Informed Trading Intensity*

Vincent Bogousslavsky
Carroll School of Management
Boston College
vincent.bogousslavsky@bc.edu

Vyacheslav Fos
Carroll School of Management
Boston College
CEPR, ECGI
fos@bc.edu

Dmitriy Muravyev
Eli Broad College of Business
Michigan State University
CDI
muravyev@msu.edu

Journal of Finance, forthcoming

*We appreciate the feedback from Shuaiyu Chen (discussant), Tarun Chordia (discussant), Kevin Crotty, Jefferson Duarte, Jiacui Li (discussant), George Malikov (discussant), Stefan Nagel (the Editor), Andriy Shkilko, Elvira Sojli (discussant), the Associate Editor, and two anonymous referees for many helpful comments. We also thank seminar participants at NBER Big Data and High-Performance Computing for Financial Economics Conference, CFEA, 5th SAFE Market Microstructure Conference, Microstructure Asia Pacific Online Seminar, SFS Cavalcade, 5th FFIC, Arizona State University, Boston College, Florida State University, Michigan State University, Morgan Stanley Quantitative Research Colloquium, and Rice University for their helpful comments and suggestions. We thank Daniel Goodman for excellent research assistance. We also thank Kevin Crotty for sharing his informed trading measure and Patrick Augustin for sharing data on illegal insider trading. Conflict-of-interest disclosure statement: None of the authors has anything to disclose.

Informed Trading Intensity

Abstract

We train a machine learning method on a class of informed trades to develop a new measure of informed trading, the Informed Trading Intensity ("ITI"). ITI increases before earnings, M&A, and news announcements, and has implications for return reversal and asset pricing. ITI is effective because it captures nonlinearities and interactions between informed trading, volume, and volatility. This data-driven approach can shed light on the economics of informed trading, including impatient informed trading, commonality in informed trading, and models of informed trading. Overall, learning from informed trading data can generate an effective informed trading measure.

1 Introduction

Informed trading is an important ingredient of financial markets and is embedded in various academic literatures. For example, informed investors help keep security prices close to fundamental values and thus are crucial for theories of market efficiency (e.g., Fama, 1970, Grossman and Stiglitz, 1980). Informed trading also can affect asset prices by changing a firm's information structure or by increasing liquidity risk (e.g., O'Hara, 2003, Kelly and Ljungqvist, 2012). However, informed trading is hard to identify empirically because investors' information sets are not directly observable and because informed investors usually hide behind uninformed order flow. To overcome these hurdles, the literature has developed several theory-based measures of informed trading and/or adverse selection. For example, among the most well-known measures are the price impact in Kyle (1985), the bid-ask spread in Glosten and Milgrom (1985), and the probability of informed trading in Easley and O'Hara (1987). Whereas these measures have been extensively used in empirical finance, economics, and accounting, a series of recent studies shows that they perform poorly in capturing realized informed trading.[1]

In this paper, we introduce a novel data-driven approach to construct a measure of realized informed trading. We train a Gradient Boosted Trees (GBT) algorithm on informed trading data. The algorithm is trained to identify days with informed trading. In this standard classification problem, a daily indicator for informed trading is predicted by a set of same-day variables related to liquidity, return, volatility, and volume. After the model is estimated on the training data of observed informed trades, we extrapolate it to the entire stock-day universe, where informed trading is not directly observed. This procedure produces a new measure of informed trading, the Informed Trading Intensity ("ITI"). We next explain the role of the two key ingredients in the construction of ITI: informed trading data and a machine learning (ML) method.

Initially, we focus on ITI trained on Schedule 13D trades to study the basic properties of our methodology. The reason is that we expect the Schedule 13D sample to have a higher "signal-to-noise ratio" than our other informed trading samples (i.e., opportunistic insiders and short sellers) that we use later.[2] As part of Schedule 13D filing, an investor must disclose all trades made in

[1] Collin-Dufresne and Fos (2015), Kacperczyk and Pagnotta (2019), Augustin, Brenner, and Subrahmanyam (2019), Duarte, Hu, and Young (2020), Ahern (2020).

[2] Cohen et al. (2012) show that nonroutine, or opportunistic, insider trades are informed in the sense that they predict future returns. An extensive literature shows that short sellers are informed and that their trading predicts

the 60 calendar days before the filing date. Collin-Dufresne and Fos (2015) are the first to collect these trades and show that they are informed. Our methodology does not require all trades to be informed, but we refer to these trades as informed for simplicity.[3] Similarly, informed trading may occur on days without Schedule 13D trading. Our methodology does assume that more informed trading occurs on days with Schedule 13D trading than on the other nearby days, on average.

The second key ingredient to construct the ITI measure is the GBT algorithm. We train the model with a set of 41 concurrent daily variables motivated by microstructure theory that capture liquidity, return, volatility, and volume (listed in the Appendix). The results are robust to using a Random Forest algorithm instead of GBT. In contrast, linear regression and the Lasso algorithm are subsumed by the Random Forest and GBT algorithms. Thus, nonlinearities and interactions are important to detect days with informed trading. We therefore use the GBT algorithm.

We establish several main results. We first ask whether ITI detects days with Schedule 13D trading. We show that ITI is a much stronger out-of-sample detector of days with Schedule 13D trading than existing measures of liquidity and informed trading. As in most of our tests, we control for stock turnover, return, realized volatility, order imbalance, and absolute order imbalance, as well as standard measures of liquidity such as effective spread, price impact, depth, and Kyle's lambda. Moreover, while markets have changed over time, ITI keeps performing well. The explanatory power of ITI for Schedule 13D trading is stable across the sample period, whereas the explanatory power of standard liquidity variables declines. Also, though the algorithm is not trained on Schedule 13D trading volume, ITI increases with the volume traded by the Schedule 13D filer, which suggests that ITI indeed captures informed trading intensity.

What contributes to ITI's effectiveness? While most input variables contribute significantly to ITI, volume-related variables are the most important. However, even when we match days with informed trading to days without informed trading on volume, ITI strongly continues to detect days with informed trading, which indicates that the measure's effectiveness cannot be attributed

future stock returns (for example, Senchack and Starks, 1993, Boehmer et al., 2008).

[3] An investor is required to file a Schedule 13D if she becomes a beneficiary owner of at least 5% of any class of equity securities in a publicly traded company and intends to influence the management (i.e., engage in activism). Several studies document large average positive abnormal returns around Schedule 13D filings (e.g., Holderness and Sheehan, 1985, Brav, Jiang, Partnoy, and Thomas, 2008, Klein and Zur, 2009). Consistent with these results, Section 2 shows an abnormal cumulative return of about 3% in the 60 calendar days before the filing date, followed by a two-day jump in excess return of about 2% around the filing date. Importantly, Collin-Dufresne and Fos (2015) show that the positive pre-filing abnormal returns happen primarily on days when Schedule 13D filers trade, suggesting that trades by Schedule filers transmit information into prices.

solely to trading volume. We examine nonlinearities with partial dependence plots and find that ITI is increasing and concave in volume and decreasing and convex in volatility. Surrogate trees, a popular method to interpret ML models, indicate that variable interactions are also important for ITI. Specifically, ITI is particularly high if volume is high but volatility is low. Similarly, ITI is particularly low if volume is low and illiquidity is high.

After we present ITI's key properties, we extrapolate the measure to the full sample of U.S. common stocks from 1993 to 2019 (about 17 million stock-days). The model parameters are estimated on about 60,000 stock-days with data on Schedule 13D trades. The estimated model then computes ITI for each stock-day with same-day input variables. The full sample excludes the training sample of Schedule 13D trades (0.35% of the sample). The extrapolation assumes that the relations between intraday variables and realized informed trading learned by ITI largely hold in the full sample. The "signal-to-noise" ratio could be lower outside the restricted sample of Schedule 13D trades. Thus, we acknowledge that in addition to capturing informed trading, ITI could also reflect uninformed trading when applied to larger unrestricted samples. The issue of false positives applies to all ML problems, but it is more challenging to evaluate here since informed trading is unobservable even ex post. Instead, we validate ITI in a variety of tests and show that it outperforms existing measures.

We present several additional tests. Using the full sample of U.S. common stocks, we show that ITI predicts the strength of price reversal. A fundamental difference between price changes due to realized informed and uninformed trading is that price changes due to uninformed trading are expected to be transient (Hasbrouck, 1988, 1991). We therefore expect returns on days with high realized informed trading to exhibit less reversal than returns on other days. We find support for this prediction, in line with ITI capturing informed trading. This result is robust to controlling for interactions between return and turnover (Campbell, Grossman, and Wang, 1993), return and volatility, and return and effective spread. Hence, ITI is not simply a proxy for turnover, volatility, or liquidity.

Next, we show that ITI increases around several types of informational events, even when controlling for standard liquidity, volume, and volatility variables. ITI increases before earnings announcements. Furthermore, ITI predicts large abnormal returns on earnings announcement dates, suggesting that ITI increases when informed trading is more likely prior to earnings surprises.

ITI also increases ahead of unscheduled informational events: prior to unscheduled news releases and prior to M&A announcements. Informed trading increases in the days following announcements. The disclosure of news could increase information asymmetry due to heterogeneity in information-processing capacity (Kim and Verrecchia, 1994). Informed investors could also take advantage of increased volume to camouflage their trades.[4]

One may be tempted to conclude that ITI is an aggregated liquidity measure. The increase in ITI ahead of earnings announcement disproves this conclusion as liquidity tends to worsen before earnings announcements (e.g., Lee et al., 1993). In addition, recall that ITI increases with the size of the activist trade, controlling for liquidity variables. Hence, while we expect ITI to correlate with liquidity variables because informed investors time liquidity (Collin-Dufresne and Fos, 2016), ITI is not equivalent to a liquidity measure.

After we establish the effectiveness of ITI in detecting informed trading, we explain how our data-driven approach can shed light on the economics of informed trading. First, we show an important distinction between patient and impatient trading, in line with theory. We take advantage of a useful feature of Schedule 13D trading data. After crossing the 5% ownership threshold, a Schedule 13D filer must file with the SEC within 10 days. Therefore, Schedule 13D filers trade more aggressively closer to the filing (Collin-Dufresne and Fos, 2015). We use this feature of the data to decompose ITI into a "patient" ITI and an "impatient" ITI, where ITI(patient) and ITI(impatient) are trained on the first 40 days and the last 20 days of the 60-day filing window, respectively.

We find that impatient informed trading is easier to detect. This result supports theories such as Kaniel and Liu (2006) where informed traders tends to use more aggressive orders as the horizon of their information shortens. Whereas ITI(impatient) and ITI(patient) are both positively correlated with turnover, the relation for ITI(impatient) is much stronger. ITI(impatient) is positively associated with realized volatility, whereas ITI(patient) displays the opposite pattern. Moreover, in the two days before an earnings announcement, ITI(impatient) increases strongly whereas ITI(patient) does not. These findings are consistent with ITI(impatient) detecting days with aggressive informed trading. Finally, ITI(impatient) is able to detect illegal insider trades (Ahern, 2020).

Second, our methodology sheds light on commonality in informed trading. We show that ITI is

[4] A post-announcement increase in informed trading is consistent with the results of Lee, Mucklow, and Ready (1993), Back, Crotty, and Li (2018), Brennan, Huh, and Subrahmanyam (2018), among others.

effective in detecting other classes of informed trading, as proxied by opportunistic insider trades and spikes in short selling. Fluctuations in the trading environment are likely to generate commonality in informed trading, but controlling for standard liquidity measures only partly explains the commonality. Moreover, we show that ITI continues to detect opportunistic insider trades and spikes in short selling even when we control for ITI measures trained on these datasets. Hence, incremental information can be gained about one type of informed trading from studying other types of informed trading. It is not obvious how strong commonalities in informed trading are ex ante. Our methodology provides a first pass at this question by comparing the ITI measure trained on Schedule 13D data to ITI measures trained on other datasets.

Third, our data-driven measure indicates that a fruitful avenue to shed light on the economics of informed trade is to develop models where volume and absolute order imbalance are both positively related to realized informed trading, even when controlling for one another, but where volatility is negatively related to realized informed trading. Based on simulated data from a range of informed trading models, we show that this dimension of informed trading is challenging to capture. Across all models that we consider, none is able to jointly match these three relations. Finally, we also show that ITI measures remain significant detectors of days with informed trading when we control for five well-known measures of informed trading.

To conclude, we provide a simple application of ITI by examining whether informed trading predicts future stock returns. Increased realized informed trading, as measured by higher ITI, is associated with higher future monthly returns in panel regressions. In portfolio sorts, the Fama-French four-factor alpha for the difference between the top and bottom decile portfolios based on ITI is 52 basis points (bps) per month, or 6.4% annualized, with a t-statistic of 6.2. Other ITI measures, except for ITI trained on short selling data, also positively predict returns. Our results imply a positive relation between realized informed trading and future returns, which supports theories in which stock purchases are more informed than stock sales. The results lend only limited support to theories where information risk is priced.

The paper contributes to four strands of literature. First, an extensive microstructure literature develops measures of stock liquidity and adverse selection (e.g., Glosten and Milgrom (1985), Easley and O'Hara (1987), Hasbrouck (1988, 1991), Amihud (2002)). Our key contribution to this literature is to show that a data-driven approach performs remarkably well in detecting various types of informed trading. Furthermore, information learned from one type of informed trading helps detect other types of informed trading. Finally, we show that our approach generates new insights about informed trading that can guide theory.

Second, our paper contributes to a series of recent papers that use informed trading data to evaluate the performance of stock liquidity and adverse selection measures as well as the trading choices of informed investors (e.g., Collin-Dufresne and Fos (2015), Gantchev and Jotikasthira (2018), Augustin and Subrahmanyam (2020), Cookson, Fos, and Niessner (2021), Akey, Gregoire, and Martineau (2022)). Recently, Back, Crotty, and Li (2018) combine return and order flow information to detect informed trading and show that their measure increases when Schedule 13D filers trade. In this paper, we use three classes of informed trades not as a laboratory to evaluate the performance of a particular measure, but instead as a training set for a ML method. Our contribution is to show that the obtained measures perform remarkably well in detecting informed trading in various settings. Moreover, we show that volume is the most important feature for all of the ITI measures, implying that the informed trading literature has perhaps focused too much on order imbalance relative to trading volume.

Third, we contribute to the active debate in the asset pricing literature on whether informed trading affects asset prices (e.g., Easley, Hvidkjaer, and O'Hara, 2002, Duarte and Young, 2009, Yang, Zhang, and Zhang, 2020). For example, Easley and O'Hara (2004) find a positive relation between PIN and future stock returns and justify it theoretically. Hughes, Liu, and Liu (2007) debate whether information risk is priced. Duarte and Young (2009) distinguish between the liquidity and information asymmetry components of PIN. We contribute to this debate by showing that ITI measures of realized informed trading are positively priced in the cross-section. We also suggest asymmetry in the informational content between stock purchases and sales as a likely explanation.

Finally, only a handful of studies apply ML techniques to market microstructure. Easley, de Prado, O'Hara, and Zhang (2021) show that microstructure-based measures predict changes in liquidity and volatility. Kwan, Philip, and Shkilko (2021) use reinforcement learning to study price discovery. In contrast, we use ML techniques to identify informed trading. We borrow some ideas from a more developed literature on ML in asset pricing (see Gu, Kelly, and Xiu (2020), Karolyi and Van Nieuwerburgh (2020), Goldstein, Spatt, and Ye (2021), Nagel (2021) for reviews).

2 Data

We use three data sources on informed trading: trades by Schedule 13D filers, opportunistic insider trades, and short sales. Table IA.2 in the Internet Appendix provides an overview of the datasets that we use and how we construct them.

Data on trades by Schedule 13D filers come from Schedule 13D filings available on SEC EDGAR. Our sample construction procedure closely follows Collin-Dufresne and Fos (2015).[5] The 13D sample includes the 60-day disclosure period up to the filing date of 1,593 Schedule 13D filings between 1994 and 2018. We extract the following information from each Schedule 13D filing: CUSIP of the underlying security, date of every trade, trade type (purchase or sell), size, and price. In this sample, Schedule 13D filers trade on about 36% of days.

Schedule 13D filers know they can increase the value of the firm they invest in by their own effort. Their effort level is, of course, conditional on them achieving a large stake in the firm (Back et al., 2018). In our setting, their very actions and share-holdership constitute the "private" information. Only when they reach the 5% threshold, does the information, due to the disclosure requirement, become public. Announcement returns allow us to measure the extent to which the market believes their future actions have value over and above what is already reflected in prices. Figure IA.1 in the Appendix plots the average buy-and-hold return, in excess of the buy-and-hold return on the CRSP value-weighed index, from 40 days prior to the filing date to ten days after. Like Collin-Dufresne and Fos (2015), we find a run-up of about 3% from forty days to one day prior to the filing date. The two-day jump in excess return at the filing date is around 2%.

Cohen, Malloy, and Pomorski (2012) show that opportunistic insider trades are informed.[6] We follow their methodology and construct a sample of opportunistic insider trades (both buys and sells). Specifically, the data are from Table 1 of the Thomson Reuters Insider Database. Transactions associated with derivative securities and observations containing cleanse indicators

[5]First, using an automatic search script, we identify all Schedule 13D filings from 1994 to 2018. Next, we check the sample manually and identify events with information on trades. Since the trading characteristics of ordinary equities might differ from those of other assets, we retain only U.S. common stocks (CRSP share code 10 or 11). We exclude stocks with price below $5 or market capitalization below $100 million at the beginning of the filing window. Moreover, we exclude events that involve derivatives, such as options, warrants, and swaps (see Collin-Dufresne et al., 2020). Finally, we exclude Schedule 13D/A filings (i.e., amendments to previously submitted filings) that are mistakenly classified as original Schedule 13D filings.

[6]Other recent studies that identify informed insider trading include Akbas, Jiang, and Koch (2020), Alldredge and Cicero (2015), Biggerstaff, Cicero, and Wintoki (2020), Cziraki and Gider (2021).

"S" and "A" are dropped from the sample. The (non)routine trades follow the classification defined in Cohen et al. (2012): "We define a routine trader as an insider who placed a trade in the same calendar month for at least a certain number of years in the past. We then define opportunistic traders as everyone else, that is, those insiders who have traded in the same years as the routine insiders, but for whom we cannot detect an obvious discernible pattern in the past timing of their trades." Our sample of insider trades starts in January 1993 and ends in December 2012.

We use short interest data from Markit, which include daily data on securities borrowing and lending activity. Markit obtains the information from more than 100 equity loan market participants, who together account for approximately 85% of US securities loans (Markit, 2012). The data start in July 2006 (the daily updates begin) and end in December 2019. Short interest is defined as the quantity on loan from Markit divided by shares outstanding from CRSP. We identify large spikes in short interest by comparing daily changes in these variables to their 90th percentiles over the entire sample. Markit reports the date when short sales are settled, which is three days (two days post 9/4/2017) after trades take place (Richardson, Saffi, and Sigurdsson (2017)). We adjust for this date shift before merging the short sale data with CRSP and TAQ. The Markit data are widely used in academic research on short selling. Furthermore, a large literature documents that short sellers are informed and that their trading predicts future stock returns (for example, Senchack and Starks (1993), Boehmer et al. (2008)). Reed (2013) reviews this literature.

We use several additional data sources. Earnings announcement dates are obtained from Compustat. We adjust for after-hours announcements by comparing trading volume on the reported announcement day to volume on the following business day. The announcement date is set to the day with highest volume. Data on M&A announcements are obtained from Thomson Reuters Securities Data Company (SDC) database, and we require that both target and acquirer are public U.S. companies. Stock returns, volume, and prices come from the Center for Research in Security Prices (CRSP). Intraday transactions data (trades and quotes) come from the Trade and Quote (TAQ) database. Table IA.3 in the Appendix lists data sources for the set of variables used to train the ML model.

Control variables are defined in Table IA.4 in the Appendix and their descriptive statistics are reported in Table 1. We report descriptive statistics for the full sample, which includes common stocks from 1/1993 to 7/2019. We focus on common stocks and exclude stocks with a price below

$5 or a market capitalization lower than $100 million at the beginning of each month to limit the influence of microcaps and penny stocks on the results. Table 1 also reports descriptive statistics for trading indicators in the 13D, insider, and short samples, as well as for the informed trading measures constructed based on these samples. We discuss these numbers after we introduce our methodology.

[Insert Table 1 about here.]

3 Informed Trading Intensity

In this section, we introduce the ITI approach to detecting informed trading. Empirical measures of informed trading are widely used in many contexts. But unlike liquidity measures such as the bid-ask spread, informed trading is directly observed only in limited cases. Informed investors also usually hide behind uninformed order flow to avoid being detected. To overcome these limitations, empirical measures are typically motivated by theories of informed trading. For example, the PIN measure of Easley et al. (1996) is a nonlinear function of order imbalance and a sufficient statistic for identifying informed trading in their sequential trading model.

We propose a data-driven approach to detect realized informed trading. This approach learns from the data how days on which informed investors trade differ from days on which they do not trade. We use ML techniques to account for non-linearities and interactions between variables, while cross-validation and regularization, which are standard in ML, prevent overfitting the data. After the model is estimated on a training sample of informed trades, we extrapolate it to the entire stock universe and compute ITI by applying the model parameters to the set of input variables for a given stock-day. ITI is bounded between zero and one and is higher when informed trading is more likely.

3.1 Estimating the Informed Trading Intensity

Economists primarily rely on the linear regression, a simple model that identifies individual effects and their significance. We also rely on classic regression models after the ITI measure is computed, but we show that more flexible methods are required to detect informed trading because the underlying relations are likely complex and include nonlinearities and variable interactions. For example, informed trading could be more likely if trading volume is above, say, the 90^{th} percentile and stock volatility is below the median. Of course, nonlinearities and interactions can be hard-wired into a linear regression, but ML methods discover them naturally.

We have to make several design choices. First, we pick a dataset where informed trading is directly observed (a training sample in ML terminology) so that a ML method can learn from contrasting model predictions with observed outcomes (supervised learning). We begin with trades disclosed in Schedule 13D filings, which report activist trading dates and trade sizes in a 60-day window prior to the filing. Once we develop the ITI measure based on Schedule 13D trades, we introduce ITI measures based on opportunistic insider trades and spikes in short selling. This variety of informed trading data allows us to study commonalities across various types of informed trading in Section 4.2. We focus on ITI trained on Schedule 13D trades because this dataset is likely to have the highest "signal-to-noise ratio." In particular, when Schedule 13D filers trade in our sample, they trade an average of 28% of daily volume.

Second, we pick a ML method. We outline a method's main idea and refer to the classic textbook of Hastie et al. (2017) for details. Lasso is popular in economics because it resembles a linear regression except that it shrinks coefficients, making many of them exactly zero (Tibshirani (1996), see Chapter 3 in Hastie et al. (2017)). Lasso is easy to interpret but only allows for pre-specified nonlinearities and interactions, and can behave poorly if predictors are highly correlated. The other two methods rely on decision trees (see Chapters 9 and 10 in Hastie et al. (2017)). Consider a simple tree example, if volume is above the 90^{th} percentile, split on whether the bid-ask spread is above or below the median, otherwise split on order imbalance's 30^{th} percentile. Each of the four leaves is then assigned an expected frequency of activist trading (the historical average).

Decision trees have many desirable features; they are invariant to variable scaling and robust to outliers. Random Forest takes an average over many random decision trees (Breiman (2001)). Each of these many trees is constructed on a sample bootstrapped from a random subset of all predictors from the original dataset (e.g., 10 our of 100, see Chapter 15 in Hastie et al. (2017)). Finally, our preferred method is eXtreme Gradient Boosting (XGBoost, Chen and Guestrin (2016)), which efficiently implements Gradient Tree Boosting (see Chapter 10 in Hastie et al. (2017)). While Random Forest averages over random trees ("bagging"), in Gradient Tree Boosting, each new tree focuses on examples that previous trees find problematic ("boosting"). In general, boosting produces better forecasts than bagging but is slower to estimate.

XGBoost makes Gradient Tree Boosting almost as fast as Random Forest. It also recognizes that trees are prone to overfitting and penalizes trees with many leaves in favor of simpler shorter trees (i.e., regularization). Regularization makes the models perform worse in-sample, but improves out-of-sample performance, which is the goal. We use the scikit-learn package in Python that implements Lasso and Random Forest and provides an XGBoost interface. While we focus on XGBoost because it yields the best performance, our main results also hold with Random Forest.[7]

Third, we pick 41 predictors (or "features") described in Table IA.3. Four predictors are from CRSP daily files: stock price, return, absolute return, and trading volume. The remaining predictors are based on intraday data from TAQ. The WRDS Intraday Indicators database aggregates TAQ data to the stock by day level into 289 variables (many of which are near duplicates) out of which we select 21 unique variables motivated by the microstructure literature. We supplement them with 16 variables that are missing from WRDS Indicators, such as depth, morning and afternoon returns, and realized volatility. We pick a limited number of predictors motivated by microstructure theories instead of directly estimating a ML method on every quote update and trade in TAQ. We prefer fewer predictors because the statistical power is limited in our setting with a training sample of about 60,000 stock-days and with informed investors actively trying to avoid detection. Also, having interpretable predictors, such as volume and volatility, helps us explain why ITI measures work. All features are standardized by subtracting their average and dividing by their standard deviation over the prior year. The standardization makes features comparable across stocks and makes them easier to interpret: if today's volume is three standard deviations above the average, what does it mean for (today's) informed trading? Note that since our focus is on detecting realized informed trading rather than estimating expected informed trading, we use contemporaneous rather than lagged predictors of informed trading.[8]

[7] Chen and Guestrin (2016) introduced XGBoost and accumulated over 10,000 Google citations in just a few years. When it comes to small-to-medium structured/tabular data, XGBoost and similar decision tree-based algorithms are considered best-in-class at the moment. About 500 developers in XGBoost's GitHub community translated the method to all major programming languages.

[8] Predicting informed trading is likely much harder than detecting it ex-post. Informed traders want to hide their trades. As a result, the set of variables that predicts informed trading likely varies over time. To illustrate, we construct ITI to predict today's informed trading using only yesterday's data (i.e., $t-1$ data instead of t data). We find that ITI$(t-1)$'s explanatory power to detect Schedule 13D trades on day t is about 13 times lower than that of ITI(t) . This is not surprising, by analogy, one can easily tell ex-post whether a day had high volatility, but the ex-ante volatility forecast will do much worse.

Cross-validation helps us avoid overfitting and keeps the analysis out-of-sample. Specifically, we follow a standard approach in ML and split the Schedule 13D events into five non-overlapping parts in calendar time and set one part (20% of the data) aside for evaluation.[9] On the remaining 80% of the data, we use standard cross-validation to find the set of parameters that balances in-sample and out-of-sample performance; i.e., we again split the 80% into five parts, estimate a model on four parts, and evaluate its performance on the remaining part. Once we settle on the model, we return to the 20% of the data set aside at the beginning and evaluate the model on a set of observations that it never touched before. We then rotate the last part of the data that we set aside and repeat the above analysis. This allows us to evaluate model performance out-of-sample on the entire Schedule 13D sample.

In the second part of our analysis, we estimate the model on the full training dataset and extrapolate it to the entire cross-section of stocks between January 1993 and July 2019. Variables are processed exactly as for the training sample and then are supplied to the estimated model that in turn produces ITI for a given stock and day. This extrapolation exercise makes implicit assumptions that we discuss in Section 3.3.

3.2 Properties of Informed Trading Intensity

In this section, we discuss the properties of ITI trained on trades by Schedule 13D filers. We regress an indicator for Schedule 13D trading on ITI and other liquidity measures. All of our specifications include filing fixed effects. Most specifications control for the effective bid-ask spread, price impact, Kyle's lambda, market depth, realized volatility, turnover, order imbalance, and absolute order imbalance, which are described in Table IA.4 in the Appendix. We control for order imbalance in addition to absolute order imbalance to capture simple "lean against the wind" effects. For example, activists mostly buy and thus could mostly trade against sell imbalances. Table IA.5 in the Appendix reports descriptive statistics for the 13D sample.

The first column of Table 2 shows that ITI predicts Schedule 13D trades, with a t-statistic greater than 43 and R^2 of 9.86%. While this number is much lower than 100%, it is large in this context. Indeed, many microstructure models assume that informed investors hide behind uninformed trading to avoid detection, and a high R^2 would directly contradict this basic assumption.

[9] Other alternatives for splitting the sample yield similar results.

Thus, ITI measures should be benchmarked against other alternative measures.[10] ITI is benchmarked against alternative measures in the second column of Table 2, which reports estimates of the same regression with our set of common liquidity variables. The adjusted R^2 of the regression is 4.61%, less than half that achieved by ITI alone. Effective spreads, Kyle's lambda, depth, and realized volatility all suggest improved liquidity on days with Schedule 13D trading. Thus, these common liquidity variables are not able to detect Schedule 13D trades (Collin-Dufresne and Fos, 2015). Again, a natural explanation for this result is that Schedule 13D traders strategically time their trades (Collin-Dufresne and Fos, 2015, 2016). The third column shows that the inclusion of common liquidity variables to ITI increases the R^2 from 9.86% (with ITI alone) to 10.73%. The coefficient on ITI decreases only slightly when we control for common liquidity measures.

[Insert Table 2 about here.]

We use ITI as a measure of informed trading intensity, but does it pick up Schedule 13D trading intensity? To answer this question, we use Schedule 13D trading volume data to compute the share turnover of the Scheduler 13D filer (share volume traded divided by total shares outstanding); i.e., Schedule 13D turnover. We then regress Schedule 13D turnover on ITI and control variables. To avoid restating the results in Columns (1)-(3), we restrict the sample to days when Schedule 13D filers trade. Hence, we focus on the intensive margin.

The last three columns of Table 2 report the results. Column (4) shows that ITI is strongly positively associated with Schedule 13D turnover, with a t-statistic greater than 22. The positive and significant relationship between ITI and Schedule 13D turnover remains when we control for common liquidity measures. This result is not a mechanical result because we do not incorporate Schedule 13D trading volume information when training the model. While we could use this information, we opt to keep the classification problem as simple as possible. Overall, ITI is associated with informed trading on both the extensive and intensive margins, even when controlling for common liquidity, volume, and volatility measures.

In summary, Table 2 shows that ITI captures informed trading not detected by standard liquidity measures. We discuss additional measures of informed trading in Section 4.3 and reach similar

[10]To provide another perspective, if we consider an accuracy metric such as ROC AUC, ITI has a 71% probability of ranking a randomly selected day with 13D trading above a randomly selected day without 13D trading (reported in Table IA.6 in the Appendix). A random classifier has a 50% ROC AUC.

conclusions. Also, even if part of ITI is subsumed by, say, turnover, this is a feature of our methodology. ITI can be interpreted as a measure of informed trading intensity, whereas it is difficult to interpret turnover in a similar way (e.g., Duarte et al., 2020).

Markets are changing over our sample period due to technological innovations such as algorithmic trading, lower transaction costs, and higher competition. Is the explanatory power of ITI mostly driven by specific parts of our sample? To address this concern, Figure 1 plots the out-of-sample R^2 when predicting Schedule 13D trades with ITI year by year.[11] The figure compares the R^2 produced by ITI with the R^2 produced by a set of common liquidity variables. ITI's ability to predict this class of informed trades is relatively stable over time and does not display any trend. In contrast, common variables' ability to predict activist trades slowly declines over the sample period. Hence, the "predictability gap" between ITI and common measures tends to be larger post 2001. While several factors may be at play, one interesting possibility is that this trend originates from the rise of algorithmic trading post decimalization. Algorithmic trading makes it easier for informed traders to camouflage their trades, which are therefore harder to detect with a simple linear specification. Overall, the predictive power of the ITI measure is not driven by a part of our sample period.

[Insert Figure 1 about here.]

It is often difficult to explain why ML methods work (e.g., Nagel (2021)). To address this concern, we identify ITI's most important components. We split the model's features, listed in Table IA.3, into four groups: liquidity, return, volatility, and volume variables. We then train the model using only each subset of the variables to predict Schedule 13D trades. We therefore obtain four subset-specific ITI variables. Table 3 reports the results. The volume grouping dominates all other groupings. Return and volatility variables have low explanatory power. On their own, liquidity variables such as spread, price impact, and depth achieve an out-of-sample explanatory power roughly half that of ITI constructed from all the variables. Nevertheless, the last column of Table 3 shows that none of the groupings is subsumed by the other groupings. This result suggests that one type of information cannot fully capture the occurrence of Schedule 13D trades.

[11] Figure IA.2 in the Appendix shows the distribution of observations over time in the dataset of Schedule 13D filings.

In particular, we show below that the *interaction* between volume and volatility plays an important role to detect Schedule 13D trades.

[Insert Table 3 about here.]

We also confirm the above results by performing traditional ML procedures to rank features by importance. First, we use XGB's internal procedure that ranks features based on their gain (a default option based on the average gain across all splits where a feature was used). Figure IA.3 in the Internet Appendix reports the ranking and shows that volume is by far the most important variable followed by volatility. Other variables are not far behind. Another popular ranking method, the SHAP (SHapley Additive exPlanations, Lundberg and Lee (2017)) is inspired by cooperative game theory. For each observation x and feature f, a SHAP value measures a weighted average gain from adding f to all possible feature subsets. A separate model must be trained for each possible subset of features, which is computationally expensive. Figure IA.4 ranks features according to SHAP and plots the distribution of SHAP values over observations in the 13D sample. Volume is again the top variable, but other variables are not far behind.

As explained in Section 3.1, our ITI measure relies on a Gradient Boosted Tree (GBT) algorithm. Table IA.6 in the Appendix shows that GBT outperforms the Lasso and Random Forests when predicting Schedule 13D trades. The out-of-sample R^2 increases from 6.44% with Lasso to 9.74% with Random Forests, and further increases to 13.68% with GBT. Standard linear regression performs very similarly to Lasso in our sample. As mentioned above, our main results are robust to using Random Forests to construct ITI. Table IA.6 also shows that Lasso is subsumed by Random Forests and GBT, which suggests that nonlinearities matter for detecting informed trading.

In Figure 2, we examine nonlinearities with partial dependence plots (see Section 10.13 in Hastie et al. (2017)). These plots show how ITI depends on a variable of interest marginalizing over the values of all other input variables. We pick volume and volatility for the plots because the XGBoost algorithm ranks them as the most important determinants of ITI in Figure IA.3. ITI is increasing and concave in volume and is decreasing and convex in volatility.

[Insert Figure 2 about here.]

We next examine how variable interactions affect ITI with the help of a surrogate tree, another standard ML technique (Craven and Shavlik (1995)). ITI is determined by a complex model. However, ITI can be approximated by a simple three-level tree, a piece-wise constant function with just eight values (leaves). Figure 3 shows the results. The tree aims to capture ITI's most important properties and explains about one-third of ITI's variation. It produces two insights. First, how often a variable enters a tree and how high it is can help assess its importance. Figure 3 shows that volume enters the tree three times including at the top level. Volatility and illiquidity determine the second level. Second, ITI is particularly high if volume is high but volatility is low. Similarly, ITI is particularly low if volume is low and illiquidity is high. This illustrates how ITI takes into account various nonlinearities and interactions between input variables.

As another way to assess the importance of interactions, we regress ITI on the ITI "subset measures" that we use in Table 3. The results are reported in Table IA.7 in the Internet Appendix. ITI(volume) explains about 35% of the (within-filing) variation in ITI. Furthermore, regressing ITI on all the subset measures achieves an adjusted R2 of about 47%. Thus, interactions across types of explanatory variables are likely important to explain variation in ITI.

[Insert Figure 3 about here.]

We conclude this section by providing an additional piece of evidence regarding the role of trading volume. Duarte et al. (2020) show that trading volume is a key driver of several existing measures of informed trading. While the results in Table 3 show that volume-related variables are not the only drivers of ITI, we next report results of a matching test based on Schedule 13D sample. Specifically, within each 13D filing, each day with a Schedule 13D trade ("treated") is matched to a non-treated observation based on turnover. Only paired observations whose absolute difference in turnover is less than 0.00002 are kept. This threshold is selected such that the within-filing difference in turnover between treated and matched days is statistically insignificant.

Table 4 presents the results. The first column of Panel (a) shows that treated and matched days have identical levels of turnover. Yet, the positive and significant coefficient of ITI in column 2 indicates that ITI is higher on days with informed trading even across days matched on turnover. In Panel (b), an indicator for days with Schedule 13D trading is regressed on a set of liquidity variables and filing fixed effects in the matched sample. ITI detects days with informed trading across days matched on turnover whereas none of the control variables is statistically significant.

We therefore conclude that ITI carries information not captured by trading volume.

[Insert Table 4 about here.]

3.3 Extrapolation to the Full Sample of Stocks

The main advantage of our approach is that ITI can be extrapolated beyond the training sample of informed trades to the full sample of U.S. common stocks. Specifically, ITI can be computed from the model estimated on a particular class of informed trades as long as the input variables are observed, which is limited only by the availability of TAQ data. We next perform this extrapolation exercise for ITI trained on Schedule 13D trades. In all of the results for the full sample of stocks, Schedule 13D trading periods are excluded since they are used to train the model. This step drops 0.35% of the sample.

The extrapolation assumes that the relations between intraday variables and realized informed trading learned by ITI largely hold in the full sample. Since ITI is trained on a restricted sample of Schedule 13D trades, a natural concern is that ITI could capture uninformed trading when applied to larger unrestricted samples. The issue of false positives applies to all ML problems, but it is more challenging to evaluate here since informed trading is unobservable even ex post. Thus, we acknowledge that, in addition to capturing informed trading, ITI could also reflect uninformed trading. Instead, we validate ITI in a variety of tests and show that it outperforms existing measures.

Column (1) of Table IA.8 reports how ITI correlates with standard liquidity measures for the full sample of common stocks. ITI is negatively related to the effective spread. ITI is negatively related to lambda and positively related to depth, turnover and absolute order imbalance. These relations echo the relations between Schedule 13D trades and liquidity measures in Table 2. Also, the R^2 of the full sample regression is around 12%. Hence, within-stock variation in ITI is not well captured by standard liquidity measures.

3.4 Informed Trading ahead of Informational Events

Next, we study the dynamics of ITI ahead of informational events. We begin with the investigation of the behavior of ITI around earnings announcements, an example of a prescheduled informational event. ITI is regressed on indicator variables for each of the 10 days before and 10 days after earnings announcement, and stock fixed effects. The coefficients on the indicator variables with 95% confidence intervals are plotted in Panel (a) of Figure 4 and can be interpreted as average changes in the value of ITI on the days around earnings announcement relative to the average stock-specific ITI value outside of these days.

[Insert Figure 4 about here.]

Panel (a) of Figure 4 shows that ITI is statistically higher two days ahead of an earnings announcement, spikes on the day of the announcement, and then remains elevated for several days. In Table 2, ITI is negatively associated with most liquidity measures. One may be tempted to conclude that ITI is simply an "aggregated" liquidity measure. The increase in ITI ahead of earnings announcement disproves this conclusion. As is well-known, market makers increase spreads and lower depth before earnings announcements (e.g., Lee et al., 1993). Hence, while ITI picks up the properties of informed trading, it is not equivalent to a liquidity measure.

Campbell et al. (2009) measure daily institutional trading from trade size combined with a buy-sell classification algorithm and find increased institutional trading in the direction of future earnings surprises already 40 days prior to earnings announcements. This is not inconsistent with our results. Figure 4 measures the difference in realized informed trading across days for a given stock, which suggests that informed trading intensity remains on average constant until a few days prior to earnings announcements. This is broadly consistent with Campbell et al. (2009), who show that cumulative institutional flows increase linearly in the direction of future earnings surprises.

The pattern in ITI is consistent with informed trading prior to the announcement but also following the announcement. This result is consistent with the model of Kim and Verrecchia (1994) in which the disclosure of public news increases information asymmetry among investors since some investors can better process new information than other investors. Another possibility is that informed traders are better able to camouflage their trades following earnings announcements thanks to the higher post-announcement trading activity, as in the model of Collin-Dufresne and Fos (2016). Nevertheless, since the pattern is robust to controlling for turnover, ITI does not simply pick up higher turnover around earnings announcements. Empirically, an increase in informed trading following earnings announcement is consistent with the results of Lee et al. (1993), Back et al. (2018), Brennan et al. (2018), among others.

We next investigate whether the dynamics of ITI depend on the content of earnings announcements and consider the announcement day return. Panel (b) plots the results for announcements that are in the top and bottom quartile of absolute announcement day return. ITI increases most strongly ahead of earnings announcements associated with large absolute surprises. Higher informed trading could be associated with a smaller announcement surprise as more information is incorporated in the price ahead of the announcement. Our results suggest otherwise. As a robustness check, we obtain qualitatively similar results with a measure of earnings surprise based on median analyst forecast (which therefore does not depend on the announcement return).

Earnings announcements are pre-scheduled and may not be representative of informational events in general. We next examine the days prior to news releases, the timing of which is generally unknown to uninformed investors. News data are obtained from Boudoukh et al. (2019) and covers S&P 500 common stocks from 2000 to 2015.[12] ITI is regressed on indicator variables for days around news, and stock fixed effects. To make sure that we are not picking up the results in Panel (a), we exclude five-day windows centered on earnings announcement days. Panel (c) in Figure 4 plots the coefficients with 95% confidence intervals. ITI increases before news and this increase is statistically significant. ITI spikes on the day of the announcement and remain elevated for several days.

We find similar evidence using M&A announcements. Figure IA.5 in the Internet Appendix reports ITI in the ten days leading to M&A announcements. ITI starts increasing five days before M&A announcements. Like Brennan et al. (2018), we also find a dramatic increase in informed trading after M&A announcements. Overall, ITI increases ahead of both scheduled and unscheduled informational events. This is intuitive and supports the view that ITI is effective in detecting informed trading

To conclude this section, we use one of the most fundamental intuitions on how informed trades and liquidity trades affect prices. The price impact of informed trades should be permanent, whereas the price impact of liquidity-motivated trades should be transient (Hasbrouck, 1988, 1991). Liquidity providers set prices to reflect the expected intensity of informed trading. If the realized intensity of informed trading (as measured by ITI) is below the expected level, then prices will

[12]We thank these authors for making this dataset available.

partially revert. To illustrate, in the seminal model of Glosten and Milgrom (1985) with a risk neutral market maker, price changes are uncorrelated, independently of the probability of informed trading *expected* by the market maker. However, conditional on knowing that a specific price change is associated with an informed (uninformed) trade, the following price change exhibits on average continuation (reversal).[13] Hence, we expect returns on days with high realized informed trading to reverse less than returns on other days, which we can test with ITI in the following panel regression:

$$r_{i,t+1} = a_t + b_1 * r_{i,t} + b_2 * \text{ITI}_{i,t} + b_3 * \text{ITI}_{i,t} * r_{i,t} + \text{controls} + e_{i,t+1}, \tag{1}$$

where $r_{i,t}$ is the return of stock i on date t.[14] We expect $b_3 > 0$ if there is less reversal on days with higher ITI. Table 5 reports the results. Column (1) shows that a higher return on day t is followed by a lower return on day $t + 1$. In Columns (2) and (3), we interact ITI with the stock return. Higher ITI is associated with lower return reversal. This result is consistent with ITI capturing days with informed trading. In terms of economic magnitude, the within-date standard deviation of ITI is 0.15. Using the values in Column (2), a two (within) standard deviation increase in ITI reduces return reversal by about two thirds.

[Insert Table 5 about here.]

In Column (3), we add interactions between return and turnover, return and volatility, and return and effective bid-ask spread. According to inventory models, return reversal should increase with turnover and volatility (e.g., Campbell et al. (1993), Llorente et al. (2002), Nagel (2012)). To further control for liquidity effects such as the bid-ask bounce, we also include an interaction between return and effective spread. Importantly, the ITI-return interaction is not affected by including these liquidity controls. Moreover, the ITI-return interaction is positive, while the other interactions are zero or negative. These results suggest that ITI captures informed trading rather than inventory shocks or microstructure effects. ITI does not simply proxy for volume, volatility, or liquidity.[15] Hence, the results in Table 5 support the view that ITI is effective in detecting days

[13]This is formally shown in Section IA.A in the Internet Appendix.

[14]Duarte et al. (2020) conduct a similar test with other informed trading measures.

[15]Table IA.9 in the Internet Appendix uses as dependent variable the future return up to ten (trading) days after the initial return. Coefficients are positive and statistically significant for the first two days, positive but generally insignificant between three and six days, and a mix of positive and negative without statistical significance after six days. This pattern does not support the idea that persistent liquidity shocks drive the results.

with informed trading.

4 The Economics of Informed Trading Intensity

In this section, we explain how our data-driven approach can shed light on the economics of informed trading. We show an important distinction between patient and impatient trading, in line with theory. We also show how the methodology can shed light on commonality in informed trading. In particular, ITI is effective in detecting various types of informed trading beyond what can be achieved with standard measures. Last, we discuss what ITI suggests standard models of informed trading are missing and compare ITI to existing measures of informed trading.

4.1 Patient and Impatient Trading

An important theoretical distinction can be made between impatient trading and patient trading. Foucault, Kadan, and Kandel (2005) show that traders' patience is a key determinant of limit order book dynamics in a model without asymmetric information. Kaniel and Liu (2006) show that the horizon of private information is crucial for informed traders' choice between market orders and limits orders. As the horizon increases, the probability of informed traders using limit orders increases, which affects liquidity measures such as the bid-ask spread. Caldentey and Stacchetti (2010) study a version of the Kyle (1985) model in which the asset value is publicly disclosed at a random time. In this model, the insider trades more aggressively when the expected time until disclosure decreases.[16]

In the 13D setting, after an investor reaches 5% ownership in a security, the investor must disclose their positions within 10 days. The first part of the 13D window can be thought of as insider trading with an (almost) infinite expected horizon of information disclosure. We therefore expect 13D filers to trade more aggressively after reaching the 5% ownership threshold. Indeed, the cumulative return pattern in Figure IA.1 is consistent with this idea as the bulk of the abnormal return is earned close to the filing date. We use this feature of the data to decompose ITI(13D) into a "patient" ITI and an "impatient" ITI. To do so, we estimate ITI(patient) using the first 40 days

[16] Bolandnazar et al. (2020) provide evidence consistent with the Caldentey and Stacchetti (2010) model using information accidentally disclosed to some investors a few seconds to a few minutes ahead of the public.

of the filing window and ITI(impatient) using the last 20 days of the filing window.[17] Consistent with greater impatience, activists trade on average 37% of daily volume on 49% of days in the last 20 days of the filing window versus 20% of daily volume on 30% of days in the first 40 days.

Both ITI(impatient) and ITI(patient) are effective in detecting days with Schedule 13D trading, as reported in Table IA.10 in the Appendix. Moreover, both measures are higher when Schedule 13D filers purchase a larger number of shares on a given day. Importantly, ITI(patient) and ITI(impatient) are not the same measure. Their unconditional correlation is 0.47. Our set of control variables explains 19.38% of the variation in ITI(impatient) but only 8.68% of the variation in ITI(patient), as reported in Table IA.8. This is consistent with the idea that patient (strategic) trading is less dependent on market conditions.

In Table 6, we contrast the relation of ITI(impatient) and ITI(patient) with different variables and informational events. In Columns (1) and (2), we regress turnover and realized volatility on ITI(impatient) and ITI(patient). Whereas ITI(impatient) and ITI(impatient) are both positively correlated with turnover, the relation for ITI(impatient) is stronger. Further, ITI(impatient) is positively associated with realized volatility, whereas ITI(patient) displays the opposite pattern. These findings are consistent with ITI(impatient) detecting days with aggressive informed trading. In line with the theory of Caldentey and Stacchetti (2010), an impatient informed trader trades more aggressively and thus generate increased volatility relative to a patient informed trader.

[Insert Table 6 about here.]

We also conduct ML interpretability procedures that generally confirm the above results.[18] First, XGB ranks volume and volatility as the top two features for both ITI(impatient) and ITI(patient). Absolute order imbalance ranks third for ITI(impatient), while number of non-trading half-hours ranks third for ITI(patient), which is consistent with impatient investors being more likely to take liquidity. Second, the surrogate trees that approximate ITI(impatient) and ITI(patient) have a lot in common and resemble the ITI(13D) tree. However, the second level of the tree uses illiquidity for ITI(impatient) whereas it uses volume for ITI(patient).

[17] We use the last 20 days instead of the last 10 days to make sure that we have enough data points to train the measure.

[18] Features importance according to XGB's internal ranking and SHAP values for ITI(impatient) and ITI(patient) are reported in Figures IA.3 and IA.4 in the Internet Appendix. Surrogate trees for ITI(impatient) and ITI(patient) are reported in Figure IA.6.

Next, we consider how ITI(impatient) and ITI(patient) change ahead of earnings announcements. Columns (3) and (4) in Table 6 report the results. In the two days before an earnings announcement, ITI(impatient) increases strongly whereas ITI(patient) does not. Furthermore, Columns (5) and (6) show that ITI(impatient) is significantly higher one day before unscheduled news events but no similar increase is observed for ITI(patient). These results support our methodology as ITI(patient) and ITI(impatient) behave in ways that are consistent with economic intuition.

We conclude by studying whether ITI measures detect *illegal* insider trades. We use a dataset of 417 illegal insider trades and follow closely the specification in Ahern (2020).[19] Column (7) shows that ITI(impatient) is significantly associated with illegal insider trading. The next section further illustrates the power of ITI(impatient) in detecting other classes of informed trading.

Overall, the above results indicate that the distinction between patient trading and impatient trading is important in the data, in line with theory. We argue that our data-driven approach allows us to effectively measure the distinction between patient trading and impatient trading.

4.2 Commonality in Informed Trading

Does ITI trained on a class of informed trading detect other classes of informed trading? In this section, we answer this question to assess the external validity of our measure.

We consider two other types of informed trades: opportunistic insider trades and short sales. Prior work suggests that on average these trades are informed. Cohen et al. (2012) find that nonroutine, or opportunistic, insider trades are informed in the sense that they predict future returns. We follow Cohen et al. (2012) and use a sample of nonroutine insider trades, which spans 1993 to 2012, to build ITI(insider). For each insider trade, we pick the day before and the day after the trade as comparisons. This short window is motivated by the idea that broad market conditions change relatively little over three days. The final insider sample includes 779,007 stock-day observations.[20] Short sellers specialize in trading on negative information and tend to be informed. For example, Boehmer et al. (2008) show that highly shorted stocks underperform less

[19]Part of this dataset is obtained from Kenneth Ahern's website and is described in Ahern (2020). We thank him for making it available. We also thank Patrick Augustin for sharing his insider trading data.

[20]Because insider trades can occur on successive days, insider trades account for more (i.e., 44%) than a third of stock-days in this sample (Table 1).

shorted stocks by 1.16% on average over the next month. We focus on days with a large increase in short interest to capture when short sellers establish their positions. Short interest equals the total quantity on loan divided by the number of shares outstanding. The daily short-sale data are from Markit and cover July 2006 to July 2019. To identify days with substantial short selling, we create an indicator that is set to one if the daily change in short interest exceeds the 90th percentile for the full sample. We randomly select 100,000 stock-days with short interest spikes (indicator equals to one). Like for the insider sample, we add to the dataset the day before and the day after the spike as comparisons.

Though opportunistic insider trades and spikes in short interest are considered to be driven by informed trading, these measures do not perfectly capture informed trading. For example, some short selling is motivated by hedging motives. Similarly, informed trading can occur on days around spikes in short interest. Our methodology assumes that realized informed trading is *on average* higher on days with insider trades and days with spikes in short interest than on adjacent days.

In our main specification, we regress indicators for insider trades and spikes in short selling on ITI(13D) and stock fixed effects. The results are reported in Table 7. Columns (1) and (6) show that ITI(13D) detects opportunistic insider trades and spikes in short selling.[21] This result provides an external validation test of the measure.[22] To give some additional perspective, the relation between liquidity variables and trading indicators is not always consistent across indicators. As shown in Table IA.13 in the Internet Appendix, both effective spread and realized volatility are positively related to insider trading but negatively related to Schedule 13D trading. Hence, ITI captures a commonality across trading indicators that is not obvious from the relation between trading indicators and liquidity variables.

[Insert Table 7 about here.]

Fluctuations in the trading environment are likely to generate commonality in informed trading. Kadan and Manela (2020) show that the value of information to a trader is a function of volatility and liquidity. Hence, informed investors can trade in a common way simply due to fluctuations

[21] In Table IA.11 in the Internet Appendix, we show that ITI(13D) is able to separately detect insider purchases and insider sales.

[22] In Table IA.12 in the Internet Appendix, we show that ITI(13D) detects Schedule 13D trades and insider trades in the full stock-day sample, which is much noisier than the subsamples considered in Tables 2 and 7. We thank an anonymous referee for suggesting this test.

in liquidity. To assess the importance of this channel, we add liquidity controls to the regression. Columns (2) and (7) show that these liquidity controls only explain about one-quarter to one-third of ITI(13D)'s ability to detect insider and short trades. For example, the estimated ITI coefficient when detecting insider trades declines from 0.150 without controls to 0.116 with controls.

ITI(13D)'s ability to detect insider and short trades beyond control variables could stem from the fact that ITI(13D) accounts for nonlinearities in the relation between informed trading and liquidity variables, which are not picked up by the linear specification. It is also possible that ITI(13D) detects insider trades because the way activists trade contains incremental information that can help detect other types of informed trading. To test these explanations, we repeat the procedure described in Section 3 to obtain two additional ITI measures: ITI(insider) and ITI(short). Tables IA.14 and IA.15 in the Appendix show that the methodology works on these other datasets: ITI(insider) (ITI(short)) strongly detects insider trades (spikes in short selling) out of sample even when controlling for standard liquidity variables.[23] We note, however, that R^2 are lower in these alternative datasets than in the Schedule 13D dataset. This is not surprising. We expect the Schedule 13D data to have a higher signal-to-noise ratio since activist investors typically trade much larger quantities than insiders.

Columns (3) reports the results of regressing the insider trade indicator on ITI(13D), liquidity controls, and ITI(insider). Even when controlling for ITI(insider), ITI(13D) strongly detects insider trades with an estimated coefficient of 0.073 and a t-statistic of 19.47. A 10 percentage points increase in ITI(13D) thus leads to 0.7 percentage points increase in the likelihood of opportunistic insider trading. Columns (8) shows a similar result in the short selling regression: even when controlling for ITI(short), ITI(13D) is a strong detector of spikes in short selling. Therefore, ITI(13D) contains information that can help detect insider and short trades beyond what can be achieved with ITI measures estimated on these trades. This finding implies that incremental information can be gained about one type of informed trading from studying other types of informed trading.

ITI(patient) and ITI(impatient) also detect opportunistic insider trades and spikes in short selling (Columns (4-5) and (9-10)) even with the full set of controls. In all cases, the coefficient

[23] Table 1 reports summary statistics for the informed trading measures. The average level of a given ITI measure is close to the average level of informed trading in the corresponding sample. Hence, one should be careful not to overinterpret the *level* of the measure for a given stock when it is averaged over a long period of time.

of ITI(impatient) is more than five times as large as the coefficient of ITI(patient). Since these variables have about the same standard deviation (Table 1), opportunistic insider trading and spikes in short selling seem better characterized as impatient. When we split insider trades into purchases and sales, we find that both ITI(patient) and ITI(impatient) detect insider purchases, but only ITI(impatient) detects insider sales, as shown in Table IA.11.

It is not obvious how strong commonalities in informed trading are *ex ante*. Our data-driven approach allows us to provide a first pass at this question. Table IA.16 in the Appendix reports estimates of regressions of each ITI measure on the other two ITI measures, control variables, and stock fixed effects. ITI measures remain correlated even when we control for various liquidity measures and shut down cross-stock variation with stock fixed effects. For example, the coefficient on ITI(short) is 0.608, indicating a high partial correlation between ITI(13D) and ITI(short). Moreover, explanatory power increases by more than 50% when ITI(insider) and ITI(short) are included in the regression relative to the specification with controls only. At the same time, the highest R^2 that we achieve is only about 31%, which indicates a sizable fraction of unexplained variation in each specific ITI measure.[24]

In summary, ITI(13D) detects other types of informed trading, which does not appear to be explained by fluctuations in the trading environment. The results indicate a common pattern in the trading of three different groups of informed investors.

4.3 Models of Informed Trading

This section explains how our data-driven approach informs theories of informed trading and compares ITI to existing theory-based informed trading measures.

The groupings analysis in Table 3 shows that volume-related variables are the most important group of variables to understand fluctuations in informed trading intensity. To gain more intuition, we specifically consider turnover and absolute order imbalance. Table IA.17 in the Internet Appendix reports regressions of ITI on turnover and absolute order imbalance. There are two key takeaways. First, both variables are strongly positively associated with ITI. Second, these variables

[24] The univariate correlation between ITI(13D) and ITI(short) is 0.35. The correlations between ITI(13D) and ITI(insider) as well as ITI(insider) and ITI(short) are 0.16 and 0.15, respectively. The highest univariate correlation between ITI(13D) and any of the other liquidity variables is 0.26 with turnover. Similarly, the two most highly correlated variables with ITI(insider) are in fact ITI(13D) and ITI(short), in front of turnover, which has a correlation of 0.10 with ITI(insider).

do not subsume each other.

In the multi-period strategic trade model of Kyle (1985), volume in a given period is mostly driven by noise trading. Hence, a measure of informed trading intensity—informed volume over total volume—is negatively associated with volume. In the multi-period sequential trade model of Easley and O'Hara (1992), volume is strongly positively associated with informed trading. Conditional on an informational event occurring, informed traders always trade whereas uninformed traders sometimes do not trade. The data, however, indicate that informed trading is related to volume not only because informed traders drive volume but also because they react to uninformed volume. Table IA.18 in the Internet Appendix shows that a 1% increase in non-13D volume is associated with increased informed trading on both extensive and intensive margins. For example, a 1% increase in non-13D volume is associated with an increase of 0.54% in 13D turnover. Hence, the endogeneity of informed volume relative to uninformed volume is an important point for models to consider.

To make the above settings more realistic, we consider simulations of informed trading models where we assume that one realization of a model occurs over the course of a day. Model values can then be aggregated to the daily level to match our regressions. This procedure allows us to vary model parameters across days to better understand what drives informed trading. We present a detailed description of the simulations in Section IA.B in Internet Appendix. In a nutshell, we use simulated data to estimate regressions of informed trading intensity on daily volume, absolute daily order imbalance, and absolute daily return. Table 8 reports the results for various models of informed trading. For each model, we employ a range of different calibrations. The label < 0 (> 0) indicates that the median estimated coefficient across calibrations is lower (greater) than zero and statistically significant at the level of 1%.

[Insert Table 8 about here.]

In both sequential trade and strategic trade models, daily absolute order imbalance is strongly related to informed trading intensity. Variation in the difference between the initial price (p_0) and the fundamental value ($\tilde{v}$) across days drives this relation. When $|\tilde{v} - p_0|$ is high, informed investors trade actively to benefit from their information, which leads to $|OI| \propto |\tilde{v} - p_0|$. Informed trading drives absolute imbalance since noise trading tends to average out in these models. This strong relation implies that regressing informed trading intensity on volume and absolute order imbalance results in large R^2; e.g., in the 70% to 97% range for the PIN-type models that we consider.

Controlling for the absolute daily return strongly weakens the role of absolute order imbalance. The reason is that the daily return is close to $\tilde{v} - p_0$ as most private information is incorporated into the price by the end of the day. This conclusion holds in all extensions of the sequential and strategic trade models that we consider, such as stochastic noise trading volatility in the strategic trade model (Collin-Dufresne and Fos (2016)), and is inconsistent with our empirical results since ITI has a much stronger association with absolute order imbalance than with any volatility measure. Table IA.17 shows that the absolute daily return is negatively related to ITI and barely affects the relation between absolute order imbalance and ITI. Intuitively, order imbalance is unlikely to be solely driven by informed trading since this would imply that informed traders are always on the aggressive side of a trade. In fact, Schedule 13D filers' order imbalance is negatively associated with Lee and Ready (1991) order imbalance.

To summarize, the data suggest that a fruitful avenue is to develop models where volume and absolute order imbalance are both positively related to realized informed trading but where volatility is negatively related to it. This dimension of informed trading appears challenging for extant models to capture, as shown in Table 8 in the Internet Appendix. Across all models and calibrations that we consider, none is able to jointly match these three relations.

In our main specification, we control for several measures that are likely associated with informed trading (e.g., spread, Kyle's lambda). Next, we compare the performance of ITI to that of several existing measures of informed trading. We consider the conditional probability of informed trading obtained from the following models: the PIN model (PIN) of Easley et al. (1996); the adjusted PIN model (APIN) of Duarte and Young (2009); the generalized PIN model (GPIN) of Duarte et al. (2020); the Odders-White and Ready (2008) model (OWR); and the Back et al. (2018) model (BCL). For simplicity, we refer to these measures as PIN measures below. We perform this analysis in this subsection rather than across all specifications because PIN measures are not available in the full sample. Specifically, the availability of PIN measures restricts our sample to NYSE-listed stocks from 1994 to 2012.[25] Each of these models is estimated for every stock-year. Then, for

[25]These measures are obtained from Edwin Hu's website and described in Duarte et al. (2020). We thank him for making the data available.

each stock-day, the conditional probability of informed trading is computed as the probability of an informational event given the estimated structural parameters and specific variables for the stock-day (e.g., the number of buy trades and the number of sell trades in the case of PIN). We follow Duarte et al. (2020) and add an additional control variable that equals one for days with above-average number of trades and zero otherwise.

Table 9 reports how the various measures detect patient Schedule 13D trades, impatient Schedule 13D trades, insider trades, and spikes in short selling. As before, we control for ITI(insider) in the insider sample and for ITI(short) in the short sale sample. ITI(impatient) consistently detects informed trading across all specifications even when controlling for PIN measures. It is the only measure to achieve a significance level of 1% in all four samples. In contrast, ITI(patient) is in general insignificant.[26] Perhaps, insider and short trades tend to be impatient.

[Insert Table 9 about here.]

Some of the PIN conditional probabilities also detect informed trades in Table 9. Among PIN measures, APIN performs best in that it detects impatient trades, insider trades, and spikes in short selling at a significance level of 5%. This is encouraging since it suggests that different approaches provide complementary information to detect informed trading. The results indicate, however, substantial differences across PIN models in detecting informed trading. For completeness, we also report the regression results with one PIN measure at a time in Tables IA.19 and IA.20 in the Internet Appendix since PIN measures are correlated with each other. Again, APIN stands out, followed by BCL and PIN.

To conclude, we note two key differences between ITI and PIN measures. First, by construction, PIN measures are re-estimated for each stock-year whereas ITI is estimated on a single sample of informed trades and then extrapolated to the full sample with parameters that are constant across all stocks and years. Second, PIN's calculation is computationally challenging for actively-traded securities (e.g., Griffin et al., 2021), whereas ITI is easy to calculate for the full cross-section of stocks once model parameters are estimated on the desired set of informed trades.

[26] One may wonder why ITI(impatient) appears to be a stronger detector of patient trades than ITI(patient). This is due to the inclusion of filing fixed effects in the regression. Filing/stock fixed effects do not materially affect other results in Table 9 and the paper.

5 Application: Informed Trading and Asset Prices

ITI measures can be useful in many settings. In this section, we provide an asset pricing application by examining the relation between ITI and future stock returns. We first distinguish predictions that concern expected informed trading from predictions that concern realized informed trading.

Several theories imply conflicting predictions about how *expected* intensity of informed trading affects expected returns. Easley and O'Hara (2004) study how the total amount of information in the market affects asset prices. They show that if a higher fraction of information is private, the risks for uninformed investors increase, and thus expected returns *increase*. In contrast, Hughes et al. (2007), Lambert et al. (2007) argue that the effect of asymmetric information on returns is diversifiable, and thus is *not priced*, in a competitive market. Empirically, Duarte and Young (2009) show that PIN reflects not only asymmetric information but also illiquidity and only the component of PIN related to illiquidity is priced. Finally, if informed trading reduces information uncertainty, more efficient prices should lead to higher stock valuations and *lower* future returns. E.g., Roll et al. (2010) argue that option trading makes the underlying stock more informationally efficient and show that higher option volume predicts lower future stock returns. In these theories, prices are set conditional on expected informed trading intensity, and expected returns are not affected if investors are risk neutral. ITI can help test these theories' conflicting return predictions assuming that ITI averaged over a period of time proxies for expected informed trading.

Deviations of *realized* informed trading intensity from its expected level can also affect expected returns. Intuitively, stock prices increase after informed buys and decrease after informed sells. If buy and sell trades are equally likely to be informed, then unsigned trading intensity (such as ITI) should not predict average returns. However, several studies suggest that stock purchases are more informed than stock sales (Kraus and Stoll (1972), Chan and Lakonishok (1993), and Campbell et al. (2009)). Also, the long-only focus of most institutional investors and short-selling costs limit these investors' ability to exploit negative information. If informed purchases outnumber informed sales, high *realized* informed trading intensity leads to higher average future returns.[27]

Motivated by these hypotheses, we estimate a (stock-by-week) panel regression of next-month

[27] To see why, consider an extreme case with informed buying but no informed selling. A market maker sets the price knowing this difference in expected informed intensities of buys and sells. If no informed buying occurs, realized informed intensity is below the expected level, the price will subsequently decrease. Similarly, if most buys turn out informed, the price will increase.

stock returns on ITI measures averaged over a prior week, standard return predictors, and week fixed effects to account for stock return commonality. Stock returns are computed from CRSP daily returns and adjusted for delistings as in Shumway (1997). We skip a day between predictors and returns to avoid confounding effects as today's closing price is an input to tomorrow's return. Other predictors include idiosyncratic volatility (computed from abnormal daily returns from the Fama-FrenchCarhart four-factor model over the prior month), momentum (stock returns from six months to one month prior to the date), monthly reversal (previous month return), log market capitalization, CAPM beta, Amihud (2002) illiquidity measure, Kyle's lambda, and the effective bid-ask spread. Predictors are winsorized at the 0.05% and 99.95% levels to avoid outliers. We rely on a stock-by-week panel, which results in overlapping monthly return observations, to avoid turn-of-the-month effects documented by Etula et al. (2020). To account for overlapping returns and cross-stock dependencies, standard errors are clustered by stock and week.[28] Table IA.21 reports summary statistics for the panel used in this section.

We find several results. Higher realized informed trading over the prior week, as measured by higher ITI(13D), is associated with higher realized returns over the next month. Column (1) in Table 10 reports that ITI(13D) has a t-statistic of 6.82 for predicting monthly returns. A one standard deviation increase in ITI(13D) raises next-month return by 14 bps on average. ITI's ability to predict returns is not affected by controlling for standard return predictors.

[Insert Table 10 about here.]

After studying ITI(13D), we proceed to study other ITI measures. As shown in Section 4.2, ITI measures have a common component. To the extent that this common component identifies general trading patterns by informed investors, we expect other ITI measures to also positively predict returns in individual regressions. In untabulated results that we confirm with portfolio sorts below, all ITI measures positively and significantly predict stock returns except for ITI(short), which is not a significant predictor. Moreover, Column (2) in Table 10 shows that when we include ITI(13D), ITI(insider) and ITI(short) in the same regression, ITI(13D) and ITI(insider) are strong predictors

[28]The main results remain robust if a Fama-MacBeth regression is estimated instead of a panel regression, if non-overlapping monthly returns are used, or if alternative specifications are used; for example, if monthly return, order imbalance, or turnover are included as controls. We also find similar results for weekly returns and report them in Internet Appendix Table IA.22.

of future returns with similar t-statistics, and ITI(short) is negatively but insignificantly related to future returns. Hence, ITI measures are ordered in line with the directional information that they are estimated on. Schedule 13D filers trade on positive private information, corporate insiders' buys are on average more informed than their sells (e.g., Jeng et al. (2003)), while short-sellers trade on negative information.

The result that ITI is positively associated with future returns is consistent the buy-sell asymmetry as well as some theories of informed trading risk, such as Easley and O'Hara (2004). However, the result that ITI(short) is a weaker (and sometimes even negative) predictor of returns compared to other ITI measures favors the buy-sell asymmetry explanation. ITI(short) reflects general informed trading that positively predicts returns but also captures unique features of short selling that are likely to negatively predict returns. Combined, these two effects tend to cancel out. In contrast, expected informed trading risk theories do not distinguish between informed buying and selling and focus on the combined amount of informed trading.

We perform another test that further explores the asymmetry between the information content of buy and sell trades. Specifically, we train ITI(insider) on insider purchases and sales separately and then test the prediction of the buy-sell asymmetry that ITI(insider buy) should predict returns more positively than ITI(insider sell). Panel B of Table 10 confirms this prediction. In univariate regressions, ITI(insider buy) has a t-statistics of 2.9 versus -1.6 for ITI(insider sell). In a joint regression, ITI(insider buy) continues to strongly predict returns while ITI(insider sells) is not significant. The difference between the two is economically and statistically significant (p-value of 2% for the zero coefficient difference hypothesis). This test further supports the buy-sell asymmetry explanation.

Next, we study ITI(patient) and ITI(impatient). Intuitively, informed investors trade more aggressively and impatiently if they are more certain about their private signal. Consistent with this hypothesis, Column (3) shows that when the regression includes ITI(patient) and ITI(impatient), ITI(impatient) is a strong predictor of future returns, whereas ITI(patient) is not significantly related to future returns. This result is consistent with results in Section 4.1, where ITI(impatient) performs better than ITI(patient) in various validation tests.

In Column (4), we include all five ITI measures in the regression. ITI(13D), ITI(insider), and ITI(impatient) are strong positive predictors of future returns, ITI(short) is a negative predictor of future returns, and ITI(patient) is not significant. The results remain robust once we add control variables to this regression in Column (5), except that the coefficient on ITI(short) becomes statistically insignificant.

Controlling for PIN does not affect the results. We use annual PIN estimates over 1993 to 2010 from Brown and Hillegeist (2007). In a univariate regression, PIN positively predicts returns with a t-statistic of 2.13. However, Column (6) of Table 10 shows that the coefficient for PIN becomes insignificant once we control for other variables. These results are consistent with Duarte and Young (2009), who show in their Table 10 that PIN is a positive but not significant return predictor. In contrast, ITI measures continue to predict returns in the PIN sample.

Portfolio sorts complement panel regressions. Table 11 reports returns (Panel (a)) and alphas from the Fama-French four-factor (FF4) model that includes the momentum factor (Panel (b)) for equally-weighted decile portfolios sorted on ITI measures averaged over the prior week. Portfolio sorts confirm the results from panel regressions. When we sort on ITI(13D), the FF4 alpha for the difference between the top and bottom decile portfolios is 0.52% per month, or 6.42% annualized, with a t-statistic of 6.2. Alpha increases gradually from -3 bps for the bottom portfolio to 49 bps for the high portfolio. Consistent with the regression results, ITI(impatient) predict returns more strongly than ITI(patient), with the FF4 monthly alphas for the ten-minus-one portfolio of 0.54% and 0.34%, respectively. ITI (insider) has a FF4 alpha of 0.22%, or 2.67% annualized, lower than for ITI(13D), while ITI(short) yields an insignificant 0.05% alpha. finally, the results are qualitatively similar for value-weighted portfolio sorts and alternative subsamples.[29]

[Insert Table 11 about here.]

The horizon of return predictability can further help us distinguish between the buy-sell asymmetry and informed trading risk explanations. To the extent that stock-specific information risk is persistent, the informed trading risk explanation implies that ITIs should predict not only short-

[29] First, in untabulated results, ITI(13D) generates a value-weighted monthly alpha of 0.25% with a t-statistic of 2.2. Second, Table IA.24 in the Appendix shows that ITI(13D) robustly predicts returns across various sub-samples. For each splitting variable, we split the full sample into two equal parts based on the median value of the splitting variable. We then sort stocks into decile portfolios based on ITI(13D) within each part and compute monthly alphas. We consider the following splitting variables: market capitalization, stock turnover, idiosyncratic volatility, effective spread, Kyle's lambda, and PIN. ITI(13D) remains a positive and significant predictor of returns in all sub-samples with monthly alphas for the difference between the top and bottom decile portfolios ranging from 0.31% to 0.85%. As expected, the predictability is stronger for stocks for which informed trading is likely to be more important, such as smaller stocks, stocks with higher trading costs, and stocks with higher volatility.

term returns but also long-term returns. In contrast, according to the buy-sell asymmetry, stock prices underreact to realized informed trading, which suggests short-term return predictability. We repeat portfolio sort analysis but instead of the next month return, we use returns from the second next month. Table IA.23 shows that none of ITI measures predicts returns beyond the next month. This lack of long-term predictability is not consistent with the informed trading risk channel. This conclusion must be interpreted with caution, however, because it assumes that ITI, which reflects current realized informed trading, correlates with expected informed trading in the future.

Overall, ITI measures are positively associated with next month returns, and this predictability is most consistent with a buy-sell asymmetry explanation.

6 Conclusion

In this paper, we develop a new measure of informed trading by directly learning from informed trading data. We use a ML algorithm to identify days with informed trading. In this standard classification problem, a daily indicator for informed trading is predicted by a set of same-day variables related to volume, volatility, and liquidity. After the model is estimated on the training data of observed informed trades, we extrapolate it to the entire stock-day universe, where informed trading is not directly observed. This procedure produces a new measure of informed trading, the Informed Trading Intensity ("ITI").

We show that ITI significantly predicts informed trading out-of-sample and is a significant predictor of various informational events. In particular, ITI increases before earnings announcements, M&A announcements, and unscheduled news releases. Moreover, returns on days with high ITI exhibit less reversal than returns on other days, in line with the intuition that the price impact of informed trades is permanent. All of these results validate our use of ITI as a measure of informed trading intensity.

We show that our data-driven approach can shed light on the economics of informed trading. First, we show a strong distinction between impatient trading and patient trading, consistent with theory. Second, we show that incremental information can be gained about one type of informed trading from studying other types of informed trading. Third, our methodology highlights specific features of informed trading that existing models struggle to capture. Indeed, ITI is not subsumed by existing theory-based measures. These stylized facts provide a fruitful avenue for future research.

ITI can be applied in many settings. We provide an application to the asset pricing literature and ask whether informed trading is priced in the cross-section of stock returns, as prior work documents conflicting results. We show that an increase in ITI is associated with higher future monthly returns in the cross-section, but this predictability is most consistent with an asymmetry in the informational content of purchases and sales.

The main implication of this paper is that a data-driven ML approach combined with data on informed trading can generate an effective measure of informed trading and improve our understanding of the economics of informed trading.

References

Ahern, K. R. (2020). Do proxies for informed trading measure informed trading? Evidence from illegal insider trades. *Review of Asset Pricing Studies 10*(3), 397–440.

Akbas, F., C. Jiang, and P. D. Koch (2020). Insider investment horizon. *The Journal of Finance 75*(3), 1579–1627.

Akey, P., V. Gregoire, and C. Martineau (2022). Price revelation from insider trading: Evidence from hacked earnings news. *Journal of Financial Economics 143*(3), 1162–1184.

Alldredge, D. M. and D. C. Cicero (2015). Attentive insider trading. *Journal of Financial Economics 115*(1), 84–101.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets 5*(1), 31–56.

Augustin, P., M. Brenner, and M. G. Subrahmanyam (2019). Informed options trading prior to takeover announcements: Insider trading? *Management Science 65*(12), 5697–5720.

Augustin, P. and M. G. Subrahmanyam (2020). Informed options trading before corporate events. *Annual Review of Financial Economics 12*, 327–355.

Back, K., P. Collin-Dufresne, V. Fos, T. Li, and A. Ljungqvist (2018). Activism, strategic trading, and liquidity. *Econometrica 86*, 1431–1643.

Back, K., K. Crotty, and T. Li (2018). Identifying information asymmetry in securities markets. *Review of Financial Studies 31*(6), 2277–2325.

Biggerstaff, L., D. Cicero, and M. B. Wintoki (2020). Insider trading patterns. *Journal of Corporate Finance 64*, 101654.

Boehmer, E., C. M. Jones, and X. Zhang (2008). Which shorts are informed? *Journal of Finance 63*(2), 491–527.

Bogousslavsky, V. (2021). The cross-section of intraday and overnight returns. *Journal of Financial Economics 141*(1), 172–194.

Bolandnazar, M., R. J. Jackson Jr, W. Jiang, and J. Mitts (2020). Trading against the random expiration of private information: A natural experiment. *Journal of Finance 75*(1), 5–44.

Boudoukh, J., R. Feldman, S. Kogan, and M. Richardson (2019). Information, trading, and volatility: Evidence from firm-specific news. *Review of Financial Studies 32*(3), 992–1033.

Brav, A., W. Jiang, F. Partnoy, and R. Thomas (2008). Hedge fund activism, corporate governance, and firm performance. *Journal of Finance 63*(4), 1729–1775.

Breiman, L. (2001). Random forests. *Machine learning 45*(1), 5–32.

Brennan, M. J., S.-W. Huh, and A. Subrahmanyam (2018). High-frequency measures of informed trading and corporate announcements. *Review of Financial Studies 31*(6), 2326–2376.

Brown, S. and S. A. Hillegeist (2007). How disclosure quality affects the level of information asymmetry. *Review of accounting studies 12*(2), 443–477.

Caldentey, R. and E. Stacchetti (2010). Insider trading with a random deadline. *Econometrica 78*(1), 245–283.

Campbell, J. Y., S. J. Grossman, and J. Wang (1993). Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics 108*(4), 905–939.

Campbell, J. Y., T. Ramadorai, and A. Schwartz (2009). Caught on tape: Institutional trading, stock returns, and earnings announcements. *Journal of Financial Economics 92*(1), 66–91.

Chan, L. K. and J. Lakonishok (1993). Institutional trades and intraday stock price behavior. *Journal of Financial Economics 33*(2), 173–199.

Chen, T. and C. Guestrin (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Cohen, L., C. Malloy, and L. Pomorski (2012). Decoding inside information. *Journal of Finance 67*(3), 1009–1043.

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.

Collin-Dufresne, P., V. Fos, and D. Muravyev (2020). Informed trading in the stock market and option-price discovery. *Journal of Financial and Quantitative Analysis*, 1–40.

Cookson, A., V. Fos, and M. Niessner (2021, January). Does disagreement facilitate informed trading? Evidence from activist investors. working paper.

Craven, M. and J. Shavlik (1995). Extracting tree-structured representations of trained networks. *Advances in neural information processing systems 8*, 24–30.

Cziraki, P. and J. Gider (2021). The dollar profits to insider trading. *Review of Finance 25*(5), 1547–1580.

Duarte, J., E. Hu, and L. Young (2020). A comparison of some structural models of private information arrival. *Journal of Financial Economics 135*(3), 795–815.

Duarte, J. and L. Young (2009). Why is PIN priced? *Journal of Financial Economics 91*(2), 119–138.

Easley, D., M. L. de Prado, M. O'Hara, and Z. Zhang (2021). Microstructure in the machine age. *Review of Financial Studies*.

Easley, D., S. Hvidkjaer, and M. O'Hara (2002). Is information risk a determinant of asset returns? *Journal of Finance 57*(5), 2185–2221.

Easley, D., N. M. Kiefer, M. O'Hara, and J. B. Paperman (1996). Liquidity, information, and infrequently traded stocks. *Journal of Finance 51*(4), 1405–1436.

Easley, D. and M. O'Hara (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics 19*(1), 69–90.

Easley, D. and M. O'Hara (1992). Time and the Process of Security Price Adjustment. *Journal of Finance 47*(2), 577–605.

Easley, D. and M. O'Hara (2004). Information and the cost of capital. *Journal of Finance 59*(4), 1553–1583.

Etula, E., K. Rinne, M. Suominen, and L. Vaittinen (2020). Dash for cash: Monthly market impact of institutional liquidity needs. *Review of Financial Studies 33*(1), 75–111.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance 25*(2), 383–417.

Foucault, T., O. Kadan, and E. Kandel (2005). Limit Order Book as a Market for Liquidity. *Review of Financial Studies 18*(4), 1171–1217.

Gantchev, N. and C. Jotikasthira (2018). Institutional trading and hedge fund activism. *Management Science 64*(6), 2930–2950.

Glosten, L. R. and P. R. Milgrom (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics 14*(1), 71–100.

Goldstein, I., C. S. Spatt, and M. Ye (2021). Big data in finance. *The Review of Financial Studies 34*(7), 3213–3225.

Griffin, J., J. Oberoi, and S. D. Oduro (2021). Estimating the probability of informed trading: A bayesian approach. *Journal of Banking and Finance 125*, 106045.

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

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies 33*(5), 2223–2273.

Hasbrouck, J. (1988). Trades, quotes, inventories, and information. *Journal of Financial Economics 22*(2), 229–252.

Hasbrouck, J. (1991). Measuring the Information Content of Stock Trades. *Journal of Finance 46*(1), 179–207.

Hastie, T., R. Tibshirani, and J. Friedman (2017). *The elements of statistical learning: data mining, inference, and prediction.* Springer Science & Business Media.

Holden, C. W. and S. Jacobsen (2014). Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *Journal of Finance 69*(4), 1747–1785.

Holderness, C. G. and D. P. Sheehan (1985). Raiders or saviors? the evidence on six controversial investors. *Journal of Financial Economics 14*(4), 555–579.

Hughes, J. S., J. Liu, and J. Liu (2007). Information asymmetry, diversification, and cost of capital. *The Accounting Review 82*(3), 705–729.

Jeng, L. A., A. Metrick, and R. Zeckhauser (2003). Estimating the returns to insider trading: A performance-evaluation perspective. *Review of Economics and Statistics 85*(2), 453–471.

Kacperczyk, M. and E. Pagnotta (2019). Chasing private information. *Review of Financial Studies 32*(12), 4997–5047.

Kadan, O. and A. Manela (2020). Liquidity and the Strategic Value of Information. *SSRN Electronic Journal*.

Kaniel, R. and H. Liu (2006). So what orders do informed traders use? *Journal of Business 79*(4), 1867–1913.

Karolyi, G. A. and S. Van Nieuwerburgh (2020). New methods for the cross-section of returns. *The Review of Financial Studies 33*(5), 1879–1890.

Kelly, B. and A. Ljungqvist (2012). Testing Asymmetric-Information Asset Pricing Models. *The Review of Financial Studies 25*(5), 1366–1413.

Kim, O. and R. E. Verrecchia (1994). Market Liquidity and Volume around Earnings Announcements. *Journal of Accounting and Economics 17*(1-2), 41–67.

Klein, A. and E. Zur (2009). Entrepreneurial shareholder activism: Hedge funds and other private investors. *The Journal of Finance 64*(1), 187–229.

Kraus, A. and H. R. Stoll (1972). Price impacts of block trading on the new york stock exchange. *Journal of Finance 27*(3), 569–588.

Kwan, A., R. Philip, and A. Shkilko (2021). The conduits of price discovery: A machine learning approach. working paper.

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

Lambert, R., C. Leuz, and R. E. Verrecchia (2007). Accounting information, disclosure, and the cost of capital. *Journal of accounting research 45*(2), 385–420.

Lee, C. and M. J. Ready (1991). Inferring trade direction from intraday data. *Journal of Finance 46*(2), 733–746.

Lee, C. M. C., B. Mucklow, and M. J. Ready (1993). Spread, Depths, and the Impact of Earnings Information: An Intraday Analysis. *Review of Financial Studies 6*(2), 345–374.

Llorente, G., R. Michaely, G. Saar, and J. Wang (2002). Dynamic volume-return relation of individual stocks. *Review of Financial Studies 15*(4), 1005–1047.

Lundberg, S. M. and S.-I. Lee (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems 30.*

Nagel, S. (2012). Evaporating Liquidity. *Review of Financial Studies 25*(7), 2005–2039.

Nagel, S. (2021). *Machine Learning in Asset Pricing.* Princeton University Press.

Odders-White, E. R. and M. J. Ready (2008). The probability and magnitude of information events. *Journal of Financial Economics 87*(1), 227–248.

O'Hara, M. (2003). Presidential address: Liquidity and price discovery. *The Journal of Finance 58*(4), 1335–1354.

Reed, A. V. (2013). Short selling. *Annu. Rev. Financ. Econ. 5*(1), 245–258.

Richardson, S., P. A. Saffi, and K. Sigurdsson (2017). Deleveraging risk. *Journal of Financial and Quantitative Analysis 52*(6), 2491–2522.

Roll, R., E. Schwartz, and A. Subrahmanyam (2010). O/s: The relative trading activity in options and stock. *Journal of Financial Economics 96*(1), 1–17.

Senchack, A. J. and L. T. Starks (1993). Short-sale restrictions and market reaction to short-interest announcements. *Journal of Financial and quantitative analysis 28*(2), 177–194.

Shumway, T. (1997). The delisting bias in CRSP data. *Journal of Finance 52*(1), 327–340.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological) 58*(1), 267–288.

Yang, Y. C., B. Zhang, and C. Zhang (2020). Is information risk priced? evidence from abnormal idiosyncratic volatility. *Journal of Financial Economics 135*(2), 528–554.

Figure 1. Stability of the algorithm over time in the Schedule 13D sample. This figure reports the adjusted R^2 estimated each year from the following two specifications: An indicator for Schedule 13D trading over the filing windows (60 days before the filing date to the filing date) is regressed on either informed trading intensity trained on Schedule 13D data (solid line) or a set of control variables (dashed line). Control variables are effective spread, price impact, lambda, depth, realized volatility, turnover, order imbalance, absolute order imbalance, and return. The sample consists of 1,593 Schedule 13D filings between 1994 and 2018.

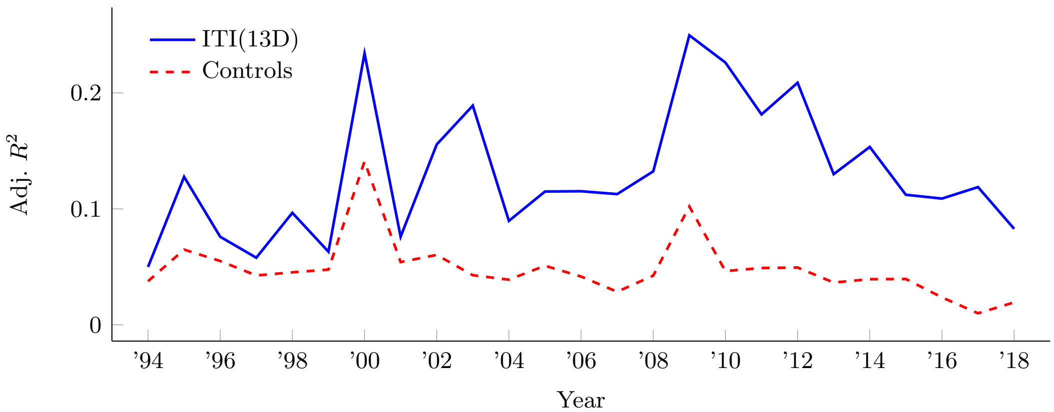


Figure 2. Partial dependence plots of ITI on volume and volatility. The top panel shows that ITI is increasing and concave in volume. The bottom panel shows that ITI is decreasing and convex in volatility. Left Y-axis reports ITI, which is shown in solid line. Right Y-axis reports the distribution of a given variable, which is shown in gray bars. The range of values in the X-axis spans the distribution of a given variable. Variables are standardized before computing ITI; hence, zero volume means that volume on this day matches the previous one-year average. Partial dependence for ITI if volume $= x$ is computed in two steps. First, volume is set to x for each observation in the Schedule 13D sample, and then ITI is computed. Second, ITI is averaged over all observations. Partial dependence for ITI if volatility $= x$ is computed in a similar way. Variables are defined in Table IA.3.

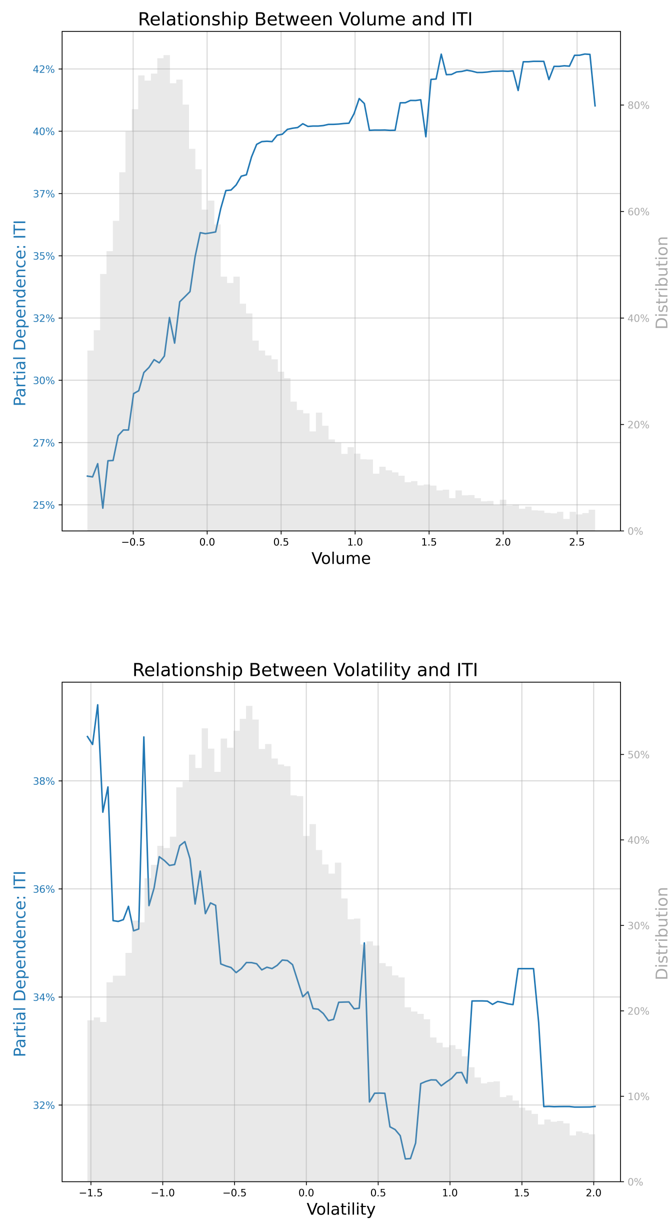


Figure 3. This figure presents a simple tree that approximates ITI. A regression tree with three levels predicts ITI for the sample of Schedule 13D trades. At each step, a tree picks a variable and a split level that minimizes mean-squared error (a greedy algorithm) while encouraging equally-sized splits. Each node reports the splitting variable and criteria followed by a measure of fit (MSE), number of observations, and the predicted value of ITI. The tree explains about one-third of the total variation in ITI. To apply the tree, we start at its top and move to the right if a splitting criterion is satisfied until we reach one of the terminal nodes. Variables are defined in Table IA.3.

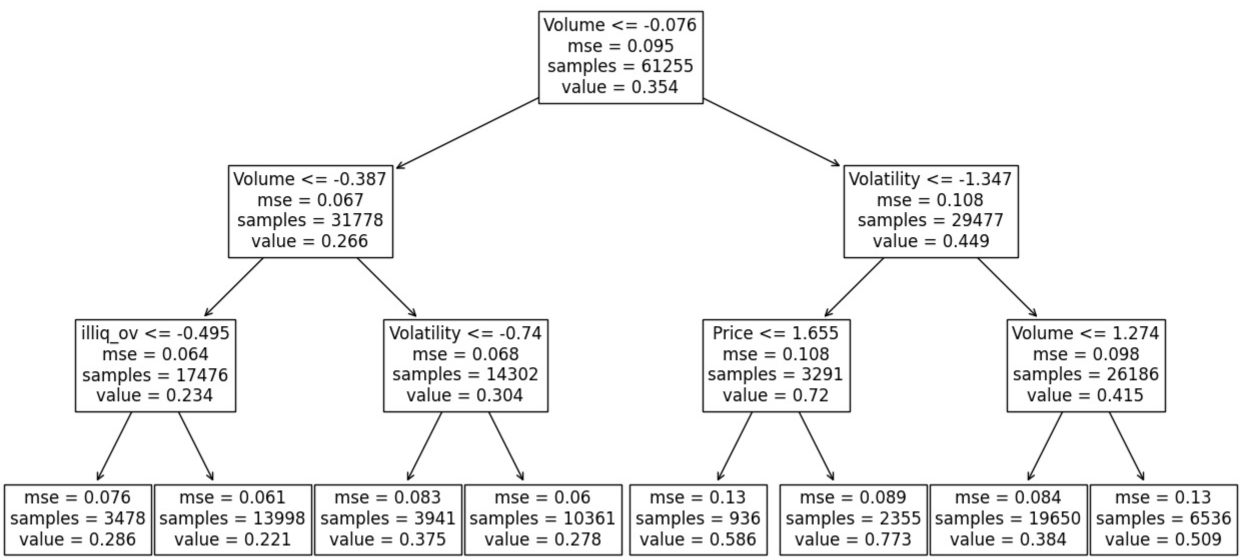
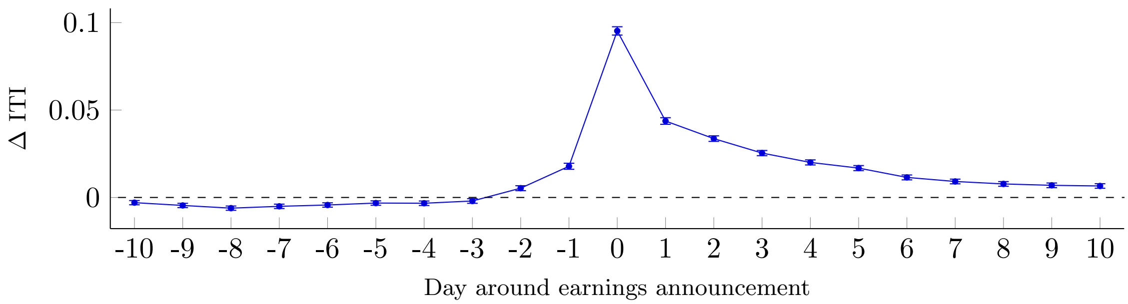
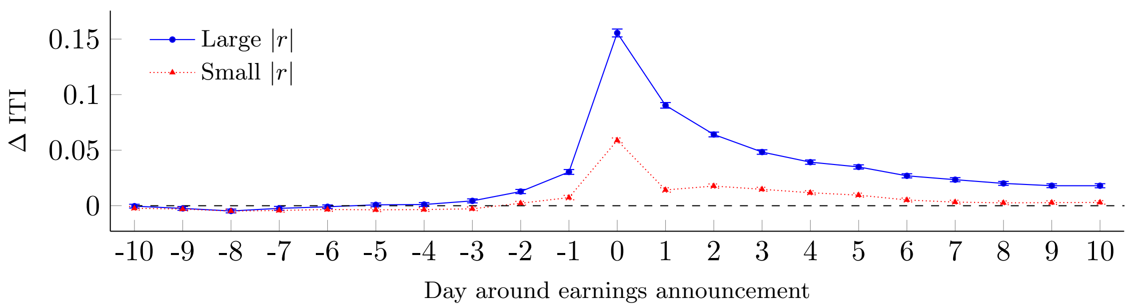


Figure 4. ITI around earnings announcements and news events. In Panel (a), informed trading intensity is regressed on indicator variables for days around earnings announcements and stock fixed effects. Panel (b) reports the results separately for announcements that are in the top and bottom quartile of absolute announcement day return. The sample includes common stocks from 1/1993 to 7/2019. In Panel (c), informed trading intensity is regressed on indicator variables for days around news events (excluding earnings announcements) and stock fixed effects. News data is obtained from Boudoukh et al. (2019) and covers S&P 500 common stocks from 2000 to 2015. The figure reports 95% confidence intervals based on standard errors that are double-clustered by stock and date.

(a) ITI around earnings announcements



(b) Top quartile and bottom quartile absolute announcement return



(c) ITI around news event (excluding earnings announcements)

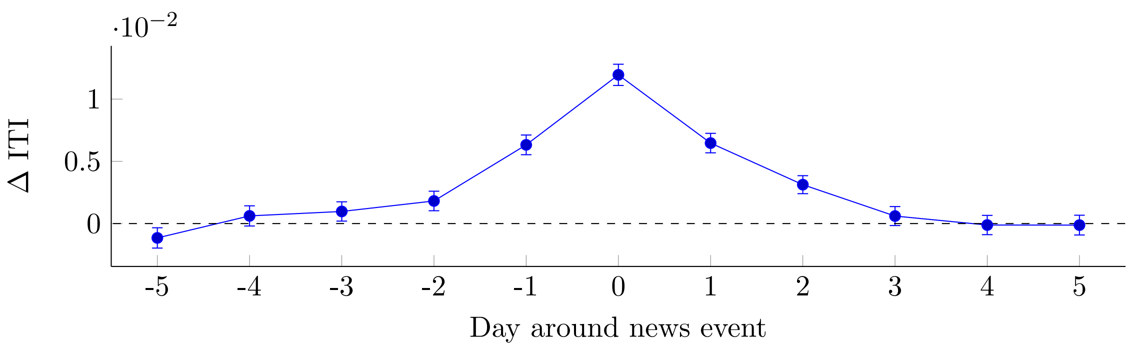


Table 1. Descriptive statistics. The table reports the mean, standard deviation (SD), within-stock standard deviation (SD_w), and 1st, 5th, 25th, 50th, 75th, 95th and 99th percentiles for the main set of variables. Control variables are described in Table IA.4 and include effective spread (ES), lambda, depth, realized volatility (rvol), turnover (turn), order imbalance (OI), absolute order imbalance (|OI|), and return (ret). These variables are winsorized at 0.05% and 99.95%. 13D trade is an indicator variable that takes the value one on days with Schedule 13D trades. Insider trade is an indicator variable that takes the value one on days with opportunistic insider trades. Δ short is an indicator variable that takes the value one on days when the daily change in short interest exceeds the 90th percentile. Short interest is defined as the total short sale demand divided by the number of shares outstanding. The full sample consists of common stocks from 1/1993 to 7/2019, excluding any stock-day with a missing value for one of the control variables and excluding any stock-day in the 13D sample. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. The 13D sample consists of the filing period for 1,593 13D filings between 1994 and 2018 (58,197 observations). For this sample, SD_w denotes the within-filing standard deviation. The insider sample consists of two days surrounding opportunistic insider trades between 1993 and 2012, which includes 95,464 days with at least one opportunistic buy trade and 260,366 days with at least one opportunistic sell trade (779,007 observations in total). The short sample consists of two days surrounding 100,000 randomly-selected spikes in short interest from 6/2006 to 12/2010 (216,979 observations).

Variable name	Mean	SD	SD_w	1%	5%	25%	50%	75%	95%	99%	N
ES	0.0042	0.0054	0.0046	0.0002	0.0004	0.0009	0.0022	0.0052	0.0144	0.0256	16,823,151
lambda	0.0036	0.0101	0.0097	-0.0159	-0.0036	0.0001	0.0014	0.0046	0.0184	0.0421	16,823,151
depth ($*10^2$)	0.0082	0.0187	0.0172	0.0002	0.0004	0.0016	0.0035	0.0079	0.0286	0.0743	16,823,151
rvol	0.0227	0.0198	0.0180	0.0026	0.0059	0.0113	0.0175	0.0275	0.0563	0.0997	16,823,151
turn	0.0088	0.0134	0.0132	0.0004	0.0008	0.0025	0.0052	0.0102	0.0273	0.0601	16,823,151
OI	0.0001	0.0026	0.0026	-0.0069	-0.0025	-0.0005	0.0000	0.0006	0.0029	0.0073	16,823,151
|OI|	0.0012	0.0024	0.0024	0.0000	0.0000	0.0002	0.0006	0.0013	0.0042	0.0103	16,823,151
ret	0.0005	0.0307	0.0279	-0.0849	-0.0432	-0.0123	0.0000	0.0126	0.0456	0.0942	16,823,151
ITI(13D)	0.2745	0.1595	0.1558	0.0313	0.0660	0.1539	0.2472	0.3696	0.5883	0.7540	16,823,151
ITI(patient)	0.2038	0.1452	0.1427	0.0156	0.0360	0.0967	0.1712	0.2802	0.5014	0.6845	16,823,151
ITI(impatient)	0.4213	0.1375	0.1322	0.1375	0.1998	0.3033	0.3953	0.4999	0.6492	0.7653	16,823,151
ITI(insider)	0.4786	0.1578	0.1569	0.1271	0.2110	0.3647	0.4732	0.5868	0.7467	0.8356	16,103,808
ITI(short)	0.4460	0.0787	0.0670	0.2433	0.3082	0.3648	0.4003	0.4447	0.5284	0.6126	16,664,285
13D trade (13D sample)	0.3603	0.4801	0.4191	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	58,197
Insider trade (insider sample)	0.4390	0.4963	0.4901	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	779,007
Δ short (short sample)	0.4323	0.4954	0.4908	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	216,979

Table 2. Does ITI detect Schedule 13D trading days? In Columns (1)-(3), an indicator for days with Schedule 13D trading over the filing windows (60 days before the filing date to the filing date) is regressed on a set of liquidity variables and filing fixed effects. In Columns (4)-(6), 13D filer turnover (share volume traded by the filer divided by total shares outstanding), is regressed on a set of liquidity variables and filing fixed effects, conditional on Schedule 13D trading on that specific stock-day. ITI is the informed trading intensity. Effective spread (ES), lambda, depth, realized volatility (rvol), order imbalance (OI), absolute order imbalance (|OI|), and return are winsorized at 0.05% and 99.95%. The sample consists of 1,593 Schedule 13D filings between 1994 and 2018. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Standard errors are clustered by filing, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

Dep. variable:	Day with Schedule 13D trading			Schedule 13D turnover		
	(1)	(2)	(3)	(4)	(5)	(6)
ITI	0.697***		0.604***	0.007***		0.004***
	(43.572)		(36.648)	(22.184)		(13.657)
ES		-3.412***	-0.807		0.099***	0.108***
		(-4.053)	(-1.098)		(3.179)	(3.505)
lambda		-1.186***	-0.589***		-0.024***	-0.020***
		(-4.822)	(-2.599)		(-5.275)	(-4.673)
depth		34.360***	21.323***		-0.096	-0.132
		(8.996)	(8.037)		(-1.135)	(-1.603)
rvol		-0.832***	-0.361***		0.000	0.001
		(-6.149)	(-3.475)		(0.111)	(0.329)
turn		1.377***	0.640***		0.039***	0.037***
		(9.845)	(6.110)		(8.284)	(7.885)
OI		0.379	0.294		0.073***	0.071***
		(0.748)	(0.703)		(3.065)	(3.092)
|OI|		6.732***	2.613***		0.180***	0.161***
		(10.534)	(4.781)		(6.329)	(5.811)
ret		-0.083	-0.030		-0.008***	-0.007***
		(-1.258)	(-0.511)		(-3.375)	(-3.064)
Filing FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.0986	0.0461	0.1073	0.0692	0.2579	0.2756
Obs.	58,197	58,197	58,197	20,969	20,969	20,969

Table 3. What variables help detect informed trading? An indicator for days with Schedule 13D trading over the filing window (60 days before the filing date to the filing date) is regressed on informed trading intensity (ITI) measured from a subset of the variables in Table IA.3 and filing fixed effects. Schedule 13D trade is an indicator variable that takes the value one on days when the filer trades. ITI(liquidity), ITI(return), ITI(volatility), and ITI(volume) are versions of ITI that are trained using a subset of the explanatory variables. The sample consists of 1,593 Schedule 13D filings between 1994 and 2018. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Standard errors are clustered by filing, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

Dep. variable:	Day with Schedule 13D trading					
	(1)	(2)	(3)	(4)	(5)	(6)
ITI	0.697***					
	(43.572)					
ITI(liquidity)		0.596***				0.289***
		(24.561)				(17.945)
ITI(return)			0.510***			0.239***
			(18.963)			(14.256)
ITI(volatility)				0.498***		0.208***
				(19.266)		(12.912)
ITI(volume)					0.671***	0.526***
					(39.235)	(36.533)
Filing FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.0986	0.0412	0.0264	0.0278	0.0769	0.1043
Obs.	58,197	58,197	58,197	58,197	58,197	58,197

Table 4. Turnover matching in the 13D sample. Within each 13D filing, each treated observation (day with Schedule 13D trade) is matched to a non-treated observation based on turnover. Only paired observations whose absolute difference in turnover is ≤ 0.00002 are kept. This threshold is selected such that the within-filing difference in turnover between stock-days with 13D trade and stock-days without 13D trade is statistically insignificant, as shown in Panel (a). In Panel (b), an indicator for days with Schedule 13D trading over the filing windows (60 days before the filing date to the filing date) is regressed on a set of liquidity variables and filing fixed effects in this restricted sample. ITI is the informed trading intensity. Effective spread (ES), lambda, depth, realized volatility (rvol), order imbalance (OI), absolute order imbalance (|OI|), and return are winsorized at 0.05% and 99.95%. The unrestricted sample consists of 1,593 Schedule 13D filings between 1994 and 2018. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Standard errors are clustered by filing, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

(a) Difference in turnover and ITI

Dep. variable:	Turnover	ITI
Day with Schedule 13D trading	0.000	0.033***
	(1.027)	(5.253)
Filing FE	Yes	Yes
Adj. R^2	-0.0004	0.0163
Obs.	2,388	2,388

(b) What variables help detect informed trading?

Dep. variable:	Day with Schedule 13D trading		
ITI	0.509***		0.507***
	(5.445)		(5.284)
ES		-2.987	-0.380
		(-0.549)	(-0.072)
lambda		-0.767	-0.505
		(-0.689)	(-0.448)
depth		23.136	24.205
		(0.549)	(0.770)
rvol		-2.099	-1.398
		(-1.211)	(-0.827)
OI		-2.444	2.237
		(-0.200)	(0.187)
|OI|		2.532	-11.174
		(0.156)	(-0.697)
ret		-0.319	-0.422
		(-0.441)	(-0.596)
Filing FE	Yes	Yes	Yes
Adj. R^2	0.0163	-0.0003	0.0149
Obs.	2,388	2,388	2,388

Table 5. Return reversal. Daily return is regressed on lagged return, ITI, control variables, date fixed effects, and interactions of lagged returns with ITI, turnover, realized volatility, and effective spread. Control variables are effective spread, price impact, lambda, depth, realized volatility, turnover, order imbalance, and absolute order imbalance. Control variables are winsorized at 0.05% and 99.95%. Standard errors are double-clustered by stock and date, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level. The sample includes common stocks from 1/1993 to 7/2019. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month.

Dep. Variable		ret(t+1)	
	(1)	(2)	(3)
ret	-0.004*	-0.014***	-0.006**
	(-1.774)	(-4.477)	(-1.977)
ITI(13D)		0.002***	0.002***
		(23.553)	(20.001)
ret*ITI(13D)		0.031***	0.034***
		(6.445)	(5.830)
ret*turn			0.008
			(0.541)
ret*rvol			-0.029*
			(-1.652)
ret*ES%			-0.756***
			(-4.936)
Controls	No	No	Yes
Date FE	Yes	Yes	Yes
Adj. R^2	0.0000	0.0002	0.0004
Obs.	16,814,168	16,814,168	16,814,168

Table 6. ITI(patient) and ITI(impatient). Different variables are regressed on ITI(patient) and ITI(impatient) and control variables. ITI(patient) (ITI(impatient)) is the informed trading intensity estimated using the first 40 days (last 20 days) of the 60-day Schedule 13D filing window. News-2 and News-1 are indicators for the two days before a news announcement, excluding earnings announcements. News data is obtained from Boudoukh et al. (2019) and covers S&P 500 common stocks from 2000 to 2015. EA-2 and EA-1 are indicators for the two days before an earnings announcement. The earnings announcement day and the 10 days after the announcement are excluded from the regression. Illegal trade is an indicator variable for days with illegal insider trades ahead of informational events. We consider a window of 120 days before the event, excluding the day before the event and the day of the event. This regression includes event fixed effects and event day fixed effects as in (Ahern, 2020). The sample includes 417 insider trades (column 5). The sample includes common stocks from 1/1993 to 7/2019. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Control variables are winsorized at 0.05% and 99.95% and include effective spread, price impact, lambda, depth, realized volatility, turnover, order imbalance, absolute order imbalance, and return. Standard errors are double-clustered by stock and date, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

	Turnover	Realized volatility	EA-2	EA-1	News-2	News-1	Illegal trade
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ITI(patient)	0.006***	-0.003***	0.001***	-0.001*	-0.004	0.002	-0.010
	(56.727)	(-11.955)	(2.581)	(-1.920)	(-0.847)	(0.463)	(-1.249)
ITI(impatient)	0.028***	0.014***	0.007***	0.021***	0.006	0.027***	0.021*
	(98.857)	(37.125)	(9.693)	(21.614)	(0.871)	(3.700)	(1.803)
Controls	No	No	Yes	Yes	Yes	Yes	Yes
Fixed effects	Stock	Stock	Stock	Stock	Stock	Stock	Event, event-day
Adj. R^2	0.1394	0.0100	0.0002	0.0020	0.0056	0.0092	0.0010
Obs.	18,062,838	18,022,699	14,159,702	14,159,702	1,652,880	1,652,880	18,695

Table 7. Do ITI measures detect various classes of informed trading? Indicator variables for days with 13D trading, days with opportunistic insider trading, or days with a spike in short interest are regressed on informed trading intensity measures and control variables. ITI(13D) is trained on Schedule 13D data. ITI(patient) (ITI(impatient)) is trained using the first 40 days (last 20 days) of the 60-day Schedule 13D filing window. ITI(insider) is trained on opportunistic insider trading data. ITI(short) is trained on short selling data. The 13D sample consists of the filing period for 1,593 13D filings between 1994 and 2018. The insider sample consists of two days surrounding opportunistic insider trades between 1993 and 2012. The short sample consists of two days surrounding 100,000 randomly-selected spikes in short interest from 6/2006 to 12/2010. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Control variables are winsorized at 0.05% and 99.95% and include effective spread, price impact, lambda, depth, realized volatility, turnover, order imbalance, absolute order imbalance, and return. Standard errors are clustered by filing for the Schedule 13D sample and by stock for the other samples, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

Dep. Variable	Insider trade					Δ Short interest				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ITI(13D)	0.150***	0.116***	0.073***			0.178***	0.118***	0.069***		
	(42.182)	(30.674)	(19.470)			(27.422)	(16.758)	(9.634)		
ITI(insider)			0.354***		0.325***					
			(47.882)		(43.406)					
ITI(short)								0.452***		0.387***
								(25.921)		(20.973)
ITI(patient)				0.032***	0.019***				0.025***	0.016**
				(7.325)	(4.503)				(3.173)	(2.001)
ITI(impatient)				0.178***	0.117***				0.211***	0.134***
				(34.951)	(23.038)				(22.194)	(13.360)
Controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.0025	0.0046	0.0082	0.0058	0.0086	0.0033	0.0077	0.0108	0.0094	0.0114
Obs.	779,007	779,007	779,007	779,007	779,007	216,979	216,979	216,979	216,979	216,979

Table 8. Statistics from simulations of informed trading models. We use simulated data to estimate regressions of informed trading intensity (ITI) on daily volume (V_t), absolute daily order imbalance ($|OI_t|$), and absolute daily return ($|r_t|$). The methodology and associated calibrations are described in Section IA.B in the Internet Appendix. The label < 0 (> 0) indicates that the median estimated coefficient across calibrations is lower (greater) than zero and statistically significant at the level of 1%. Avg R-squared is the average adjusted R^2 across simulations. T denotes the number of trading periods within a day. Kyle-tv refers to the strategic trade model where we allow noise trading volatility to vary across days. Kyle-stochastic refers to the strategic trade model with stochastic noise trading volatility.

	$\text{ITI}_t = a + b_1 V_t + b_2 \vert OI_t \vert + e_t$			$\text{ITI}_t = c + d_1 V_t + d_2 \vert OI_t \vert + d_3 \vert r_t \vert + u_t$			
	b_1	b_2	Avg R-squared	d_1	d_2	d_3	Avg R-squared
Data	> 0	> 0	10.2%	> 0	> 0	< 0	10.4%
PIN ($T = 40$)	> 0	> 0	76.5%	0	> 0	> 0	81.6%
PIN ($T = 100$)	> 0	> 0	87.8%	0	> 0	> 0	92.2%
PIN ($T = 400$)	> 0	> 0	96.6%	0	> 0	> 0	99.2%
APIN ($T = 40$)	0	> 0	71.1%	< 0	> 0	> 0	78.2%
APIN ($T = 100$)	< 0	> 0	83.6%	< 0	> 0	> 0	90.2%
APIN ($T = 400$)	< 0	> 0	93.2%	< 0	> 0	> 0	95.9%
Kyle ($T = 40$)	< 0	> 0	75.6%	-	-	-	-
Kyle ($T = 100$)	< 0	> 0	82.3%	-	-	-	-
Kyle ($T = 400$)	< 0	> 0	91.1%	-	-	-	-
Kyle-tv ($T = 40$)	< 0	> 0	57.0%	0	0	> 0	74.2%
Kyle-tv ($T = 100$)	< 0	> 0	63.6%	0	0	> 0	78.3%
Kyle-tv ($T = 400$)	< 0	> 0	67.8%	0	0	> 0	83.0%
Kyle-stochastic	0	> 0	32.5%	> 0	0	> 0	37.4%

Table 9. ITI measures and other measures of informed trading. This table compares ITI measures to the conditional probability of informed trading obtained from several models: the PIN model (PIN); the adjusted PIN model of Duarte and Young (2009) (APIN); the generalized PIN model (GPIN) of Duarte et al. (2020); the Odders-White and Ready (2008) model (OWR); and the Back et al. (2018) model (BCL). ITI(patient) (ITI(impatient)) is the informed trading intensity estimated using the first 40 days (last 20 days) of the 60-day Schedule 13D filing window. ITI(insider) is trained on opportunistic insider trading data. ITI(short) is trained on short selling data. These datasets are described in Table IA.2. In Columns (1) and (2), indicators for days with Schedule 13D trading in the first 40 days (patient trade) and last 20 days (impatient trade) of the 60-day Schedule 13D filing window are regressed on the above measures, control variables, and stock fixed effects (filing fixed effects in the Schedule 13D sample). In Columns (3) and (4), the dependent variables are indicators for days with opportunistic insider trades and spikes in short selling. Control variables are winsorized at 0.05% and 99.95% and include effective spread, price impact, lambda, depth, realized volatility, turnover, order imbalance, absolute order imbalance, return, and an indicator variable that takes the value one for days with above-average number of trades (Duarte et al. (2020)). Adj. R^2 ITI(x) is the adjusted R^2 from a regression that only includes the ITI measure trained on the specific dataset and fixed effects. The sample consists of NYSE-listed stocks from 1994 to 2012. To be included, a stock must have a price greater than \$5 and a market capitalization greater than \$100 million at the end of the previous month. Standard errors are clustered by filing and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

Dep. Variable	13D patient trade (1)	13D impatient trade (2)	Insider trade (3)	Δ Short (4)
ITI(patient)	0.104**	0.062	-0.015*	0.018
	(2.133)	(0.988)	(-1.767)	(1.240)
ITI(impatient)	0.330***	0.386***	0.081***	0.127***
	(5.732)	(5.945)	(8.318)	(6.778)
ITI(insider)			0.184***	
			(12.011)	
ITI(short)				0.320***
				(9.637)
PIN	0.015	0.044	-0.001	0.005
	(0.839)	(1.341)	(-0.266)	(0.729)
APIN	0.013	0.053**	0.010***	0.014***
	(0.936)	(2.458)	(3.974)	(3.294)
GPIN	-0.021*	0.005	-0.002	-0.001
	(-1.727)	(0.276)	(-0.626)	(-0.157)
OWR	0.001	0.068	0.002	0.014*
	(0.030)	(1.141)	(0.496)	(1.677)
BCL	-0.002	0.043**	0.009***	0.007
	(-0.113)	(2.018)	(3.823)	(1.638)
Controls and FE	Yes	Yes	Yes	Yes
Adj. R^2	0.0391	0.0575	0.0042	0.0109
Adj. R^2 ITI(x)	0.0173	0.0430	0.0023	0.0081
Obs.	6,250	2,979	199,109	71,243

Table 10. Informed trading intensity and future returns. Monthly returns are regressed on ITI measures (weekly averages), other predictors, and weekly fixed effects. ITI(13D) is trained on Schedule 13D data. ITI(insider) is trained on opportunistic insider trading data. ITI(insider, buy) and ITI(insider, sell) are trained on insider buy and sell trades, respectively. ITI(short) is trained on short selling data. ITI(patient) (ITI(impatient)) is trained on the first 40 days (last 20 days) of the 60-day Schedule 13D trading window. Column (5) in Panel (a) controls for other predictors including log market capitalization, CAPM beta, last month return, two-to-six month return, idiosyncratic volatility, Amihud illiquidity, the effective bid-ask spread, and Kyle's lambda. The coefficient estimates for these predictors are reported in columns 4 through 6 of Table IA.22. Column (6) in Panel (a) adds PIN. The PIN sample is from Brown and Hillegeist (2007) and covers 1993 to 2010. The main sample includes common stocks from 1/1993 to 7/2019. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million. t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level. Standard errors are clustered by stock and month.

(a) ITIs and stock returns

Dep. variable:	Monthly return					
	(1)	(2)	(3)	(4)	(5)	(6)
ITI(13D)	0.0126***	0.0117***		0.0053**	0.0061***	0.0096***
	(6.82)	(4.67)		(2.22)	(3.23)	(4.06)
ITI(insider)		0.0119***		0.0107***	0.0094***	0.0092***
		(4.82)		(4.42)	(5.69)	(4.96)
ITI(short)		-0.0084		-0.0201***	-0.0065	0.0036
		(-1.29)		(-2.79)	(-1.26)	(0.60)
ITI(patient)			0.0006	-0.003	-0.0038*	-0.0049*
			(0.28)	(-1.43)	(-1.85)	(-1.82)
ITI(impatient)			0.0158***	0.0163***	0.0164***	0.0177***
			(6.51)	(5.83)	(5.25)	(4.67)
PIN						0.0019
						(0.35)
Controls	No	No	No	No	Yes	Yes
Fixed effects	Week	Week	Week	Week	Week	Week
R^2	0.0001	0.0002	0.0002	0.0002	0.0011	0.0014
Obs.	3,484,923	3,484,923	3,484,923	3,484,923	3,484,923	2,338,675

(b) ITI estimated on insider buys or sells

	(1)	(2)	(3)	(4)
ITI(insider, buy)	0.0120***		0.0112***	0.0070***
	(2.91)		(3.45)	(3.46)
ITI(insider, sell)		-0.0065	-0.0018	0.0029
		(-1.62)	(-0.60)	(1.35)
Controls	No	No	No	Yes
R^2	0.0001	0.0000	0.0001	0.0012
Obs.	3,484,923	3,484,923	3,484,923	3,484,923

Table 11. Portfolio sorts based on ITI measures. For each ITI measure, we sort stocks into decile portfolios based on their average ITI measures during the prior week. We compute equally-weighted average return during next month for each decile and the top-minus-bottom difference. We report raw returns and alphas from three factor Fama-French models with a momentum factor. We consider separately five ITI measures. ITI(13D) is trained on Schedule 13D data. ITI(insider) is trained on opportunistic insider trading data. ITI(short) is trained on short selling data. These datasets are described in Table IA.2. ITI(patient) (ITI(impatient)) is trained on the first 40 days (last 20 days) of the 60-day Schedule 13D trading window. "0.0071" corresponds to 0.71% per month. The sample includes U.S. common stocks from 1/1993 to 7/2019. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level. Standard errors are computed using Newey-West adjustment with eight lags.

(a) Average monthly returns

	Low	2	3	4	5	6	7	8	9	High	H-L
ITI(13D)	0.0071**	0.0079***	0.0083***	0.0088***	0.0089***	0.0087***	0.0099***	0.0099***	0.0106***	0.0115***	0.0044***
	(2.4)	(2.8)	(2.9)	(3.1)	(3.2)	(3.1)	(3.5)	(3.5)	(3.7)	(4.2)	(4.9)
ITI(insider)	0.0072**	0.0078***	0.0085***	0.0090***	0.0089***	0.0096***	0.0094***	0.0103***	0.0103***	0.0109***	0.0037***
	(2.5)	(2.8)	(3.0)	(3.2)	(3.2)	(3.4)	(3.4)	(3.6)	(3.6)	(3.6)	(3.3)
ITI(short)	0.0085***	0.0086***	0.0086***	0.0089***	0.0093***	0.0088***	0.0095***	0.0098***	0.0099***	0.0098***	0.0014
	(3.2)	(3.2)	(3.1)	(3.2)	(3.3)	(3.1)	(3.3)	(3.4)	(3.3)	(3.1)	(1.0)
ITI(patient)	0.0079***	0.0082***	0.0085***	0.0088***	0.0090***	0.0091***	0.0097***	0.0095***	0.0101***	0.0109***	0.0030***
	(2.7)	(2.9)	(3.0)	(3.1)	(3.2)	(3.2)	(3.5)	(3.4)	(3.5)	(3.9)	(3.9)
ITI(impatient)	0.0064**	0.0077***	0.0080***	0.0088***	0.0093***	0.0094***	0.0099***	0.0103***	0.0104***	0.0115***	0.0051***
	(2.2)	(2.7)	(2.8)	(3.1)	(3.2)	(3.4)	(3.5)	(3.6)	(3.7)	(4.2)	(5.1)

(b) Four-factor Fama-French monthly alphas

	Low	2	3	4	5	6	7	8	9	High	H-L
ITI(13D)	-0.0003	0.0005	0.0007	0.0014***	0.0017***	0.0013***	0.0027***	0.0027***	0.0035***	0.0049***	0.0052***
	(-0.5)	(1.0)	(1.4)	(2.9)	(3.6)	(2.6)	(5.0)	(4.6)	(5.2)	(7.0)	(6.2)
ITI(insider)	0.0003	0.0010*	0.0016***	0.0020***	0.0018***	0.0024***	0.0021***	0.0028***	0.0025***	0.0026***	0.0022***
	(0.6)	(1.6)	(2.8)	(3.5)	(3.2)	(4.4)	(4.1)	(5.9)	(5.1)	(4.4)	(2.9)
ITI(short)	0.0017***	0.0016**	0.0015**	0.0017***	0.0020***	0.0014***	0.0021***	0.0024***	0.0025***	0.0022***	0.0005
	(2.6)	(2.5)	(2.5)	(3.1)	(3.7)	(2.8)	(3.9)	(4.5)	(4.4)	(3.2)	(0.5)
ITI(patient)	0.0008	0.0007	0.0011**	0.0014***	0.0016***	0.0018***	0.0024***	0.0023***	0.0027***	0.0041***	0.0034***
	(1.1)	(1.5)	(2.3)	(2.9)	(3.1)	(3.7)	(4.4)	(4.3)	(4.9)	(6.2)	(4.2)
ITI(impatient)	-0.0006	0.0004	0.0006	0.0015***	0.0018***	0.0021***	0.0025***	0.0030***	0.0032***	0.0048***	0.0054***
	(-1.0)	(0.8)	(1.2)	(3.2)	(3.6)	(4.6)	(4.9)	(5.3)	(4.8)	(6.4)	(5.7)

Internet Appendix to "Informed Trading Intensity"

This appendix provides additional results to supplement the main text.

Appendix IA.A. Reversal and Informed Trading

This section discusses the link between price change reversal and expected/realized informed trading in the Glosten and Milgrom (1985) framework.

The model features three dates and an asset whose value at time 2 is denoted by V, which equals V_L with probability δ and V_H with probability $1 - \delta$. A risk-neutral market maker quotes a bid and ask at time 1. With probability μ, an informed trader who knows the final value of the asset arrives. Hence, μ is the expected informed trading intensity. With probability $1 - \mu$, the trader is uninformed and buys or sells with equal probabilities.

At time 0, there is no informed trading and the price of the asset equals its expected value: $P_0 = \delta V_L + (1 - \delta)V_H = V_H - \delta(V_H - V_L)$.

We focus on the occurrence of a sell order at time 1. The case of a buy order is symmetric. As shown by Glosten and Milgrom (1985), the market maker quotes the following bid:

$$\text{Bid}_1 = E[V|\text{Sell}] = V_H - \delta_1(V_H - V_L),$$

where $\delta_1 = \frac{\delta(1+\mu)}{1-(1-2\delta)\mu}$. Note that $\delta_1 > \delta$ as long as $\mu > 0$. Hence, $P_1 - P_0 = -(V_H - V_L)(\delta_1 - \delta) < 0$.

At time 2, there are two possibilities:

1. If the sell order at time 1 was informed, then $V = V_L$ and $P_2 - P_1 < 0$. Therefore, conditional on realized informed trading at time 1, there is continuation (i.e., price changes are positively correlated).

2. If the order was uninformed, then $E[P_2 - P_1|\text{Uninformed Sell}] = \delta(V_L - P_1) + (1 - \delta)(V_H - P_1) = P_0 - P_1$ since the sell order was uninformative. On average, the initial price change reverses.

Therefore, high realized informed trading (relative to the market maker's expectation) is associated with continuation. In contrast, expected informed trading (μ) does not affect continuation and reversal. The price is determined by the information set of the market maker, which includes μ), and is a martingale in this model.

Appendix IA.B. Simulations of Informed Trading Models

We consider simulations of multi-period informed trading models from two main classes:

1. Models with sequential trade like the PIN model of Easley and O'Hara (1992) and its extension with symmetric order flow shocks: the APIN model of Duarte and Young (2009).

2. Model with strategic trade as in Kyle (1985) and its extension with stochastic noise trading volatility (Collin-Dufresne and Fos (2016)).

The purpose of these simulations is to understand how measures of informed trading relate to variables that can be measured in the data according to these theories.

Methodology

We simulate sequences of multi-period models where each sequence is assumed to represent one trading day. This procedure assumes that each day is independent from the previous one. In the main text, we discuss the case where one day represents one period of a multi-period model.

We simulate 250 days for each of model. To obtain daily variables, we aggregate intraday variables as follows. In the sequential trade model, daily volume is the sum of volume executed in each period. Absolute imbalance is the sum of buy volume minus sell volume. The daily price change is equal to the final transaction price minus the market maker's prior at the beginning of the day. In the strategic trade model, order flow is the sum of noise flow and informed flow, and volume is equal to $\frac{1}{2}(|\text{informed flow}| + |\text{noise flow}| + |\text{order flow}|)$. Absolute order imbalance is equal to $|\text{order flow}|$. To measure informed trading intensity (ITI), we divide daily informed volume by daily volume. In the sequential trade model, we also consider an indicator variable for whether there was informed trading on a specific day since there can be days without informational events. In the strategic trade model, this indicator always equals one since the informed investor always trades.

We estimate the following two regressions using data simulated from the models:

$$\text{ITI}_t = a + b_1 V_t + b_2 |OI_t| + e_t, \qquad \text{and} \tag{IA.B1}$$

$$\text{ITI}_t = c + d_1 V_t + d_2 |OI_t| + d_3 |r_t| + u_t, \tag{IA.B2}$$

where V_t denotes daily volume and OI_t denotes daily order imbalance. We are interested in the estimated coefficients and the variables' explanatory power for ITI. To summarize results across calibrations, which are described below, we report the average adjusted R^2 from estimating (IA.B1) and (IA.B2) and whether the median estimated coefficient for a variable is statistically lower or greater than zero at a significance level of 1%. Finally, we also consider whether any of the specific calibration is able to match the pattern that we observe in the data, which is reported in the first row of Table 8 in the main text.

Results

In the sequential trade model, we try the following combinations of values: probability of informational event to be any of [0.2, 0.4, 0.6, 0.8]; probability of bad event (low signal) to be 0.5; proportion of informed traders to be any of [0.2, 0.4, 0.6, 0.8]; probability that uninformed trader trades (buy or sell) to be any of [0.2, 0.4, 0.6, 0.8]; number of periods to be any of [40, 100, 400];

value of the asset if good news to be 10; value of the asset if bad news to be 5. The above therefore implies $4 \times 4 \times 4 \times 3 = 192$ models. In the Duarte and Young (2009) extension of the PIN model, we take the probability of a symmetric order flow shock to be any of [0.2, 0.4, 0.6, 0.8]. In the case of no order flow shock, we take the probability that uninformed trader trades to be any of [0.2, 0.4]. In the case of an order flow shock, we take the probability that uninformed traders trade to be any of [0.6, 0.8]. The above therefore implies $4 \times 4 \times 3 \times 4 \times 2 \times 2 = 768$ models.

In the above sequential models, the number of periods (T) and the proportion of informed traders have the strongest effect on the quantities of interest in (IA.B1) and (IA.B2). In Table 8, we report summary simulation results for our three values of T. Volume is not positively related to ITI when controlling for absolute order imbalance and absolute daily return. Though $|OI|$ retains statistical significance, its estimated coefficient tends to be reduced by a factor of 2 to 10 with absolute daily return included in the regression. This stems from the strong positive relation between informed trading and absolute daily return. Adjusted R^2 are also on average extremely large. In Table IA.1, we report similar results using an indicator variable for informed trading as dependent variable. None of the calibrations can qualitatively match the data.

In the strategic trade model (Kyle (1985)), we set the initial price to 2, the market maker's prior on the fundamental variance to 0.04; noise trading volatility to 0.2; the number of trading periods to 100. Except for the number of trading periods (T), these variables do not affect the quantities of interest in (IA.B1) and (IA.B2). Hence, we only report simulation results for different T in Table 8. Volume is negatively related to ITI when controlling for absolute order imbalance. Even with only 40 trading periods, volume-related variables achieve an R^2 of about 76%. Since volume is always negatively related to ITI, none of the calibrations can qualitatively match the data. We also cannot estimate (IA.B2) because of perfect collinearity between absolute order imbalance and absolute daily return. To break this collinearity, we vary noise trading volatility across days to be either 0.1 or 0.3 with equal probability. As shown in Table 8 (Kyle-tv), the results are similar.

Last, we simulate data from the stochastic noise trading volatility model of Collin-Dufresne and Fos (2016). This model is based on the continuous-time version of the Kyle (1985) model. To simulate it, we set time steps of $1/1,000$. We set Poisson jump intensities from the low (high) volatility state to the high (low) volatility state to either 2 or 10, with noise trading volatility in the high (low) state set to 0.4 (0.1). We also set the market maker's prior the market maker's prior on the fundamental variance to be either 0.04 or 0.16. As shown in Table 8 (Kyle-stochastic), stochastic noise trading volatility weakens the link between informed trading and volume. Controlling for the absolute return renders absolute order imbalance statistically insignificant. None of the calibrations can qualitatively match the data.

Table IA.1. Statistics from simulations of informed trading models (indicator). The methodology is described in the text. We use simulated data to estimate regressions of an indicator for days with informed trading (I_t) on daily volume (V_t), absolute daily order imbalance ($|OI_t|$), and absolute daily return ($|r_t|$). < 0 (> 0) indicates that the median estimated coefficient across calibrations is lower (greater) than zero and statistically significant at the level of 1%. avg R-squared is the average adjusted R^2 across simulations. T denotes the number of trading periods.

	$\mathrm{I}_t = a + b_1 V_t + b_2 \lvert OI_t \rvert + e_t$			$\mathrm{I}_t = c + d_1 V_t + d_2 \lvert OI_t \rvert + d_3 \lvert r_t \rvert + u_t$			
	b_1	b_2	R-squared	d_1	d_2	d_3	R-squared
PIN ($T = 40$)	> 0	> 0	73.4%	0	0	> 0	81.2%
PIN ($T = 100$)	> 0	> 0	86.6%	0	0	> 0	92.9%
PIN ($T = 400$)	> 0	> 0	96.2%	0	0	> 0	99.8%
APIN ($T = 40$)	0	> 0	68.2%	0	0	> 0	76.7%
APIN ($T = 100$)	0	> 0	83.0%	0	< 0	> 0	91.3%
APIN ($T = 400$)	0	> 0	94.8%	0	0	> 0	99.1%

Appendix IA.C. Additional Figures and Tables

Figure IA.1. Cumulative abnormal return around Schedule 13D filing. The cumulative abnormal return is obtained by subtracting the cumulative market return from the cumulative stock return. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million. The final sample consists of 1,593 Schedule 13D filings between 1994 and 2018.

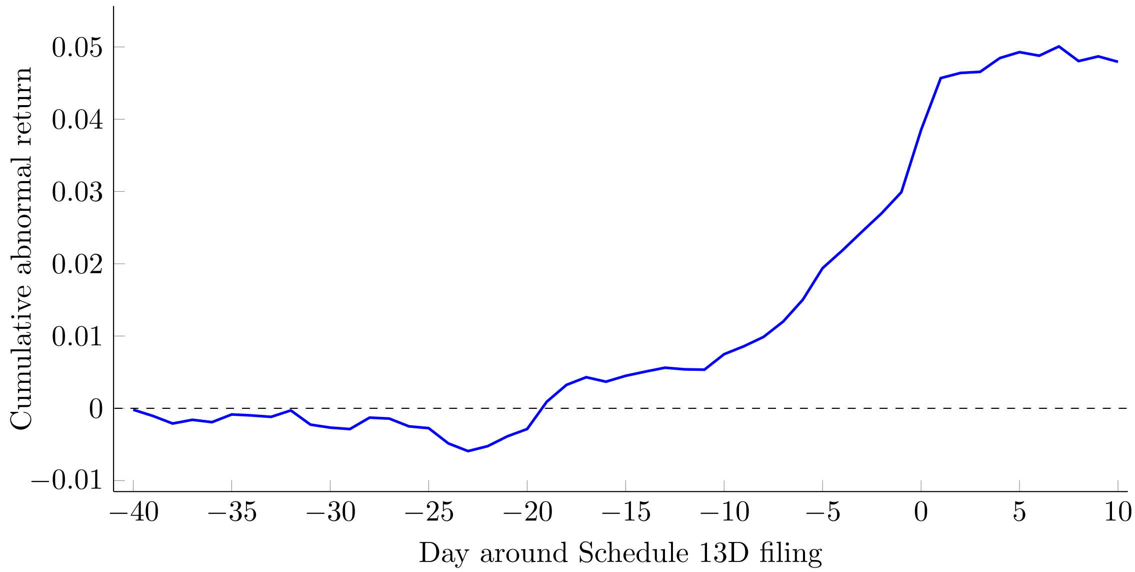


Figure IA.2. This figure reports the total number of observations each year in the 13D sample. The dashed line indicates the number of observations with a trade by a 13D filer. The sample consists of 1,593 13D filings between 1994 and 2018.

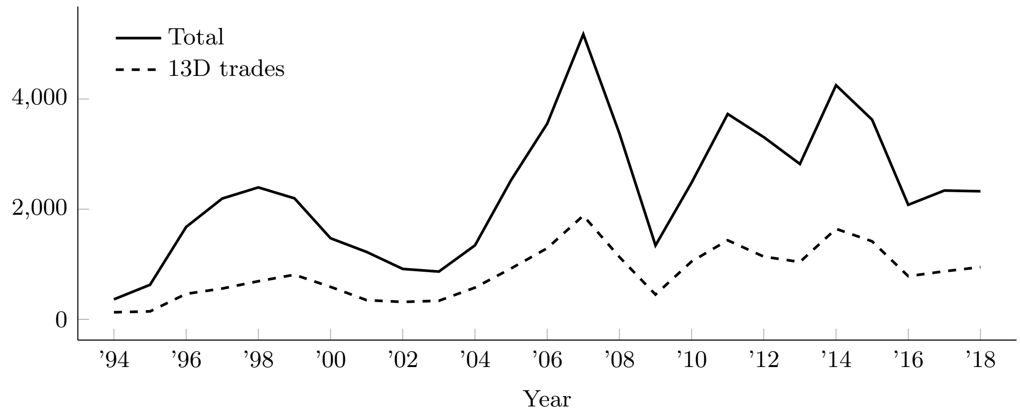
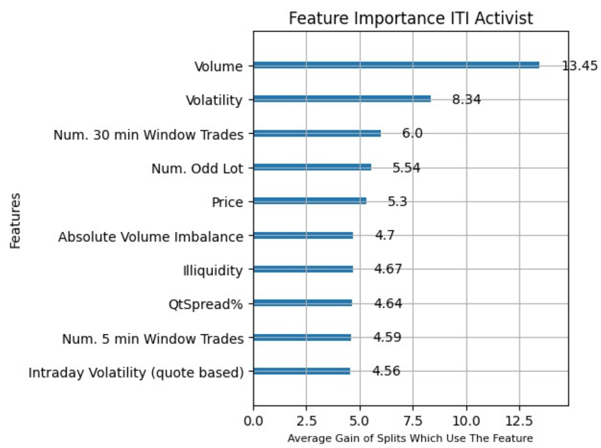
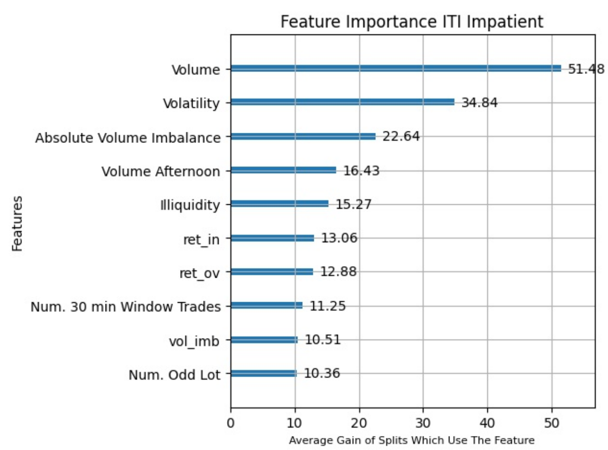


Figure IA.3. This figure ranks features by their importance according to XGB's internal procedure that ranks features based on their gain (a default option based on the average gain across all splits where a feature was used). Variables are defined in Table IA.3.

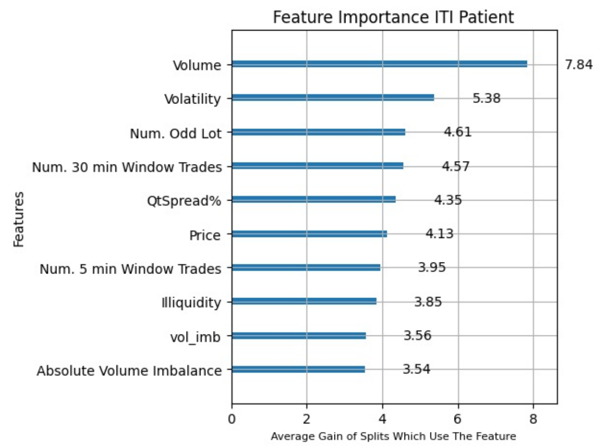
(a) ITI(13D)



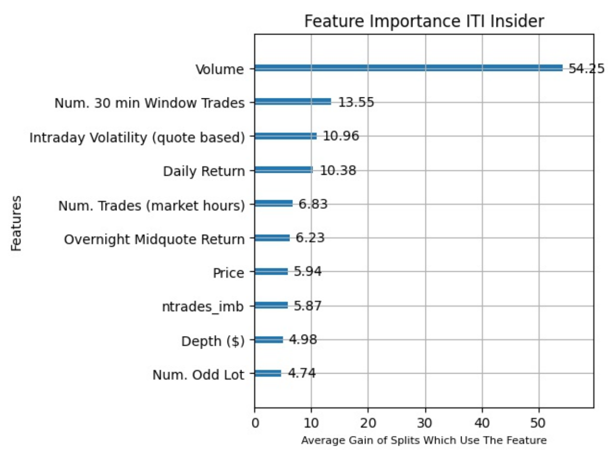
(b) ITI(impatient)



(c) ITI(patient)



(d) ITI(insider)



(e) ITI(short)

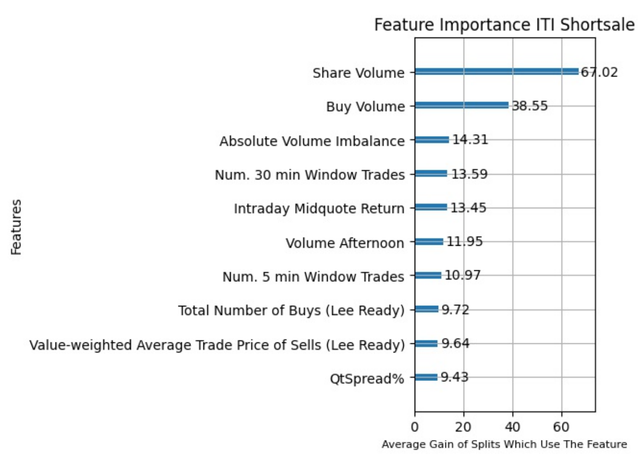
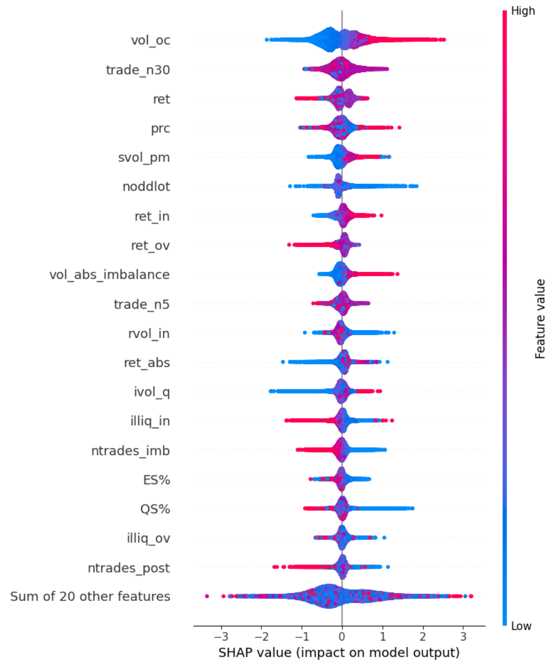
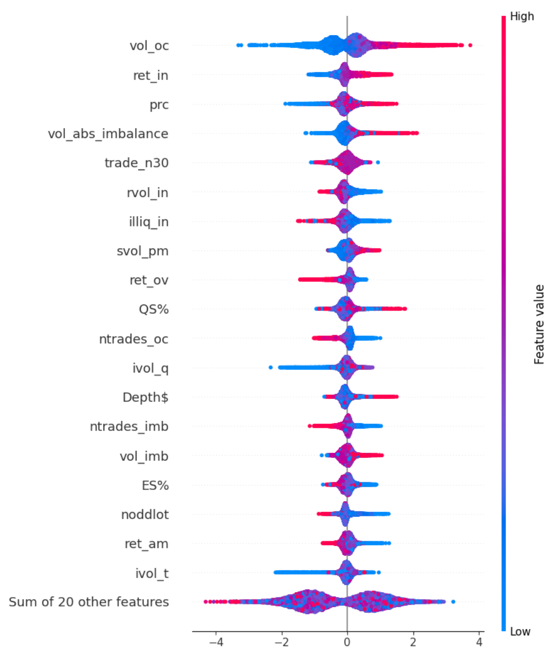


Figure IA.4. This figure ranks features by their importance according to SHAP values (SHapley Additive exPlanations, Lundberg and Lee (2017)). Variables are defined in Table IA.3.

(a) ITI(13D)



(b) ITI(impatient)



(c) ITI(patient)

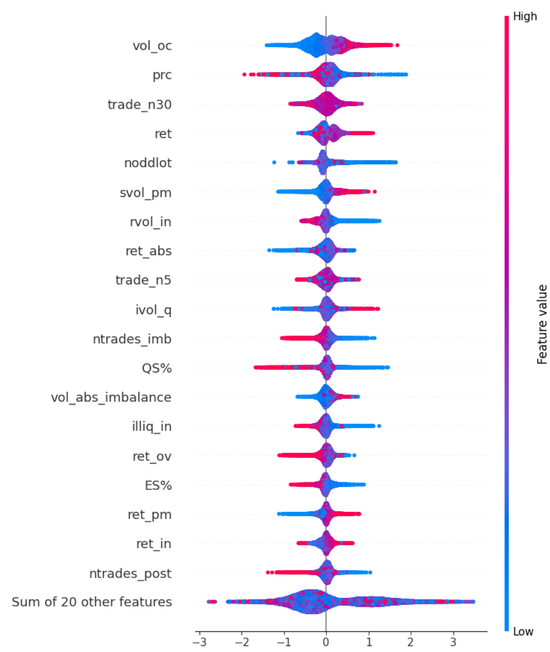
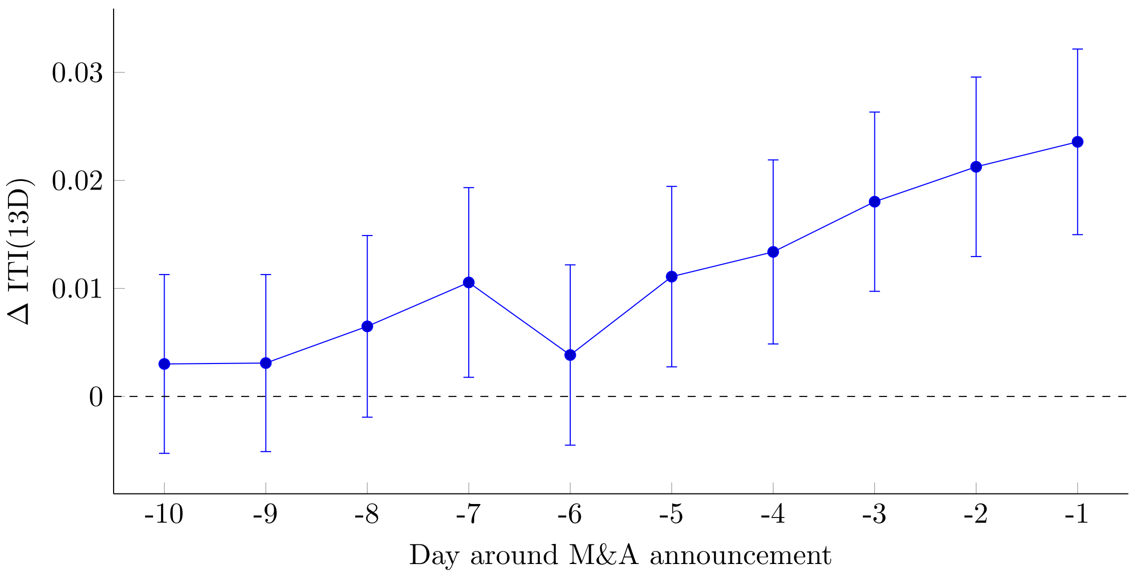


Figure IA.5. ITI(13D) before and after M&A announcements. Informed trading intensity is regressed on indicator variables for days before and after M&A announcements for targeted firms and stock fixed effects. The sample includes common stocks from 1/1996 to 12/2010, and the number of M&A events is 1,577. The figure reports 95% confidence intervals based on standard errors that are double-clustered by stock and date.

(a) Pre announcement



(b) Post announcement

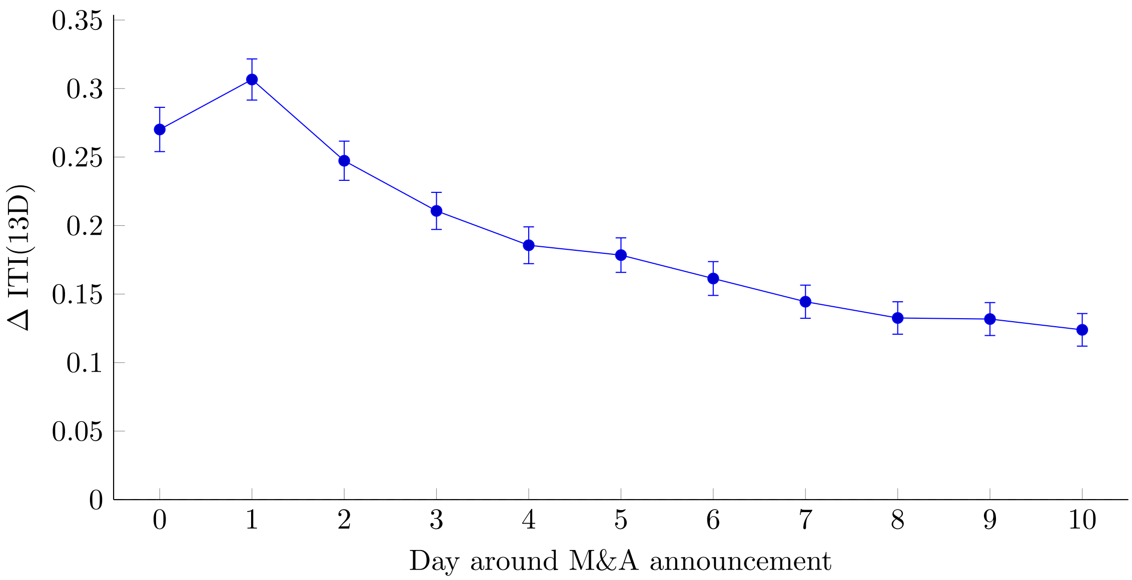
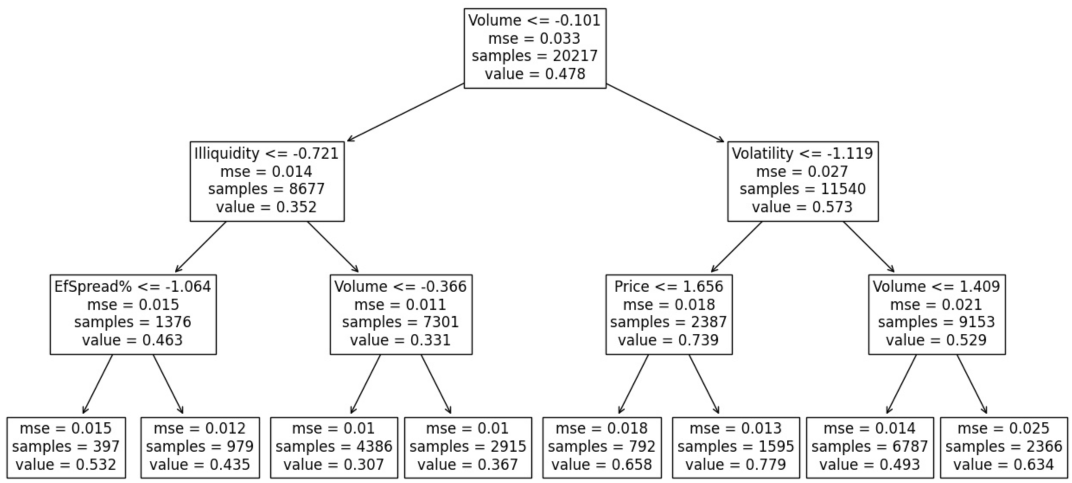
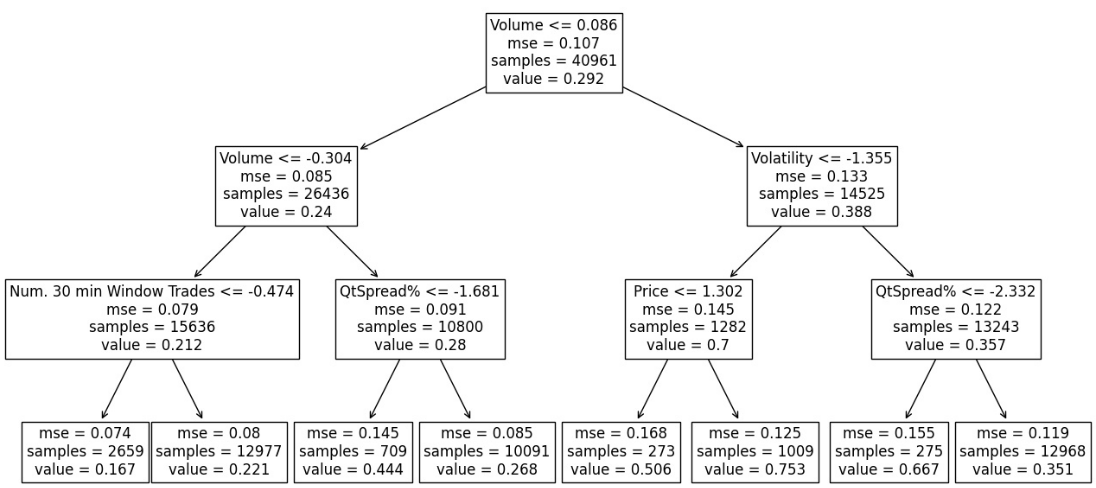


Figure IA.6. This figure reports surrogate trees. Variables are defined in Table IA.3.

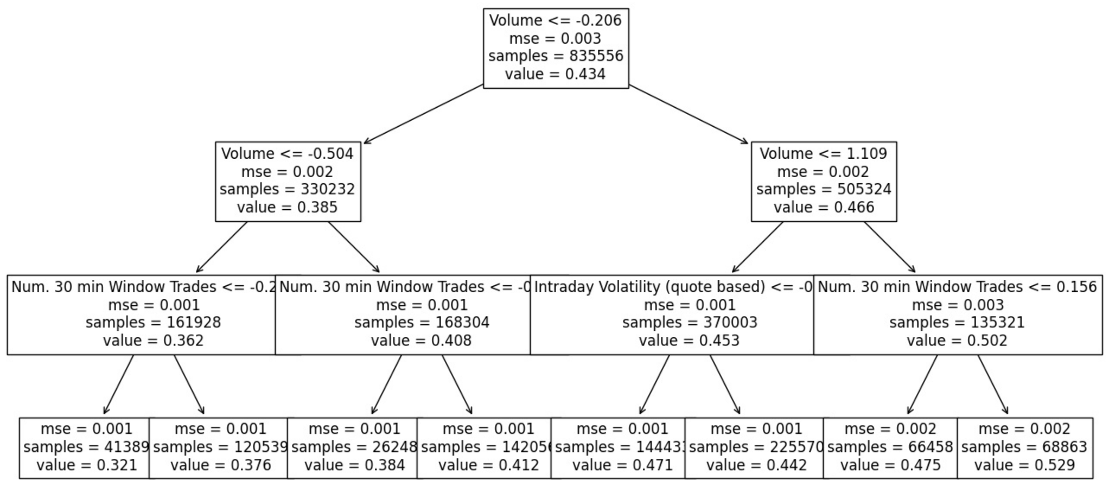
(a) ITI(impatient)



(b) ITI(patient)



(c) ITI(insider)



(d) ITI(short)

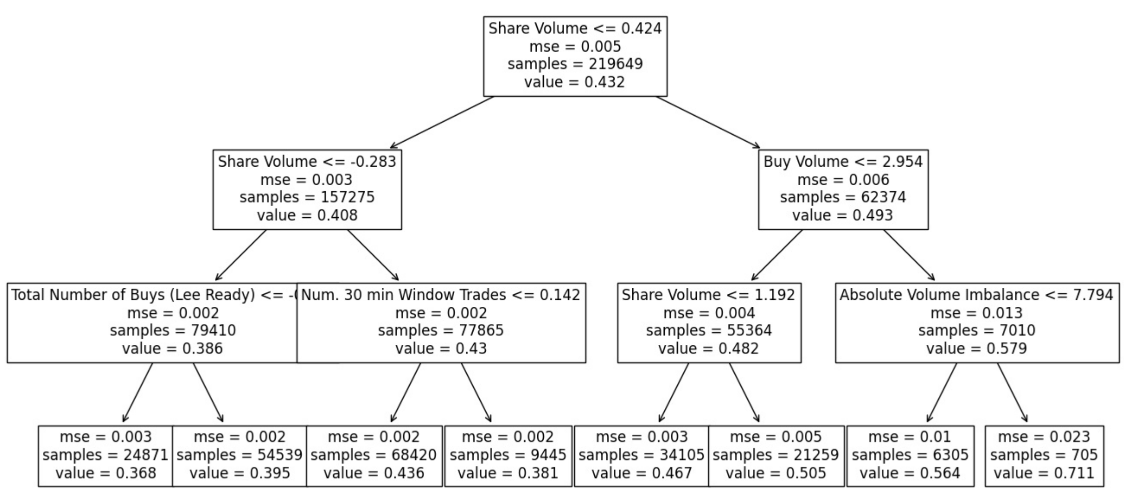


Table IA.2. Overview of datasets used in the analysis. In all datasets, stocks with a price below \$5 or a market capitalization lower than \$100 million at the end of the previous month are excluded.

Dataset Name	Source	Methodology	Obs.
Schedule 13D	EDGAR	The dataset includes the 60-day disclosure period up to the filing date of 1,593 Schedule 13D filings between 1994 and 2018. The sample construction procedure follows Collin-Dufresne and Fos (2015).	58,197
Insider	TR	Data are from Table 1 of the Thomson Reuters (TR) Insider Database. Transactions associated with derivative securities and observations containing cleanse indicators "S" and "A" are dropped from the sample. The (non)routine trades follow the classification defined in Cohen et al. (2012). The data cover 1/1993-12/2012. The final insider dataset includes the day of each insider trade as well as the day before and the day after the trade as comparisons.	779,007
Short	Markit, CRSP	Short interest is defined as the quantity on loan from Markit divided by shares outstanding from CRSP. We identify large spikes in short interest by comparing daily changes in these variables to their 90th percentiles over the entire sample. We adjust for the fact that Markit reports the date when short sales are settled. The Markit data cover 7/2006-12/2019. To create the final short dataset, we randomly select 100,000 stock-days with short interest spikes (as defined above). Like for the insider sample, the dataset includes the day before and the day after the spike as comparisons.	216,979
Main	CRSP, TAQ	Dataset of return, liquidity, volume, and volatility variables listed in Table IA.4, which covers U.S. common stocks over 1/1993 to 7/2019 excluding observations in the Schedule 13D dataset.	16,823,151

Table IA.3. List of variables. This table lists the variables that are used to train the machine learning model. CRSP denotes the Center for Research in Security Prices (CRSP). TAQ denotes Trade and Quote database WRDSii denotes the WRDS Intraday Indicators dataset.

Description	Data source	Abbrev.
Return and price variables		
Daily return	CRSP	ret
Daily price	CRSP	prc
Overnight midquote return (previous close to 10am; see Bogousslavsky (2021) for details)	CRSP, TAQ	ret_ov
Intraday midquote return (10am to close)	TAQ	ret_in
Morning midquote return (10am-12pm)	TAQ	ret_am
Afternoon midquote return (2pm-4pm)	TAQ	ret_pm
Liquidity variables		
Effective spread (dollar-weighted, in percent)	WRDSii	ES%
Realized spread (dollar-weighted, in percent)	WRDSii	RS%
Price impact (dollar-weighted, in percent)	WRDSii	PI%
Quoted spread (time-weighted, in percent)	WRDSii	QS%
Lambda (price impact coefficient) with intercept	WRDSii	lambda1
Lambda (price impact coefficient) without intercept	WRDSii	lambda2
Intraday price impact (Average |30-minute return| divided by 30-minute volume over the day (10am-4pm), excluding zero-volume intervals)	TAQ	illiq_in
Overnight price impact, (|overnight return (previous close to 10am)| divided by share volume (9:30-10am), = 0 if zero volume)	CRSP, TAQ	illiq_ov
Dollar depth at best bid and ask, time-weighted (as in Holden and Jacobsen (2014))	TAQ	Depth$
Difference between dollar depth at best ask and best bid	TAQ	Depth_imb
Volatility and autocorrelation variables		
Daily absolute return	CRSP	ret_abs
Absolute overnight (midquote) return	CRSP, TAQ	ret_ov_abs
Absolute intraday (midquote) return	TAQ	ret_in_abs
Intraday realized volatility (square root of sum of squared 30-minute return 10am-4pm)	TAQ	rvol_in
Realized volatility (square root of squared overnight return plus intraday realized variance)	CRSP, TAQ	rvol

Description	Data source	Abbrev.
Intraday Volatility, second-by-second, quote-based	WRDSii	ivol_q
Intraday Volatility, second-by-Second, trade-based	WRDSii	ivol_t
Variance ratio 1 (15-second/3*5-second)	WRDSii	var_ratio15s
Variance ratio 2 (1-min/4*15-second)	WRDSii	var_ratio1mn
Intraday 30-minute return autocorrelation	TAQ	autocorr_in
Product of overnight return and 3:30-4:00pm return	TAQ	ov_last
Volume and imbalance variables		
Daily share volume	CRSP	svol
Morning share volume	TAQ	svol_am
Afternoon share volume	TAQ	svol_pm
Number of 30-minute time intervals with trade	WRDSii	trade_n30
Number of 5-minute time intervals with trade	WRDSii	trade_n5
Number of trades (market hours)	WRDSii	ntrades_oc
Number of trades (post close)	WRDSii	ntrades_post
Number of odd lot trades (market hours)	WRDSii	noddlot
Herfindahl index calculated across 30 minute time units	WRDSii	hindex
Buy trade volume minus sell trade volume signed using (Lee and Ready, 1991)	WRDSii	svol_imb
Absolute value of buy trade volume minus sell trade volume	WRDSii	svol_imb_abs
Total number of buys minus total number of sells, signed using (Lee and Ready, 1991)	WRDSii	ntrades_imb
Absolute value of total number of buys minus sells	WRDSii	ntrades_imb_abs
Value-weighted average trade price of buys minus sells	WRDSii	vwap_imb

Table IA.4. Control variables. This table details the construction of the main control variables. Each control variable is computed daily for each stock.

Name	Abbreviation	Methodology
Effective spread	ES	$\text{ES}_k = \frac{2D_k(P_k - M_k)}{M_k}$, where D_k takes the value 1 (-1) if the trade is classified as a buy (sell) based on the Lee and Ready (1991) algorithm, P_k denotes the transaction price, and M_k denotes the midpoint of the best quote available immediately preceding the transaction (the quote is lagged five seconds before 1998). Effective spread is computed by dollar-weighting ES_k across transactions during trading hours.
Price impact	PI	$\text{PI}_k = \frac{2D_k(M_{k+5} - M_k)}{M_k}$, where M_{k+5} denotes the midpoint five minutes after a trade. Price impact is computed by dollar-weighting PI_k across transactions during trading hours.
Realized spread	RS	$\text{RS}_k = \frac{2D_k(P_k - M_{k+5})}{M_k}$. Realized spread is computed by dollar-weighting PI_k across transactions during trading hours.
Lambda	lambda	Slope (multiplied by 1,000) from the following regression estimated each day for each stock: $\ln \Delta M_{i,\tau} = \alpha + \lambda \left(\text{signed}\sqrt{(ShareVolume_{i,\tau})}\right)$, where τ indicates a five-minute interval.
Depth	depth	Sum of share depth at the best ask and best bid, divided by total shares outstanding.
Realized volatility	rvol	Square root of the sum of squared overnight return and squared 30-minute returns (10am-4pm).
Order imbalance	OI	Buy share volume minus sell share volume, divided by total shares outstanding.

Table IA.5. Descriptive statistics (13D sample). The table reports the mean, standard deviation (SD), within-stock standard deviation (SD_w), and 1st, 5th, 25th, 50th, 75th, 95th and 99th percentiles for the main set of variables. Control variables are described in Table IA.4 and include effective spread (ES), lambda, depth, realized volatility (rvol), turnover (turn), order imbalance (OI), absolute order imbalance (|OI|), and return (ret). These variables are winsorized at 0.05% and 99.95%. 13D trade is an indicator variable that takes the value one on days with Schedule 13D trades. 13D turn is the 13D filer turnover (share volume traded by the filer divided by total shares outstanding). 13D $turn_{tr}$ refers to 13D turn conditional on trading by the filer. The full sample consists of common stocks from 1/1993 to 7/2019, excluding any stock-day with a missing value for one of the control variables. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. The 13D sample consists of the filing period for 1,593 13D filings between 1994 and 2018 (58,197 observations). SD_w denotes the within-filing standard deviation.

Variable name	Mean	SD	SD_w	1%	5%	25%	50%	75%	95%	99%	N
ES	0.0036	0.0049	0.0029	0.0002	0.0004	0.0009	0.0019	0.0040	0.0130	0.0226	58,197
lambda	0.0030	0.0093	0.0081	-0.0157	-0.0046	-0.0001	0.0010	0.0044	0.0164	0.0380	58,197
depth ($*10^2$)	0.0287	0.1301	0.1051	0.0006	0.0011	0.0026	0.0050	0.0125	0.0911	0.5543	58,197
rvol	0.0231	0.0310	0.0284	0.0004	0.0017	0.0101	0.0171	0.0271	0.0573	0.1305	58,197
turn	0.0153	0.0363	0.0336	0.0005	0.0012	0.0038	0.0076	0.0153	0.0455	0.1282	58,197
OI	-0.0007	0.0070	0.0067	-0.0231	-0.0063	-0.0011	-0.0001	0.0007	0.0041	0.0121	58,197
|OI|	0.0025	0.0067	0.0064	0.0000	0.0001	0.0004	0.0009	0.0022	0.0089	0.0275	58,197
ret	0.0009	0.0386	0.0380	-0.0909	-0.0431	-0.0110	0.0000	0.0106	0.0456	0.1111	58,197
13D trade	0.3603	0.4801	0.4191	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	58,197
13D turn ($*10^2$)	0.1005	0.3935	0.3773	0.0000	0.0000	0.0000	0.0000	0.0496	0.4744	1.5091	58,197
13D $turn_{tr}$ ($*10^2$)	0.2755	0.6130	0.4874	0.0005	0.0034	0.0341	0.1055	0.2701	1.0287	2.7766	20,969

Table IA.6. Model Comparison. In Panel (a), an indicator for 13D trading over the filing windows (60 days before filing to filing) is regressed on informed trading intensity computed from different machine learning models. LASSO denotes the LASSO model. RF denotes the random forest model. XGB denotes the boosted trees model. The final sample consists of 1,593 13D filings between 1994 and 2018. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Standard errors are clustered by filing, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level. Panel (b) reports additional evaluation metrics for each model.

(a) Regression

Dep. variable:	13D trade			
	(1)	(2)	(3)	(4)
LASSO	0.981***			-0.049
	(32.498)			(-1.424)
RF		1.140***		0.286***
		(41.439)		(6.258)
XGB			0.803***	0.690***
			(62.651)	(31.021)
Adj. R^2	0.0643	0.0974	0.1367	0.1386
Obs.	58,197	58,197	58,197	58,197

(b) Additional metrics

	LASSO	RF	XGB
	(1)	(2)	(3)
mean absolute error	0.4307	0.4210	0.3763
log loss	0.6510	0.6053	0.5940
ROC AUC	0.6469	0.6670	0.7084

Table IA.7. What variables help explain ITI? ITI is the informed trading intensity estimate on Schedule 13D trades. ITI is regressed on informed trading intensity (ITI) measured from a subset of the variables in Table IA.3 and filing fixed effects. ITI(liquidity), ITI(return), ITI(volatility), and ITI(volume) are versions of ITI that are trained using a subset of the explanatory variables. The sample consists of 1,593 Schedule 13D filings between 1994 and 2018. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Standard errors are clustered by filing, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

Dep. variable:			ITI(13D)		
	(1)	(2)	(3)	(4)	(5)
ITI(liquidity)	0.578***				0.286***
	(40.739)				(48.982)
ITI(return)		0.476***			0.212***
		(28.528)			(35.316)
ITI(volatility)			0.481***		0.204***
			(30.715)		(35.314)
ITI(volume)				0.645***	0.505***
				(73.998)	(106.199)
Filing FE	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.1909	0.1135	0.1276	0.3498	0.4736
Obs.	58,197	58,197	58,197	58,197	58,197

Table IA.8. ITI measures and control variables. ITI measures are regressed on control variables and stock fixed effects. ITI(13D) is the informed trading intensity estimated using the 60-day Schedule 13D filing window. ITI(patient) (ITI(impatient)) is the informed trading intensity estimated using the first 40 days (last 20 days) of the 60-day Schedule 13D filing window. The sample includes common stocks from 1/1993 to 7/2019, excluding Schedule 13D filing windows. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Control variables are winsorized at 0.05% and 99.95% and include effective spread, price impact, lambda, depth, realized volatility, turnover, order imbalance, absolute order imbalance, and return. Standard errors are double-clustered by stock and date, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

Dep. variable:	ITI(13D) (1)	ITI(patient) (2)	ITI(impatient) (3)
ES	-3.604***	-3.342***	-3.514***
	(-58.925)	(-70.975)	(-49.597)
lambda	-0.567***	-0.371***	-0.639***
	(-33.088)	(-28.435)	(-34.730)
depth	55.001***	64.608***	28.871***
	(41.054)	(50.185)	(25.813)
rvol	-0.173***	-0.400***	-0.196***
	(-8.027)	(-24.822)	(-7.963)
turn	2.671***	2.217***	2.719***
	(46.343)	(48.027)	(45.732)
OI	-2.810***	-3.434***	-2.006***
	(-37.627)	(-55.283)	(-24.088)
|OI|	11.661***	7.478***	14.238***
	(37.184)	(32.797)	(53.139)
ret	0.015**	-0.080***	0.219***
	(2.121)	(-14.051)	(31.829)
Stock FE	Yes	Yes	Yes
Adj. R^2	0.1192	0.0868	0.1935
Obs.	16,823,151	16,823,151	16,823,151

Table IA.9. Return reversal (extended). Daily return is regressed on past return, ITI, control variables, date fixed effects, and interactions of lagged returns with ITI, turnover, realized volatility, and effective spread. Control variables are effective spread, price impact, lambda, depth, realized volatility, turnover, order imbalance, and absolute order imbalance. Control variables are winsorized at 0.05% and 99.95%. Standard errors are double-clustered by stock and date, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level. The sample includes common stocks from 1/1993 to 7/2019. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month.

Dep. Variable	ret(t+1)	ret(t+2)	ret(t+3)	ret(t+4)	ret(t+5)	ret(t+6)	ret(t+7)	ret(t+8)	ret(t+9)	ret(t+10)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ret(t)	-0.004	-0.016***	-0.007***	-0.007**	-0.008***	-0.008***	-0.001	-0.003	-0.002	0.003
	(-1.399)	(-5.673)	(-2.616)	(-2.311)	(-3.013)	(-3.177)	(-0.313)	(-1.307)	(-0.640)	(1.161)
ITI(13D)(t)	0.002***	0.001***	0.001***	0.001***	0.001***	0.000***	0.000***	0.000***	0.000***	0.000***
	(20.255)	(9.127)	(9.039)	(6.537)	(7.298)	(4.756)	(5.082)	(4.369)	(5.600)	(4.088)
ret(t)*ITI(13D)(t)	0.028***	0.010**	0.006	0.004	0.006	0.007**	-0.001	-0.001	0.001	-0.003
	(5.189)	(2.569)	(1.537)	(1.125)	(1.569)	(2.165)	(-0.278)	(-0.164)	(0.490)	(-0.887)
ret(t)*turn(t)	0.007	-0.002	0.014	-0.010	0.011	-0.005	0.002	-0.000	-0.000	0.011*
	(0.434)	(-0.196)	(1.586)	(-1.235)	(1.465)	(-0.870)	(0.442)	(-0.001)	(-0.039)	(1.901)
ret(t)*rvol(t)	-0.027	-0.007	-0.002	0.015***	-0.001	0.004	0.003	0.010*	-0.001	-0.005
	(-1.582)	(-0.873)	(-0.243)	(3.199)	(-0.124)	(0.849)	(0.732)	(1.864)	(-0.214)	(-0.875)
ret(t)*ES%(t)	-0.545***	-0.498***	-0.313***	-0.138	-0.119	-0.073	0.017	0.051	-0.003	-0.132*
	(-3.718)	(-3.108)	(-3.092)	(-1.525)	(-1.421)	(-0.902)	(0.235)	(0.742)	(-0.030)	(-1.815)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.0004	0.0004	0.0001	0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000
Obs.	16,708,897									

Table IA.10. 13D filings: patient trading and impatient trading. An indicator for days with Schedule 13D trading and 13D filer turnover (share volume traded by the filer divided by total shares outstanding) are regressed on ITI(patient), ITI(impatient), control variables and filing fixed effects. Effective spread (ES), lambda, depth, realized volatility (rvol), order imbalance (OI), absolute order imbalance (|OI|), and return are winsorized at 0.05% and 99.95%. The sample consists of 1,593 Schedule 13D filings between 1994 and 2018. The regressions with 13D filer turnover are conditional on Schedule 13D trading on that specific stock-day. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Standard errors are clustered by filing, and the associated *t*-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

Dep. variable:	Patient sample		Impatient sample	
	13D trade	13D turnover	13D trade	13D turnover
	(1)	(2)	(3)	(4)
ITI(patient)	0.405***	0.002***		
	(19.127)	(10.571)		
ITI(impatient)			0.589***	0.006***
			(22.188)	(9.661)
ES	-2.045***	0.090***	0.932	0.110**
	(-2.764)	(3.086)	(0.843)	(2.154)
lambda	-0.192	-0.011***	-0.885**	-0.028***
	(-0.739)	(-4.156)	(-2.160)	(-2.859)
depth	29.169***	-0.013	8.210*	-0.198*
	(9.279)	(-0.152)	(1.782)	(-1.707)
rvol	-0.096	-0.003	-0.563***	0.000
	(-0.765)	(-1.072)	(-3.462)	(0.033)
turn	0.709***	0.023***	0.799***	0.063***
	(5.396)	(6.331)	(4.682)	(6.462)
OI	-0.510	0.015	1.042**	0.073**
	(-0.932)	(0.665)	(2.104)	(2.358)
|OI|	2.323***	0.092***	3.644***	0.149***
	(3.354)	(3.331)	(5.707)	(3.794)
ret	-0.078	-0.002	0.054	-0.010***
	(-1.171)	(-1.041)	(0.514)	(-2.721)
Filing FE	Yes	Yes	Yes	Yes
Adj. R^2	0.0629	0.2633	0.0738	0.3166
Adj. R^2 (controls only)	0.0367	0.2485	0.0359	0.3001
Obs.	38,846	11,533	19,274	9,376

Table IA.11. Do ITI measures detect insider purchases and insider sales? Indicator variables for days with opportunistic insider buys or days with opportunistic insider sales are regressed on informed trading intensity measures and control variables. ITI(13D) is trained on Schedule 13D data. ITI(patient) (ITI(impatient)) is trained using the first 40 days (last 20 days) of the 60-day Schedule 13D filing window. ITI(insider) is trained on opportunistic insider trading data. The insider sample consists of two days surrounding opportunistic insider trades between 1993 and 2012. For insider buys, the sample is restricted to days with an insider buy but without an insider sale or days without an insider trade. For insider sales, the sample is restricted to days with an insider sale but without an insider buy or days without an insider trade. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Control variables are winsorized at 0.05% and 99.95% and include effective spread, price impact, lambda, depth, realized volatility, turnover, order imbalance, absolute order imbalance, and return. Standard errors are clustered by filing for the Schedule 13D sample and by stock for the other samples, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

Dep. Variable	Insider purchase				Insider sale			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ITI(13D)	0.110***	0.104***			0.073***	0.024***		
	(28.565)	(26.743)			(17.439)	(5.779)		
ITI(insider)		0.048***		0.027***		0.403***		0.381***
		(7.265)		(4.044)		(50.083)		(47.344)
ITI(patient)			0.048***	0.047***			0.009**	-0.006
			(11.228)	(10.962)			(2.014)	(-1.235)
ITI(impatient)			0.112***	0.107***			0.144***	0.072***
			(22.063)	(20.581)			(25.651)	(13.081)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.0133	0.0134	0.0136	0.0137	0.0067	0.0117	0.0077	0.0120
Obs.	529,740	529,740	529,740	529,740	683,542	683,542	683,542	683,542

Table IA.12. ITI, 13D trades, and insider trades in the full sample. In this table, we examine the ability of ITI to detect 13D trades and insider trades in the full stock-day sample. ITI is the informed trading intensity estimated using the 60-day Schedule 13D filing window. 13D trade is an indicator for days with Schedule 13D trades. Insider trade is an indicator for days with opportunistic insider trades. In Columns (1)-(2), the sample includes common stocks from 1/1993 to 7/2019. In Column (3)-(4), the sample includes common stocks from 1/1993 to 12/2012. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Control variables are winsorized at 0.05% and 99.95% and include effective spread, price impact, lambda, depth, realized volatility, turnover, order imbalance, absolute order imbalance, and return. Standard errors are double-clustered by stock and date, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

Dep. variable:	13D trade (1)	13D trade (2)	Insider trade (3)	Insider trade (4)
ITI(13D)	0.008***	0.005***	0.026***	0.017***
	(21.490)	(18.114)	(46.775)	(31.352)
ES%		-0.112***		0.538***
		(-13.407)		(16.889)
lambda		0.008***		-0.066***
		(4.754)		(-8.938)
depth		9.739***		1.045*
		(10.568)		(1.761)
rvol		-0.019***		-0.030***
		(-9.170)		(-3.269)
turn		0.056***		0.308***
		(11.373)		(19.374)
OI		-0.193***		-0.000
		(-12.795)		(-0.010)
|OI|		0.135***		0.566***
		(8.083)		(9.965)
ret		0.004***		0.059***
		(7.716)		(16.371)
Stock FE	Yes	Yes	Yes	Yes
Adj. R^2	0.0014	0.0045	0.0006	0.0017
Obs.	16,881,293	16,881,293	12,776,809	12,776,809

Table IA.13. Do control variables detect various classes of informed trading? Indicator variables for days with 13D trading, days with opportunistic insider trading, or days with a spike in short interest are regressed on control variables. The 13D sample consists of the filing period for 1,593 13D filings between 1994 and 2018. The insider sample consists of two days surrounding opportunistic insider trades between 1993 and 2012. The short sample consists of two days surrounding 100,000 randomly-selected spikes in short interest from 6/2006 to 12/2010. To be included, a stock must have a price greater than \$5 and a market capitalization greater than \$100 million at the end of the previous month. Control variables are winsorized at 0.05% and 99.95% and include effective spread (ES), price impact (PI), lambda, depth, realized volatility (rvol), turnover (turn), order imbalance (OI), absolute order imbalance (|OI|), and return (ret). Standard errors are clustered by filing for the Schedule 13D sample and by stock for the other samples, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

Dep. variable:	Schedule 13D trade (1)	Opp. insider trade (2)	Δ Short interest (3)
ES	-3.412***	1.002***	-1.773***
	(-4.053)	(7.017)	(-2.915)
lambda	-1.186***	-0.582***	-0.566***
	(-4.822)	(-8.712)	(-3.160)
depth	0.344***	0.003***	-0.311***
	(8.996)	(7.733)	(-3.729)
rvol	-0.832***	0.141***	0.137**
	(-6.149)	(4.044)	(2.037)
turn	1.377***	0.854***	1.551***
	(9.845)	(15.694)	(13.473)
OI	0.379	-0.346**	1.325***
	(0.748)	(-2.050)	(4.196)
|OI|	6.732***	3.250***	5.346***
	(10.534)	(14.985)	(14.618)
ret	-0.083	0.366***	0.269***
	(-1.258)	(20.149)	(8.356)
Stock/Filing FE	Yes	Yes	Yes
R^2	0.0462	0.0033	0.0065
Obs.	58,197	779,007	216,979

Table IA.14. Does ITI(insider) detect insider trades? An indicator for days with opportunistic insider trades is regressed on a set of liquidity variables and stock fixed effects. ITI(insider) is the informed trading intensity trained on the dataset of insider trades. Effective spread (ES), lambda, depth, realized volatility (rvol), order imbalance (OI), absolute order imbalance (|OI|), and return are winsorized at 0.05%. The sample covers 1993 to 2012 and includes 95,464 days with at least one opportunistic buy trade and 260,366 days with at least one opportunistic sell trade. For each day with an insider trade, two days around the trade (one before and one after) without an insider trade are selected and included in the sample. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Standard errors are double-clustered by stock and date, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

Dep. variable:	Day with opportunistic insider trade		
	(1)	(2)	(3)
ITI(insider)	0.444***		0.383***
	(62.052)		(52.317)
ES%		1.002***	0.767***
		(7.017)	(5.459)
lambda		-0.582***	-0.389***
		(-8.712)	(-5.877)
depth		0.299***	0.156***
		(7.733)	(4.217)
rvol		0.141***	0.113***
		(4.044)	(3.295)
turn		0.854***	0.548***
		(15.694)	(10.651)
OI		-0.346**	-0.363**
		(-2.050)	(-2.204)
|OI|		3.250***	2.050***
		(14.985)	(9.631)
ret		0.366***	0.239***
		(20.149)	(13.233)
Stock FE	Yes	Yes	Yes
Adj. R^2	0.0064	0.0033	0.0077
Obs.	779,007	779,007	779,007

Table IA.15. Does ITI(short) detect spikes in short selling? An indicator for days with spikes in short interest is regressed on a set of liquidity variables and stock fixed effects. ITI(short) is the informed trading intensity trained on the dataset of spikes in short interest. Effective spread (ES), lambda, depth, realized volatility (rvol), order imbalance (OI), absolute order imbalance (|OI|), and return are winsorized at 0.05%. The spikes in short interest indicator variable takes the value one if the daily change in short interest exceeds the 90th percentile over July 2006 to July 2019. Short interest is defined as the total short sale demand divided by the number of shares outstanding. The short sample consists of two days surrounding 100,000 randomly-selected spikes in short interest from 6/2006 to 12/2010. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Standard errors are double-clustered by stock and date, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

Dep. variable:	Day with spike in short interest		
	(1)	(2)	(3)
ITI(short)	0.640***		0.494***
	(43.900)		(29.173)
ES%		-1.773***	-1.615***
		(-2.915)	(-2.691)
lambda		-0.566***	-0.333*
		(-3.160)	(-1.866)
depth		-0.311***	-0.159*
		(-3.729)	(-1.945)
rvol		0.137**	0.056
		(2.037)	(0.870)
turn		1.551***	0.850***
		(13.473)	(8.079)
OI		1.325***	0.617**
		(4.196)	(2.024)
|OI|		5.346***	2.759***
		(14.618)	(7.751)
ret		0.269***	0.143***
		(8.356)	(4.463)
Stock FE	Yes	Yes	Yes
Adj. R^2	0.0091	0.0064	0.0104
Obs.	216,979	216,979	216,979

Table IA.16. Commonality in ITI measures. This table examines commonalities among ITI measures trained on Schedule 13D trades (ITI(13D), ITI(patient), ITI(impatient)), opportunistic insider trades (ITI(insider)), and on spikes in short-selling (ITI(short)), which are described in Table IA.2. ITI measures are regressed on each other, a set of control variables, and stock fixed effects. The sample includes common stocks from 1/1993 to 7/2019. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Control variables are winsorized at 0.05% and 99.95% and include effective spread, price impact, lambda, depth, realized volatility, turnover, order imbalance, absolute order imbalance, and return. Standard errors are double-clustered by stock and date, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

Dep. variable:	ITI(13D)	ITI(insider)		ITI(short)	
	(1)	(2)	(3)	(4)	(5)
ITI(insider)	0.076***			0.038***	0.027***
	(78.320)			(72.802)	(41.057)
ITI(short)	0.608***	0.260***	0.205***		
	(141.716)	(77.792)	(44.730)		
ITI(13D)		0.088***		0.102***	
		(73.485)		(176.459)	
ITI(patient)			0.046***		0.023***
			(28.465)		(34.067)
ITI(impatient)			0.117***		0.179***
			(43.289)		(144.797)
Controls	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes
Adj. R^2 (controls only)	0.1197	0.0228	0.0228	0.1817	0.1817
Adj. R^2	0.1843	0.0441	0.0485	0.2441	0.3100
Obs.	15,839,674	15,839,674	15,839,674	15,839,674	15,839,674

Table IA.17. ITI measures, volume, absolute order imbalance, and absolute return. ITI measures are regressed on volume, absolute order imbalance, absolute return, and stock fixed effects. ITI(13D) is the informed trading intensity estimated using the 60-day Schedule 13D filing window. ITI(patient) (ITI(impatient)) is the informed trading intensity estimated using the first 40 days (last 20 days) of the 60-day Schedule 13D filing window. The sample includes common stocks from 1/1993 to 7/2019, excluding Schedule 13D filing windows. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Control variables are winsorized at 0.05% and 99.95% and include effective spread, price impact, lambda, depth, realized volatility, turnover, order imbalance, absolute order imbalance, and return. Standard errors are double-clustered by stock and date, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

(a) ITI, turnover, and absolute order imbalance

	ITI(13D)			ITI(patient)			ITI(impatient)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
turn	3.910***		2.692***	2.847***		2.039***	4.214***		2.750***
	(69.980)		(52.152)	(67.777)		(50.820)	(71.566)		(53.103)
|OI|		18.855***	11.516***		13.195***	7.636***		21.343***	13.846***
		(56.946)	(37.454)		(53.351)	(33.307)		(74.168)	(54.288)
Adj. R^2	0.0826	0.0745	0.1024	0.0526	0.0438	0.0631	0.1321	0.1314	0.1715
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	16,823,151	16,823,151	16,823,151	16,823,151	16,823,151	16,823,151	16,823,151	16,823,151	16,823,151

(b) With absolute return

	ITI(13D)	ITI(patient)	ITI(impatient)
	(1)	(2)	(3)
turn	2.901***	2.321***	2.982***
	(51.226)	(51.442)	(51.536)
|OI|	11.512***	7.631***	13.842***
	(36.724)	(32.189)	(52.859)
|ret|	-0.295***	-0.397***	-0.327***
	(-23.427)	(-40.014)	(-25.811)
Adj. R^2	0.1039	0.0663	0.1740
Stock FE	Yes	Yes	Yes
Obs.	16,823,151	16,823,151	16,823,151

Table IA.18. Sensitivity of informed trade to uninformed trade. In Column (1), an indicator for days with Schedule 13D trading over the filing windows (60 days before the filing date to the filing date) is regressed on the logarithm of turnover that excludes trading by the 13D filer (non13Dturn) and filing fixed effects. In Column (2), the logarithm of 13D filer turnover (13Dturn) is regressed on the logarithm of non13Dturn and filing fixed effects, conditional on Schedule 13D trading on that specific stock-day. The sample consists of 1,593 Schedule 13D filings between 1994 and 2018. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Standard errors are clustered by filing, and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

	(1)	(2)
Dep. variable:	13D trade	log 13Dturn
log non13Dturn	0.093***	0.546***
	(19.639)	(28.139)
Filing FE	Yes	Yes
Adj. R^2	0.0265	0.0882
Obs.	58,053	20,818

Table IA.19. ITI measures and other measures of informed trading. This table compares ITI measures to the conditional probability of informed trading obtained from several models: the PIN model (PIN); the adjusted PIN model of Duarte and Young (2009) (APIN); the generalized PIN model (GPIN) of Duarte et al. (2020); the Odders-White and Ready (2008) model (OWR); and the Back et al. (2018) model (BCL). ITI(patient) (ITI(impatient)) is the informed trading intensity estimated using the first 40 days (last 20 days) of the 60-day Schedule 13D filing window. Indicators for days with Schedule 13D trading in the the first 40 days (patient trade) and last 20 days (impatient trade) of the 60-day Schedule 13D filing window are regressed on the above measures and control variables. Control variables are winsorized at 0.05% and 99.95% and include effective spread, price impact, lambda, depth, realized volatility, turnover, order imbalance, absolute order imbalance, return, and an indicator variable that takes the value one for days with above-average number of trades (Duarte et al. (2020)). The sample consists of NYSE-listed stocks from 1994 to 2012. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Standard errors are clustered by filing and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

Dep. Variable	Schedule 13D patient trade					Schedule 13D impatient trade				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ITI(patient)	0.115**	0.115**	0.115**	0.115**	0.104**	0.091	0.092	0.100*	0.097*	0.068
	(2.561)	(2.570)	(2.568)	(2.565)	(2.132)	(1.608)	(1.614)	(1.753)	(1.701)	(1.085)
ITI(impatient)	0.335***	0.332***	0.343***	0.342***	0.343***	0.414***	0.419***	0.433***	0.433***	0.415***
	(6.521)	(6.541)	(6.778)	(6.749)	(6.105)	(6.892)	(7.060)	(7.203)	(7.205)	(6.404)
PIN	0.017					0.077***				
	(1.176)					(2.822)				
APIN		0.020*					0.060***			
		(1.735)					(3.123)			
GPIN			-0.015					0.026		
			(-1.289)					(1.469)		
OWR				0.000					0.029	
				(0.009)					(0.538)	
BCL					-0.003					0.051**
					(-0.232)					(2.435)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Filing FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.0421	0.0424	0.0422	0.0419	0.0389	0.0557	0.0565	0.0536	0.0531	0.0533
Obs.	7,651	7,651	7,651	7,651	6,250	3,642	3,642	3,642	3,642	2,979

Table IA.20. ITI measures and other measures of informed trading. This table compares ITI measures to the conditional probability of informed trading obtained from several models: the PIN model (PIN); the adjusted PIN model of Duarte and Young (2009) (APIN); the generalized PIN model (GPIN) of Duarte et al. (2020); the Odders-White and Ready (2008) model (OWR); and the Back et al. (2018) model (BCL). ITI(patient) (ITI(impatient)) is the informed trading intensity estimated using the first 40 days (last 20 days) of the 60-day Schedule 13D filing window. ITI(insider) is trained on opportunistic insider trading data. ITI(short) is trained on short selling data. These datasets are described in Table IA.2. Indicators for days with opportunistic insider trades and spikes in short selling are regressed on the above measures and control variables. Control variables are winsorized at 0.05% and 99.95% and include effective spread, price impact, lambda, depth, realized volatility, turnover, order imbalance, absolute order imbalance, return, and an indicator variable that takes the value one for days with above-average number of trades (Duarte et al. (2020)). The sample consists of NYSE-listed stocks from 1994 to 2012. To be included, a stock must have a price greater than \$5 and a market capitalization greater than \$100 million at the end of the previous month. Standard errors are clustered by stock and the associated t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level.

Dep. Variable	Insider trade					Δ Short interest				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ITI(patient)	-0.000	-0.001	-0.000	-0.001	-0.015*	0.025*	0.025*	0.026*	0.025*	0.019
	(-0.048)	(-0.079)	(-0.034)	(-0.075)	(-1.744)	(1.798)	(1.750)	(1.834)	(1.803)	(1.297)
ITI(impatient)	0.083***	0.081***	0.084***	0.084***	0.084***	0.114***	0.109***	0.118***	0.118***	0.138***
	(9.773)	(9.520)	(9.841)	(9.867)	(8.612)	(6.497)	(6.157)	(6.740)	(6.737)	(7.441)
ITI(insider)	0.225***	0.223***	0.225***	0.225***	0.186***					
	(17.805)	(17.607)	(17.831)	(17.831)	(12.145)					
ITI(short)						0.317***	0.311***	0.320***	0.319***	0.330***
						(9.903)	(9.712)	(10.050)	(10.014)	(10.039)
PIN	0.003					0.011*				
	(0.982)					(1.888)				
APIN		0.010***					0.017***			
		(5.035)					(4.331)			
GPIN			0.004*					0.004		
			(1.702)					(1.173)		
OWR				0.006*					0.013*	
				(1.753)					(1.687)	
BCL					0.010***					0.008**
					(4.052)					(2.034)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.0050	0.0051	0.0050	0.0050	0.0041	0.0106	0.0108	0.0105	0.0105	0.0107
Obs.	265,154	265,154	265,154	265,154	199,109	76,241	76,241	76,241	76,241	71,243

Table IA.21. Descriptive statistics for return predictability. The table reports the mean, standard deviation, and 1st, 5th, 25th, 50th, 75th, 95th and 99th percentiles for the main set of variables for the stock-by-week sample in Section 5. We first report statistics for ITI measures. ITI(13D) is trained on Schedule 13D data. ITI(insider) is trained on opportunistic insider trading data. ITI(short) is trained on short selling data. These datasets are described in Table IA.2. ITI(patient) (ITI(impatient)) is trained on the first 40 days (last 20 days) of the 60-day Schedule 13D trading window. Other variables include log market capitalization, CAPM beta, last month return, two-to-six month return, idiosyncratic volatility, Amihud illiquidity, the effective bid-ask spread, and Kyle's lambda, and PIN. The PIN sample is from Brown and Hillegeist (2007) and covers 1993 to 2010. The main sample includes common stocks from 1/1993 to 7/2019. To be included, a stock must have a price greater than \$5 and a market capitalization greater than \$100 million at the end of the previous month.

Variable name	Mean	SD	1%	5%	25%	50%	75%	95%	99%	N
ITI(13D)	0.273	0.104	0.084	0.125	0.198	0.260	0.334	0.461	0.568	3,486,194
ITI(insider)	0.473	0.091	0.259	0.324	0.412	0.473	0.533	0.621	0.686	3,486,194
ITI(short)	0.406	0.045	0.310	0.342	0.376	0.402	0.433	0.485	0.527	3,486,194
ITI(patient)	0.204	0.092	0.051	0.081	0.139	0.190	0.253	0.371	0.484	3,486,194
ITI(impatient)	0.402	0.096	0.209	0.256	0.333	0.395	0.465	0.568	0.643	3,486,194
log(MCap)	6.968	1.536	4.656	4.928	5.772	6.723	7.904	9.890	11.309	3,486,194
Beta	1.035	0.470	0.026	0.355	0.756	1.000	1.276	1.818	2.389	3,486,194
Reversal	0.012	0.138	-0.355	-0.195	-0.054	0.009	0.073	0.223	0.433	3,486,194
Momentum	0.094	0.374	-0.569	-0.352	-0.089	0.054	0.211	0.631	1.388	3,486,194
Idio. Volatility	0.023	0.016	0.005	0.007	0.013	0.019	0.029	0.054	0.084	3,486,194
Illiquidity	0.031	0.186	0.000	0.000	0.001	0.004	0.019	0.117	0.407	3,486,194
ES %	0.005	0.006	0.000	0.000	0.001	0.003	0.006	0.016	0.026	3,486,194
Lambda	0.004	0.008	-0.009	-0.001	0.000	0.002	0.005	0.017	0.036	3,486,194
PIN	0.163	0.094	0.000	0.020	0.088	0.158	0.228	0.317	0.391	2,490,763

Table IA.22. Informed trading intensity and future returns. Returns are regressed on ITI (a weekly average), other predictors, and weekly fixed effects. Other predictors include log market capitalization, CAPM beta, last month return, two-to-six month return, idiosyncratic volatility, and Amihud illiquidity. The fourth and last columns also include measures of informed trading (weekly averages): PIN, the effective bid-ask spread, and Kyle's lambda. We control for a "high turnover" indicator, which equals one for days with above-average number of trades. The PIN sample is from Brown and Hillegeist (2007) and covers 1993 to 2010. The main sample includes common stocks from 1/1993 to 7/2019. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month. Future returns skip a day between signal and return. t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level. Standard errors are clustered by stock and week.

Return	Weekly	Weekly	Weekly	Monthly	Monthly	Monthly
ITI	0.0060***	0.0064***	0.0082***	0.0126***	0.0148***	0.0206***
	(7.85)	(7.88)	(8.32)	(6.82)	(7.33)	(8.37)
log(MCap)		-0.0002*	-0.0003*		-0.0010***	-0.0014***
		(-1.81)	(-2.18)		(-3.90)	(-4.31)
Beta		-0.0006	-0.01		-0.0024**	-0.003
		(-1.35)	(-1.78)		(-2.36)	(-2.96)
Reversal		-0.0054*	-0.0054*		0.0003	-0.001
		(-1.70)	(-1.62)		(0.04)	(-0.14)
Momentum		0.0016	0.0024		0.0048	0.0084**
		(0.94)	(1.27)		(1.51)	(2.22)
Idio. Volatility		-0.0316	-0.0263		-0.1990***	-0.1942**
		(-1.08)	(-0.76)		(-2.95)	(-2.40)
Illiquidity		-0.0003	-0.0014		-0.001	-0.0042
		(-0.75)	(-1.22)		(-1.06)	(-1.60)
Eff. Spread		0.0069	0.0239		-0.0894	-0.0673
		(0.18)	(0.49)		(-1.03)	(-0.60)
Lambda		-0.0442***	-0.0538***		-0.1384***	-0.1564***
		(-3.63)	(-3.86)		(-4.90)	(-4.90)
PIN			0.0005			0.0018
			(0.25)			(0.34)
R^2	0.010%	0.040%	0.050%	0.010%	0.120%	0.140%
Num. Obs.	3,484,919	3,484,919	2,338,672	3,484,923	3,484,923	2,338,675

Table IA.23. Portfolio sorts based on ITI measures: second month after portfolio formation. This table reports average monthly returns in the second month after portfolio formation. For each ITI measure, we sort stocks into decile portfolios based on their average ITI measures during the prior week. We compute equally-weighted average return during the second month (skipping the next month) for each decile and the top-minus-bottom difference. We report raw returns and consider separately five ITI measures. ITI(13D) is trained on Schedule 13D data. ITI(insider) is trained on opportunistic insider trading data. ITI(short) is trained on short selling data. These datasets are described in Table IA.2. ITI(patient) (ITI(impatient)) is trained on the first 40 days (last 20 days) of the 60-day Schedule 13D trading window. "0.0071" corresponds to 0.71% per month. The sample includes U.S. common stocks from 1/1993 to 7/2019. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million. t-statistic are reported in parentheses, where *, **, and *** denote significance at the 10%, 5%, and 1% level. Standard errors are computed using Newey-West adjustment with eight lags.

	Low	2	3	4	5	6	7	8	9	High	H-L
ITI(13D)	0.0093***	0.0091***	0.0093***	0.0095***	0.0093***	0.0093***	0.0094***	0.0093***	0.0090***	0.0088***	-0.0005
	(3.2)	(3.2)	(3.2)	(3.4)	(3.3)	(3.3)	(3.4)	(3.3)	(3.2)	(3.1)	(-0.6)
ITI(insider)	0.0089***	0.0091***	0.0094***	0.0090***	0.0095***	0.0096***	0.0091***	0.0093***	0.0097***	0.0091***	0.0001
	(3.2)	(3.2)	(3.3)	(3.2)	(3.4)	(3.4)	(3.2)	(3.3)	(3.4)	(3.0)	(0.1)
ITI(short)	0.0093***	0.0092***	0.0096***	0.0096***	0.0096***	0.0092***	0.0091***	0.0092***	0.0091***	0.0086***	-0.0007
	(3.6)	(3.4)	(3.4)	(3.4)	(3.4)	(3.2)	(3.2)	(3.2)	(3.1)	(2.1)	(-0.5)
ITI(patient)	0.0092***	0.0095***	0.0094***	0.0095***	0.0094***	0.0095***	0.0091***	0.0091***	0.0089***	0.0090***	-0.0002
	(3.2)	(3.3)	(3.3)	(3.3)	(3.3)	(3.3)	(3.2)	(3.2)	(3.1)	(3.2)	(-0.2)
ITI(impatient)	0.0086***	0.0091***	0.0097***	0.0091***	0.0097***	0.0095***	0.0094***	0.0092***	0.0091***	0.0091***	0.0004
	(3.0)	(3.2)	(3.3)	(3.2)	(3.4)	(3.4)	(3.3)	(3.3)	(3.2)	(3.2)	(0.5)

Table IA.24. Informed trading intensity predicts returns in subsamples. In the first row, we split the full sample into two equal parts ("High" and "Low") based on median stock's market capitalization. We then sort stocks into decile portfolios based on their average ITI during the prior week independently within each part. For each decile portfolio, we compute monthly alpha from a four factor Fama-French model with a momentum factor. To save space, we report the first and last decile portfolio and their difference."0.0085" corresponds to a 0.85% alpha per month. Subsequent rows repeat the analysis for other splitting variables including stock turnover, idiosyncratic volatility, the effective bid-ask spread, Kyle's lambda, and the PIN. The median is computed separately for each week. t-statistic are reported in parentheses. Standard errors are computed using Newey-West adjustment with eight lags. The sample includes common stocks from 1/1993 to 7/2019. To be included, a stock must have a price greater than $5 and a market capitalization greater than $100 million at the end of the previous month.

	Low, Variable<Median			High, Variable>Median			
Split Variable	1	10	10-1	1	10	10-1	H-L
Size	-0.0005	0.0079***	0.0085***	0.0002	0.0039***	0.0037***	-0.0048
	(-0.7)	(7.9)	(8.1)	(0.5)	(5.9)	(4.2)	
Turnover	-0.0003	0.0058***	0.0061***	0.0002	0.0051***	0.0050***	-0.0011
	(-0.4)	(6.6)	(7.0)	(0.2)	(6.6)	(4.9)	
Idio. Volatility	0.0028***	0.0065***	0.0038***	-0.0027***	0.0043***	0.0070***	0.0032
	(3.7)	(8.2)	(6.0)	(-3.3)	(4.7)	(6.9)	
Eff. Spread	0.0016**	0.0047***	0.0031***	-0.0012	0.0069***	0.0081***	0.0050
	(2.4)	(6.6)	(4.1)	(-1.5)	(6.4)	(7.4)	
Lambda	0.0008	0.0058***	0.0050***	-0.001	0.0053***	0.0062***	0.0012
	(1.4)	(8.5)	(5.8)	(-1.3)	(5.3)	(5.8)	
PIN	0.001	0.0067***	0.0057***	-0.0004	0.0052***	0.0056***	-0.0001
	(0.9)	(6.1)	(5.9)	(-0.6)	(6.2)	(6.1)	