26	0.25	651.11	164.38	0.24	0.30	0.01	32678.92	303.31	0.51	0.59	54.15	14.10	763.81	0.63
7	0.34	0.39	499.83	194.43	0.21	0.40	0.02	11161.75	205.93	0.26	0.76	74.70	11.75	877.82	0.59
8	0.48	0.47	809.68	377.89	0.28	0.55	0.03	17906.00	523.89	0.41	0.98	102.84	9.65	992.81	0.52
9	0.74	1.00	649.57	647.47	0.30	0.83	0.06	14232.04	811.25	0.39	1.36	169.37	9.10	1540.69	0.60
10	1.71	2.09	896.96	1871.49	0.39	1.65	0.27	6405.78	1708.07	0.38	2.40	490.72	7.80	3827.12	0.76
1-10				-1871.26***					-1700.9***					-3818.32***	

Table E.1 (continuing)

Decile	Households					TAQ Small trades					TAQ Improved trades				
	Median value of sorting variable	SD of trades (x1M)	Regression coef. *b*	Lower bound on SD (x1M)	Lower bound on SD (% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD (x1M)	Lower bound on SD (% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD (x1M)	Lower bound on SD (% of SD of total trades)
By stock bid-ask spread															
1	73.50	0.98	448.19	440.13	0.17	83.68	0.03	30781.63	998.65	0.57	2.10	82.52	13.12	1083.00	0.71
2	143.18	0.63	807.38	506.84	0.32	134.04	0.04	3499.10	153.78	0.20	3.47	119.17	11.11	1324.49	0.69
3	200.70	0.72	481.32	346.68	0.29	183.75	0.14	-180.99	-25.07	-0.03	5.22	136.13	10.29	1400.56	0.67
4	267.66	0.80	324.87	260.16	0.26	238.56	0.13	497.05	63.40	0.08	7.53	159.56	10.61	1693.36	0.75
5	348.40	0.99	223.64	222.10	0.22	299.27	0.09	1858.06	166.19	0.21	11.33	176.56	9.30	1642.67	0.71
6	438.33	0.83	177.77	146.82	0.16	373.97	0.16	330.07	52.03	0.06	18.65	257.16	7.69	1978.86	0.69
7	555.63	1.15	155.35	178.80	0.20	468.22	0.24	-1.38	-0.32	0.00	34.74	210.43	5.01	1053.68	0.54
8	723.49	1.19	86.09	102.24	0.14	607.51	0.39	-71.90	-28.03	-0.03	76.46	250.40	5.15	1289.11	0.62
9	1001.06	0.97	86.97	84.40	0.12	858.99	2.19	116.13	254.44	0.31	158.51	263.08	5.62	1477.80	0.77
10	1501.78	0.84	88.33	74.23	0.15	1442.26	1.72	117.40	202.32	0.29	363.73	270.05	4.55	1228.14	0.92
1-10				365.9***					796.33***					-145.14***	
By stock Amihud illiquidity measure															
1	0.00	0.25	1157.69	285.52	0.33	0.00	0.01	44742.70	376.25	0.49	0.00	85.51	14.71	1257.72	0.76
2	0.01	0.33	741.40	248.08	0.29	0.01	0.04	11445.15	420.95	0.47	0.00	78.46	14.03	1101.14	0.64
3	0.03	0.49	626.51	307.07	0.35	0.02	0.08	4122.48	315.09	0.33	0.00	78.87	12.62	995.64	0.56
4	0.08	0.60	366.15	220.81	0.30	0.06	0.14	1803.64	248.68	0.29	0.00	78.78	10.49	826.65	0.53
5	0.21	0.54	297.79	159.90	0.23	0.15	0.27	1187.27	314.65	0.42	0.00	83.70	8.59	719.12	0.48
6	0.53	0.52	192.68	100.11	0.18	0.34	0.32	589.04	187.22	0.34	0.01	96.96	5.74	556.63	0.40
7	1.26	0.51	125.14	63.59	0.15	0.82	0.81	195.17	159.03	0.34	0.03	108.53	4.45	483.28	0.35
8	3.02	0.48	103.92	49.94	0.16	1.98	0.66	392.16	259.76	0.44	0.09	137.10	4.53	621.40	0.49
9	8.04	0.54	65.21	35.52	0.13	5.34	0.91	197.72	179.23	0.41	0.52	151.12	4.08	616.70	0.66
10	38.17	3.35	9.53	31.89	0.11	28.85	1.82	74.11	135.09	0.24	5.83	100.88	4.05	408.92	0.73
1-10				253.63***					241.16***					848.8***	

Table E.1 (continuing)

	Households					TAQ Small trades					TAQ Improved trades				
Decile	Median value of sorting variable	SD of trades (x1M)	Regression coef. *b*	Lower bound on SD (x1M)	Lower bound on SD (% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD (x1M)	Lower bound on SD (% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD (x1M)	Lower bound on SD (% of SD of total trades)
By stock return volatility															
1	1.02	0.19	288.21	54.64	0.14	1.03	0.01	24117.12	177.07	0.45	0.90	62.01	13.26	822.45	0.67
2	1.47	0.20	552.30	113.12	0.21	1.51	0.01	52911.33	284.09	0.58	1.22	117.41	8.37	982.63	0.66
3	1.89	0.34	697.49	234.39	0.31	1.94	0.01	36409.21	355.44	0.54	1.45	140.03	7.60	1063.74	0.61
4	2.34	0.72	428.84	308.16	0.22	2.37	0.02	32577.72	577.65	0.56	1.72	164.33	8.29	1361.71	0.68
5	2.81	0.98	472.50	462.58	0.26	2.84	0.03	24322.01	736.30	0.46	2.00	181.67	7.83	1421.96	0.64
6	3.34	1.15	717.97	826.36	0.33	3.41	0.06	15889.72	952.97	0.43	2.30	272.02	6.58	1790.48	0.68
7	3.99	1.63	691.67	1128.69	0.35	4.05	0.12	5123.88	590.36	0.20	2.65	380.49	5.87	2234.59	0.67
8	4.89	2.05	628.85	1291.12	0.30	4.95	0.22	2358.78	517.42	0.12	3.13	552.50	6.15	3399.38	0.78
9	6.32	3.50	572.00	2004.77	0.39	6.41	0.59	984.25	581.74	0.11	3.93	740.22	6.48	4798.19	0.81
10	10.13	7.03	557.45	3916.48	0.43	10.21	3.26	802.21	2614.83	0.22	5.84	1234.91	6.38	7873.91	0.70
1-10				-3861.84***					-2437.76***					-7051.46***	
By stock return autocovariance															
1	-30.64	3.18	463.37	1474.34	0.32	-31.61	2.47	338.47	834.39	0.11	-5.90	934.80	5.78	5400.00	0.63
2	-10.65	1.95	521.11	1018.51	0.29	-9.94	0.54	409.12	219.82	0.06	-2.19	468.41	6.10	2857.36	0.73
3	-5.15	1.40	815.03	1139.30	0.36	-4.83	0.29	626.14	184.06	0.07	-1.12	346.67	6.80	2358.30	0.78
4	-2.79	0.93	580.66	539.78	0.25	-2.58	0.10	4049.12	411.36	0.25	-0.61	181.89	8.32	1514.06	0.70
5	-1.53	0.64	773.62	497.36	0.33	-1.37	0.03	16811.02	505.90	0.45	-0.34	110.88	10.16	1126.21	0.62
6	-0.73	0.42	431.85	181.49	0.20	-0.64	0.02	25556.35	394.85	0.53	-0.16	123.72	9.09	1124.27	0.69
7	-0.26	0.22	655.54	147.01	0.25	-0.21	0.01	43515.93	339.39	0.51	0.00	117.36	10.30	1208.37	0.75
8	0.04	0.24	710.39	172.27	0.28	0.10	0.01	41983.20	381.33	0.55	0.21	179.53	8.68	1558.88	0.77
9	0.55	0.39	1454.13	568.20	0.54	0.72	0.03	32352.17	811.50	0.60	0.65	274.92	7.65	2102.10	0.76
10	2.90	1.66	873.22	1448.38	0.46	3.24	0.36	3936.36	1400.60	0.38	2.43	459.01	7.18	3294.41	0.71
1-10				25.96					-566.21***					2105.6***	

Table D.2: Estimating the intensity of noise trading for stock groups at weekly frequency

This table reports bounds on the standard deviation of noise trades estimated across groups of stocks sorted on various characteristics, using weekly data. For each month, we sort stocks into deciles according to their capitalization (reported in $ millions), turnover, closing bid–ask spread (reported in basis points), Amihud illiquidity ratio (reported multiplied by 1 million), return volatility, and return autocovariance (reported in basis points). All variables are estimated every month from daily observations. For a stock's capitalization, turnover, 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), or price-improved TAQ trades (columns 12–16). This table reports estimates from weekly regressions. 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. The row labelled "1-10" at the bottom of each stock characteristic panel reports the difference in the estimate of the standard deviation of noise trading between deciles 1 and 10, with stars indicating the results of a standard t-test for this difference being significatively different from zero. We use *, **, and *** to indicate statistical significance at (respectively) the 10%, 5%, and 1% level.

Decile	Households					TAQ Small trades					TAQ Improved trades				
	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD of (x1M)	Lower bound on SD of (% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD of (x1M)	Lower bound on SD of (% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD of (x1M)	Lower bound on SD of (% of SD of total trades)
By stock size															
1	5.12	9.70	70.28	681.79	0.12	5.54	27.56	105.20	2,898.89	0.34	20.04	3,276.06	3.74	12,238.20	0.88
2	12.07	3.74	359.65	1,343.88	0.24	14.24	9.40	292.90	2,753.29	0.40	53.31	2,417.39	4.13	9,987.10	0.94
3	21.05	3.97	259.75	1,032.03	0.18	26.01	5.24	424.23	2,224.24	0.37	111.67	942.74	4.96	4,674.94	0.69
4	34.35	3.56	547.73	1,951.30	0.35	43.24	4.17	632.55	2,638.24	0.36	207.03	739.91	5.84	4,320.11	0.69
5	54.91	4.36	706.40	3,080.50	0.51	69.91	2.64	1,046.57	2,760.24	0.39	359.81	729.53	7.28	5,310.76	0.75
6	88.15	3.14	974.82	3,059.59	0.49	113.89	1.20	2,166.85	2,602.17	0.37	662.18	671.01	7.60	5,097.33	0.66
7	148.42	2.57	1,157.73	2,972.00	0.50	189.31	1.10	1,779.45	1,955.32	0.30	1,157.74	683.75	9.54	6,522.46	0.70
8	281.43	2.38	1,299.69	3,086.94	0.54	351.24	0.38	4,691.16	1,768.56	0.31	2,162.21	507.18	12.92	6,555.31	0.69
9	707.42	1.24	1,636.46	2,035.93	0.44	823.39	0.17	12,206.56	2,092.62	0.44	4,705.68	444.54	15.41	6,852.07	0.70
10	6,255.78	0.68	1,985.80	1,344.30	0.40	3,381.67	0.04	40,183.26	1,619.48	0.52	18,507.94	400.40	13.54	5,421.21	0.80
1-10				-662.51***					1279.41***					6817***	
By stock turnover															
1	0.02	0.29	33.30	9.80	0.04	0.03	0.04	952.70	36.44	0.16	0.05	29.35	2.15	63.03	0.13
2	0.06	0.42	112.98	47.59	0.05	0.07	0.08	- 1,246.19	- 102.40	- 0.15	0.13	91.01	6.84	622.80	0.52
3	0.10	0.40	435.19	173.32	0.13	0.11	0.07	- 724.25	- 52.79	- 0.05	0.23	133.24	10.24	1,364.46	0.63
4	0.14	0.47	649.46	305.81	0.17	0.16	0.03	16,947.44	578.47	0.39	0.34	171.57	13.36	2,292.80	0.73
5	0.19	0.45	1,327.54	592.10	0.25	0.22	0.03	33,634.29	1,014.04	0.53	0.45	197.56	15.78	3,117.53	0.76
6	0.26	0.71	1,439.66	1,025.10	0.33	0.30	0.05	35,054.59	1,627.33	0.60	0.59	258.24	14.04	3,625.90	0.70
7	0.34	1.11	1,111.16	1,235.76	0.30	0.40	0.09	13,453.05	1,196.35	0.33	0.76	367.43	11.72	4,307.29	0.66
8	0.47	1.37	1,675.52	2,289.11	0.38	0.55	0.14	19,040.10	2,731.05	0.49	0.98	483.39	10.27	4,963.38	0.60
9	0.74	3.08	1,210.04	3,723.79	0.39	0.83	0.27	15,710.60	4,257.11	0.46	1.36	773.90	9.29	7,188.98	0.64
10	1.71	6.96	1,455.47	10,124.86	0.47	1.65	1.15	8,539.49	9,813.21	0.50	2.40	2,368.95	8.00	18,952.63	0.83
1-10				-10115.06***					-9776.76***					-18889.6***	

Table D.2 (continuing)

	Households					TAQ Small trades					TAQ Improved trades- Turnover x 1M				
				Lower bound on SD of					Lower bound on SD of					Lower bound on SD of	
Decile	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	(x1M)	(% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	(x1M)	(% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	(x1M)	(% of SD of total trades)
By stock bid-ask spread															
1	73.03	2.79	855.04	2,383.96	0.20	83.85	0.15	30,779.70	4,604.64	0.58	2.10	393.41	12.85	5,053.98	0.76
2	142.94	1.99	1,718.75	3,423.73	0.48	134.08	0.19	3,072.38	590.86	0.17	3.47	561.27	11.07	6,211.93	0.74
3	200.66	2.04	1,152.88	2,352.44	0.45	183.80	0.56	- 498.79	- 280.83	- 0.08	5.22	645.20	10.78	6,952.20	0.74
4	267.74	1.97	1,201.56	2,361.46	0.52	238.59	0.61	575.41	351.97	0.10	7.63	751.63	10.83	8,137.36	0.81
5	349.04	2.57	772.42	1,982.99	0.45	299.36	0.42	1,958.50	812.93	0.24	11.43	844.67	9.60	8,105.73	0.82
6	438.99	2.23	616.31	1,375.56	0.32	375.56	0.70	383.21	269.67	0.06	18.64	1,107.46	7.76	8,592.82	0.77
7	561.00	3.10	406.65	1,260.89	0.31	468.22	1.06	- 43.14	- 45.82	- 0.01	34.74	928.37	5.35	4,965.71	0.70
8	734.27	3.08	277.86	855.58	0.25	607.65	1.76	- 90.81	- 160.08	- 0.04	76.46	1,071.29	5.41	5,797.76	0.80
9	1,035.65	2.42	210.63	509.94	0.17	859.46	8.64	112.83	974.29	0.29	158.68	1,020.26	5.14	5,239.12	0.78
10	1,538.46	2.05	195.38	400.66	0.20	1,442.26	6.96	126.95	883.55	0.33	363.49	1,096.68	4.47	4,901.99	0.95
1-10				1983.3***					3721.09***					151.98***	
By stock Amihud illiquidity measure															
1	-	0.76	1,962.03	1,496.38	0.39	-	0.04	42,576.58	1,763.75	0.52	-	419.43	13.87	5,815.58	0.80
2	0.01	1.03	1,578.29	1,625.02	0.41	0.01	0.17	11,829.20	2,065.03	0.51	-	414.46	13.77	5,707.36	0.72
3	0.03	1.52	1,366.56	2,081.80	0.52	0.02	0.35	4,240.02	1,462.88	0.34	-	409.46	13.12	5,373.05	0.68
4	0.08	1.66	1,049.83	1,744.27	0.52	0.06	0.61	1,609.18	978.19	0.26	-	384.51	10.87	4,178.30	0.67
5	0.21	1.55	783.73	1,215.79	0.37	0.15	1.11	1,092.54	1,210.60	0.39	-	399.96	9.45	3,779.45	0.68
6	0.54	1.36	588.26	798.51	0.31	0.34	1.36	549.01	748.57	0.31	0.01	453.24	6.42	2,911.14	0.65
7	1.28	1.28	439.58	564.12	0.31	0.83	3.15	220.47	694.91	0.37	0.03	467.10	4.79	2,237.79	0.58
8	3.03	1.28	243.41	312.13	0.25	1.98	2.52	369.67	931.13	0.40	0.09	529.64	4.14	2,190.95	0.64
9	8.33	1.34	191.18	255.38	0.24	5.35	3.79	159.17	603.28	0.36	0.51	558.02	3.87	2,158.49	0.69
10	39.22	6.44	27.10	174.55	0.16	28.86	7.33	68.99	505.34	0.28	5.85	344.49	4.44	1,530.17	0.79
1-10				1321.83***					1258.41***					4285.41***	

Table D.2 (continuing)

Decile	Households					TAQ Small trades					TAQ Improved trades				
	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD of (x1M)	Lower bound on SD of (% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD of (x1M)	Lower bound on SD of (% of SD of total trades)	Median value of sorting variable	SD of trades (x1M)	Regression coef. b	Lower bound on SD of (x1M)	Lower bound on SD of (% of SD of total trades)
By stock return volatility															
1	1.04	0.47	874.42	413.11	0.25	1.03	0.04	24,053.00	843.62	0.52	0.90	291.08	13.63	3,967.06	0.74
2	1.47	0.57	1,088.34	619.96	0.27	1.51	0.03	52,850.57	1,366.88	0.64	1.22	527.73	8.26	4,356.83	0.70
3	1.89	1.00	1,197.61	1,199.17	0.37	1.94	0.05	34,558.29	1,627.42	0.57	1.45	657.35	7.84	5,153.79	0.67
4	2.34	1.78	1,084.88	1,927.47	0.32	2.37	0.09	32,571.40	2,820.15	0.62	1.71	760.72	8.06	6,131.88	0.71
5	2.81	2.95	822.45	2,426.81	0.32	2.85	0.14	24,325.60	3,485.57	0.50	1.99	823.59	8.01	6,599.64	0.67
6	3.34	3.54	1,211.60	4,285.95	0.40	3.41	0.28	16,199.82	4,568.54	0.48	2.30	1,160.51	6.63	7,696.84	0.69
7	3.99	4.86	1,362.73	6,621.94	0.48	4.07	0.55	5,013.72	2,768.04	0.23	2.65	1,822.26	6.00	10,941.18	0.75
8	4.88	6.27	1,258.21	7,886.46	0.42	4.95	1.00	2,770.38	2,774.01	0.16	3.13	2,505.75	5.99	15,021.47	0.81
9	6.29	10.60	1,072.47	11,371.11	0.52	6.41	2.72	981.51	2,664.83	0.12	3.93	3,157.20	6.36	20,079.07	0.84
10	10.03	24.33	803.15	19,538.46	0.52	10.22	14.18	844.97	11,978.77	0.25	5.84	5,127.46	6.01	30,828.88	0.74
1-10				-19125.36***					-11135.15***					-26861.82***	
By stock return autocovariance															
1	- 30.73	9.19	905.47	8,325.60	0.45	- 31.64	10.70	273.65	2,929.25	0.10	- 5.90	3,980.22	5.56	22,149.60	0.71
2	- 10.55	5.45	1,168.22	6,364.75	0.41	- 10.06	2.36	423.44	998.16	0.06	- 2.19	2,114.98	6.12	12,950.47	0.80
3	- 5.18	4.40	1,510.74	6,651.84	0.48	- 4.88	1.35	740.87	1,001.36	0.09	- 1.12	1,574.54	6.52	10,267.55	0.80
4	- 2.77	2.52	1,275.34	3,215.67	0.34	- 2.58	0.47	4,509.41	2,125.58	0.30	- 0.61	817.12	8.46	6,913.35	0.73
5	- 1.54	1.95	1,394.01	2,714.63	0.40	- 1.36	0.14	16,482.44	2,365.20	0.49	- 0.34	502.34	10.21	5,127.07	0.65
6	- 0.73	1.06	1,007.68	1,066.54	0.27	- 0.64	0.07	25,765.57	1,905.79	0.58	- 0.16	557.07	9.39	5,231.22	0.74
7	- 0.26	0.62	1,510.54	939.05	0.36	- 0.20	0.04	42,198.73	1,602.91	0.54	-	540.60	10.33	5,585.27	0.80
8	0.04	0.75	1,366.47	1,031.41	0.38	0.10	0.04	43,372.16	1,866.26	0.61	0.20	815.02	8.76	7,143.04	0.80
9	0.55	1.37	2,303.63	3,166.10	0.67	0.72	0.12	32,644.76	3,818.87	0.63	0.65	1,219.17	7.73	9,428.29	0.78
10	2.84	6.22	1,266.15	7,880.95	0.57	3.29	1.48	4,838.60	7,167.73	0.45	2.44	1,964.77	7.26	14,266.12	0.75
1-10				444.65					-4238.48***					7883.48***	

Internet Appendix E: Results for the net number of households buying shares

In the paper, our analyses of household at from a large discount broker (from January 1991 through November 1996) uses the considers households' net turnover (defined as the aggregate value of their buys, minus the aggregate value of their sells, divided by the market's total value). Here, we report estimates based on the net number of households buying shares at the large discount broker: the number of households buying minus the number of households selling.

E.1: Descriptive statistics

Table E.1: Descriptive statistics for the net number of households buying shares

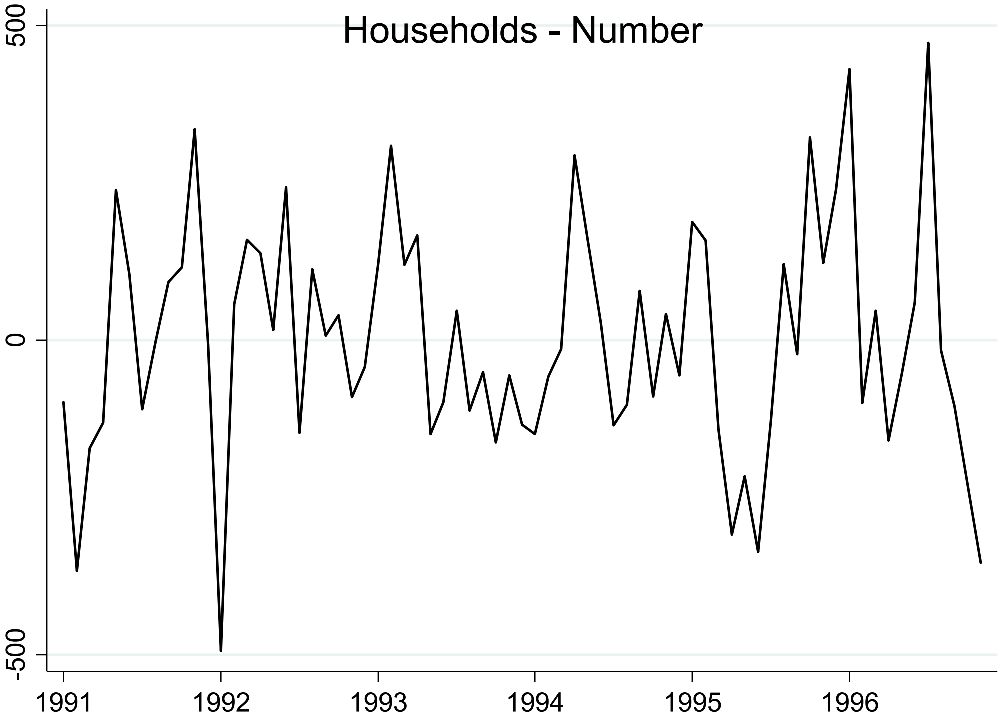
This table presents summary statistics for the time series of net number of households buying shares at a large discount broker (from January 1991 through November 1996). The household data are for those holding common stock positions for 71 consecutive months. The net number of households buying shares equals the number of households buying minus the number of households selling. 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.

Frequency	Obs.	min.	mean	median	max.	std. dev.	skewness	kurtosis
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								

E.2: Time series of the number of households buying shares

Figure E.2: Time series

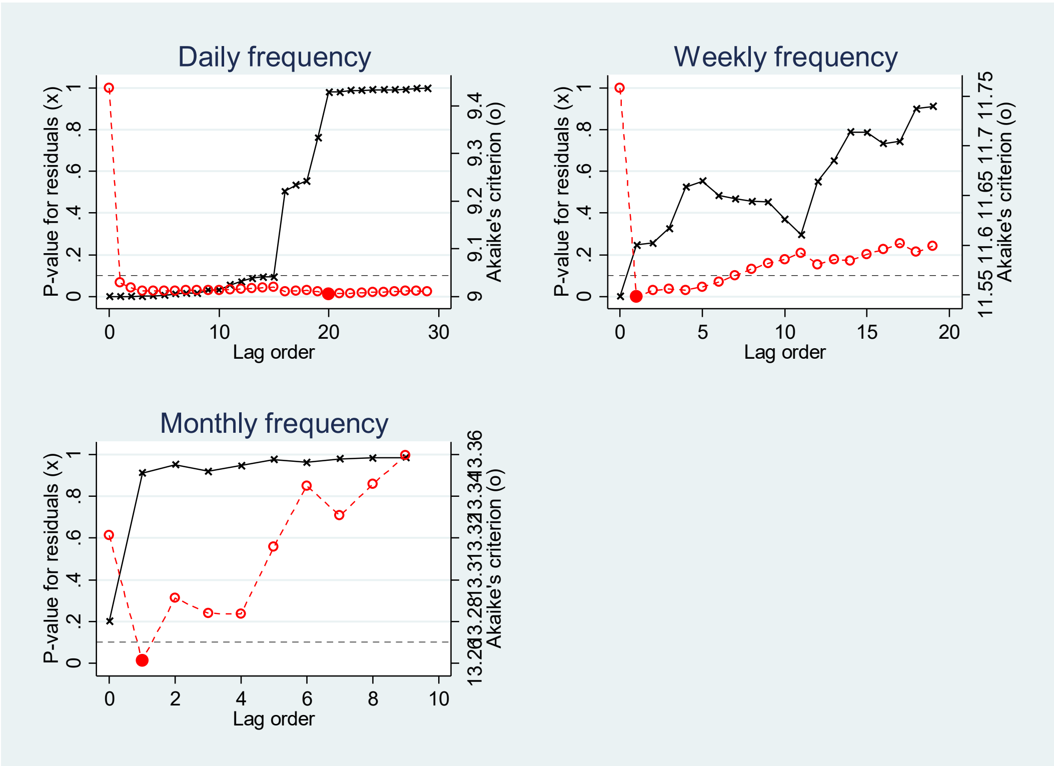
This figure displays the monthly time series of net number of households buying shares at a large discount broker (from January 1991 through November 1996). The household data are for those holding common stock positions for 71 consecutive months. The net number of households buying shares equals the number of households buying minus the number of households selling. 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.



E.2: Lag-order selection

Figure E.2: Lag-order selection for the net number of households buying shares

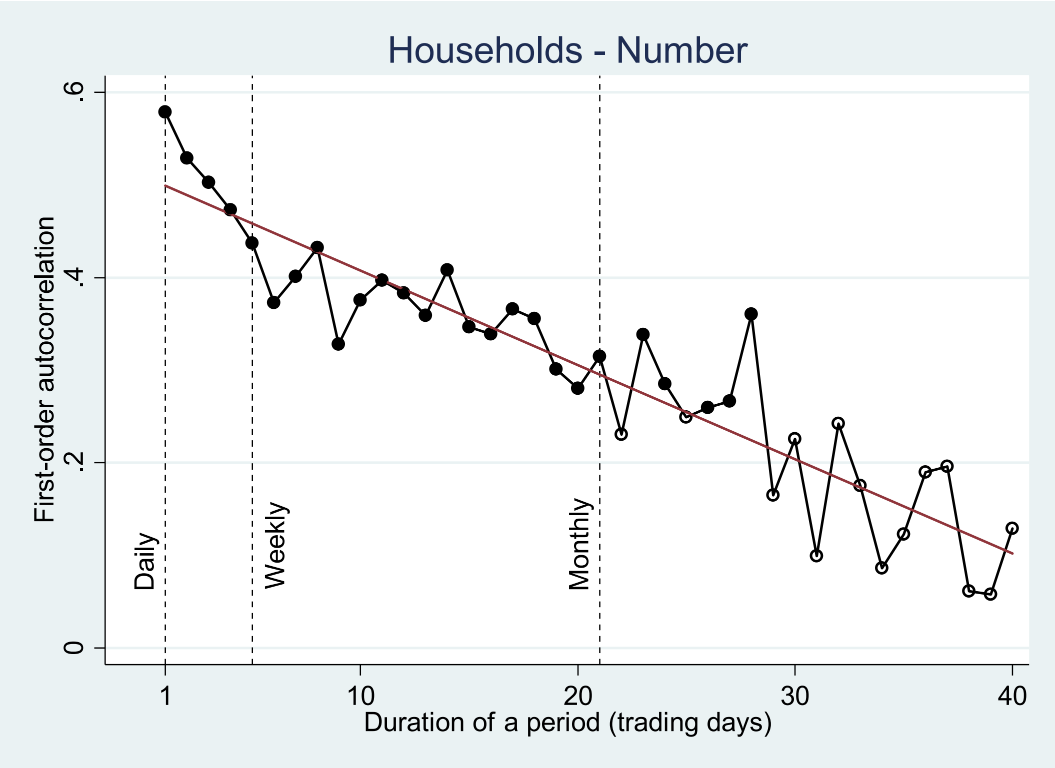
This figure displays the performance of autoregressive models fitted to the net number of households buying shares as a function of the number of lags. The household data are those holding common stock positions for 71 consecutive months from January 1991 through November 1996. The net number of households buying shares equals the number of households buying minus the number of households selling. The number of lags ranges from 0 to 30, 0 to 20, and 0 to 10 at (respectively) daily (top left panel), weekly (top right panel), and monthly (bottom left panel) 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.



E.3: Fitting an AR(1) process

Figure E.3: Fitting an AR(1) process to the net number of households buying shares

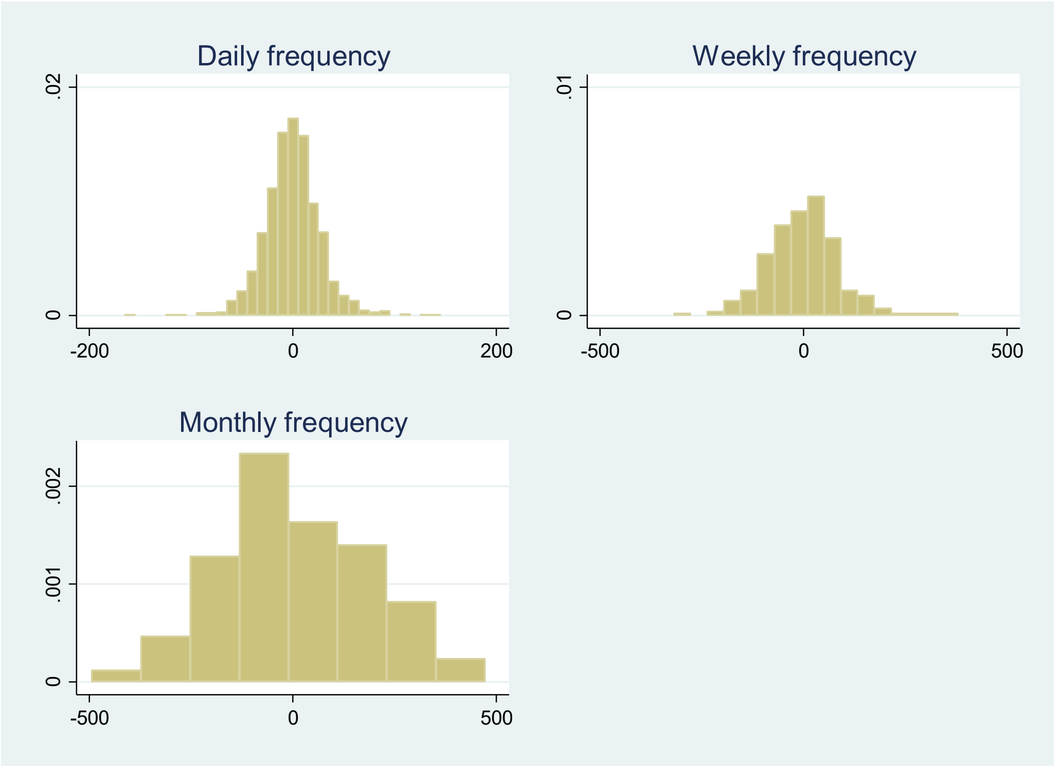
This figure plots the first-order autocorrelation coefficient of the net number of households buying shares as a function of the duration of a time period in days. The household data are those holding common stock positions for 71 consecutive months from January 1991 through November 1996. The net number of households buying shares equals the number of households buying minus the number of households selling. Solid circles mark coefficients that are statistically significant at the 10% level. We adjust the variables for seasonality and time trends by regressing them on dummies for month of the year and year, and then taking the residuals.



E.4: Histograms

Figure E.4: Histograms of the net number of households buying shares

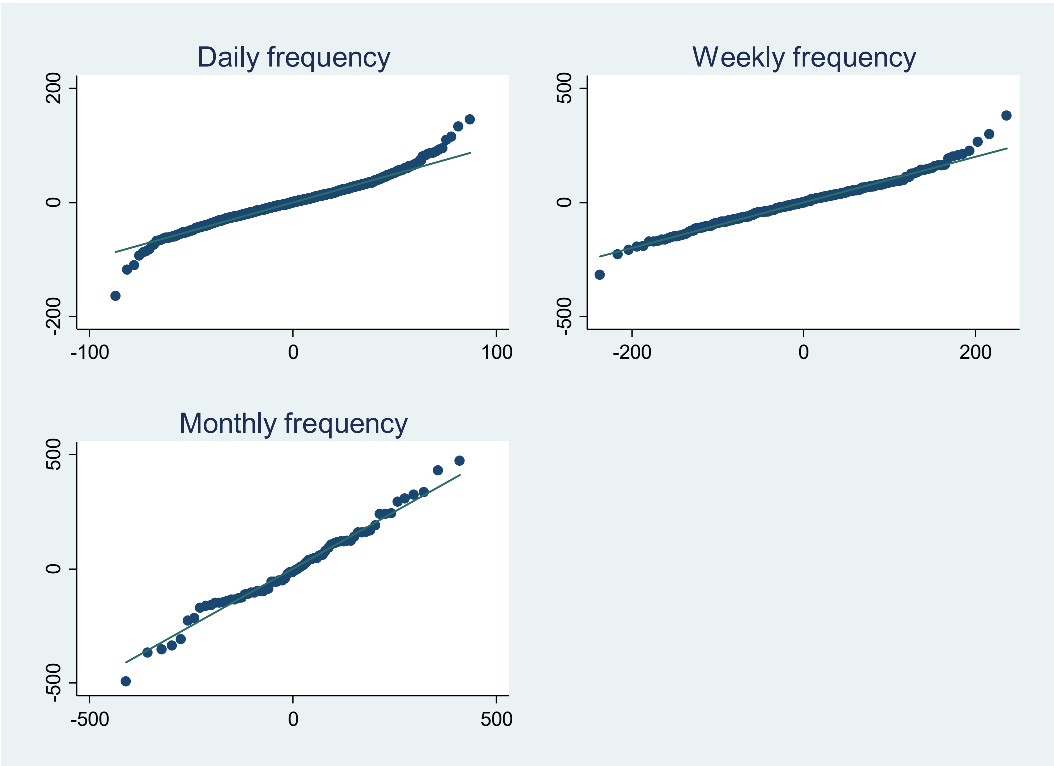
The graphs in this figure are histograms of the net number of households buying shares. The top left, top right, and bottom left panels consider (respectively) daily, weekly, and monthly data. The household data are those holding common stock positions for 71 consecutive months from January 1991 through November 1996. The net number of households buying shares equals the number of households buying minus the number of households selling. We adjust the variables for seasonality and time trends by regressing them on dummies for month of the year and year, and then taking the residuals.



E.5: Probability (Q-Q) plots

Figure E.5: Probability (Q-Q) plots of the net number of households buying shares

The graphs in this figure plot quantiles of the net number of households buying shares against quantiles of a normal distribution at various frequencies. The top left, top right, and bottom left panels consider (respectively) daily, weekly, and monthly data. The household data are those holding common stock positions for 71 consecutive months from January 1991 through November 1996. The net number of households buying shares equals the number of households buying minus the number of households selling. We adjust the variables for seasonality and time trends by regressing them on dummies for month of the year and year, and then taking the residuals.



E.6: Shapiro–Wilk test for normality

Table E.6: Shapiro–Wilk test for normality for the net number of households buying shares

This table reports results of a Shapiro–Wilk test that the net number of households buying shares (columns 2 and 3) and its residuals from a fitted AR(1) process (columns 4 and 5) 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 household data are those holding common stock positions for 71 consecutive months from January 1991 through November 1996. The net number of households buying shares equals the number of households buying minus the number of households selling. We adjust the variables for seasonality and time trends by regressing them on dummies for month of the year and year, and then taking the residuals.

	Variables		Residuals from fitted AR(1)	
	Test Statistic	p-value	Test Statistic	p-value
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

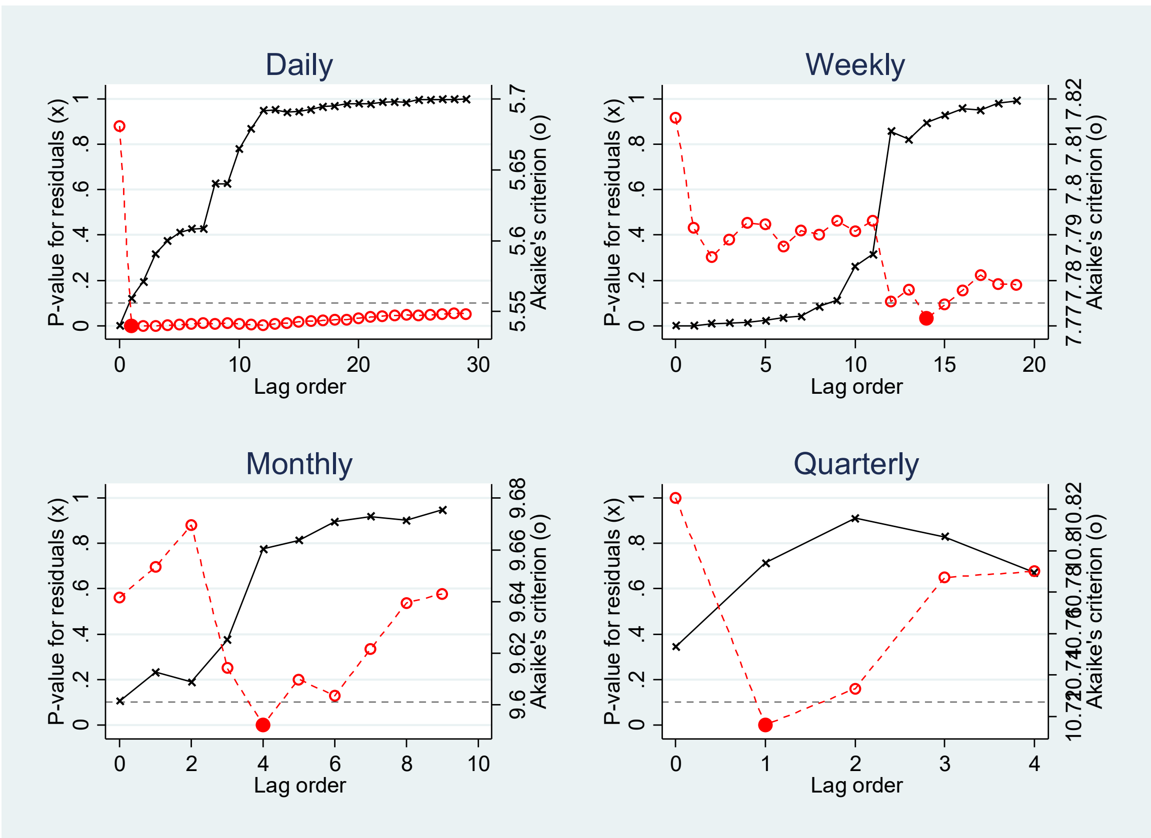
Internet Appendix F: Falsification tests

F.1: Large TAQ trades

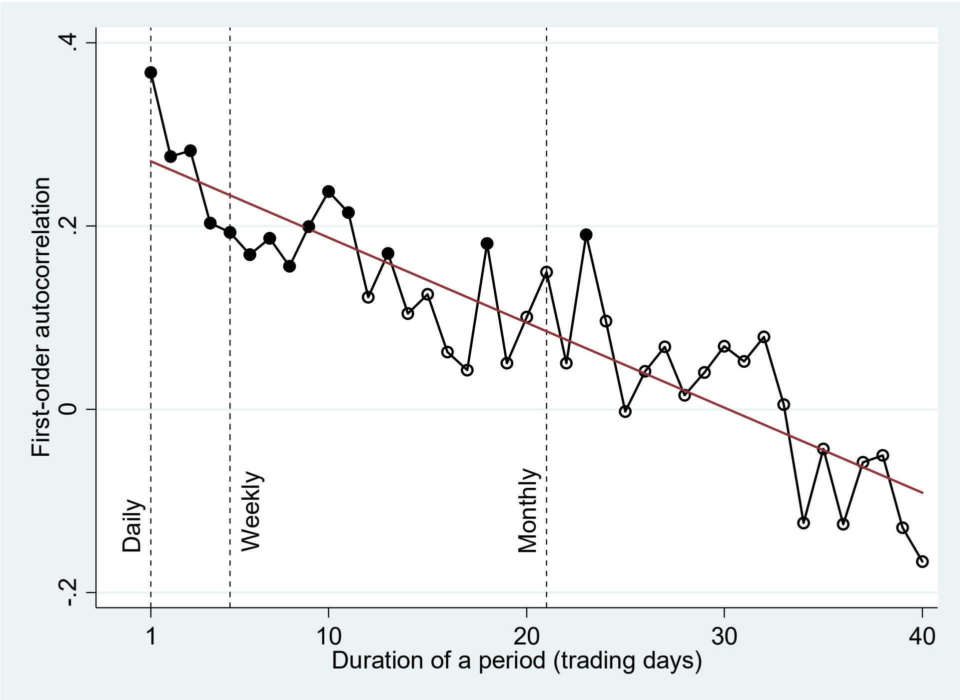
Figure F1: Fitting an autoregressive process and parametric form for TAQ large trades

This figure displays estimates for the net turnover for large trades in the TAQ dataset trades (from January 1991 through December 2000). Net turnover is measured as the aggregate value of large buys, minus the aggregate value of large sells, divided by the market's total value. We adjust the series 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 displays the performance of autoregressive models fitted to net turnover at daily (top left panel), weekly (top right panel), and monthly (bottom left panel) frequencies 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, weekly, and monthly 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. Panel B plots the first-order autocorrelation coefficient of net turnover as a function of the duration of a time period in days. Solid circles mark coefficients that are statistically significant at the 10% level. Panel C displays histograms of net turnover at daily (top left panel), weekly (top right panel), and monthly (bottom left panel) frequencies. Panel D displays probability (Q-Q) plots of net turnover at daily (top left panel), weekly (top right panel), and monthly (bottom left panel) frequencies. The graphs plot quantiles of net turnover against quantiles of a normal distribution.

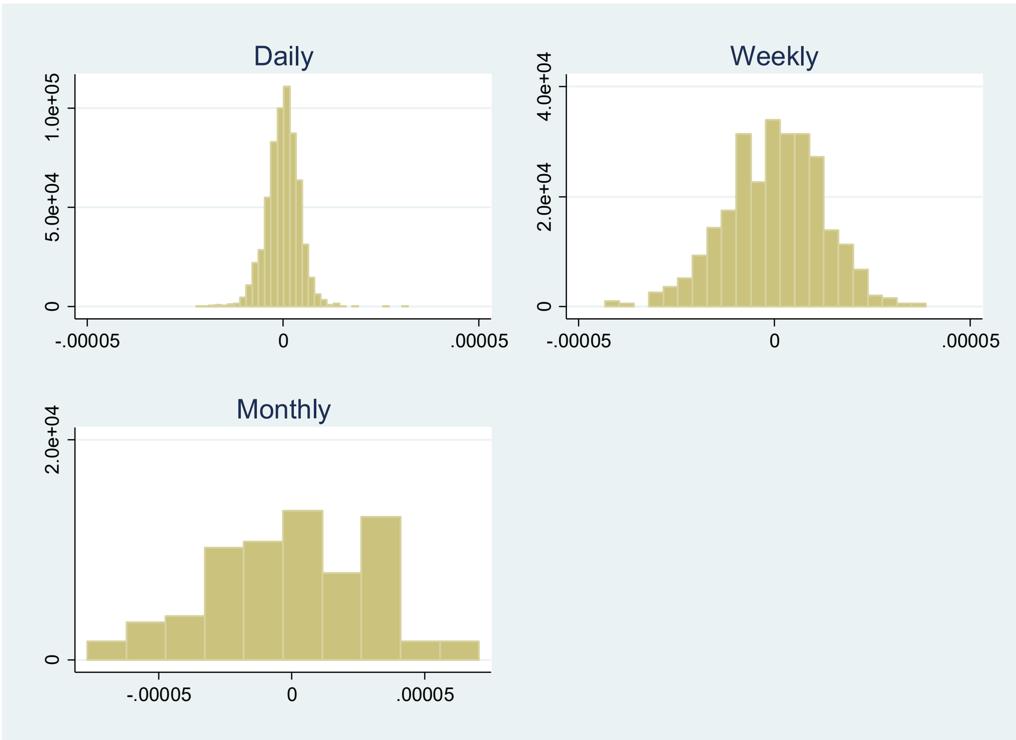
Panel A: Lag-order selection



Panel B: Fitting an AR(1) process



Panel C: Histograms



Panel D: Probability (Q-Q) plots

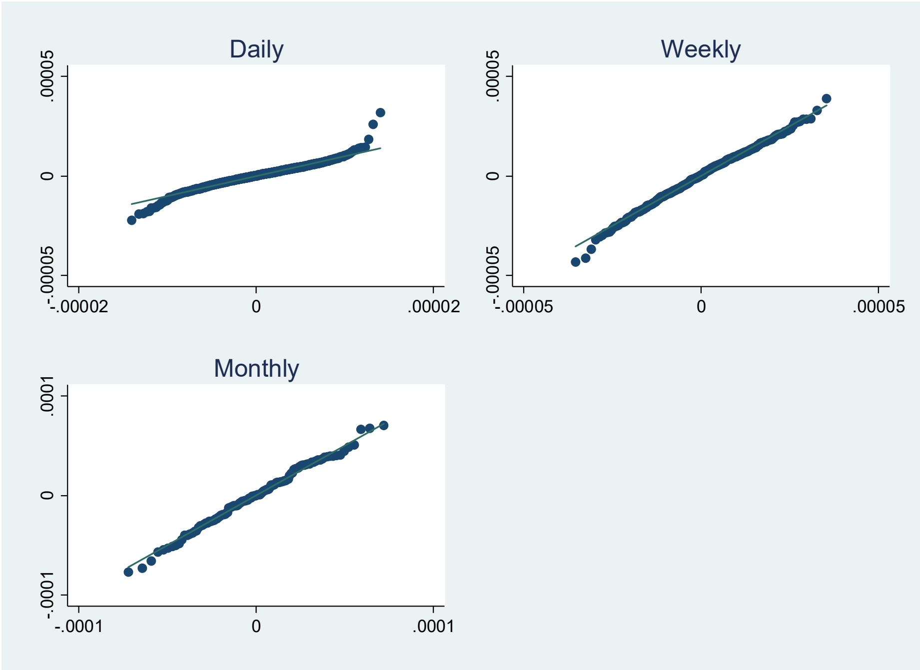


Table F1: Shapiro–Wilk test for normality for TAQ large trades.

This table reports results of a Shapiro–Wilk test that the net turnover for large TAQ trades (from January 1991 through December 2000) and their residuals from a fitted AR(1) process (columns 4 and 5) 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 net turnover for large TAQ trades is measured as the aggregate value of large buys, minus the aggregate value of large sells, divided by the market's total value. We adjust the series 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.

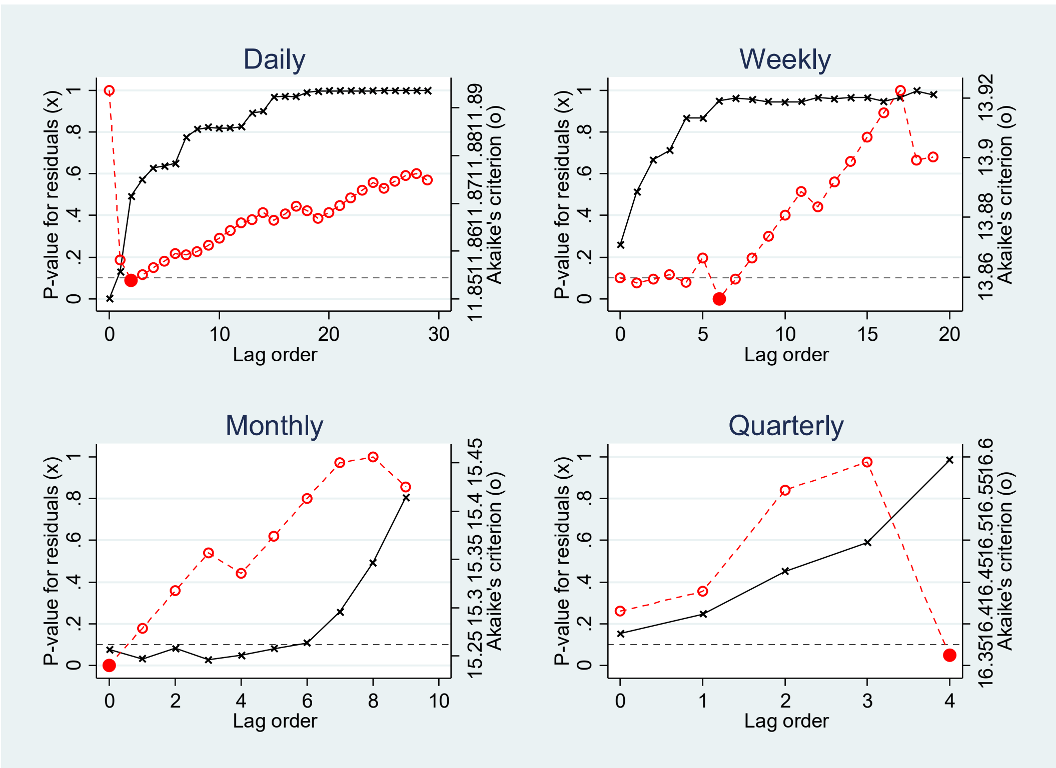
	TAQ Large trades		Residuals from fitted AR(1)	
	Test Statistic	p-value	Test Statistic	p-value
Daily	9.443	0.000	9.280	0.000
Weekly	0.334	0.369	0.882	0.189
Monthly	-0.241	0.595	-0.607	0.728
Quarterly	0.021	0.492	0.531	0.298

F.2: TAQ trades without price improvement

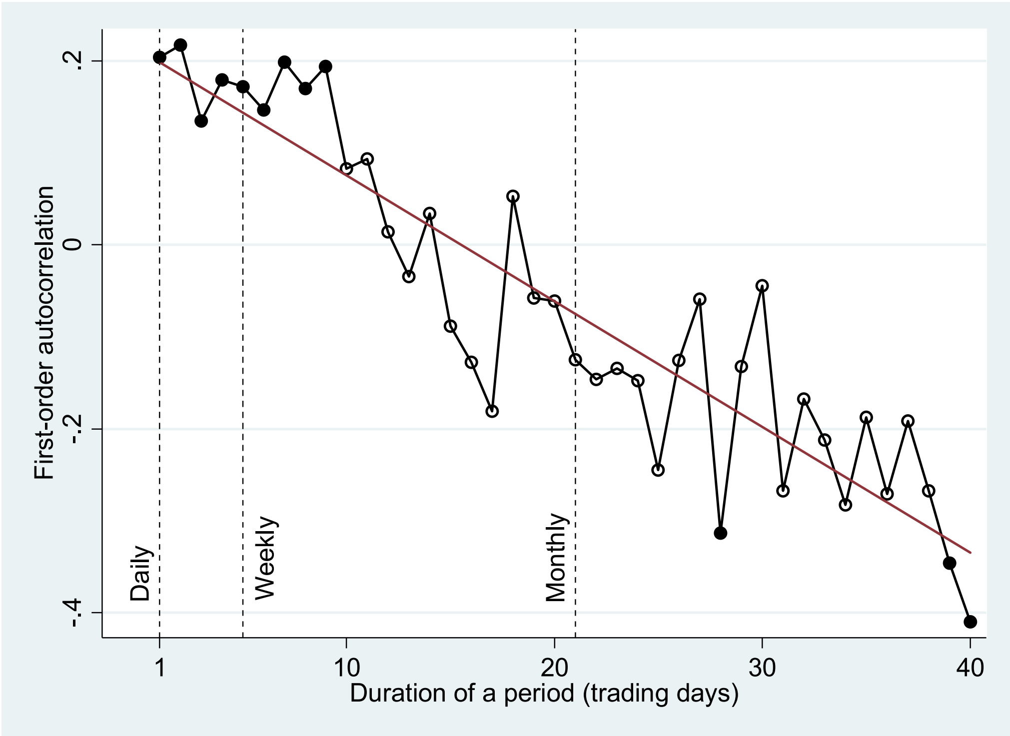
Figure F2: Fitting an autoregressive process and parametric form for TAQ trades without price improvement

This figure displays estimates for the net turnover for trades without price improvement in the TAQ dataset (from January 2010 through December 2014). Net turnover is measured as the aggregate value of buys with a transaction price that is not slightly below the round penny, minus the aggregate value of sells with a transaction price that is not slightly above the round penny, divided by the market's total value. We adjust the series 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 displays the performance of autoregressive models fitted to net turnover at daily (top left panel), weekly (top right panel), and monthly (bottom left panel) frequencies 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, weekly, and monthly 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. Panel B plots the first-order autocorrelation coefficient of net turnover as a function of the duration of a time period in days. Solid circles mark coefficients that are statistically significant at the 10% level. Panel C displays histograms of net turnover at daily (top left panel), weekly (top right panel), and monthly (bottom left panel) frequencies. Panel D displays probability (Q-Q) plots of net turnover at daily (top left panel), weekly (top right panel), and monthly (bottom left panel) frequencies. The graphs plot quantiles of net turnover against quantiles of a normal distribution.

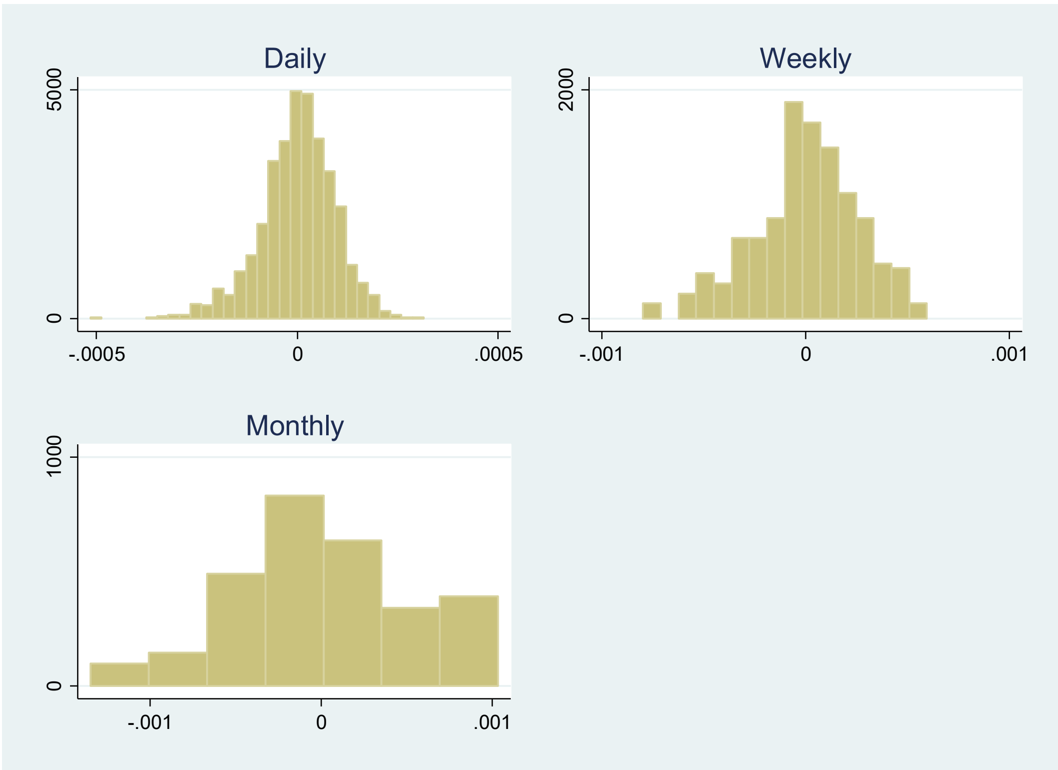
Panel A: Lag-order selection



Panel B: Fitting an AR(1) process



Panel C: Histograms



Panel D: Probability (Q-Q) plots

Table F2: Shapiro–Wilk test for normality for TAQ trades without price improvement

This table reports results of a Shapiro–Wilk test that the net turnover for trades without price improvement in the TAQ dataset (from January 2010 through December 2014), and their residuals from a fitted AR(1) process (columns 4 and 5) 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 net turnover for trades without price improvement is measured as the aggregate value of buys with a transaction price that is not slightly below the round penny, minus the aggregate value of sells with a transaction price that is not slightly above the round penny, divided by the market's total value. We adjust the series 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.

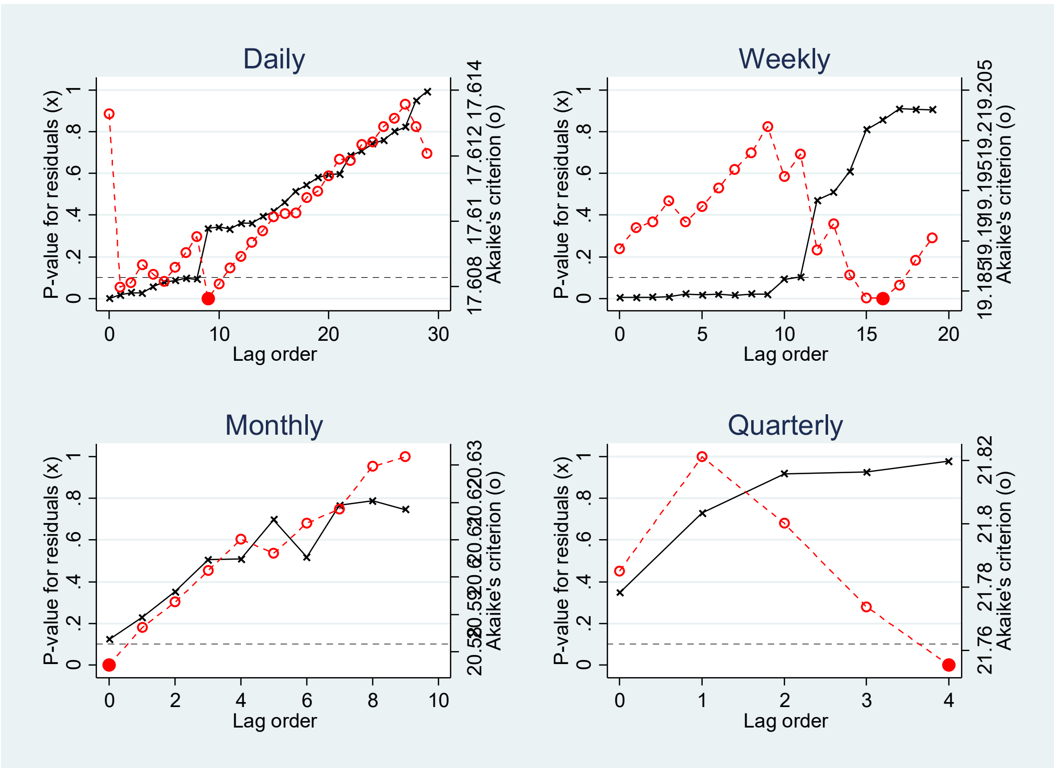
	TAQ Non-improved trades		Residuals from fitted AR(1)	
	Test Statistic	p-value	Test Statistic	p-value
Daily	6.453	0.000	6.648	0.000
Weekly	1.553	0.060	2.288	0.011
Monthly	-0.214	0.585	0.022	0.491
Quarterly	-2.242	0.988	-0.345	0.635

F.3: Flows to non-retail mutual funds

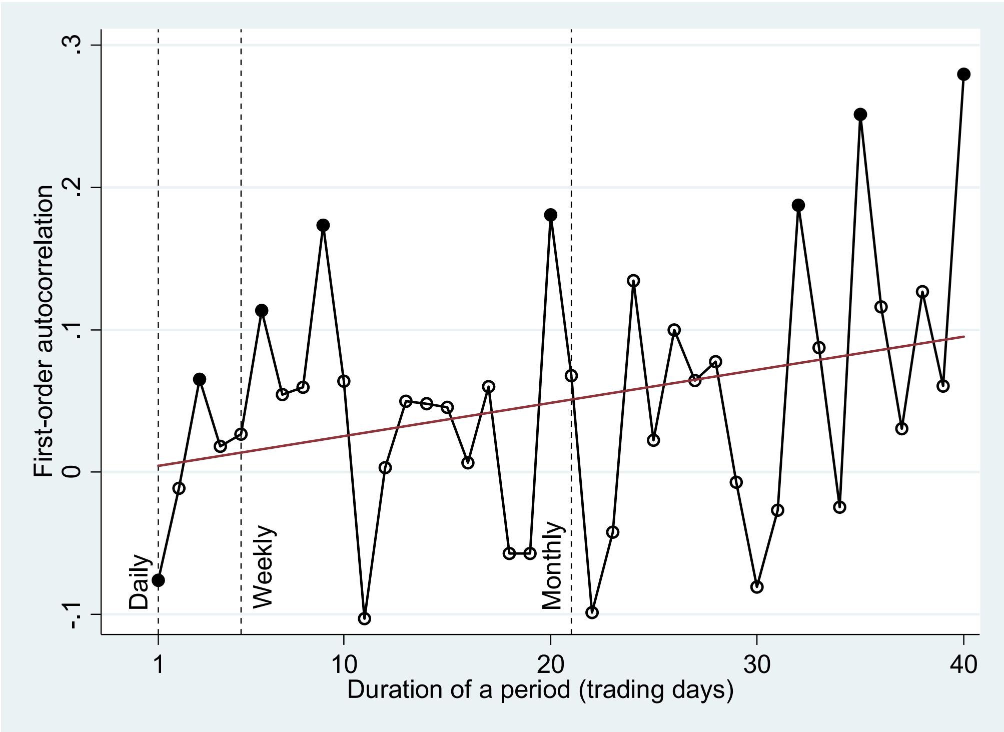
Figure F3: Fitting an autoregressive process and parametric form for flows to non-retail equity mutual funds

This figure displays estimates for the net turnover for flows to non-retail equity mutual funds in the TrimTabs dataset (from January 1999 through August 2013). Net turnover is measured as the aggregate value of purchases, minus redemptions, of equity mutual funds divided by these funds' aggregate TNA. We adjust the series 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 displays the performance of autoregressive models fitted to net turnover at daily (top left panel), weekly (top right panel), and monthly (bottom left panel) frequencies 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, weekly, and monthly 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. Panel B plots the first-order autocorrelation coefficient of net turnover as a function of the duration of a time period in days. Solid circles mark coefficients that are statistically significant at the 10% level. Panel C displays histograms of net turnover at daily (top left panel), weekly (top right panel), and monthly (bottom left panel) frequencies. Panel D displays probability (Q-Q) plots of net turnover at daily (top left panel), weekly (top right panel), and monthly (bottom left panel) frequencies. The graphs plot quantiles of net turnover against quantiles of a normal distribution.

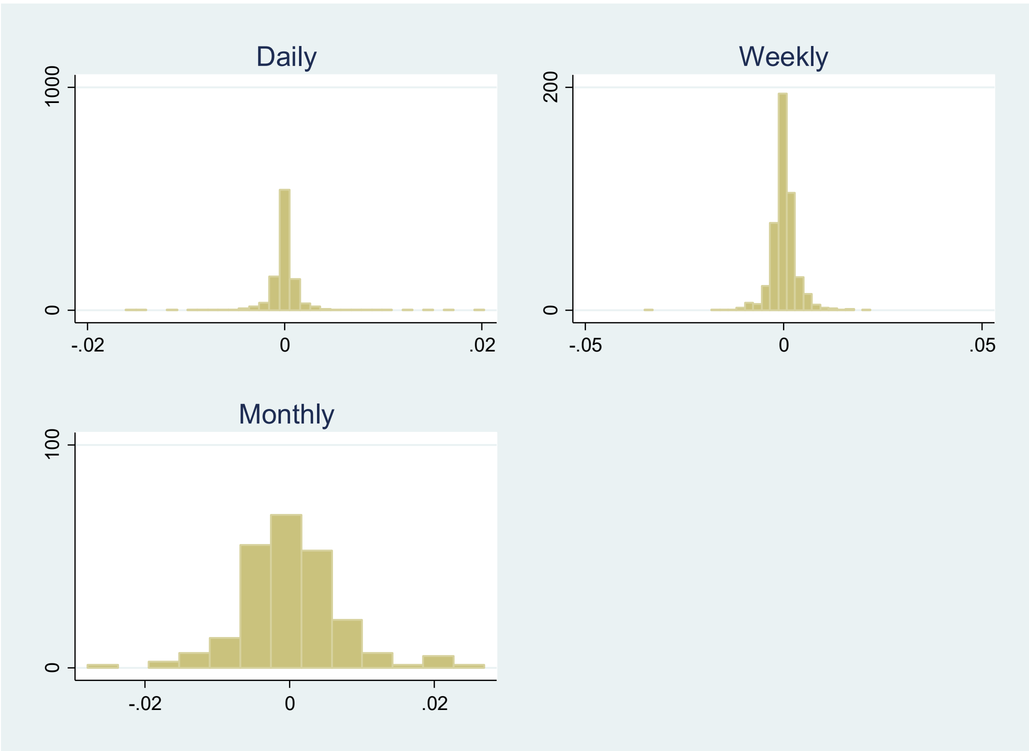
Panel A: Lag-order selection



Panel B: Fitting an AR(1) process



Panel C: Histograms



Panel D: Probability (Q-Q) plots

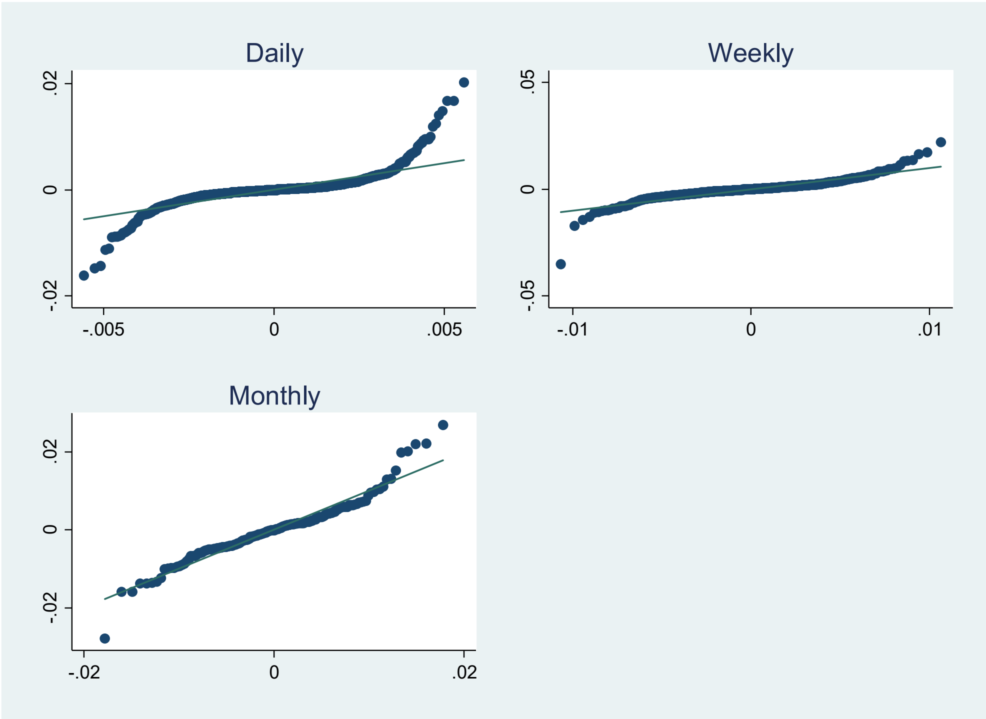


Table F3: Shapiro–Wilk test for normality for flows to non-retail equity mutual funds

This table reports results of a Shapiro–Wilk test that the net turnover for flows to non-retail equity mutual funds (from January 1999 through August 2013), and their residuals from a fitted AR(1) process (columns 4 and 5) 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 net turnover for flows to non-retail equity mutual funds is measured as the aggregate value of purchases, minus redemptions, of equity mutual funds divided by these funds' aggregate TNA. We adjust the series 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.

	Non-retail mutual funds		Residuals from fitted AR(1)	
	Test Statistic	p-value	Test Statistic	p-value
Daily	16.510	0.000	16.564	0.000
Weekly	10.679	0.000	10.676	0.000
Monthly	4.619	0	4.466	0.000
Quarterly	-0.493	0.689	-0.043	0.517

F.4: Hedge funds' trades (proxying for informed trades)

We check that some characteristics of noise trades are specific to these investors and do not mechanically extend to informed speculators. We proxy for their trades using hedge fund trades. We use data from Abel Noser Solutions (commonly known as ANcerno), which cover institutional stock transactions from 1999 to 2011. We restrict our analysis to trades by asset managers classified as hedge funds by Russell Jame and as described in Jame (2018).[11] This classification can be downloaded from http://russelljame.com/research.html. We aggregate these trades across traders and stocks, and normalize by the capitalization of the market to obtain a net turnover series, as we do with noise (household) trades. Then we carry out the same estimations on this time series.

The findings for hedge funds differ from those we obtained for households, along two important dimensions: their 1) serial correlation, and 2) parametric form.

Starting with serial correlation, in Figure F4, we fit hedge funds' net turnover to autoregressive models with up to 30 lags. At the daily frequency, 7 lags are needed to eliminate serial dependence in the residuals. At the weekly frequency, none are needed (i.e., trades are serially uncorrelated). At the monthly frequency, 8 lags are needed. These findings contrast sharply with those we obtained for households, along two dimensions. First, for hedge funds, the required number of lags doesn't decrease with the frequency, whereas it does for households. Second, hedge fund trades remain autocorrelated at monthly frequency, whereas households' trades are not.

Taking a closer look at the performance of AR(1) processes (the focus of many theoretical papers), we find, in Figure F5, no relation between the first-order autocorrelation coefficient and the duration of a period. This is clear visually and confirmed by a least-square regression of the autocorrelation coefficient on the duration of a time period, which yields a statistically insignificant coefficient estimate of -.00052 (p-value = 0.24). This again distinguishes hedge funds' trades from households' trades, for which the autocorrelation coefficient decreases with the duration of a period, as argued by theorists

Turning to the parametric form of hedge funds' net turnover, we display their histograms and quantile-to-quantile (Q-Q) plots in Figures F6 and F7 , respectively. We also report in Table F4 the results of the Shapiro–Wilk test of the normality of trades and their residuals from the fitted AR(1) process. Consistent with a visual inspection of the figures, the table indicates that the null hypothesis of normality is rejected at the daily, weekly and monthly frequencies, and that test statistics decrease in frequency. These results are similar to those for households. However, in contrast to households, hedge funds' trades a not normally distributed at the quarterly frequency. They continue to display fat tails, particularly for the AR(0) specification.

To summarize, while some of the findings we derived for households' trades extend to hedge funds' trades, many do not. Specifically, we had reported that, for households, i) the first-order autocorrelation coefficient decreases with the duration of a period (consistent with theorists' assertion about noise trading), that ii) monthly trades are i.i.d. and iii) that quarterly trades are normally distributed. For hedge funds in contrast, we find that i) the first-order autocorrelation coefficient is unrelated to the duration of a period, ii) monthly trades are serially correlated with multiple lags needed, and iii) quarterly trades are fat-tailed.

[11] Jame, Russell (2018), Liquidity Provision and the Cross Section of Hedge Fund Returns, *Management Science* 64, 3288-3312.

Figure F4: Lag-order selection for hedge funds' aggregate trades

This figure displays the performance of autoregressive models fitted to hedge funds' aggregate trades as a function of the number of lags for different at daily (top left panel), weekly (top right panel), and monthly (bottom left panel) frequencies. The number of lags ranges from 0 to 30, 0 to 20, and 0 to 10 at (respectively) daily, weekly, and monthly 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. Hedge funds' net turnover is the aggregate value of their buys, minus the aggregate value of their sells, divided by the market's total value, where buys and sells are those recorded in the Abel Noser Solutions (commonly known as ANcerno) dataset from 1999 to 2011 for asset managers classified as hedge funds according to Jame (2018). 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.

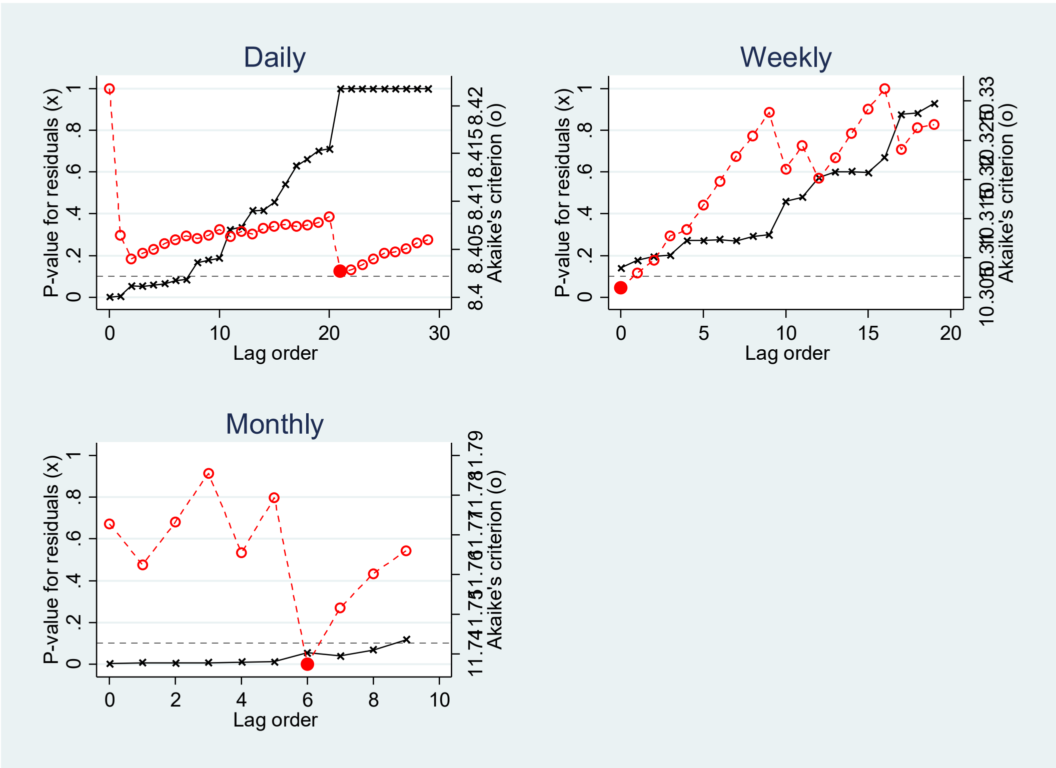
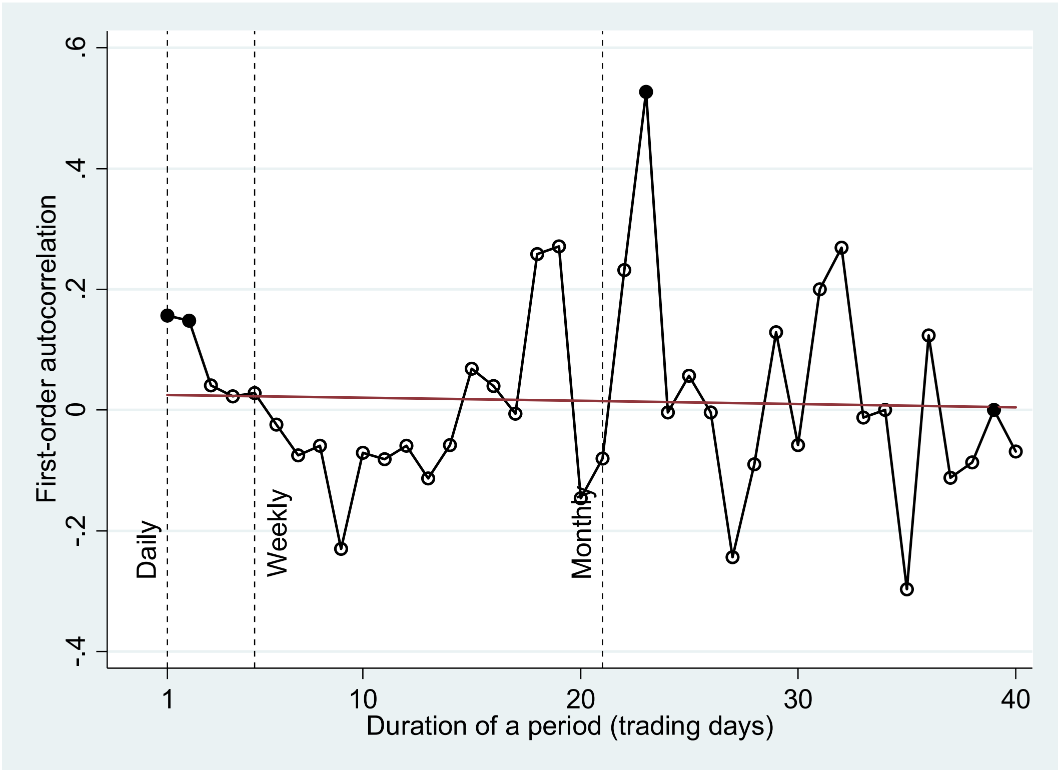


Figure F5: Fitting an AR(1) process to hedge funds' aggregate trades

The graphs in this figure plot the first-order autocorrelation coefficient of hedge funds' 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. Hedge funds' net turnover is the aggregate value of their buys, minus the aggregate value of their sells, divided by the market's total value, where buys and sells are those recorded in the Abel Noser Solutions (commonly known as ANcerno) dataset from 1999 to 2011 for asset managers classified as hedge funds according to Jame (2018). 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. A least-square regression of the first-order autocorrelation coefficient on the duration of a time period yields a statistically insignificant coefficient estimate of -.00052 (p-value = 0.24).



.

Figure F6: Histograms of hedge funds' aggregate trades

The graphs in this figure are histograms of hedge funds' aggregate trades, estimated at daily (top left panel), weekly (top right panel), and monthly (bottom left panel) frequencies. Hedge funds' net turnover is the aggregate value of their buys, minus the aggregate value of their sells, divided by the market's total value, where buys and sells are those recorded in the Abel Noser Solutions (commonly known as ANcerno) dataset from 1999 to 2011 for asset managers classified as hedge funds according to Jame (2018). 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 F7: Probability (Q-Q) plots of hedge funds' aggregate trades

The graphs in this figure plot quantiles of hedge funds' aggregate trades against quantiles of a normal distribution, estimated at daily (top left panel), weekly (top right panel), and monthly (bottom left panel) frequencies. Hedge funds' net turnover is the aggregate value of their buys, minus the aggregate value of their sells, divided by the market's total value, where buys and sells are those recorded in the Abel Noser Solutions (commonly known as ANcerno) dataset from 1999 to 2011 for asset managers classified as hedge funds according to Jame (2018). 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.

Table F8: Shapiro–Wilk test for normality for hedge funds' aggregate trades

This table reports results of a Shapiro–Wilk test that hedge funds' aggregate trades (columns 2 and 3), and their residuals from a fitted AR(1) process (columns 4 and 5) 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. Hedge funds' net turnover is the aggregate value of their buys, minus the aggregate value of their sells, divided by the market's total value, where buys and sells are those recorded in the Abel Noser Solutions (commonly known as ANcerno) dataset from 1999 to 2011 for asset managers classified as hedge funds according to Jame (2018). 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
	AncernoHF		AncernoHF	
Daily	15.283	0.000	15.328	0.000
Weekly	9.104	0.000	9.127	0.000
Monthly	2.334	0.010	2.138	0.016
Quarterly	1.014	0.155	1.411	0.079

Internet Appendix G: A (non-exhaustive list) of published articles that carry out a calibration or a simulation of a NREE model

1. Albagli, E. (2015). Investment horizons and asset prices under asymmetric information. *Journal of Economic Theory*, *158*, 787-837.
2. Andrei, D., & Cujean, J. (2017). Information percolation, momentum and reversal. *Journal of Financial Economics*, *123*(3), 617-645.
3. Banerjee, S., 2011, Learning from Prices and the Dispersion in Beliefs, *Review of Financial Studies* 24, 3025-3068.
4. Begenau, J., Farboodi, M., & Veldkamp, L. (2018). Big data in finance and the growth of large firms. *Journal of Monetary Economics*, *97*, 71-87.
5. 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.
6. Biais, B., Bossaerts, P., & Spatt, C., 2010, Equilibrium asset pricing and portfolio choice under asymmetric information. *Review of Financial Studies*, 23(4), 1503-1543.
7. Brennan, M. J., & Cao, H. H., 1996. Information, trade, and derivative securities. *Review of Financial Studies,* 9(1), 163-208.
8. 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.
9. Campbell, John Y., and Albert S. Kyle, 1993, Smart money, noise trading and stock price behaviour, *Review of Economic Studies* 60.
10. Cespa, G., & Vives, X., 2015, The Beauty Contest and Short-Term Trading, *Journal of Finance*, 70(5), 2099-2154.
11. Cho, J. W., & Krishnan, M. (2000). Prices as aggregators of private information: evidence from S&P 500 Futures data. *Journal of Financial and Quantitative Analysis*, *35*(1), 111-126.
12. Farboodi, M. and Veldkamp, L., 2019, Long Run Growth of Financial Technology. *American Economic Review,* forthcoming
13. Guidolin, M. (2005). Home Bias and High Turnover in an Overlapping-generations Model with Learning. *Review of international Economics*, *13*(4), 725-756.
14. Manela, A., 2014, The value of diffusing information. *Journal of Financial Economics,* 111 (1), 181–199.
15. Peress, J., 2004, Wealth, information acquisition, and portfolio choice, *Review of Financial Studies*, 17(3), 879-914.
16. Spiegel, M. (1998). Stock price volatility in a multiple security overlapping generations model. *The Review of Financial Studies*, *11*(2), 419-447.
17. Watanabe, M., 2008, Price Volatility and Investor Behavior in an Overlapping Generations Model with Information Asymmetry, *Journal of Finance* 63, 229–272.
18. Yuan, K. (2005). Asymmetric price movements and borrowing constraints: A rational expectations equilibrium model of crises, contagion, and confusion. *The Journal of Finance*, *60*(1), 379-411.

References

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.

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.

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

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.

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.

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.

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.

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

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.

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.

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.

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.

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

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.

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

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

Peress, J. and D. Schmidt, 2020, Glued to the TV: Distracted Noise Investors and Stock Market Liquidity, . *Journal of Finance,* forthcoming.

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.

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

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.