

Internet Appendix to “Informed Trading Intensity”

VINCENT BOGOUSSLAVSKY, VYACHESLAV FOS, AND DMITRIY MURAVYEV*

This appendix provides additional results to supplement the main text.

- Section I discusses the link between price change reversal and expected/realized informed trading in the [Glosten and Milgrom \(1985\)](#) framework.
- Section II examines simulations of multiperiod informed trading models.
- Section III reports additional figures and tables.

I. 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

*Citation format: Bogousslavsky, Vincent, Vyacheslav Fos, and Dmitriy Muravyev, Internet Appendix to “Informed Trading Intensity,” *Journal of Finance*, DOI: 10.1111/jofi.13320. Please note: Wiley is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the authors of the article.

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.

II. Simulations of Informed Trading Models

We consider simulations of multiperiod 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 multiperiod 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 multiperiod 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:

$$\text{ITI}_t = a + b_1 V_t + b_2 |OI_t| + e_t, \quad \text{and} \quad (\text{IA.1})$$

$$\text{ITI}_t = c + d_1 V_t + d_2 |OI_t| + d_3 |r_t| + u_t, \quad (\text{IA.2})$$

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.1) and (IA.2) 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 VIII 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 [\(IA.1\)](#) and [\(IA.2\)](#). In Table VIII, 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.I](#), 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 [\(IA.1\)](#) and [\(IA.2\)](#). Hence, we only report simulation results for different T in Table ???. 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 [\(IA.2\)](#) 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 VIII (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 VIII (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.I. 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%. R^2 is the average adjusted R^2 across simulations. T denotes the number of trading periods.

	$I_t = a + b_1 V_t + b_2 OI_t + e_t$			$I_t = c + d_1 V_t + d_2 OI_t + d_3 r_t + u_t$			
	b_1	b_2	R^2	d_1	d_2	d_3	R^2
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%

III. Additional Figures and Tables

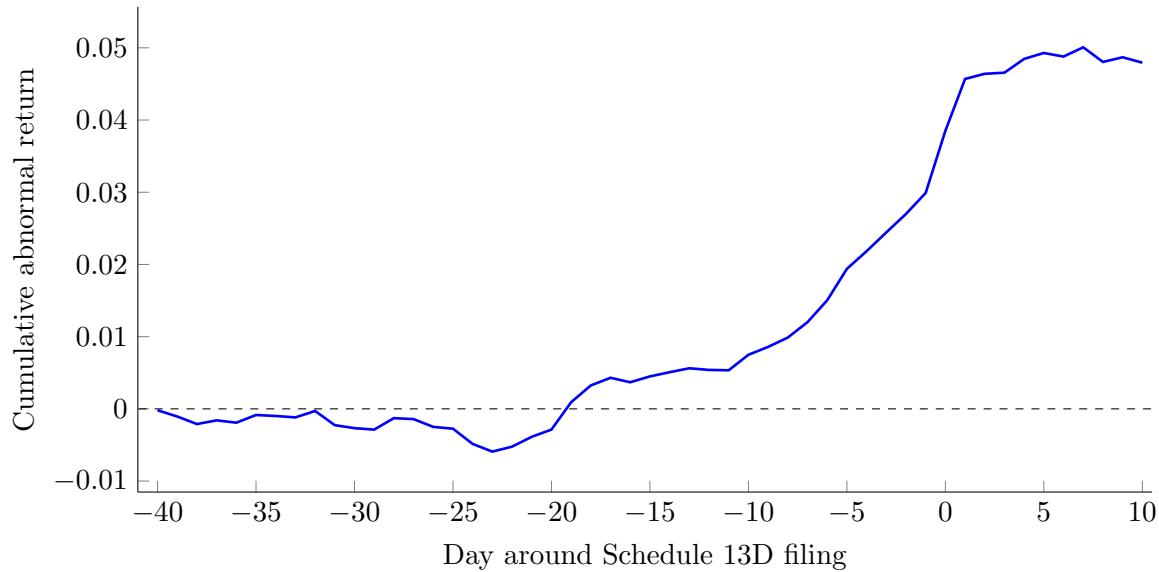


Figure IA.1. Cumulative abnormal return around Schedule 13D filing. The cumulative abnormal return is obtained by subtracting the cumulative market return from the cumulative stock return. To be included, a stock must have a price greater than \$5 and a market capitalization greater than \$100 million.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

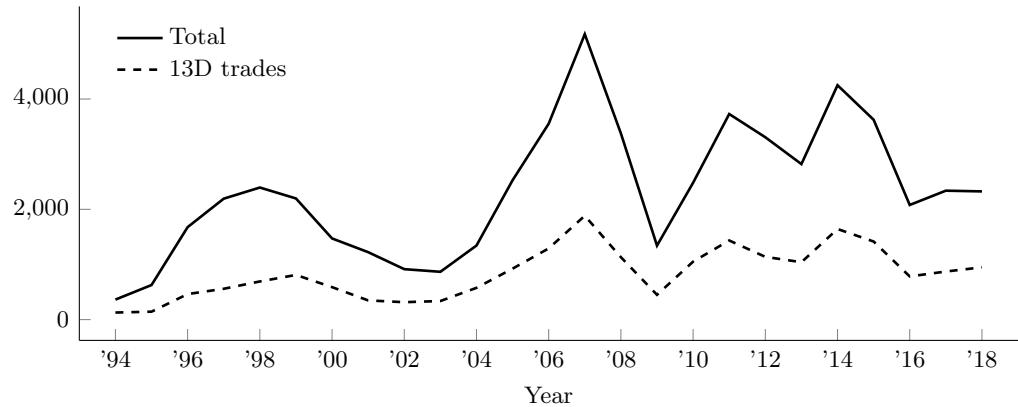
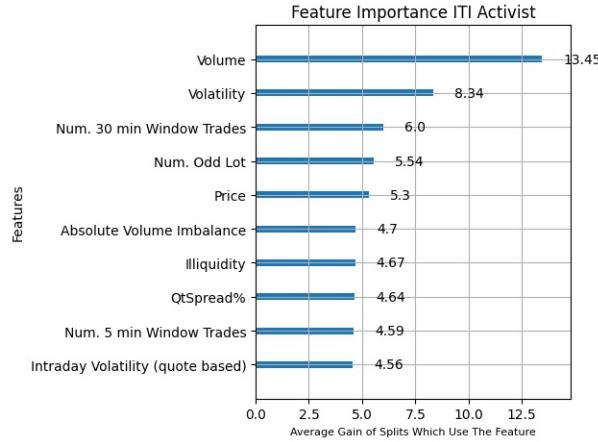


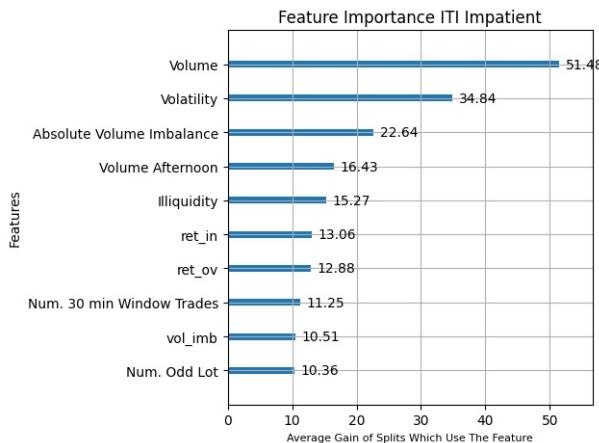
Figure IA.2. Total number of observations each year in the 13D sample.

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.

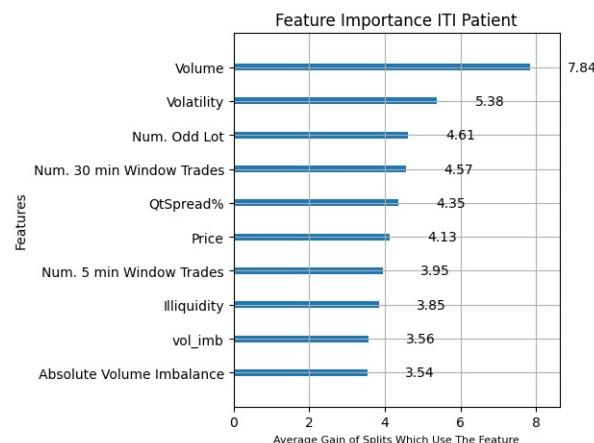
Panel A. ITI(13D)



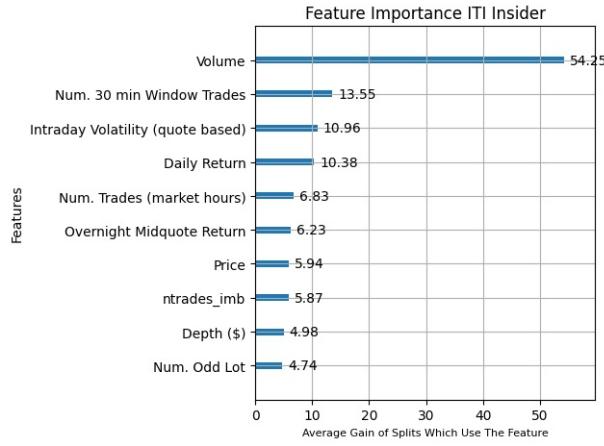
Panel B. ITI(impatient)



Panel C. ITI(patient)



Panel D. ITI(insider)

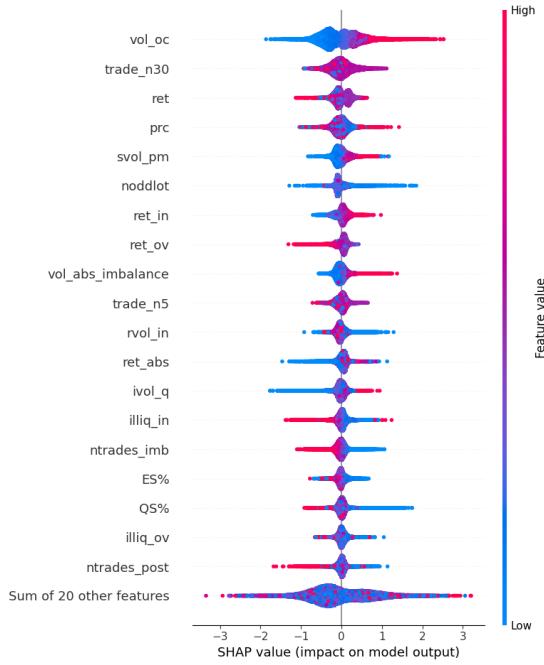


Panel E. ITI(short)

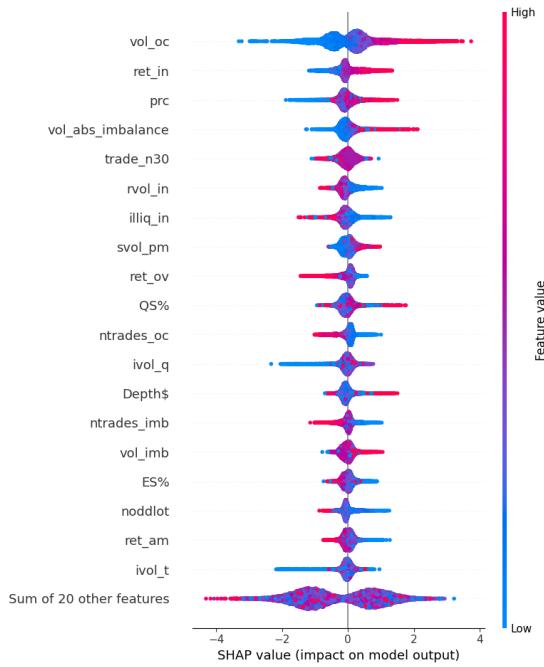


Figure IA.3. Features importance. This figure ranks features by their importance according to XGB's internal procedure that ranks features based on their gain (a default option based on the average gain across all splits where a feature was used). Variables are defined in Table IA.III.

Panel A. ITI(13D)



Panel B. ITI(impatient)



Panel C. ITI(patient)

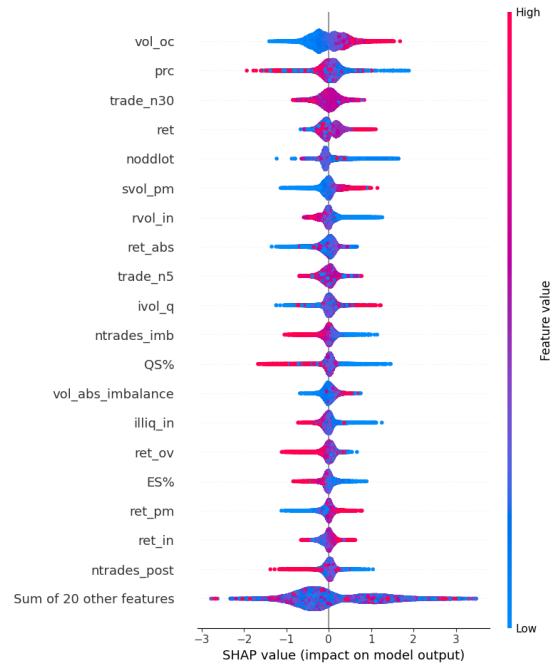


Figure IA.4. SHAP values. This figure ranks features by their importance according to SHapley Additive exPlanations (SHAP) values (Lundberg and Lee (2017)). Variables are defined in Table IA.III.

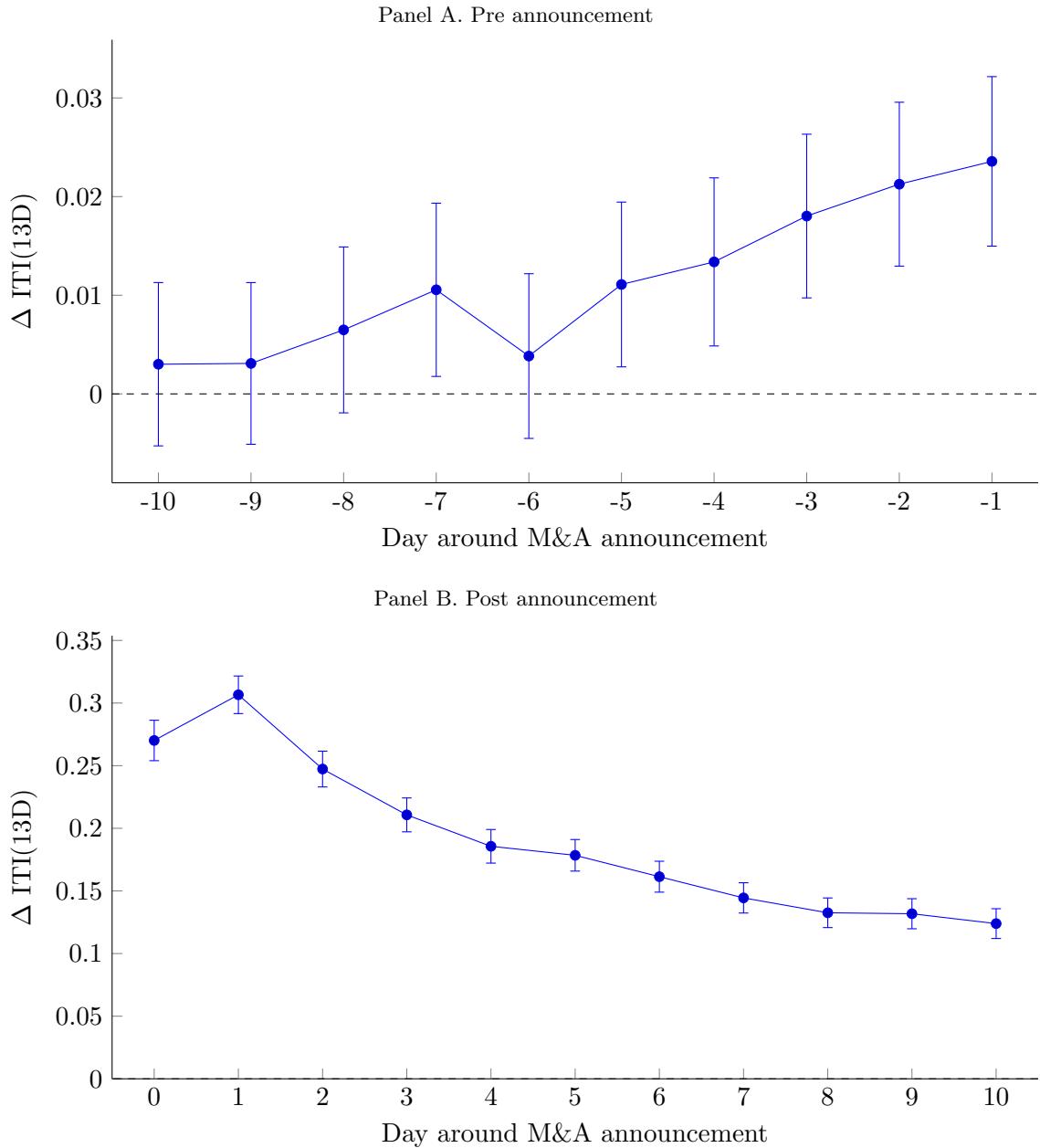
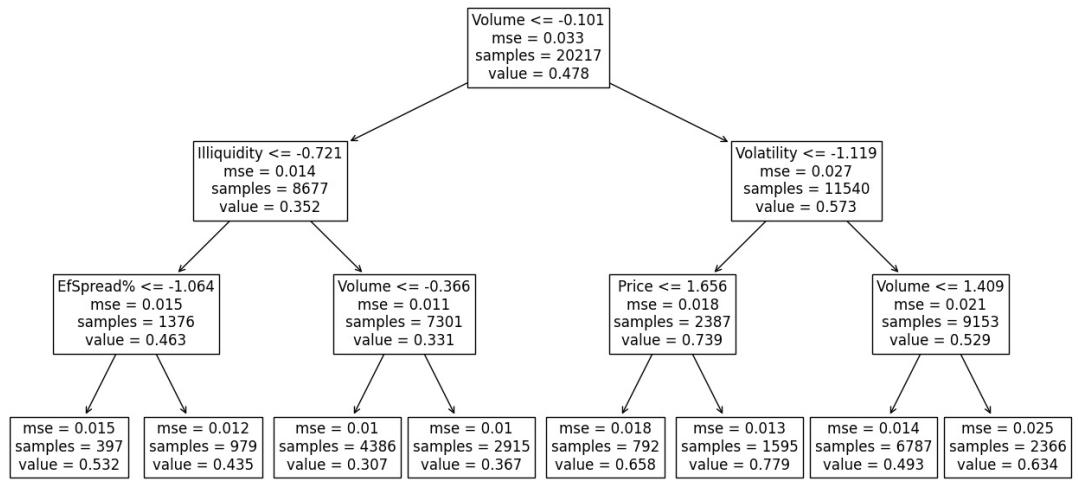
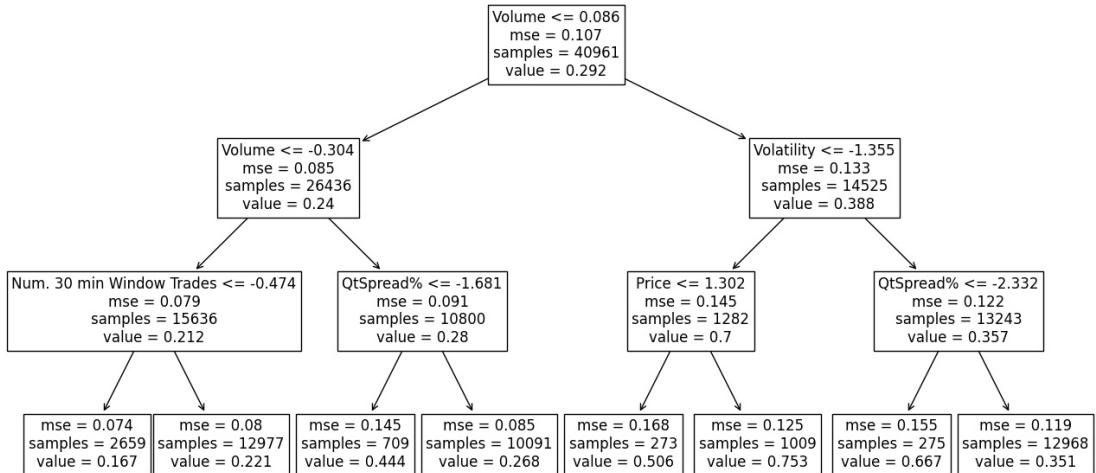


Figure IA.5. ITI(13D) before and after M&A announcements. Informed trading intensity is regressed on indicator variables for days before and after M&A announcements for targeted firms and stock fixed effects. The sample includes common stocks from January 1996 to December 2010, and the number of M&A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

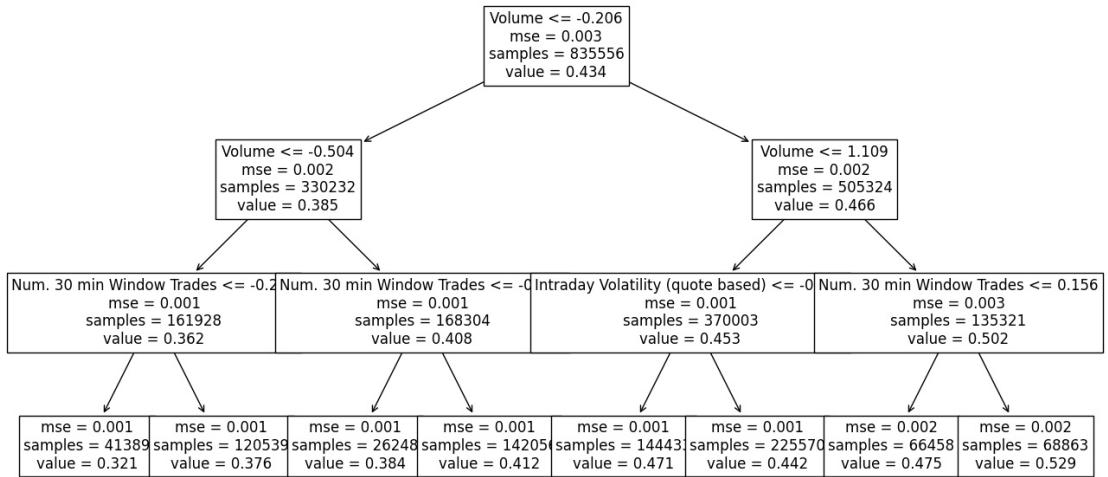
Panel A. ITI(impatient)



Panel B. ITI(patient)



Panel C. ITI(insider)



Panel D. ITI(short)

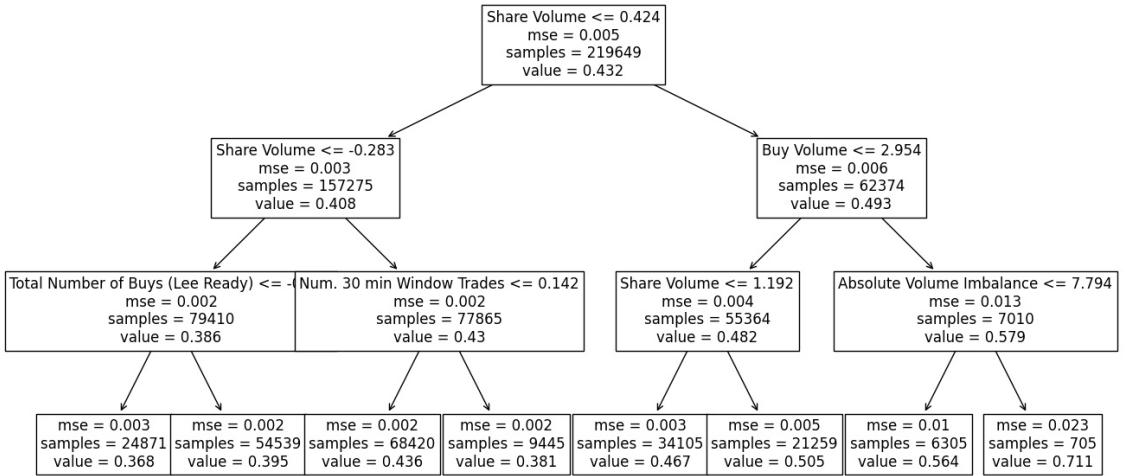


Figure IA.6. Surrogate trees. Variables are defined in Table IA.III.

Table IA.II. Overview of Data Sets Used in the Analysis

In all data sets, 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 January 1993 to December 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 July 2006 to December 2019. To create the final short data set, 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	Data set of return, liquidity, volume, and volatility variables listed in Table IA.IV , which covers U.S. common stocks over January 1993 to July 2019 excluding observations in the Schedule 13D data set.	16,823,151

Table IA.III. 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 data set.

Description	Data source	Abbrev.
<i>Return and price variables</i>		
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
<i>Liquidity variables</i>		
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
<i>Volatility and autocorrelation variables</i>		
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
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

Description (Table IA.III continued)	Data source	Abbrev.
<i>Volume and imbalance variables</i>		
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.IV. 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	$ES_k = \frac{2D_k(P_k - M_k)}{M_k}$, 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	$PI_k = \frac{2D_k(M_{k+5} - M_k)}{M_k}$, 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	$RS_k = \frac{2D_k(P_k - M_{k+5})}{M_k}$. 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: $\ln \Delta M_{i,\tau} = \alpha + \lambda (\text{signed} \sqrt{(\text{ShareVolume}_{i,\tau})})$, 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.V. 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.IV](#) 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 January 1993 to July 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.VI. 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.

Panel A. Regression

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

Panel B. Additional metrics

	LASSO (1)	RF (2)	XGB (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.VII. 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.III 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*** (40.739)				0.286*** (48.982)
ITI(return)		0.476*** (28.528)			0.212*** (35.316)
ITI(volatility)			0.481*** (30.715)		0.204*** (35.314)
ITI(volume)				0.645*** (73.998)	0.505*** (106.199)
Filing FE	Yes	Yes	Yes	Yes	Yes
Adj. <i>R</i> ²	0.1909	0.1135	0.1276	0.3498	0.4736
Obs.	58,197	58,197	58,197	58,197	58,197

Table IA.VIII. 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 January 1993 to July 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 associated *t*-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

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

Table IA.IX. 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 January 1993 to July 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 (-1.399)	-0.016*** (-5.673)	-0.007*** (-2.616)	-0.007** (-2.311)	-0.008*** (-3.013)	-0.008*** (-3.177)	-0.001 (-0.313)	-0.003 (-1.307)	-0.002 (-0.640)	0.003 (1.161)
ITI(13D)(t)	0.002*** (20.255)	0.001*** (9.127)	0.001*** (9.039)	0.001*** (6.537)	0.001*** (7.298)	0.000*** (4.756)	0.000*** (5.082)	0.000*** (4.369)	0.000*** (5.600)	0.000*** (4.088)
ret(t)*ITI(13D)(t)	0.028*** (5.189)	0.010** (2.569)	0.006 (1.537)	0.004 (1.125)	0.006 (1.569)	0.007** (2.165)	-0.001 (-0.278)	-0.001 (-0.164)	0.001 (0.490)	-0.003 (-0.887)
ret(t)*turn(t)	0.007 (0.434)	-0.002 (-0.196)	0.014 (1.586)	-0.010 (-1.235)	0.011 (1.465)	-0.005 (-0.870)	0.002 (0.442)	-0.000 (-0.001)	-0.000 (-0.039)	0.011* (1.901)
ret(t)*rvol(t)	-0.027 (-1.582)	-0.007 (-0.873)	-0.002 (-0.243)	0.015*** (3.199)	-0.001 (-0.124)	0.004 (0.849)	0.003 (0.732)	0.010* (1.864)	-0.001 (-0.214)	-0.005 (-0.875)
ret(t)*ES%(t)	-0.545*** (-3.718)	-0.498*** (-3.108)	-0.313*** (-3.092)	-0.138 (-1.525)	-0.119 (-1.421)	-0.073 (-0.902)	0.017 (0.235)	0.051 (0.742)	-0.003 (-0.030)	-0.132* (-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. <i>R</i> 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.X. 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*** (19.127)	0.002*** (10.571)		
ITI(impatient)			0.589*** (22.188)	0.006*** (9.661)
ES	-2.045*** (-2.764)	0.090*** (3.086)	0.932 (0.843)	0.110** (2.154)
lambda	-0.192 (-0.739)	-0.011*** (-4.156)	-0.885** (-2.160)	-0.028*** (-2.859)
depth	29.169*** (9.279)	-0.013 (-0.152)	8.210* (1.782)	-0.198* (-1.707)
rvol	-0.096 (-0.765)	-0.003 (-1.072)	-0.563*** (-3.462)	0.000 (0.033)
turn	0.709*** (5.396)	0.023*** (6.331)	0.799*** (4.682)	0.063*** (6.462)
OI	-0.510 (-0.932)	0.015 (0.665)	1.042** (2.104)	0.073** (2.358)
$ OI $	2.323*** (3.354)	0.092*** (3.331)	3.644*** (5.707)	0.149*** (3.794)
ret	-0.078 (-1.171)	-0.002 (-1.041)	0.054 (0.514)	-0.010*** (-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.XI. 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*** (28.565)	0.104*** (26.743)			0.073*** (17.439)	0.024*** (5.779)		
ITI(insider)		0.048*** (7.265)		0.027*** (4.044)		0.403*** (50.083)		0.381*** (47.344)
ITI(patient)			0.048*** (11.228)	0.047*** (10.962)			0.009** (2.014)	-0.006 (-1.235)
ITI(impatient)			0.112*** (22.063)	0.107*** (20.581)			0.144*** (25.651)	0.072*** (13.081)
Controls	Yes							
Stock FE	Yes							
Adj. <i>R</i> 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.XII. 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 January 1993 to July 2019. In Column (3)-(4), the sample includes common stocks from January 1993 to December 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 associated *t*-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

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

Table IA.XIII. 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 June 2006 to December 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*** (-4.053)	1.002*** (7.017)	-1.773*** (-2.915)
lambda	-1.186*** (-4.822)	-0.582*** (-8.712)	-0.566*** (-3.160)
depth	0.344*** (8.996)	0.003*** (7.733)	-0.311*** (-3.729)
rvol	-0.832*** (-6.149)	0.141*** (4.044)	0.137** (2.037)
turn	1.377*** (9.845)	0.854*** (15.694)	1.551*** (13.473)
OI	0.379 (0.748)	-0.346** (-2.050)	1.325*** (4.196)
$ OI $	6.732*** (10.534)	3.250*** (14.985)	5.346*** (14.618)
ret	-0.083 (-1.258)	0.366*** (20.149)	0.269*** (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.XIV. 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*** (62.052)		0.383*** (52.317)
ES%		1.002*** (7.017)	0.767*** (5.459)
lambda		-0.582*** (-8.712)	-0.389*** (-5.877)
depth		0.299*** (7.733)	0.156*** (4.217)
rvol		0.141*** (4.044)	0.113*** (3.295)
turn		0.854*** (15.694)	0.548*** (10.651)
OI		-0.346** (-2.050)	-0.363** (-2.204)
$ OI $		3.250*** (14.985)	2.050*** (9.631)
ret		0.366*** (20.149)	0.239*** (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.XV. 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 June 2006 to December 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*** (43.900)	0.494*** (29.173)	
ES%		-1.773*** (-2.915)	-1.615*** (-2.691)
lambda		-0.566*** (-3.160)	-0.333* (-1.866)
depth		-0.311*** (-3.729)	-0.159* (-1.945)
rvol		0.137** (2.037)	0.056 (0.870)
turn		1.551*** (13.473)	0.850*** (8.079)
OI		1.325*** (4.196)	0.617** (2.024)
$ OI $		5.346*** (14.618)	2.759*** (7.751)
ret		0.269*** (8.356)	0.143*** (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.XVI. 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.II. ITI measures are regressed on each other, a set of control variables, and stock fixed effects. The sample includes common stocks from January 1993 to July 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 associated t -statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Dep. variable:	ITI(13D) (1)	ITI(insider) (2)	ITI(short) (3)	ITI(short) (4)	ITI(short) (5)
ITI(insider)	0.076*** (78.320)			0.038*** (72.802)	0.027*** (41.057)
ITI(short)	0.608*** (141.716)	0.260*** (77.792)	0.205*** (44.730)		
ITI(13D)		0.088*** (73.485)		0.102*** (176.459)	
ITI(patient)			0.046*** (28.465)		0.023*** (34.067)
ITI(impatient)			0.117*** (43.289)		0.179*** (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.XVII. 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 January 1993 to July 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 associated *t*-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Panel 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*** (69.980)		2.692*** (52.152)	2.847*** (67.777)		2.039*** (50.820)	4.214*** (71.566)		2.750*** (53.103)
OI		18.855*** (56.946)	11.516*** (37.454)		13.195*** (53.351)	7.636*** (33.307)		21.343*** (74.168)	13.846*** (54.288)
Adj. <i>R</i> 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

Panel B. With absolute return

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

Table IA.XVIII. 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.

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

Table IA.XIX. 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 [Oders-White and Ready \(2008\)](#) model (OWR); and the [Back et al. \(2018b\)](#) 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** (2.561)	0.115** (2.570)	0.115** (2.568)	0.115** (2.565)	0.104** (2.132)	0.091 (1.608)	0.092 (1.614)	0.100* (1.753)	0.097* (1.701)	0.068 (1.085)
ITI(impatient)	0.335*** (6.521)	0.332*** (6.541)	0.343*** (6.778)	0.342*** (6.749)	0.343*** (6.105)	0.414*** (6.892)	0.419*** (7.060)	0.433*** (7.203)	0.433*** (7.205)	0.415*** (6.404)
PIN	0.017 (1.176)					0.077*** (2.822)				
APIN		0.020* (1.735)					0.060*** (3.123)			
GPIN			-0.015 (-1.289)					0.026 (1.469)		
OWR				0.000 (0.009)					0.029 (0.538)	
BCL					-0.003 (-0.232)					0.051** (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. <i>R</i> 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.XX. ITI Measures and Other Measures of Informed Trading (Insider and Short Samples)

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. \(2018b\)](#) 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.II](#). 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 (-0.048)	-0.001 (-0.079)	-0.000 (-0.034)	-0.001 (-0.075)	-0.015* (-1.744)	0.025* (1.798)	0.025* (1.750)	0.026* (1.834)	0.025* (1.803)	0.019 (1.297)
ITI(impatient)	0.083*** (9.773)	0.081*** (9.520)	0.084*** (9.841)	0.084*** (9.867)	0.084*** (8.612)	0.114*** (6.497)	0.109*** (6.157)	0.118*** (6.740)	0.118*** (6.737)	0.138*** (7.441)
ITI(insider)	0.225*** (17.805)	0.223*** (17.607)	0.225*** (17.831)	0.225*** (17.831)	0.186*** (12.145)					
ITI(short)						0.317*** (9.903)	0.311*** (9.712)	0.320*** (10.050)	0.319*** (10.014)	0.330*** (10.039)
PIN	0.003 (0.982)					0.011* (1.888)				
APIN		0.010*** (5.035)					0.017*** (4.331)			
GPIN			0.004* (1.702)					0.004 (1.173)		
OWR				0.006* (1.753)					0.013* (1.687)	
BCL					0.010*** (4.052)					0.008** (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. <i>R</i> 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.XXI. 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 V in the main text. 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.II. 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 January 1993 to July 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.XXII. 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 January 1993 to July 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*** (7.85)	0.0064*** (7.88)	0.0082*** (8.32)	0.0126*** (6.82)	0.0148*** (7.33)	0.0206*** (8.37)
log(MCap)		-0.0002* (-1.81)	-0.0003* (-2.18)		-0.0010*** (-3.90)	-0.0014*** (-4.31)
Beta		-0.0006 (-1.35)	-0.01 (-1.78)		-0.0024** (-2.36)	-0.003 (-2.96)
Reversal		-0.0054* (-1.70)	-0.0054* (-1.62)		0.0003 (0.04)	-0.001 (-0.14)
Momentum		0.0016 (0.94)	0.0024 (1.27)		0.0048 (1.51)	0.0084** (2.22)
Idio. Volatility		-0.0316 (-1.08)	-0.0263 (-0.76)		-0.1990*** (-2.95)	-0.1942** (-2.40)
Illiquidity		-0.0003 (-0.75)	-0.0014 (-1.22)		-0.001 (-1.06)	-0.0042 (-1.60)
Eff. Spread		0.0069 (0.18)	0.0239 (0.49)		-0.0894 (-1.03)	-0.0673 (-0.60)
Lambda		-0.0442*** (-3.63)	-0.0538*** (-3.86)		-0.1384*** (-4.90)	-0.1564*** (-4.90)
PIN			0.0005 (0.25)			0.0018 (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.XXIII. 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 January 1993 to July 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.

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

Table IA.XXIV. 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.II. 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 January 1993 to July 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*** (3.2)	0.0091*** (3.2)	0.0093*** (3.2)	0.0095*** (3.4)	0.0093*** (3.3)	0.0093*** (3.3)	0.0094*** (3.4)	0.0093*** (3.3)	0.0090*** (3.2)	0.0088*** (3.1)	-0.0005 (-0.6)
ITI(insider)	0.0089*** (3.2)	0.0091*** (3.2)	0.0094*** (3.3)	0.0090*** (3.2)	0.0095*** (3.4)	0.0096*** (3.4)	0.0091*** (3.2)	0.0093*** (3.3)	0.0097*** (3.4)	0.0091*** (3.0)	0.0001 (0.1)
ITI(short)	0.0093*** (3.6)	0.0092*** (3.4)	0.0096*** (3.4)	0.0096*** (3.4)	0.0096*** (3.4)	0.0092*** (3.2)	0.0091*** (3.2)	0.0092*** (3.2)	0.0091*** (3.1)	0.0086*** (2.1)	-0.0007 (-0.5)
ITI(patient)	0.0092*** (3.2)	0.0095*** (3.3)	0.0094*** (3.3)	0.0095*** (3.3)	0.0094*** (3.3)	0.0095*** (3.3)	0.0091*** (3.2)	0.0091*** (3.2)	0.0089*** (3.1)	0.0090*** (3.2)	-0.0002 (-0.2)
ITI(impatient)	0.0086*** (3.0)	0.0091*** (3.2)	0.0097*** (3.3)	0.0091*** (3.2)	0.0097*** (3.4)	0.0095*** (3.4)	0.0094*** (3.3)	0.0092*** (3.3)	0.0091*** (3.2)	0.0091*** (3.2)	0.0004 (0.5)

REFERENCES

- Back, Kerry, Kevin Crotty, and Tao Li, 2018b, Identifying information asymmetry in securities markets, *Review of Financial Studies* 31, 2277–2325.
- Bogousslavsky, Vincent, 2021, The cross-section of intraday and overnight returns, *Journal of Financial Economics* 141, 172–194.
- Brown, Stephen, and Stephen A. Hillegeist, 2007, How disclosure quality affects the level of information asymmetry, *Review of Accounting Studies* 12, 443–477.
- Cohen, Lauren, Christopher Malloy, and Lukasz Pomorski, 2012, Decoding inside information, *Journal of Finance* 67, 1009–1043.
- Collin-Dufresne, Pierre, and Vyacheslav Fos, 2015, Do prices reveal the presence of informed trading? *Journal of Finance* 70, 1555–1582.
- Collin-Dufresne, Pierre, and Vyacheslav Fos, 2016, Insider trading, stochastic liquidity and equilibrium prices, *Econometrica* 84, 1441–1475.
- Duarte, Jefferson, Edwin Hu, and Lance Young, 2020, A comparison of some structural models of private information arrival, *Journal of Financial Economics* 135, 795–815.
- Duarte, Jefferson, and Lance Young, 2009, Why is PIN priced? *Journal of Financial Economics* 91, 119–138.
- Easley, David, and Maureen O’Hara, 1992, Time and the Process of Security Price Adjustment, *Journal of Finance* 47, 577–605.
- Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.
- Holden, Craig W., and Stacey Jacobsen, 2014, Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions, *Journal of Finance* 69, 1747–1785.
- Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1335.

Lee, Charles, and Mark J. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733–746.

Lundberg, Scott M., and Su-In Lee, 2017, A unified approach to interpreting model predictions, *Advances in Neural Information Processing Systems* 30.

Odders-White, Elizabeth R., and Mark J. Ready, 2008, The probability and magnitude of information events, *Journal of Financial Economics* 87, 227–248.