Asymmetric Information and the Distribution of Trading Volume *

Matthijs Lof† Jos van Bommel‡

March 19, 2021

Abstract

We propose the Volume Coefficient of Variation (VCV), the ratio of the standard deviation to the mean of trading volume, as a new and easily computable measure of information asymmetry in security markets. We use a microstructure model to demonstrate that VCV is strictly increasing in the proportion of informed trade. Empirically, we find that VCV, computed from daily observations of trading volume, is correlated with extant firm-level measures of asymmetric information in the cross-section of US stocks. Moreover, VCV increases following exogenous reductions in analyst coverage induced by brokerage closures, and steeply decreases around earnings announcements and other information disclosures.

Keywords: VCV, Trading volume, Informed trading.
JEL classification: D82, G12, G14

*We thank Anna Bayona, Alex Boulatov, Darwin Choi, Tarun Chordia, Jonathan Cohn, David Easley, Eliezer Fich, Carlos Forner Rodríguez, Thierry Foucault, Bruce Grundy, Sam Hartzmark, Henry Jarva, Dirk Jenter, Pete Kyle, Albert Menkveld, Anna Obizhaeva, Deniz Okat, Conall O'Sullivan, Paolo Pasquariello, Jillian Popadak, Jorg Prokop, David Reiffen, Dagfinn Rime, Francesco Sangiorgi, Enrique Sentana, Andriy Shkilko, Valeri Sokolovski, Sorin Sorescu, Erik Theissen, Ian Tonks, Hugo van Buggenum, Michael Weber, Christian Westheide and participants at the ABFER, CEPR and CUHK Annual Symposium (Hong Kong), the American Economic Association (Chicago), the Eastern Finance Association (Philadelphia), the EBC network, the French Finance Association, the Future of Financial Information conference, the German Finance Association, the SAFE Market Microstructure Conference, the SFM conference (Kaohsiung, Taiwan), the Spanish Finance Association, Aalto University, the Bank of England, the Bank of Finland, Erasmus University Rotterdam, the Higher School of Economics, Hong Kong Baptist University, University of Amsterdam, University of Liverpool, University of Maastricht, University of Toulouse, University of Turku, and VU Amsterdam for constructive comments. We thank Fabian Brunner for help with the EDGAR data.

†Aalto University School of Business. P.O. BOX 21210, 00076 Aalto Finland. E-mail: matthijs.lof@aalto.fi. Phone: +358504014448.

‡University of Luxembourg. Campus Limpertsberg, Université du Luxembourg 162 A, avenue de la Faïencerie, L-1511 Luxembourg. E-mail: jos.vanbommel@uni.lu. Phone: +3524666446825

Asymmetric Information and the Distribution of Trading Volume

March 19, 2021

Abstract

We propose the Volume Coefficient of Variation (VCV), the ratio of the standard deviation to the mean of trading volume, as a new and easily computable measure of information asymmetry in security markets. We use a microstructure model to demonstrate that VCV is strictly increasing in the proportion of informed trade. Empirically, we find that VCV, computed from daily observations of trading volume, is correlated with extant firm-level measures of asymmetric information in the cross-section of US stocks. Moreover, VCV increases following exogenous reductions in analyst coverage induced by brokerage closures, and steeply decreases around earnings announcements and other information disclosures.

Keywords: VCV, Trading volume, Informed trading.
JEL classification: D82, G12, G14

1 Introduction

In this paper, we propose a novel, intuitive, versatile, and easy-to-compute measure of information asymmetry in security markets that is based on trading volume only. We provide theoretical and empirical evidence that the distribution of trading volume depends on the level of information asymmetry. Specifically, the ratio of the standard deviation to the mean of trading volume, denoted Volume Coefficient of Variation (VCV), increases monotonically in the proportion of informed trade. VCV is easy to compute and only requires observable trading volume. VCV can be computed from time-series of volume, to compare information asymmetry across assets, as well as from cross-sections, to analyze information asymmetry over time.[1]

We use a canonical market microstructure model, based on Kyle (1985), to derive the mean and standard deviation of trading volume as a function of the number of market participants, their trading intensity, and the proportion of informed trade. We show that both the mean and standard deviation of volume increase linearly in the proportion of informed trade, but that the standard deviation does so at a higher rate. The coefficient of variation of trading volume (VCV) therefore increases monotonically in the proportion of informed trade, while it is asymptotically independent of the number of market participants and their trading intensity. The intuition behind our measure is that the distribution of trading volume depends on the correlation of individual orders. When traders are uninformed and have uncorrelated liquidity needs, most orders will be netted out against each other, so that the order imbalance is relatively low compared to the observed trading volume, which in this case follows a Normal-like distribution. The correlated liquidity demand from informed traders on the other hand leads to increased trading of liquidity providers and a more skewed and dispersed distribution of trading volume.

Information asymmetry is a key concept in the accounting, finance, and economics liter-

[1] To the best of our knowledge, we are the first to relate the coefficient of variation of trading volume to asymmetric information. Chordia et al. (2001) use the coefficient of variation of trading volume when examining the relation between stock returns and the variability of trading volume, without relating this measure to asymmetric information.

ature. Within the context of capital markets, information asymmetry affects liquidity, asset prices, and financing and investment decisions. It is determined by many factors including corporate governance, disclosure standards and regulations, and information intermediation by third-parties (See Beyer et al., 2010, for a survey). Since information asymmetry is not directly observable, numerous proxies have been proposed in the literature.

Early papers use the bid-ask spread to measure information asymmetry (e.g. Amihud and Mendelson, 1986; Welker, 1995; Healey et al., 1999; Bushee and Leuz, 2005; Heflin et al., 2005) even though bid-ask spreads indicate not only asymmetric information, but also other determinants of illiquidity including risk, inventory costs and transactions costs (Huang and Stoll, 1997). Easley et al. (1996) develop a measure for the probability of informed trading, the well-known PIN measure, which is estimated from transaction-level data and requires trades to be classified as either buyer- or seller-initiated.[2] Also other recent information asymmetry measures, such as VPIN (Easley et al., 2012), order flow volatility (Chordia et al., 2019) and XPIN (Bongaerts et al., 2016), rely on signed transaction-level data. It has been recognized that such order classification, e.g. using the Lee and Ready (1991) algorithm, is not error-free and has become increasingly problematic in recent years due to continuous trading (e.g. Boehmer et al., 2007; Easley et al., 2012; Johnson and So, 2018). An alternative measure of asymmetric information is the multimarket information asymmetry (MIA) measure of Johnson and So (2018), which is based on the relative daily trading volumes in options and stocks, building on the premise that informed investors are more likely than uninformed investors to trade in options. The C2 measure by Llorente et al. (2002) estimates the effect of trading volume on return autocorrelation, following the intuition that price changes induced by uninformed trades are more likely to revert. Yang et al. (2020) suggest abnormal idiosyncratic volatility (AIV) prior to earnings announcements as an indicator of information asymmetry.

[2]The PIN measure has been widely used to study information asymmetries in security markets. Easley et al. (1997a,b) analyze the information content around trade lags and trade size. Applications of PIN include, i.a., accounting conservatism (Lafond and Watts, 2008), analyst coverage (Easley et al., 1998), disclosure quality (Vega, 2006; Brown and Hillegeist, 2007), earnings surprises (Brown et al., 2009), institutional ownership (Brockman and Yan, 2009; Boone and White, 2015), the pricing of information asymmetry (Easley et al., 2002, 2010; Mohanram and Rajgopal, 2009; Hwang et al., 2013), and trader anonymity (Gramming et al., 2001).

Our measure does not require quotes, prices, option volumes nor transaction classification. VCV estimates can be computed during monthly, daily or intraday intervals for any security for which trading volume is observable, and is therefore applicable to a much broader set of assets than incumbent proxies of information asymmetry.

To assess the power and robustness of VCV, we conduct a comprehensive Monte Carlo analysis. We find that VCV can be estimated effectively from relatively small samples even though the small-sample estimates are inevitably less precise. In addition to robustness to sample size, our Monte Carlo analysis examines various modifications of our theoretical model including stochastic variation in trading activity, the proportion of informed trade, and market maker capacity. For all these specifications we find a strictly positive relation between VCV and the average proportion of informed trade, confirming that VCV detects information asymmetry under very general conditions.

For our empirical analyses, we compute firm-level observations of VCV from daily volumes obtained from CRSP, for stocks listed on NYSE, AMEX and NASDAQ, from 1980 until 2020. We use three distinct volume measures: (*i*) trading volume in dollars, (*ii*) turnover, and (*iii*) volume market shares (dollar volume as a fraction of total market dollar volume). The coefficients of variation of these measures are nearly identical, implying that VCV is not sensitive to aggregate market-level variation in trading volume. This is important, since it is well known that other factors besides idiosyncratic firm-level information can drive variation in trading activity, such as sentiment (Kumar and Lee, 2006), or common liquidity shocks (Admati and Pfleiderer, 1988; Brogaard et al., 2018). We compare VCV to annual firm-level characteristics and other information asymmetry measures. We find that VCV is higher for smaller and younger firms that have less analyst following, lower trading turnover, higher bid-ask spreads, and are more volatile and less liquid. VCV is positively correlated with incumbent proxies of firm-level information asymmetry including PIN, MIA and C2.

We also find that, controlling for Amihud (2002) illiquidity, short-term return reversals are weaker for high-VCV stocks, consistent with the hypothesis that informed trading is predictive of future price changes (Llorente et al., 2002), demonstrating that VCV measures information asymmetry, rather than general illiquidity. Duarte and Young (2009) argue that (unadjusted) PIN is not only measuring informed trade, but also general illiquidity unrelated to information asymmetry. They derive a new measure of general illiquidity unrelated to informed trading: PSOS (Probability of Systematic Order-flow Shock), as well as a measure called Adjusted PIN, which measures asymmetric information net of unrelated illiquidity effects. We find that VCV is strongly related to Adjusted PIN, while the relationship to PSOS is weak, giving further evidence that VCV is a measure of informed trading, rather than general illiquidity.

Consistent with recent studies documenting a positive impact of institutional ownership on the firm's information environment (Boone and White, 2015; Bai et al., 2016), we find that firms with more institutional shareholders (i.e. high breadth of ownership) have on average lower VCVs. We specifically look at two types of institutional investors that can be considered relatively informed about a firm: monitoring investors, defined as those institutional investors for which the firm represents a significant allocation of the institution's portfolio (Fich et al., 2015), and dedicated investors, defined as institutional investors that predominantly make long-term investments in a selective set of stocks (Bushee and Noe, 2000; Bushee, 2001). We find that, controlling for breadth of ownership, VCV is higher for firms with monitoring and dedicated (i.e. informed) investors.

To identify exogenous changes in information asymmetry, we exploit terminations in analyst coverage induced by brokerage closures, similar to Kelly and Ljungqvist (2012), Derrien and Kecskes (2013), Li and You (2015), Bushman et al. (2017), and Chen and Lin (2017). As emphasized by Beyer et al. (2010) and Blankespoor et al. (2018), sell-side analysts play a key role as information intermediaries in the corporate information environment. Exogenous terminations in analyst coverage disrupt this information environment and are expected to increase information asymmetries. We find that the VCV of affected firms significantly increases following these events.

In addition to computing VCV at the firm-level from time-series of volumes, we also compute VCV from cross-sections of volume observations in event time, for a large sample of firms around their quarterly earnings announcement dates.A large literature starting with Ball and Brown (1968) and Beaver (1968) considers the information content of earnings announcements, widely recognizing that information asymmetries are resolved around these events. Consistent with this view, we find that the cross-sectional VCV is relatively high prior to announcements, and drops significantly in the days following the announcement. This suggests that information asymmetries build up and discourage uninformed traders to trade just before earnings announcements (See Milgrom and Stokey, 1982; Black, 1986; Wang, 1994; and Chae, 2005), while the market is more attractive for uninformed traders after these information events. We also look at the cross-sectional VCV around unscheduled form 8-K disclosures of major corporate events. We find a significant decrease in VCV already in the days prior to the filing date, because form 8-K is typically filed up to four dates after the event (Ben-Rephael et al., 2017).

We consider different time-series windows for estimating VCV: when comparing VCV to annual firm-level characteristics and other information asymmetry measures, we estimate VCV annually, from one year of daily volume observations. When we consider short-term reversals we estimate VCV on a monthly basis, and we estimate VCV quarterly when analyzing quarterly 13F filings of institutional ownership. We find that these estimates of VCV are highly correlated, but different in level. On average, annual estimates of VCVs are higher because they capture seasonal volume variation. We conclude that VCV is a very flexible measure, and that the estimation window and sampling frequency is at the discretion of the researcher, depending on the context and the availability of data. However, when comparing VCVs of different assets cross-sectionally, it is crucial that these are estimated from equal estimation windows at equal sampling frequencies, to avoid making misleading inferences. We also examine subsamples of NYSE/AMEX and NASDAQ stocks, as well as pre- and post-2000 periods, further validating the robustness of VCV as a measure of information asymmetry across different market environments.

The remainder of this paper is organized as follows: In the next section we present our model and show that the volume coefficient of variation emerges as a natural measure of the proportion of informed trade. Section 3 contains a comprehensive Monte Carlo analysis to test the robustness of VCV across different environments. Section 4 presents our empirical analysis where we show how stock-level VCVs, computed from time series of volume observations are related to firm characteristics, incumbent asymmetric information measures, institutional ownership, and analyst coverage. Section 5 illustrates how our measure taken from cross-sections of volume observations can be used to gauge asymmetric information in event time, by documenting VCV around earnings announcements and major corporate events. Section 6 concludes.

2 Theory

To analyze the distribution of trading volume, we present a simple model in which we postulate M individual liquidity seekers, who each submit Normally distributed orders with mean zero and standard deviation σ, and where competitive liquidity providers (market makers) absorb the order imbalance. A proportion η of the M liquidity seekers is informed, with ηM being an integer. We refer to σ as *trading intensity*. For ease of exposition, we first assume η to be exogenous. At the end of this section, we demonstrate that the results hold when the proportion of informed trade η realizes endogenously, e.g. when informed investors choose their trading intensity strategically.

We denote the individual demands of all (informed and uninformed) liquidity seekers y_i, for which positive values indicate buy orders and negative values indicate sell orders. The order imbalance (net order flow) is the sum of all orders, $\sum_M y_i$, which is taken up by the liquidity providers who determine the price. This imbalance is typically not publicly observable. Total trading volume can then be written as:

$$V = \tfrac{1}{2}\left(\sum_M |y_i| + \left|\sum_M y_i\right|\right). \tag{1}$$

The term inside brackets is the "double-counted transaction volume", counting both buys and sells, of the liquidity seekers (the first term) and the liquidity providers (the second term). This double-counted volume includes the trades among liquidity seekers, as well as the trades between the liquidity providers and unmatched liquidity seekers.[3]

As an example, consider five liquidity seekers whose demands are -1, 2, 2, -2, 1. The order imbalance is two, meaning that the liquidity providers end up selling two units. The observed trading volume is five: we have three units sold by liquidity seekers, five units bought by liquidity seekers and two units sold by liquidity providers. The double-counted volume is thus ten, and the commonly recorded single-counted volume is half this number.

The orders of the informed liquidity seekers are perfectly correlated, so that all ηM informed traders submit identical orders. On the other hand, the demands of the $(1-\eta)\,M$ uninformed liquidity seekers are uncorrelated (*i.i.d.*). Following these assumptions, the order imbalance follows a Normal distribution around zero, as in Kyle (1985):

$$\sum_M y_i \sim N\left(0, \sigma^2\left(\eta^2 M^2 + (1-\eta)\,M\right)\right). \tag{2}$$

The variance of the order imbalance is a nonlinear function of η, due to the different correlations of informed and uninformed demand. If most liquidity seekers are uninformed, their orders will be mostly matched to each other and the order imbalance is expected to be relatively low. When many traders are informed, their correlated demands can lead to large imbalances. As a result, the standard deviation of the order imbalance is increasing in the proportion of informed trade η.

We now derive the first two moments of the total trading volume (Eq.1) as a function of η. Using the properties of the Half Normal distribution we find:[4]

$$\begin{aligned} E[V] &= \tfrac{1}{2}\left(E\left[\textstyle\sum_M |y_i|\right] + E\left[\left|\textstyle\sum_M y_i\right|\right]\right) \\ &= \tfrac{\sigma M}{\sqrt{2\pi}}\left(1 + \sqrt{\eta^2 + (1-\eta)M^{-1}}\right). \end{aligned} \tag{3}$$

[3] This expression for trading volume is similar to that in Admati and Pfleiderer (1988) and Grundy and McNichols (1989).

[4] If $x \sim N\left(0, \sigma^2\right)$, then $|x|$ follows a *Half Normal* distribution with $E\left(|x|\right) = \frac{\sigma\sqrt{2}}{\sqrt{\pi}}$ and $Var\left(|x|\right) = \sigma^2\left(1 - \frac{2}{\pi}\right)$.

From this we see that as the number of market participants M increases, expected trading volume per capita converges to an increasing linear function of the proportion of informed trade η:

$$\lim_{M\to\infty} E\left[\frac{V}{M}\right] = \frac{\sigma}{\sqrt{2\pi}}\left(1+\eta\right). \tag{4}$$

To analyze the variance of the observed trading volume, we consider each of the three components of the double-counted volume that can be attributed to *(i)* informed liquidity seekers $\left(\sum_{1\ldots\eta M}|y_i|\right)$, *(ii)* uninformed liquidity seekers $\left(\sum_{\eta M+1\ldots M}|y_i|\right)$, and *(iii)* liquidity providers $(|\sum_M y_i|)$. The variances and covariances of these three components are derived in Internet Appendix A. For large M, we find that the variance of of the per capita trading volume increases in η^2:

$$\lim_{M\to\infty} Var\left(\frac{V}{M}\right) = \sigma^2\left(1-\frac{2}{\pi}\right)\eta^2. \tag{5}$$

We thus see that for large M, the ratio of the standard deviation to the mean (the coefficient of variation) of trading volume is strictly increasing in η and is independent of the number of market participants M and their trading intensity σ.

<u>Proposition 1</u>

Consider a market where M liquidity seeking traders submit Normally distributed market orders with mean zero and standard deviation σ, and where the order imbalance is absorbed by liquidity suppliers. If ηM of the M liquidity seeking traders are informed:

i. *The coefficient of variation of observed trading volume increases monotonically in the proportion of informed traders, η.*

ii. *For large M, the relationship converges to:*

$$\lim_{M\to\infty}\frac{\sigma_V}{\mu_V} = \sqrt{2\pi-4}\,\frac{\eta}{\eta+1}, \tag{6}$$

where μ_V and σ_V denote the expected value and standard deviation of trading volume V.

Corollary

If $\hat{\mu}_V$ and $\hat{\sigma}_V$ denote the sample average and standard deviation of a sample of trading volumes generated by trading sessions with parameters $\{\sigma, M, \eta\}$,

$$VCV \equiv \frac{\hat{\sigma}_V}{\hat{\mu}_V} \tag{7}$$

is a consistent estimator of $\frac{\sigma_V}{\mu_V}$.

The Volume Coefficient of Variation (VCV) is a measure of informed trade. $E[VCV]$ increases monotonically in η.

Our finding that VCV is asymptotically independent of σ and M is important. It means that even when σ and M are subject to exogenous variation, e.g. due to sentiment (Kumar and Lee, 2006), or correlated liquidity shocks (Admati and Pfleiderer, 1988; Brogaard et al., 2018), VCV will increase in the proportion of informed trade. We present simulations and empirical analyses in the next sections that strongly support this result.

The above analysis also shows that a direct estimator of the proportion of informed trade is implied from Eq.(6):

$$\hat{\eta} \equiv \frac{\hat{\sigma}_V}{\hat{\mu}_V\sqrt{2\pi - 4} - \hat{\sigma}_V}. \tag{8}$$

However, as our simulation results in Section 3 show, $\hat{\eta}$ is a consistent estimator of η only when demand is Normally distributed, M is large, and η is constant across observations. We find that $\hat{\eta}$ behaves particularly poorly in small samples or when we relax the assumptions of the model, primarily because its denominator can be close to zero or turn negative. On the other hand, our simulations indicate that VCV is well-behaved and increases monotonically in η under general conditions, including non-Normality and time-varying proportions of informed trade.

The earlier assumption that informed and uninformed liquidity seekers have equal trading intensity is for convenience only and without loss of generality: the distribution of trading volume would be identical if we define η and M as:

$$\eta = \frac{\sigma_m m}{\sigma_m m + \sigma_u n}; \quad M = n + \frac{\sigma_m}{\sigma_u} m, \tag{9}$$

where m and n refer to the number of informed and uninformed liquidity seekers, respectively, while σ_m and σ_n denote the trading intensity of informed and uninformed liquidity seekers. That is, M is a measure of total *liquidity demand*, in which the number of individual traders are weighted by their trading intensities, while η is the proportion of informed *trade*, rather than the proportion of informed *traders*.

In Internet Appendix section A.1, we present an extended version of our model in which the informed liquidity seekers choose their trading intensity σ_m strategically, by taking into account the strategies of the other informed traders. The results demonstrate that the relation between the *endogenized* proportion of informed trade and the coefficient of variation of trading volume is equivalent to Proposition 1.

Further enriching the model with risk aversion, long lived information, or strategic uninformed trading (as in Admati and Pfleiderer, 1988, or Foster and Viswathan, 1990) will not change Proposition 1, as there will always be a equilibrium proportion of informed trade η and a weighted number of liquidity seekers M. The intuition that correlated orders will lead to larger trading volumes than uncorrelated orders, and thus to a more dispersed distribution, also pertains in dynamic and continuous markets: liquidity providers will accumulate more inventory (positive or negative) when confronted with correlated liquidity seeking demand. As they find themselves on the wrong side of the market, they will increase their efforts to unwind these inventories, resulting in excessive sequential 'hot potato' transactions and 'intermediation chains' (see e.g. Jovanovic and Menkveld, 2016; Glode and Opp, 2016; and Rosu, 2019).[5] Our simulation analysis in the next section provides further evidence that our proposition that VCV increases in the proportion of informed trade holds for many different volume-generating models.

[5] The link between volume volatility and asymmetric information has likely increased with the advent of high frequency trading (see, e.g. Easley et al., 2011; Hendershott et al., 2011; and Kirilenko et al., 2017).

3 Simulations

In this section, we analyze the distribution of trading volume generated by our model, for different values of η (proportion of informed trade) and M (equivalent number of liquidity seekers). To do this, we draw $1 + (1 - \eta)M$ random observations from the Standard Normal distribution to simulate the individual demands (i.e. we assume $\sigma = 1$). The first observation is multiplied by ηM, and represents the aggregate informed demand. The remaining observations represent the individual uninformed demands. We compute the observed trading volume V from Eq.(1). For each (M, η) pair, we generate a sample of T volume (V) observations, from which we compute the coefficient of variation VCV.

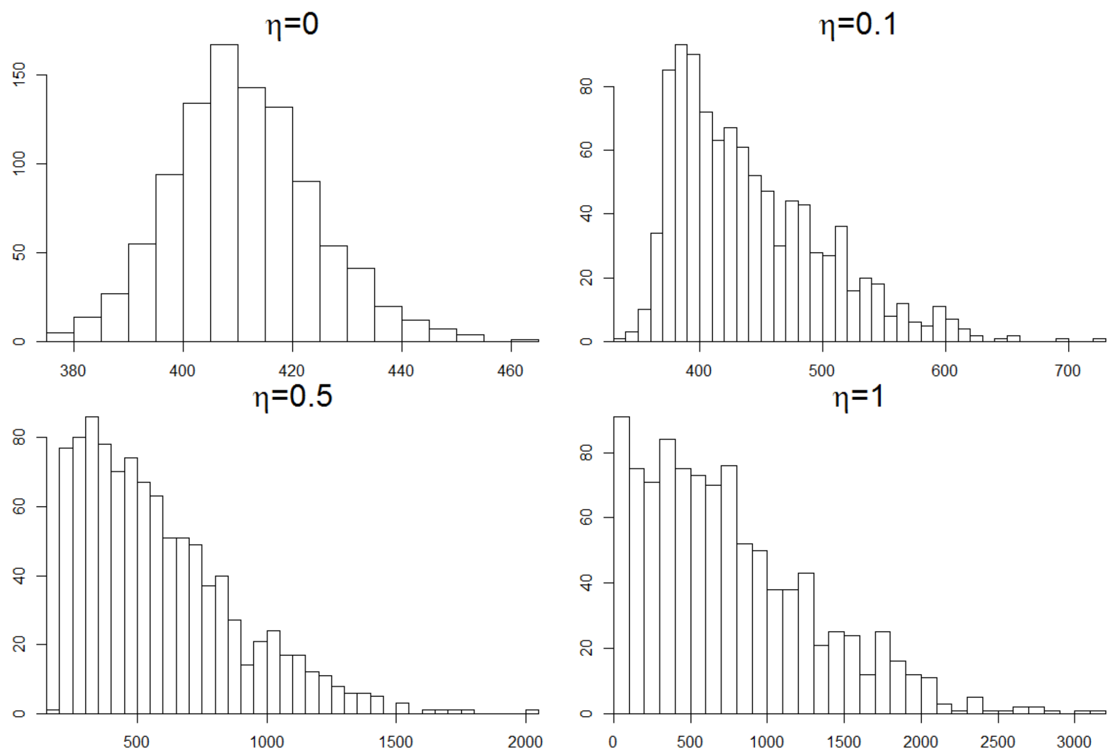

Figure 1: Histogram of T=1,000 volume realizations simulated from the model outlined in Section 2, for various values of the proportion of informed trading η. The number of liquidity seekers (M) is 1,000 and the trading intensity (σ) is fixed at unity.

Figure 1 displays four histograms of simulated volumes with $M = 1,000$ liquidity seekers, for different values of η. The sample size is $T = 1,000$ trading sessions. The simulation confirms the analysis in the previous section: in the case of no informed traders ($\eta = 0$), the volume distribution follows a slightly skewed bell-curve, while in the presence of informed traders volume is higher in level and far more dispersed. The simulated VCVs for the four panels are 0.03, 0.14, 0.48 and 0.77, respectively.[6]

Figure 2 reports the average VCV from $R = 1,000,000$ repetitions of simulating a sample of $T = 100$ trading sessions with M traders, for different values of η and M. As we can see, the average VCV only deviates substantially from its asymptotic value (Eq.6) when both M and η are low. Nevertheless, even for small M, the average VCV is strictly increasing in η. The insensitivity to M is encouraging as it implies that there is little concern for confounding a high η with a low M. The insensitivity to M is also desirable from an empirical perspective, because the number of traders (M) in markets is typically unknown.

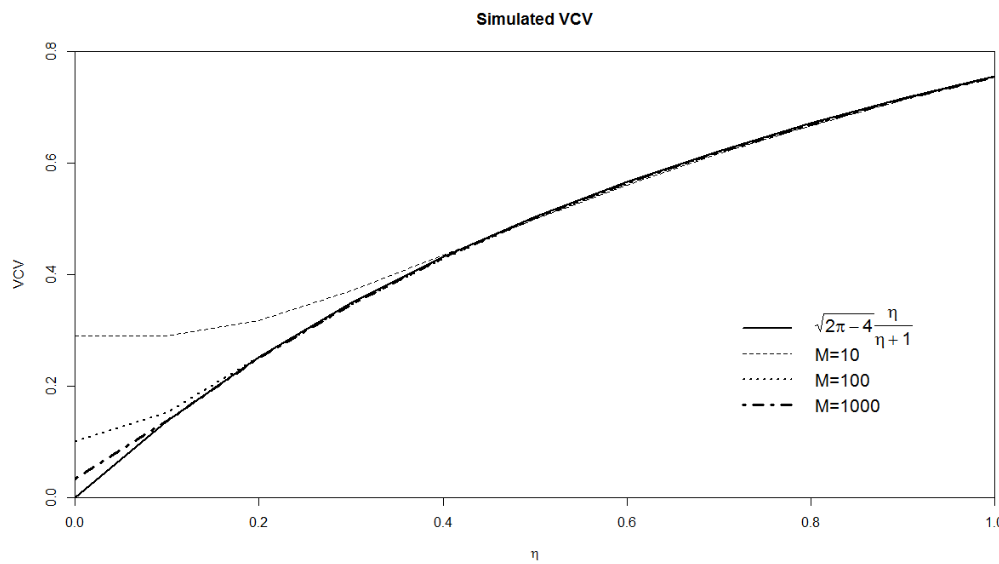

Figure 2: Average VCV obtained from $R = 1,000,000$ replications of $T = 100$ volume realizations simulated from the model outlined in Section 2, for various values of the proportion of informed trading η and number of liquidity seekers M.

In Table 1, Panel A, we report the average VCV as plotted in Figure 2 for selected values of η, as well as the standard deviations to evaluate VCV's precision. In addition to VCV, we also report these statistics on simulated values of $\widehat{\eta}$ (Eq.8). Both VCV and $\widehat{\eta}$ increase monotonically in the true proportion of informed trade (η). This is even the case for markets with low trading activity M. Also, the estimator $\widehat{\eta}$ in our simulations traces the true value of η closely, in particular when either M or η are not too low. Panel B of Table 1 reports simulation

[6]The slightly skewed bell-curved volume distribution for $\eta = 0$ converges (as $M \to \infty$) to the distribution of the maximum of two Normally distributed random variables, which was first described by Clark (1961).

Table 1: Simulation results - Benchmark model

This table reports the average and standard deviation of VCV (left) and $\widehat{\eta}$ (right) obtained from $R = 1,000,000$ replicated samples of T volume realizations, simulated from the model outlined in Section 2, for various values of the proportion of informed trade η and number of liquidity seekers M. In Panel A, the number of volume observations in each replication is $T = 100$. In panel B, $T = 10$. Detailed simulation results are reported in Internet Appendix Section A1.

			Panel A: $T = 100$							
η	0	0.2	0.5	0.8	1	0	0.2	0.5	0.8	1
			VCV					$\widehat{\eta}$		
			$M = 10$					$M = 10$		
Mean	0.29	0.32	0.5	0.67	0.75	0.24	0.27	0.50	0.80	1.01
s.d.	0.02	0.02	0.04	0.05	0.06	0.02	0.03	0.05	0.10	0.15
			$M = 100$					$M = 100$		
Mean	0.10	0.25	0.50	0.67	0.75	0.07	0.20	0.50	0.80	1.01
s.d.	0.01	0.02	0.03	0.05	0.06	0.01	0.02	0.05	0.10	0.15
			$M = 1000$					$M = 1000$		
Mean	0.03	0.25	0.50	0.67	0.75	0.02	0.20	0.50	0.80	1.01
s.d.	0.00	0.02	0.03	0.05	0.06	0.00	0.02	0.05	0.10	0.15
			Panel B: $T = 10$							
η	0	0.2	0.5	0.8	1	0	0.2	0.5	0.8	1
			VCV					$\widehat{\eta}$		
			$M = 10$					$M = 10$		
Mean	0.28	0.31	0.48	0.65	0.74	0.23	0.26	0.49	0.81	1.27
s.d.	0.07	0.07	0.11	0.15	0.17	0.07	0.08	0.18	4.12	50.5
			$M = 100$					$M = 100$		
Mean	0.10	0.24	0.48	0.65	0.74	0.07	0.19	0.49	0.83	1.11
s.d.	0.02	0.06	0.11	0.15	0.17	0.02	0.06	0.17	0.46	3.05
			$M = 1000$					$M = 1000$		
Mean	0.03	0.24	0.48	0.65	0.74	0.02	0.19	0.49	0.78	1.16
s.d.	0.01	0.06	0.11	0.15	0.17	0.01	0.06	0.17	14.28	24.67

results for smaller simulated samples, of $T = 10$ trading sessions. We still find the average VCV and $\widehat{\eta}$ to increase monotonically in η. This result implies that VCV can be applied even in small time-series samples, e.g. when estimated monthly using daily volume observations. The reported standard deviations however reveal that VCV, and more so $\widehat{\eta}$, are less precisely estimated when T is small.

To investigate the robustness of VCV as a measure of information asymmetry, we conduct simulations for various modifications of the benchmark model. The results of this extensive simulation exercise, as well as supplementary results to Table 1, are reported in Internet Appendix B. First, we relax the assumption of Normally distributed demand and allow for leptokurtic and skewed demand, to generate outliers in trading volume that are unrelated to informed trading (Table B.4). We also relax the assumption of *i.i.d* demand and allow for dynamic demand, generating persistence in trading volume (Table B.5). Next, we allow the proportion of informed trade η to be random across observations, to demonstrate that VCV increases in the *average* proportion of informed trade, either over a time-series or over a cross-section of observations (Table B.6). We allow for variation of M and σ randomly across observation (Tables B.7 and B.8), to demonstrate VCV in a setting where trading volume differs across observations for reasons unrelated to private information. We also investigate random liquidity supply (Table B.9), as opposed to full unconditional market maker capacity in our benchmark model. We consider the situation where different groups of informed investors receive distinct signals (Table B.10) and finally, we endogenize informed trading, by allowing the trading intensity of informed investors to be proportional to the uninformed order flow (Table B.11).

Overall, the simulation results in this section and Internet Appendix B demonstrate the robustness of VCV as a measure of asymmetric information. The basic result that VCV is monotonically increasing in the proportion of informed trade η holds under very general conditions and in small samples, while the standard deviation of VCV remains fairly low. These simulations also reveal, however, that the baseline level of VCV is sensitive to the underlying assumptions and thus may differ across different trading environments. It is therefore important to compare VCV only across comparable assets, and when estimated with the same number of volume observations. In our regression analyses below comparing VCV across US stocks, we always control for industry, size, liquidity, and book-to-market and year-fixed effects to adjust for any possible differences in the baseline level of VCV unrelated to private information.

We emphasize that VCV can be applied both to a time-series or a cross-section of volume observations. In particular the result that VCV increases in the average proportion of informed trade when η and M vary across observations, supports the applicability of VCV to a cross-section of volumes. We apply the cross-sectional VCV in Section 5 when investigating the patterns of VCV around earnings announcements.

The simulations also reveal that $\widehat{\eta}$ clearly does not perform well as a measure of informed trading beyond the benchmark model. The simulated observations of $\widehat{\eta}$ are more widely dispersed than VCV, while their averages are often not monotonically increasing in η, and are not always bounded by 0 and 1. This poor performance of $\widehat{\eta}$ occurs because the denominator in Eq.(8) can easily take on small or negative numbers, which makes the estimator highly erratic. In the remainder of this paper, we therefore focus on VCV as our measure of informed trade.

4 VCV in the cross-section of US stocks

After having established, from analytical and numerical analysis, a positive monotonic relation between VCV and the proportion of informed trade, we now turn to the data to analyze the empirical properties of our measure. In this section, we describe cross-sectional variation in VCV for US stocks. We compute VCV for US stocks and compare these figures with other firm-level characteristics, including indicators of informed trade and illiquidity. We obtain daily trading volumes from the CRSP daily stock file for all common stocks listed on NYSE, AMEX and NASDAQ over the period January 1980 - December 2020. We disregard the most infrequently traded stocks by only including firm-year observations for stocks with positive trading volume in at least 200 days during that year.[7]

Annual firm-level observations of VCV are computed by dividing the annual standard deviation of daily trading volumes by the annual average of daily trading volumes. The volume coefficient of variation of stock i in year τ is defined as:

[7] For NASDAQ listed firms, we adjust trading volume prior to 2004 following Gao and Ritter (2004): reported volume on NASDAQ stocks is divided by 2.0, 1.8, and 1.6 during the period prior to February 1st 2001, the period between February 1st 2001-December 31st 2001, and January 1st 2002 - December 31st 2003, respectively. Note that this adjustment does not affect VCV, in which volume is both in the nominator and denominator, but it does affect other measures that are based on volume, such as Amihud (2002) Illiquidity.

$$VCV_{i,\tau} = \frac{\hat{\sigma}_{V(i,t\in\tau)}}{\hat{\mu}_{V(i,t\in\tau)}}, \tag{10}$$

where $\hat{\mu}_{V(i,t\in\tau)}$ is the sample average and $\hat{\sigma}_{V(i,t\in\tau)}$ is the sample standard deviation of all daily trading volumes of stock *i*, $V_{i,t}$, in year τ. We compute VCV using three different measures of trading volume: (*i*) trading volume in US dollars (V_{USD}), (*ii*) volume *market shares* ($V_{\%}$), defined as daily volume in a single stock as a fraction of total market volume on the same day, to control for market-wide variation in trading-activity that is unrelated to firm-specific information, such as macro-level sentiment and liquidity shocks, and (*iii*) daily *turnover* (V_{TO}), to control for differences in market capitalization:

$$\begin{aligned} V_{USD,i,t} &= shares\ traded_{i,t} \times closing\ price_{i,t} \\ V_{\%,i,t} &= \frac{V_{USD,i,t}}{\sum_i V_{USD,i,t}} \\ V_{TO,i,t} &= \frac{shares\ traded_{i,t}}{shares\ outstanding_{i,t}}. \end{aligned} \tag{11}$$

Table 2 reports summary statistics for these three measures of VCV. The sample averages, as well as other statistics, are highly similar for the three VCV measures. The bottom rows of Table 2 show that the three different measures of VCV are highly correlated. The strong similarity between the three VCV measures offers support for the theoretical analysis of Section 2: although trading intensity (σ) and participation (M) are determinants of the level and variance of volume, VCV is independent of both σ and M (Eq.(6)). Market-wide variation in the number of market participants and their trading intensity should therefore have little impact, so that VCV derived from dollar volume, volume market shares, or turnover, should be virtually equivalent. The results in Table 2 support this premise. In the remainder of this section, our measure of informed trading VCV is defined as the coefficient of variation of daily volume market shares (VCV$_{\%}$), which controls for market-wide variation in volume that is unrelated to firm-specific information. Highly similar results are obtained when using any of

Table 2: VCV Summary Statistics

This table reports summary statistics of annual firm-level observations of the Volume Coefficient of Variation (VCV) of daily dollar trading volume in US dollars (VCV_{USD}), daily volume market shares (daily dollar volume as a percentage of total market dollar volume – $VCV_{\%}$), and turnover (dollar volume as a fraction of market capitalization – VCV_{TO}). The table reports the total number of observations, the number of distinct stocks in the sample (N), the number of time-series observations/years (T), mean, standard deviation, s.d. (CS), the time-series average of annual cross-sectional standard deviations, s.d. (TS), the cross-sectional average of stock-specific time-series standard deviations, selected quantiles (q), and the cross-sectional average of stock-specific first-order autocorrelations (ρ). The bottom two rows report the time-series averages of within-year rank (Spearman) correlations between the different VCV measures. Sample: 1980-2020.

	VCV_{USD}	$VCV_{\%}$	VCV_{TO}
Observations	148,673	148,673	148,673
N	16,571	16,571	16,571
T	41	41	41
Mean	1.353	1.333	1.302
s.d.	0.844	0.857	0.786
s.d. (CS)	0.818	0.830	0.768
s.d. (TS)	0.552	0.556	0.520
$q_{0.1}$	0.575	0.534	0.569
$q_{0.25}$	0.808	0.779	0.788
Median	1.193	1.177	1.144
$q_{0.75}$	1.638	1.627	1.573
$q_{0.9}$	2.231	2.217	2.155
ρ	0.192	0.198	0.208
Correlations			
$VCV_{\%}$	0.983		
VCV_{TO}	0.971	0.960	

the other volume definitions.[8]

In addition to the annual estimates reported in Table 2, we also consider quarterly and monthly firm-level estimates of VCV. Summary statistics on these measures are reported in Internet Appendix Section C.2. The advantage of the annual VCV is that the coefficient of variation is estimated more precisely due to a larger number of observations, while the quarterly and monthly VCV allow studying variation of information asymmetry at a higher frequency. Overall, we find the annual, quarterly and monthly VCV to be highly correlated. Interestingly, the annual VCV is on average higher than its quarterly and monthly counter-

[8] Internet Appendix Section C.1 reports summary statistics for the underlying volume measures (V_{USD}, $V_{\%}$, V_{TO}), as well as VCV summary statistics for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX, and for subsamples of observations prior to 2000 (1980-1999) and post 2000 (2000-2020), showing that the three measures of VCV behave fairly similar across these subsamples.

parts, because the annual estimates also capture within-year seasonal variation in volume. This observation stresses the importance of comparing VCV only when estimated over similar sized samples. In the remainder of this section, we use annual estimates of VCV when comparing to other annual firm characteristics, quarterly estimates when studying the relation between VCV and institutional ownership (from quarterly 13F filings), and monthly estimates when considering monthly return reversals.

4.1 VCV and other firm characteristics

Table 3 reports the correlations between VCV and other firm-level characteristics: size, book-to-market ratio, firm age, return volatility, turnover, Amihud (2002) illiquidity, bid-ask spread, Roll's (1984) estimate of the bid-ask spread, and analyst coverage. Size is defined as the log of market capitalization on the last trading day of June. Return volatility is the annual standard deviation of daily returns. Amihud (2002) illiquidity is defined as the the log of the annual average of the daily ratio $\frac{|R_{i,t}|}{V_{USD,i,t}}$. The bid-ask spread is the annual average of daily closing bid-ask spreads as a percentage of the closing price $\frac{ask_{i,t}-bid_{i,t}}{price_{i,t}}$. Roll's (1984) measure is the square root of the negative of the daily return autocovariance $\sqrt{-Cov\left(R_{i,t},R_{i,t-1}\right)}$.[9] The book-to-market ratio is the ratio of the book value of equity at the fiscal year end, obtained from COMPUSTAT, to the market value of equity at the end of the same calendar year. Firm age is proxied by the number of years passed since the firm appeared for the first time in the CRSP database. Analyst coverage is defined as the number of distinct analysts covering a stock in a given year (Source: IBES). Summary statistics of these variables and subsample analyses are provided in Internet Appendix Section C.3.

As can be seen from Table 3, VCV is negatively correlated with size and turnover and positively correlated with return volatility, Amihud illiquidity and the bid-ask spread. These results are consistent with our proposition that VCV is a measure of informed trading, since

[9]In the case of positive return autocorrelations, we set Roll's measure equal to $-\sqrt{Cov\left(R_{i,t},R_{i,t-1}\right)}$, following Roll (1984). We obtain qualitatively similar results when we either set these observations of Roll's measure to zero, or omit them from our sample.

Table 3: VCV and other firm characteristics

This table reports the correlations between annual firm-level observations of VCV (obtained from daily volume market shares) and other annual firm-level characteristics. Each entry reports the time-series average of within-year rank (Spearman) correlations. *Size* is the log of market capitalization at the last trading day of June. *BM ratio* is the ratio of the book value to the market value of equity. *Age* is the number of years since the firm's first appearance in CRSP. *Volatility* is the annual standard deviation of daily returns. *Turnover* is the annual average of daily trading volume as a percentage of market capitalization. *Illiquidity* is the log of the annual average of the daily ratio $\frac{|R_{i,t}|}{V_{USD,i,t}}$ (Amihud, 2002). *Bid-Ask spread* is the annual average of daily bid-ask spreads $\frac{ask_{i,t}-bid_{i,t}}{price_{i,t}}$. *Roll's measure* is the square root of the negative of the daily return autocovariance $\sqrt{-Cov\left(R_{i,t},R_{i,t-1}\right)}$. *Coverage* refers to the number of distinct analysts covering a stock in a given year Source: CRSP, COMPUSTAT, and IBES.

	VCV	Size	BM	Age	Vol.	Turn.	Illiq	B-A	Roll
Size	−0.66								
BM ratio	0.17	−0.23							
Age	−0.30	0.33	0.11						
Volatility	0.41	−0.61	−0.03	−0.40					
Turnover	−0.28	0.33	−0.22	−0.02	0.17				
Illiquidity	0.69	−0.95	0.23	−0.33	0.54	−0.52			
Bid-Ask spread	0.63	−0.87	0.26	−0.25	0.60	−0.41	0.91		
Roll's measure	0.25	−0.36	0.19	−0.05	0.23	−0.30	0.41	0.50	
Coverage	−0.56	0.78	−0.26	0.18	−0.30	0.48	−0.81	−0.71	−0.29

information asymmetry is likely to be stronger in smaller stocks and asymmetric information reduces liquidity. Yang et al. (2020) associate return volatility with information asymmetry. The negative correlation with firm age suggest that information asymmetry is lower for more mature firms. Analyst coverage is likely to reduce information asymmetry, which is consistent with the negative correlation with VCV. In Section 4.5, we study the impact of exogenous reductions in analyst coverage due to brokerage closures and find that reductions in analyst coverage are associated with an increase in VCV.

4.2 Return reversals

The correlation between VCV and the bid-ask spread reported in Table 3, is clearly higher than the correlation between VCV and Roll's (1984) estimate of the bid-ask spread. This result is expected, as it is well known from Huang and Stoll (1997) and others that Roll's measure underestimates the bid-ask spread in the presence of information asymmetries, since price changes due to informed trading are less likely to be reversed by the bid-ask bounce, and are thus characterized by less negative autocorrelations. To further evaluate the relationship between VCV and the bid-ask spread, we double-sort stocks within each year into quartiles based on the bid-ask spread and on Roll's measure. Table 4 shows the average VCV for each of these sixteen groups of firms. We find that VCV is monotonically increasing in the bid-ask spread but not in Roll's measure, which is consistent with the downward bias of Roll's measure in the presence of information asymmetry. Stocks with high information asymmetry are expected to have a relatively high bid-ask spread but a relatively low value of Roll's measure. We see from Table 4 that these stocks are precisely the stocks with a high VCV.

Table 4: VCV and the Bid-Ask spread

This table reports the sample average VCV for 16 groups of stocks double-sorted within each year on the *Bid-Ask spread* (the annual average of daily bid-ask spreads $\frac{ask_{i,t}-bid_{i,t}}{price_{i,t}}$) and *Roll's* (1984) measure (the square root of the negative of the daily return autocovariance $\sqrt{-Cov\left(R_{i,t},R_{i,t-1}\right)}$). The final row and column report the difference in average VCV between high and low quartiles, with significant differences at the 10%, 5%, and 1% level indicated by *, **, and ***. Source: CRSP.

	Roll: Low	2	3	High	High-Low
Bid-Ask: Low	0.931	0.778	0.725	0.922	−0.009
2	1.124	1.078	1.072	0.993	−0.132**
3	1.344	1.308	1.451	1.375	0.031*
High	2.037	1.712	1.807	2.078	0.041*
High-Low	1.106***	0.935***	1.082***	1.156***	

To further study the relation between VCV and return autocorrelation, we consider monthly return reversals. It is well known that returns on individual stocks, in particular illiquid stocks, exhibit significant short-term reversals (e.g. Jegadeesh, 1990). We double-sort stocks within each month into quartiles based on monthly estimates of Amihud's (2002) Illiquidity and VCV. We compute, for each month within each group, the correlation between the stocks' returns in the sorting month and in the following month: $cor(R_{i,t}, R_{i,t+1})$. Table 5 reports the average of these monthly return autocorrelations, for each of the 16 groups. Across all groups, we find return reversals (i.e. negative autocorrelation). These reversals are clearly stronger for the more illiquid stocks. However, within each liquidity quartile, we find that reversals are decreasing in VCV. The final row of Table 5 shows that return autocorrelation is lower for High VCV stocks than for Low VCV stocks. This result implies, similar to Table 4, that short-term reversals are in general more profound for illiquid stocks, but that these reversals are weaker when the illiquidity is associated with information asymmetry, implying that reversals are a signal of uninformed order flow.

Table 5: VCV and monthly reversals

This table reports the average correlation between stock returns in the month of sorting and the following month ($cor(R_{i,t}, R_{i,t+1})$) for 16 groups of stocks double-sorted within each month t on the monthly estimate of VCV and Amihud (2002) illiquidity. The final row and column report the difference in average VCV between high and low quartiles, with significant differences at the 10%, 5%, and 1% level indicated by *, **, and ***. Source: CRSP.

	Illiq: Low	2	3	High	High-Low
VCV: Low	−0.018	−0.015	−0.044	−0.126	−0.108***
2	−0.006	−0.002	−0.036	−0.100	−0.094***
3	−0.004	−0.013	−0.014	−0.069	−0.065***
High	−0.001	−0.008	−0.015	−0.054	−0.053***
High-Low	0.017***	0.007*	0.029***	0.072***	

To demonstrate the robustness of the results reported in this section, we report subsample results as well as analysis of weekly reversals in Internet Appendix Section C.4. Our results are broadly consistent with existing research: Llorente et al. (2002), Hameed et al. (2008), Odders-White and Ready (2008), Bongaerts et al. (2016), and Johnson and So (2018) use various measures to show that asymmetric information is associated with weaker short-term reversals.

4.3 VCV and other measures of asymmetric information

In this subsection, we compare VCV to various incumbent measures of asymmetric information. These measures include the probability of informed trade (PIN; Easley et al., 1996), C2 (Llorente et al., 2002), and the Multimarket Information Asymmetry measure (MIA; Johnson and So, 2018). PIN is estimated by fitting a structural microstructure model to signed transaction data. C2 measures the relation between daily volume and return persistence, based on the premise that prices changes due informed trading tend to persist, while price changes due to uninformed trading are more likely to revert. MIA is based on relative trading volume in options and stocks, based on the assumption that informed traders are more likely to trade in options.

Table 6: VCV and other information asymmetry measures

This table reports the correlation between the annual firm-level coefficients of variation of daily volume market shares (VCV) and various annual firm-level information asymmetry measures. Each entry reports the time-series average of within-year rank (Spearman) correlations. PIN_{BHL} is estimated by Brown, Hillegeist and Lo (2004). PIN_{BH} is estimated by Brown and Hillegeist (2007). PIN_{EHO} is estimated by Easley, Hvidkjaer, and O'Hara (2010). PIN_{DY}, Adjusted PIN, and the illiquidity measure PSOS are estimated by Duarte and Young (2009). MIA is the annual average of firm-day level observations estimated by Johnson and So (2018). C2 is estimated following Llorente et al. (2002). Sources: CRSP and cited authors' websites.

	VCV	PIN_{BHL}	PIN_{BH}	PIN_{EHO}	PIN_{DY}	Adj.PIN	PSOS	MIA
PIN_{BHL}	0.53							
PIN_{BH}	0.60	0.74						
PIN_{EHO}	0.53	0.62	0.68					
PIN_{DY}	0.57	0.65	0.69	0.86				
Adjusted PIN	0.52	0.58	0.71	0.64	0.72			
PSOS	0.46	0.45	0.44	0.62	0.71	0.39		
MIA	0.25	0.37	0.43	0.11	0.23	0.32	0.05	
C2	0.11	0.14	0.12	0.01	0.01	0.02	0.02	0.01

For our analysis, we make use of the various PIN and MIA measures that are kindly made publicly available by the authors of previous studies. These measures include MIA estimated by Johnson and So (2018) and PIN measures estimated by Easley et al. (2010 – PIN_{EHO}); Brown, Hillegeist and Lo (2004 – PIN_{BHL}); Brown and Hillegeist (2007 – PIN_{BH}); and Duarte and Young (2006 – PIN_{DY}).[10] We compute annual firm-level observations of MIA as the annual average of the available daily observations for each firm. We derive annual stock-level observations of C2 as the estimated slope coefficient from running regressions, for each firm in each year, of daily returns on the interaction of lagged returns and lagged

[10]Annual firm-level observations of PIN_{DY}, PIN_{EHO}, PIN_{BH} and PIN_{BHL} are made available by Jefferson Duarte (`http://www.owlnet.rice.edu/~jd10/`), Søren Hvidkjær (`https://sites.google.com/site/hvidkjaer/data`) and Stephen Brown (`http://scholar.rhsmith.umd.edu/sbrown/pin-data`), respectively. Daily firm-level observations of MIA are made available by Travis Johnson (`http://travislakejohnson.com/data.html`). Summary statistics of the measures employed in this section, as well as subsample analyses, are provided in Internet Appendix Section B.5.

(detrended) turnover, while controlling for daily lagged returns (see Llorente et al., 2002, for details).

Table 6 shows the correlations between VCV and various annual firm-level information asymmetry measures. Our VCV measure is positively correlated with all PIN measures. The correlation between VCV and PIN is of similar magnitude as the correlations between the various PIN measures. The correlations between VCV and the MIA and C2 measures are substantially lower, although still positive.

Compared to these incumbent measures, our VCV measure is far easier to compute. In addition, VCV does not require intraday order-level data or option volume data and covers therefore a much larger set of stocks than PIN and MIA. VCV has in particular higher coverage among small and illiquid stocks, for which the risk of information asymmetry is most relevant. Moreover, VCV is available for the full time-series sample, while the available PIN measures are not available post 2010. As discussed in the introduction, the classification into buy and sell orders has been increasingly problematic in the time of continuous trading. [11]

Duarte and Young (2009) argue that PIN does not only measure informed trading, but also other illiquidity effects. They therefore decompose PIN into *Adjusted PIN*, which is proposed as a cleaner measure of asymmetric information; and *PSOS* (probability of symmetric order-flow shock), which is a measure of illiquidity unrelated to asymmetric information. These additional variables are included in Table 6. Both Adjusted PIN and PSOS are positively correlated with VCV.

[11]The time-series samples and number of distinct stocks N in our sample with coverage of VCV, PIN_{BH}, PIN_{DY} and MIA are: 1980-2020, N=16,571 (VCV); 1993-2010, N=11,334 (PIN_{BH}); 1993-2010, N=11,316 (PIN_{BHL}); 1983-2001, N=4,365 (PIN_{EHO}); 1983-2004, N=4,634 (PIN_{DY}); 1996-2018, N=3,877 (MIA). See Internet Appendix Table C.11 for details.

Table 7: VCV and Adjusted PIN

This table shows the results from regressing annual firm-level coefficients of variation of daily volume market shares (VCV) on the measures by Duarte and Young (2009): PIN_{DY}, Adjusted PIN, and PSOS (probability of symmetric order-flow shock). All regressions include fixed effects for each year, industry, size decile, book-to-market decile and Illiquidity decile. Two-way clustered standard errors, clustered at the year and industry level, are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5%, and 1% level. Source: CRSP and the website of Jefferson Duarte (`http://www.owlnet.rice.edu/~jd10/`)

	VCV			
	(1)	(2)	(3)	(4)
PIN_{DY}	0.449***		0.963***	0.505***
	(0.127)		(0.164)	(0.113)
Adjusted PIN	0.956***	1.186***		0.925***
	(0.156)	(0.178)		(0.160)
PSOS	0.047	0.187***	−0.131**	
	(0.056)	(0.053)	(0.059)	
Observations	37,918	37,918	37,918	37,918
Adjusted R^2	0.356	0.355	0.352	0.356
Fixed effects	Yes	Yes	Yes	Yes

In Table 7, we examine the correlation between VCV and the three measures by Duarte and Young (2009) in a regression context. To control for time variation and firm characteristics unrelated to asymmetric information, we include year fixed effects, 48 Fama-French industry fixed effects, and decile fixed effects for size, book-to-market and Amihud illiquidity deciles.[12] The regression results indicate that VCV is mostly associated with adjusted PIN, while there is no robust relation between VCV and PSOS, thereby supporting our claim that VCV, like adjusted PIN, is indicative of asymmetric information rather than general illiquidity.

4.4 VCV and institutional ownership

In this subsection, we study the relationship between VCV and various indicators of institutional ownership that we obtain from quarterly 13F filings recorded in the Spectrum database.

[12] Rather than including size, book-to-market and illiquidity as control variables, we control for these characteristics using decile fixed effects, in order to accommodate nonlinearities and outliers.

Table 8 reports the results from regressing quarterly estimates of VCV on various end-of-quarter characteristics of institutional ownership. These characteristics include institutional holdings (defined as the percentage of shares of a firm held by institutional investors at the end of the quarter) and breadth of ownership (defined as the number of institutional investors holding shares in the firm, as a percentage of the total number of institutional investors reported in the Spectrum 13F database at the end of each quarter – Chen et al., 2002). Boone and White (2015) find that institutional ownership leads to an improvement in disclosure practices and therefore lower information asymmetry. The first column of Table 8 shows that VCV indeed has a significantly negative association with breadth of ownership. VCV is lower (implying lower information asymmetry) for firms that have high breadth of ownership.

Table 8: VCV and institutional ownership

This table reports the results from regressing quarterly firm-level coefficients of variation of daily volume market shares (VCV) on various measures of institutional ownership. *Holdings* is the percentage of shares of the firm held by institutional investors at the end of the quarter; *Breadth* is the percentage of all institutional investors that hold shares of the firm (Chen et al., 2002); $Monitors$ is the fraction of institutional investors in each firm for which the firm is in the top 10% of the institution's holdings (Fich et al., 2015); and *Dedicated* is the fraction of institutional investors in each firm that are classified as 'Dedicated' investors by Bushee and Noe (2000). All regressions include fixed effects for each quarter, industry, size decile, book-to-market decile and illiquidity decile. Two-way clustered standard errors, clustered at the quarter and industry level, are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5%, and 1% level. Sources: CRSP, 13F and the website of Brian Bushee `http://acct.wharton.upenn.edu/faculty/bushee/`. Sample period: 1980Q1-2018Q4.

	VCV			
	(1)	(2)	(3)	(4)
Holdings	0.0004	0.0003	0.0003	0.0002
	(0.001)	(0.001)	(0.001)	(0.001)
Breadth	-0.934^{***}	-1.296^{***}	-0.902^{***}	-1.256^{***}
	(0.253)	(0.358)	(0.243)	(0.356)
Monitors		0.624^{**}		0.608^{**}
		(0.256)		(0.265)
Dedicated			0.456^{***}	0.448^{***}
			(0.114)	(0.110)
Observations	346,858	346,858	346,679	346,679
Adjusted R^2	0.394	0.395	0.396	0.397
Fixed effects	Yes	Yes	Yes	Yes

In addition, we consider two measures that identify groups of presumably well-informed investors: monitoring investors and dedicated investors. Following Fich et al. (2015), we define an institutional investor to be a 'monitor' for a certain firm if that firm belongs to the top 10% of holdings in the institution's portfolio. These monitoring investors are likely to be better informed about the firm than non-monitoring investors. Dedicated investors are those institutional investors that Bushee and Noe (2000) and Bushee (2001) classify as 'dedicated'. They are characterized by large, stable holdings in a small number of firms, as opposed to 'quasi-indexing' investors and 'transient' investors.[13]

The variable *Monitors* in Table 8 is the percentage of institutional investors in each firm that are defined as monitoring investors. The variable *Dedicated* in Table 8 is the percentage of institutional investors in each firm that are classified as dedicated investors. Columns 2–4 of Table 8 show that these variables are both significantly positively associated with VCV, consistent with our proposition that VCV measures informed trade.

The relationship between patterns in institutional ownership and VCV reported in Table 8 reaffirms that VCV is a measure of asymmetric information. Suppose that a firm is held by only a small number of institutional investors, who each assign a relatively large fraction of their portfolio to this firm's stock (i.e. *Breadth* is low, while *Monitors* and *Dedicated* are high). Ownership of such a firm is therefore relatively concentrated in the hands of a small number of presumably well informed investors. When trading this stock, information asymmetry should be a significant concern, as it is not unlikely that the counterparty is one of these better informed investors. On the other hand, for a firm that is widely held among institutional investors, each of which holding only a relatively small share of the firm (i.e.: *Breadth* is high, while *Monitors* and *Dedicated* are low), the risk of asymmetric information should be lower, which is in accordance with the results reported in Table 8. Summary statistics of the measures employed in this section, subsample analyses and a robustness test using annual estimates of VCV are provided in Internet Appendix Section C.6.

[13]Classification into these three groups is based on a factor and cluster analysis approach (see Bushee, 2001, for details). The classification of institutional investors in the 13F Spectrum database is made available on the website of Brian Bushee `http://acct.wharton.upenn.edu/faculty/bushee/`.

4.5 VCV around brokerage closures

Various recent studies (e.g. Kelly and Ljungqvist, 2012; Derrien and Kecskes, 2013; Li and You, 2015; Bushman et al., 2017; and Chen and Lin, 2017) consider terminations of analyst coverage due to brokerage closures as exogenous shocks to the information environment of individual stocks. Kelly and Ljungqvist (2012) find that information asymmetry increases following these exogenous terminations in analyst coverage. For the 22 brokerage closures between April 2000 and January 2008 listed in Appendix A of Kelly and Ljungqvist (2012), we identify in the IBES database a treatment sample of a total of 1,764 observations of firms that experience reductions in analyst coverage due to one of these closures.

We perform a simple difference-in-differences regression, to compare the VCV of treated firms (i.e. firms that experience closure-induced coverage terminations) to non-treated firms (the control group), before and after the brokerage closure. For each brokerage closure, our control group includes all non-treated firms in our sample analyzed in Section 4, for which analyst coverage in the calendar year prior to the brokerage closure is strictly positive. The VCV before closure is defined as the coefficient of variation of daily volume market shares over a 12-month period before the closure, while the VCV after closure is calculated over a 12-month period after the closure. Following Derrien and Kecskes (2013), we impose three-month gaps between the event and the estimation windows, such that the VCV before (after) closure is calculated from trading volumes over the months -14 to -3 (+3 to +14), with the brokerage closure occurring in month 0. These observations of VCV are regressed on a dummy variable indicating observations in the *treatment* group, a dummy variable indicating the observations *after* each brokerage closure, and an interaction term.

The results of the difference-in-differences regression are reported in the first column of Table 9. The coefficient on the interaction term *After*×*Treated* is of primary interest. This interaction coefficient is positive and significant, meaning that the VCV of firms that face exogenous analyst reductions as a result of brokerage closures *increases* relative to the VCV of control firms that are not exposed to the brokerage closures. The coefficient on *After* is

Table 9: Brokerage closures

This table reports the results from difference-in-differences regressions around brokerage closure-induced terminations of analyst coverage. The treatment sample consist of 1,764 observations of firms that experience a reduction in analyst coverage due to a total of 22 distinct brokerage closures between April 2000 and January 2008. The control sample consists of 31,661 observations. For all 33,425 observations, we compute VCV over the months $[-14, -3]$, and over the months $[3, 14]$, with the brokerage closure occurring in month 0, resulting in a total of 66,850 observations of VCV. These VCVs are regressed on dummies indicating the treatment group (Treated), the post-closure window (After), and their interaction. In the second (third) column, the sample is restricted to firms with analyst coverage of less than 10 (5) in the calendar year prior to the closure. All regressions include fixed effects for each year, industry, size decile, book-to-market decile and illiquidity decile. Two-way clustered standard errors, clustered at the year and industry level, are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5%, and 1% level.

	Full sample	Coverage < 10	Coverage < 5
	VCV	VCV	VCV
After × Treated	0.035^{***}	0.051^{**}	0.113^{***}
	(0.007)	(0.020)	(0.040)
After	-0.081^{***}	-0.090^{***}	-0.097^{***}
	(0.009)	(0.010)	(0.011)
Treated	-0.024^{**}	-0.022^{*}	-0.045
	(0.011)	(0.012)	(0.031)
Observations	66,850	46,952	27,760
Adjusted R^2	0.401	0.395	0.434
Fixed effects	Yes	Yes	Yes

negative, which reflects that VCV is on average decreasing over time.[14] The *Treated* coefficient indicates that there is a minor difference between the VCV of treated and control firms, prior to the event.[15]

The second and third column of Table 9 show that the interaction coefficient becomes larger when we restrict the sample to firms with lower analyst coverage. The intuition behind this result is that the event of one analyst discontinuing coverage of a firm is a greater

[14]Internet Appendix Section C.7 reports average time-series trends of VCV, finding a declining trend post-2000. This is consistent with recent studies that document improved market transparency, which is attributed partly to regulation, such as the enactment by the SEC of Regulation Fair Disclosure (Reg FD) in 2000 and the Sarbanes-Oxley Act in 2002 (e.g. Chen et al., 2010; Petacchi, 2015; Beaver et al., 2018; and Pawlewicz, 2018).

[15]Internet Appendix Section C.8 provides results for a matched control sample. In this case, we continue to find a positive interaction term, implying a positive impact of brokerage closures on information asymmetry, while the *Treated* coefficient is insignificant.

disruption to the information environment when the firm has already low analyst coverage to begin with. Indeed, the difference-in-differences estimate is approximately doubled (tripled) when covering only firms with analyst coverage of less than 10 (5) in the calendar year prior to the event. Overall, the results in Table 9 provide strong evidence for our proposition that VCV measures information asymmetry.

5 Cross-sectional VCV

In this section we look at VCVs computed from cross-sections of volume data. In particular, we document the pattern of the cross-sectional VCV around earnings announcements. It is widely recognized that earnings announcements resolve information asymmetries (e.g. Chae, 2005; George et al., 1994). In this section we show that, consistent with this view, VCV is relatively high prior to announcements and low afterwards, suggesting that uninformed traders delay their trades until information asymmetries are resolved after the announcement.

We obtain N = 339,257 quarterly earnings announcement dates from COMPUSTAT, from a total of 13,885 distinct NYSE, AMEX, and NASDAQ listed US firms over the period 1980-2016. To analyze the evolution of information asymmetry in event time, we compute the *cross-sectional VCV* for each day $d \in [-30, 30]$ around the announcement date, using the N trading volumes recorded for each stock on d days after the firm's earning announcement: [16]

$$VCV_{XS,d} = \frac{\hat{\sigma}_{V(t=t_i+d)}}{\hat{\mu}_{V(t=t_i+d)}}, \tag{12}$$

where $\hat{\mu}_{V(t=t_i+d)}$ is the sample average and $\hat{\sigma}_{V(t=t_i+d))}$ is the sample standard deviation of N daily trading volumes on day d after the firm-specific announcement date t_i. All volumes are as before defined as volume market shares, $V_{\%i,t}$, i.e.: volumes as a percentage as total trading volume on that calendar date t.

[16]Our simulation results in Section 3 and in the Internet Appendix demonstrate that VCV can be estimated from a cross-sectional sample of volume observations, generated by stock-days with different proportions of informed trade. In this setting, VCV is increasing in the *average* proportion of informed trade.

The black line in Figure 3 shows the pattern of the cross-sectional VCV around the announcement date, while the shaded areas indicate 95% confidence bounds, computed from the asymptotic distribution of sample coefficients of variation as derived by Albrecher et al. (2010). Figure 5 clearly shows that VCV is higher in the weeks prior to the announcement, which is consistent with uninformed investors delaying their trading activity as the announcement date is approaching. After information asymmetries are resolved on the announcement date, VCV is relatively low for multiple trading days. After 30 trading days, the cross-sectional VCV is approximately equal to the cross-sectional VCV 30 trading days prior to the announcement.

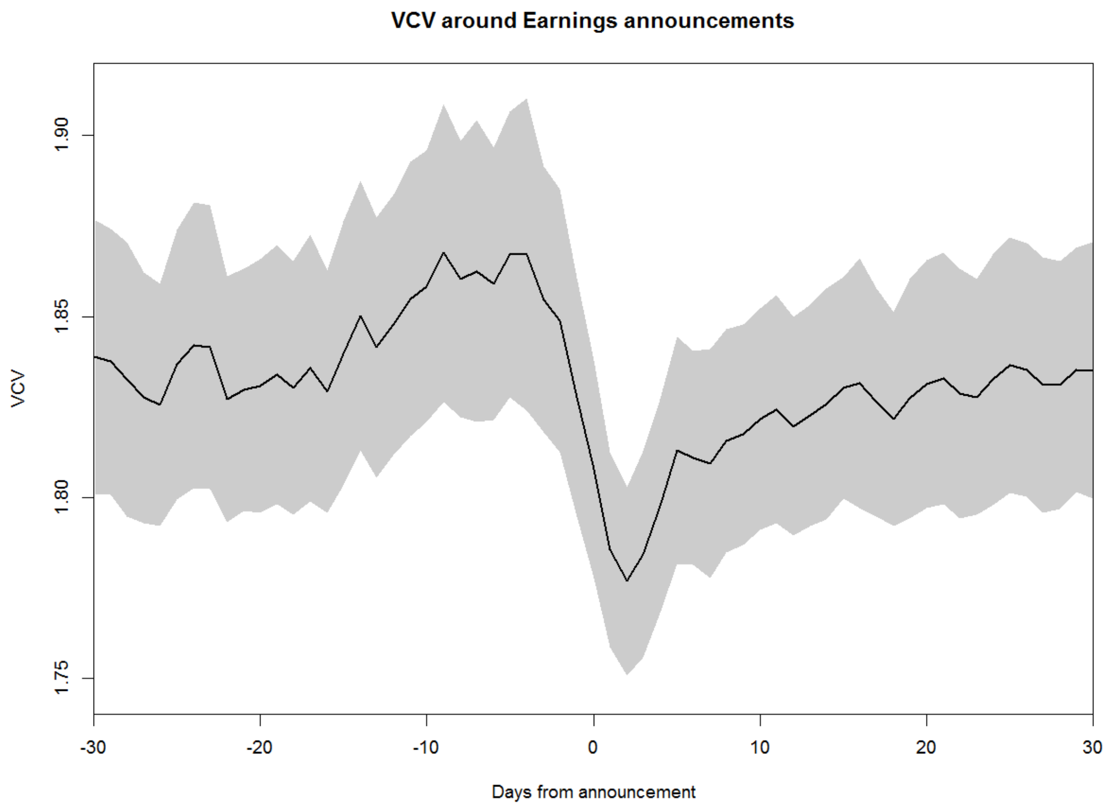

Figure 3: The black line shows the evolution of the daily cross-sectional VCV_{XS} around quarterly earnings announcements. The full sample includes all daily trading volumes over a 61 day event-window (day -30:30) around $N = 339,257$ quarterly announcements (sources: CRSP and COMPUSTAT). The reported VCV at d days after the announcement is estimated from the subsample of each stock's trading volume market shares at date d after each firm's announcement. The gray shaded areas indicate 95% confidence intervals: $VCV_{XS,d} \pm 1.96 \times S.E.(VCV_{XS,d})$. Standard errors ($S.E.$) are derived following Albrecher et al. (2010).

Table 10 reports VCV and its components: the cross-sectional mean and standard deviation of volume shares, for each day around the announcement. The level of volume is low

Table 10: Volume around earnings announcements

This table reports the cross-sectional mean $\hat{\mu}_d$, standard deviation $\hat{\sigma}_d$ (both multiplied by 1,000), and coefficient of variation $VCV_{CS,d}$ of all firms' daily trading volume shares on day d before and after $N = 339,257$ firm-specific earnings announcement dates, as well as the difference between these moments d days before and after the announcement. *, ** and *** indicate significant differences at the 10%, 5%, and 1% level.

	$\hat{\mu}_d \times 1,000$				$\hat{\sigma}_d \times 1,000$				$VCV_{XS,d}$			
d	Before	After	Diff		Before	After	Diff		Before	After	Diff	
0	0.86	0.86	0.00		1.55	1.55	0.00		1.81	1.81	0.00	
1	0.81	0.87	0.06	***	1.47	1.55	0.08	***	1.83	1.78	-0.04	**
2	0.75	0.86	0.11	***	1.39	1.52	0.13	***	1.85	1.77	-0.07	***
3	0.74	0.83	0.10	***	1.37	1.49	0.12	***	1.85	1.78	-0.07	***
4	0.73	0.82	0.09	***	1.36	1.46	0.11	***	1.87	1.80	-0.07	***
5	0.73	0.80	0.07	***	1.36	1.46	0.09	***	1.86	1.81	-0.05	**
6	0.73	0.79	0.06	***	1.36	1.43	0.07	***	1.86	1.81	-0.05	**
7	0.73	0.79	0.06	***	1.37	1.43	0.06	***	1.86	1.81	-0.05	**
8	0.74	0.79	0.05	***	1.37	1.43	0.06	***	1.86	1.81	-0.05	*
9	0.74	0.78	0.05	***	1.38	1.42	0.05	***	1.87	1.81	-0.05	**
10	0.74	0.78	0.03	***	1.38	1.42	0.04	***	1.86	1.82	-0.04	
11	0.75	0.78	0.02	***	1.39	1.41	0.02	***	1.85	1.82	-0.03	
12	0.75	0.78	0.02	***	1.39	1.41	0.02	***	1.85	1.82	-0.03	
13	0.75	0.77	0.02	***	1.39	1.41	0.02	***	1.84	1.82	-0.02	
14	0.76	0.77	0.02	***	1.40	1.41	0.01	***	1.85	1.82	-0.02	
15	0.76	0.77	0.01	*	1.40	1.41	0.01	**	1.84	1.83	-0.01	
16	0.76	0.77	0.01		1.39	1.40	0.01	***	1.83	1.83	0.00	
17	0.77	0.77	0.00		1.40	1.40	0.00		1.83	1.82	-0.01	
18	0.77	0.77	0.00		1.40	1.40	-0.00		1.83	1.82	-0.01	
19	0.77	0.77	0.00		1.40	1.40	-0.00		1.83	1.83	-0.01	
20	0.77	0.77	-0.00		1.40	1.40	-0.00		1.83	1.83	0.00	
21	0.77	0.77	-0.00		1.41	1.41	-0.00		1.83	1.83	0.00	
22	0.77	0.77	0.01	*	1.40	1.41	0.01	**	1.83	1.83	0.00	
23	0.77	0.77	0.00		1.41	1.41	-0.00		1.84	1.83	-0.01	
24	0.77	0.77	0.00		1.41	1.41	-0.00		1.84	1.83	-0.01	
25	0.77	0.77	0.00		1.41	1.41	0.00		1.83	1.83	-0.00	
26	0.77	0.77	0.00		1.40	1.41	0.01	*	1.82	1.83	0.01	
27	0.77	0.77	0.00		1.40	1.41	0.01	**	1.83	1.83	0.00	
28	0.77	0.77	0.00		1.41	1.41	0.00	*	1.83	1.83	-0.00	
29	0.77	0.77	0.00		1.41	1.41	0.00		1.84	1.83	-0.00	
30	0.77	0.77	-0.00		1.42	1.41	-0.01	*	1.84	1.83	-0.00	

prior to announcements and high following announcement, which is consistent with the patterns documented by Chae (2005) and Akbas (2016). The standard deviation of volume moves in the same direction as the mean, which could be due to be the increased illiquidity and price elasticity in the days before the announcement, as documented by George et al. (1994) and

Chae (2005). What we are most interested in is the pattern of VCV as a proxy for information asymmetry. Since the changes in the standard deviation are smaller in relative terms than the changes in the mean, VCV is high prior to the announcement and low afterwards. As Table 10 shows, the differences between VCV are statistically significant up to nine days before and after the announcement.

This pattern of VCV around earnings announcements is consistent with the hypothesis that information asymmetries are resolved around earnings announcements, and with previously documented behavior of alternative information asymmetry measures. Johnson and So (2018) report that the Multimarket Information Asymmetry (MIA) measure, calculated from the relative trading volume of options and stocks, increases in the days before earnings announcements, and rapidly declines around the announcement, similar to VCV. Chordia et al. (2019) find that the volatility of order flow, driven by correlated liquidity demand, significantly increases before earnings announcements. There is mixed evidence on the behavior of PIN around announcement dates. Benos and Jochec (2007), Back et al. (2018), and Duarte et al. (2020) find that PIN is in fact lower prior to earnings announcements and higher afterwards. Duarte et al. (2020) explain this puzzling result by demonstrating that the PIN measure mis-identifies asymmetric information when applied on a daily frequency, and instead simply indicates abnormal turnover. Easley et al. (2008), on the other hand, estimate a generalized PIN model in which the arrival rate of information is time-varying and find that PIN is high (low) before (after) earnings announcements, resembling the pattern of VCV in Figure 3.

We also consider surprising and non-surprising earnings announcements separately. We expect the S-shaped pattern around the announcement date to be more pronounced for surprising announcements, as these are more informative. Following Livnat and Mendenhall (2006), we define Standardized Unexpected Earnings (SUE) as the difference between actual reported earnings and the median analyst forecast over the 90 day-period prior to the announcement date reported in the IBES database, divided by the stock price at the end of the preceding quarter. Within each quarter, we then sort announcements into terciles based on

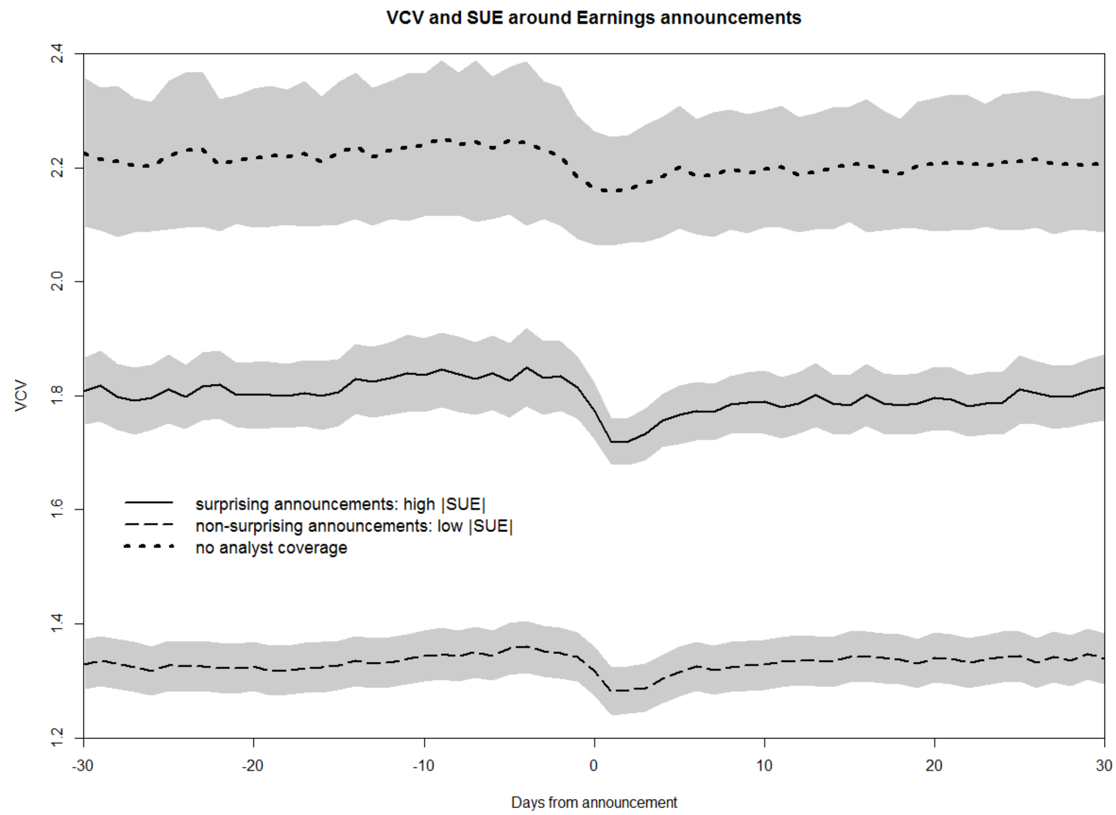

Figure 4: This figure shows the evolution of the daily cross-sectional VCV_{XS} over a 61 day event-window (day -30:30) around quarterly earnings announcements. The sample is divided into surprising announcements (solid line, $N = 55,340$), non-surprising announcements (dashed line, $N = 55,341$), and earnings announcement for which no analyst forecasts are reported (dotted line, $N = 173,235$). Surprising (non-surprising) announcements are defined as the announcements in the highest (lowest) tercile of announcements sorted on the absolute value of SUE: the difference between the actual and median analyst forecast of earnings, scaled by the price. See Figure 3 for details.

the absolute value of SUE.

Figure 4 reports the cross-sectional VCV around non-surprising announcement dates (tercile 1) and surprising announcement dates (tercile 3). The figure clearly shows a higher level of VCV for the surprising announcements, indicating a higher degree of information asymmetry. Around the announcements, the surprising announcements see a steeper drop, confirming that surprising announcements are more informative than non-surprising announcements. For both of these subsamples, the level of VCV is lower than it is for the full sample in Figure 3. This is because both the surprising and non-surprising subsamples are restricted to those firms for which analyst expectations are available. For completeness, Figure 4 also displays the cross-sectional VCV around the announcements for which no analyst forecasts are reported in IBES. The level of the cross-sectional VCV for the no-forecast sample is clearly higher, consistent with the negative relation between VCV and analyst coverage as reported in Tables 3 and 9. However, the S-shape around the announcement is flatter than for the sample of surprising announcements. Overall, the breakdown in Figure 4 reveals that both the level of the cross-sectional VCV and its dynamics around the announcement dates behave as expected, depending on the informativeness of the announcement, providing evidence that cross-sectional VCV efficiently captures information asymmetry in event time, and can be used in an event-study framework.

In Internet Appendix Section C.9, we also consider firm-level VCVs estimated from short time-series of 10 days before and after the announcement date and find on average a significant decline around the announcement. We also reproduce Figure 3 for various subsets of the data, showing a qualitatively similar pattern of VCV around earnings announcements for both NASDAQ and NYSE/AMEX stocks as well as before and after 2000. The drop in VCV around announcements has in fact become sharper post 2000. This result is consistent with Beaver et al. (2018) and Pawlewicz (2018), who find a recent increase in the information content of earnings announcements, and Weller (2018) who finds that price informativeness prior to earnings announcements has decreased, despite the presence of algorithmic trading.

We also analyze the cross sectional VCV around 8-K filings. Following the Sarbanes-Oxley Act of 2002, the SEC requires public companies since 2005 "to announce major events that shareholders should know about", by filing form 8-K (see Lerman and Livnat, 2010). We collect from SEC EDGAR all 8-K filings by US firms listed on NYSE, AMEX and NASDAQ. We exclude the event if the firm had a quarterly earnings announcement within one day of the filing, resulting in a total of $N = 162,529$ filing events between 2005 and 2015. Similar as for the earnings announcements, we compute the cross-sectional VCV for a 61 day window around the filing date, following Eq. (12).

The cross-sectional VCV in Figure 5 shows, similar to Figure 3, a drop in VCV around the disclosure of information. Interestingly, this drop already starts prior to the announcement, with a significant decline during the week before the event. This can be explained by the

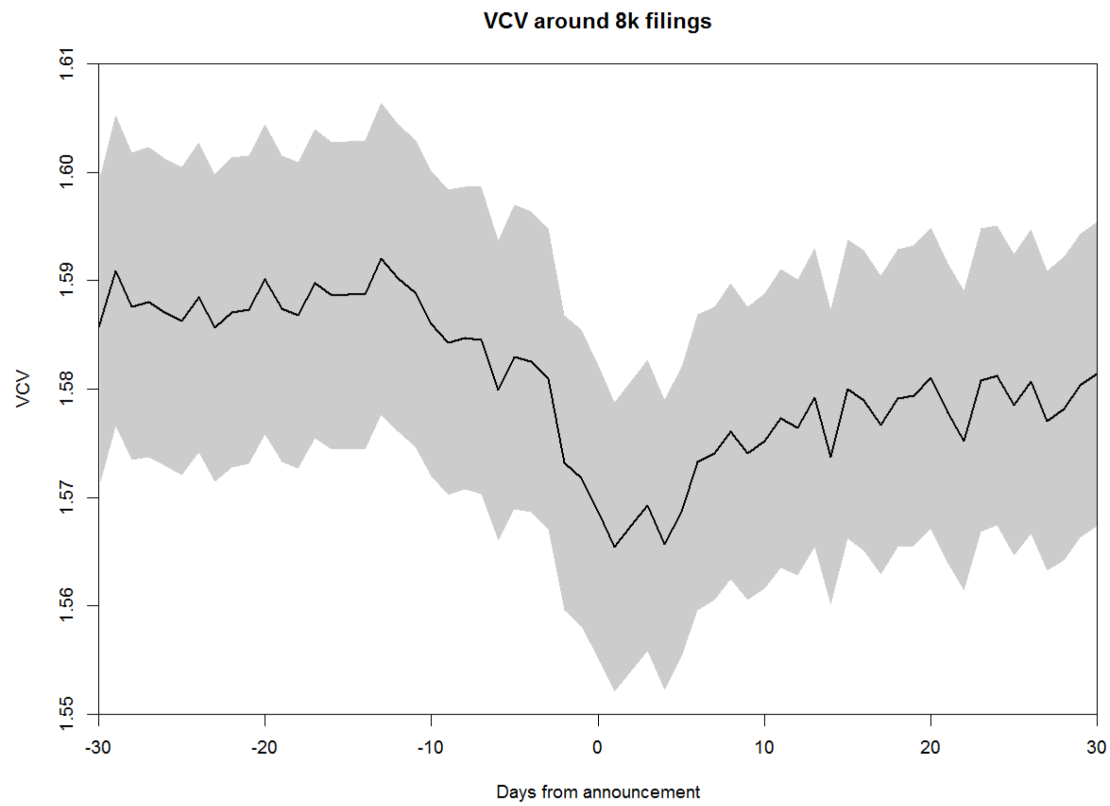

Figure 5: This figure shows the evolution of the daily cross-sectional VCV_{XS} around SEC form 8-K filings. The full sample includes all daily trading volumes over a 61 day event-window (day -30:30) around $N = 162,529$ filings (sources: CRSP and SEC EDGAR). The reported VCV at d days after the filing is estimated from the subsample of each stock's trading volume market shares at date d after each firm's filing. The gray shaded areas indicate 95% confidence intervals: $VCV_{XS,d} \pm 1.96 \times S.E.(VCV_{XS,d})$. Standard errors ($S.E.$) are derived following Albrecher et al. (2010).

requirement for companies to file form 8-K at most four days after the event that triggered the filing. As Ben-Rephael et al. (2017) point out, the event is often already public information at the time of the 8-K filing, such that the filing in itself contains little information. This is fully consistent with the decline in VCV, and thus in information asymmetry, prior to the filing date. Another difference between Figure 3 and 5 is that there is no substantial increase in VCV during the weeks prior to the 8-K filings. Unlike earnings announcements, 8-K filings are unscheduled, meaning that uninformed investors do not decrease their trading in anticipation of the 8-K filing, as they do prior to earnings announcements.

6 Conclusion

We use a microstructure model based on Kyle (1985) to demonstrate that the distribution of total trading volume depends on the proportion of informed (correlated) liquidity seeking demand. Specifically, we show that the coefficient of variation of trading volume increases in the proportion on informed trade. We therefore propose the sample coefficient of variation, VCV, as a measure of information asymmetry. Monte Carlo simulations confirm that VCV increases in the proportion of informed liquidity seekers, for a wide selection of model specifications.

Our empirical results indicate that stocks with high VCVs tend to have characteristics that are typically associated with asymmetric information (e.g.: high PIN, low breadth of institutional ownership, low analyst coverage, small size, low liquidity) and vice versa. Consistent with the hypothesis that informed trade is predictive of future price changes, we find that short-term return reversals are weaker for high VCV stocks, confirming that VCV is not just a measure of illiquidity. Our finding that VCV significantly increases following exogenous reductions in analyst overage due to brokerage closures, provides further evidence that VCV captures information asymmetry.

We introduce the cross-sectional VCV, which can be applied to evaluate information asymmetry in event time, e.g. following corporate disclosures, regulatory changes, or other information events. We apply this measure to quarterly earnings announcements and find, consistent with prior research, that asymmetric information is higher shortly before the announcement, and lower afterwards. In addition, we find that VCV decreases significantly around the disclosure of corporate events through form 8-K. Collectively, our empirical results provide broad support for the hypothesis that VCV is a measure of informed trading not only within our stylized microstructure model, but also when applied to real world data.

VCV is an appealing proxy for information asymmetry because of its simplicity: computing VCV, by dividing the sample standard deviation of trading volumes over the sample mean, is very straightforward. Unlike alternative measures of information asymmetry, estimating VCV requires only total trading volumes, and can be implemented both in cross-sections and in time-series. As a simple measure derived from widely available volume data, VCV can contribute positively to the reproducibility of research (Hail et al., 2020). VCV is applicable to any security for which trading volume is observable, including stocks, bonds, asset-backed securities, credit-default swaps, options and other derivatives. In a recent paper, Ghosh et al. (2020) apply VCV to study the effect of the adoption of the International Accounting Standard 40 (IAS 40) on information asymmetry in the market for REITS in the European Union. The potential applications of our measure are numerous. For example, VCV can be used as a control variable when there is a need to control for information asymmetry, as a sorting characteristic when studying the pricing effects of asymmetric information, or as the dependent variable of interest to compare patterns in information asymmetry across assets or over time.

References

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

Albrecher, H. , Ladoucette, S., and Teugels, J. (2010). Asymptotics of the sample coefficient of variation and the sample dispersion. *Journal of Statistical Planning and Inference*, 140(2), 358-368.

Akbas, F. (2016). The calm before the storm. *The Journal of Finance*, 71(1), 225-266.

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

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

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

Ball, R., and Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal* *of accounting research*, 159-178.

Beaver, W. H. (1968). The information content of annual earnings announcements. *Journal of accounting research*, 67-92.

Beaver, W.H., McNichols, M.F., and Wang, Z.Z. (2018). The information content of earnings announcements: new insights from intertemporal and cross-sectional behavior. *Review of Accounting Studies*, 23(1), 95-135.

Ben-Rephael, A., Da, Z., Easton, P. D., and Israelsen, R. D. (2017). Who Pays Attention to SEC Form 8-K?. Working paper.

Benos, E., and Jochec, M. (2007). Testing the PIN variable. Working paper.

Beyer, A., Cohen, D. A., Lys, T. Z., and Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. *Journal of accounting and economics*, 50(2-3), 296-343.

Black, F. (1986). Noise. *Journal of Finance*, 41, 528-543.

Blankespoor, E., Dehaan, E., Wertz, J., and Zhu, C. (2019). Why do individual investors disregard accounting information? The roles of information awareness and acquisition costs. *Journal of Accounting Research*, 57(1), 53-84.

Boehmer, E., Gramming, J., and Theissen, E. (2007). Estimating the probability of informed trading – does trade misclassification matter? *Journal of Financial Markets*, 10, 26-47.

Bongaerts, D., Rösch, D., and Van Dijk, M. A. (2016). Cross-sectional identification of informed trading. Working paper.

Boone, A., and White, J. T. (2015). The effect of institutional ownership on firm transparency and information production. *Journal of Financial Economics*, 117 , 508-533.

Brockman, P., and Yan, X.S. (2009). Block ownership and firm-specific information. *Journal of Banking and Finance*, 33, 308-316.

Brogaard, J., Carrion, A., Moyaert, T., Riordan, R., Shkilko, A., and Sokolov, K. (2018). High frequency trading and extreme price movements. *Journal of Financial Economics*, 128(2), 253-265.

Brown, S., and Hillegeist, S. A. (2007). How disclosure quality affects the level of information asymmetry. *Review of Accounting Studies*, 12(2-3), 443-477.

Brown, S., Hillegeist, S. A., and Lo, K. (2004). Conference calls and information asymmetry. *Journal of Accounting and Economics*, 37, 343-366.

Brown, S., Hillegeist, S. A., and Lo, K. (2009). The effect of earnings surprises on information asymmetry. *Journal of Accounting and Economics*, 47(3), 208-225.

Bushee, B. J. (2001). Do institutional investors prefer near-term earnings over long-run value?. *Contemporary Accounting Research*, 18(2), 207-246.

Bushee, B. J., an Leuz, C. (2005). Economic consequences of SEC disclosure regulation: Evidence from the OTC bulletin board. *Journal of accounting and economics*, 39(2), 233-264.

Bushee, B. J., and Noe, C. F. (2000). Corporate disclosure practices, institutional investors, and stock return volatility. *Journal of accounting research*, 171-202.

Bushman, R. M., Williams, C. D., and Wittenberg-Moerman, R. (2017). The informational role of the media in private lending. *Journal of Accounting Research*, 55(1), 115-152.

Chae, J. (2005). Trading volume, information asymmetry, and timing information. *Journal of Finance*, 60, 413-442.

Chen, Z., Dhaliwal, D. S., and Xie, H. (2010). Regulation fair disclosure and the cost of equity capital. *Review of Accounting Studies*, 15(1), 106-144.

Chen, J., Hong, H., and Stein, J. C. (2002). Breadth of ownership and stock returns. *Journal of Financial Economics*, 66, 171-205.

Chen, T., and Lin, C. (2017). Does Information Asymmetry Affect Corporate Tax Aggressiveness?. *Journal of Financial and Quantitative Analysis*, 52(5), 2053-2081.

Chordia, T, Subrahmanyam, A., and Anshuman, R. V. (2001). Trading activity and expected stock returns. *Journal of Financial Economics* 59, 3-32.

Chordia, T., Hu, J., Subrahmanyam, A., and Tong, Q. (2019). Order flow volatility and equity costs of capital. *Management Science* 65, 1520-1551.

Clark, Charles E. (1961). The greatest of a finite set of random variables, *Operations Research*, 9, 2, 145-162.

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

Derrien, F., and Kecskes, A. (2013). The real effects of financial shocks: evidence from exogenous changes in analyst coverage. *Journal of Finance*, 68(4), 1407-1440.

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

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

Easley, D., Engle, R. F., O'Hara, M., and Wu, L. (2008). Time-varying arrival rates of informed and uninformed trades. *Journal of Financial Econometrics*, 6(2), 171-207.

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

Easley, D., Hvidkjaer, S., and O'Hara, M. (2010). Factoring information into returns. *Journal of Financial & Quantitative Analysis*, 45, 293-309.

Easley, D., Kiefer, N.M., and O'Hara, M. (1997a). One day in the life of a very common stock. *Review of Financial Studies*, 10, 805-835.

Easley, D., Kiefer, N.M., and O'Hara, M. (1997b). The information content of the trading process. *Journal of Empirical Finance*, 4, 159-186.

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

Easley, D., De Prado, M. L., and O'Hara, M. (2011). The microstructure of the flash crash: Flow toxicity, liquidity crashes and the probability of informed trading. *Journal of Portfolio Management*, 37(2), 118-128.

Easley, D., Lopez de Prado, M. and O'Hara, M. (2012). Flow toxicity and liquidity in a high frequency world. *Review of Financial Studies*, 25, 1457-1493.

Easley, D., O'Hara, M., and Paperman, J.B. (1998). Financial analysts and information based trade. *Journal of Financial Markets*, 1, 175-201.

Fich, E.M., Harford, J., and Tran, A. L. (2015). Motivated monitors: The importance of institutional investors? portfolio weights. *Journal of Financial Economics*, 118, 21-48.

Foster, F. D., and Viswanathan, S. (1990). A theory of the interday variations in volume, variance, and trading costs in securities markets. *The Review of Financial Studies*, 3(4), 593-624.

George, T. J., Kaul, G., and Nimalendran, M. (1994). Trading volume and transaction costs in specialist markets. *The journal of Finance*, 49(4), 1489-1505.

Ghosh, C., Liang, M., and Petrova, M. T. (2020). The Effect of Fair Value Method Adoption: Evidence from Real Estate Firms in the EU. *The Journal of Real Estate Finance and Economics*, 60(1), 205-237.

Glode, V., and Opp, C. (2016). Asymmetric information and intermediation chains. *American Economic Review*, 106(9), 2699-2721.

Gramming, J., Schiereck, D., and Theissen, E. (2001). Knowing me, knowing you: trader anonymity and informed trading in parallel markets. *Journal of Financial Markets*, 4, 385-412.

Grundy, B. D., and McNichols, M. (1989). Trade and the revelation of information through prices and direct disclosure. *Review of financial Studies*, 2(4), 495-526.

Hail, L., Lang, M., and Leuz, C. (2020). Reproducibility in Accounting Research: Views of the Research Community. *Journal of Accounting Research*, 58(2), 519-543.

Hameed, A., Hong, D., and Warachka, M. (2008). Momentum and informed trading. Working paper.

Healy, P. M., Hutton, A. P., and Palepu, K. G. (1999). Stock performance and intermediation changes surrounding sustained increases in disclosure. *Contemporary accounting research*, 16(3), 485-520.

Heflin, F. L., Shaw, K. W., and Wild, J. J. (2005). Disclosure policy and market liquidity: Impact of depth quotes and order sizes. *Contemporary accounting research*, 22(4), 829-865.

Hendershott, T., Jones, C. M., and Menkveld, A. J. (2011). Does algorithmic trading improve liquidity?. *The Journal of Finance*, 66(1), 1-33.

Holden, C. W., and Subrahmanyam, A. (1992). Long-lived private information and imperfect competition. *Journal of Finance*, 47, 247-270.

Huang, R., and Stoll H.R. (1997) The components of the bid-ask spread: a general approach, *Review of Financial Studies*, 19, 995-1033.

Hwang, L. S., Lee, W. J., Lim, S. Y., and Park, K. H. (2013). Does information risk affect the implied cost of equity capital? An analysis of PIN and adjusted PIN. *Journal of Accounting and Economics*, 55(2-3), 148-167.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance*, 45(3), 881-898.

Johnson, T. L., and So, E. C. (2018). A simple multimarket measure of information asymmetry. *Management Science*, 64(3), 1055-1080.

Jovanovic, B., and Menkveld, A. J. (2016). Middlemen in limit order markets. Working paper.

Kelly, B., and Ljungqvist, A. (2012). Testing asymmetric-information asset pricing models. *The Review of Financial Studies*, 25(5), 1366-1413.

Kirilenko, A., Kyle, A. S., Samadi, M., and Tuzun, T. (2017). The flash crash: High-frequency trading in an electronic market. *The Journal of Finance*, 72(3), 967-998.

Kumar, A., and Lee, C. M. (2006). Retail investor sentiment and return comovements. *The Journal of Finance*, 61(5), 2451-2486.

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

LaFond, R., and Watts, R. L. (2008). The information role of conservatism. *The Accounting Review*, 83(2), 447-478.

Lee C. and Ready M. (1991). Inferring trade direction from intradaily data. *Journal of Finance*, 46, 733-746.

Lerman, A., and Livnat, J. (2010). The new Form 8-K disclosures. *Review of Accounting Studies*, 15(4), 752-778.

Li, K.K., and You, H. (2015). What is the value of sell-side analysts? Evidence from coverage initiations and terminations. *Journal of Accounting and Economics*, 60(2-3), 141-160.

Livnat, J., and Mendenhall, R. R. (2006). Comparing the post-earnings announcement drift for surprises calculated from analyst and time series forecasts.*Journal of accounting research*, 44(1), 177-205.

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

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

Mohanram, P., and Rajgopal, S. (2009). Is PIN priced risk?. *Journal of Accounting and Economics*, 47(3), 226-243.

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

Pawlewicz, R. (2018). The effect of regulation on the timeliness and informational role of earnings announcements. *Contemporary Accounting Research*, 35(4), 1675-1701.

Petacchi, R. (2015). Information asymmetry and capital structure: Evidence from regulation FD. *Journal of Accounting and Economics*, 59(2-3), 143-162.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39(4), 1127-1139.

Rosu, I. (2019). Fast and slow informed trading. *Journal of Financial Markets* 43, 1-30.

Vega, C. (2006). Stock price reaction to public and private information. *Journal of Financial Economics*, 82, 103-133.

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

Welker, M. (1995). Disclosure policy, information asymmetry, and liquidity in equity markets. *Contemporary accounting research*, 11(2), 801-827.

Weller, B. M. (2018). Does algorithmic trading reduce information acquisition?. *The Review of Financial Studies*, 31(6), 2184-2226.

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

Asymmetric Information and the Distribution of Trading Volume

Internet Appendix

March 19, 2021

Contents

A Proofs and model extensions

In the first subsection, we present an extension of the model in Section 2 of the paper, in which the proportion of informed trade η is determined endogenously in equilibrium. The second subsection presents the derivation of the variance of trading volume (Required for Eq.(5) and Proposition 1 in the paper), as well as the proof for Proposition A.1.

A.1 Strategic informed trading

In this subsection, we demonstrate that the insights from our simple model also hold in a setting where the proportion of informed trade is endogenously determined, as in Kyle (1985). We now consider a model where multiple informed traders, with correlated noisy signals, choose their orders strategically while taking into account the strategies of the other informed traders.

In particular, we assume that there are m informed liquidity seekers, n uninformed liquidity seekers, and, as common in the literature, sufficient liquidity providers for them to act competitively. We assume a zero interest rate and risk neutrality of all market participants. The informed traders place orders, denoted by x_j, after receiving a signal s_j equal to the liquidation value (v) plus an independent noise term: $s_j = v + \varepsilon_j$. The orders of the uninformed, denoted u_i, are *i.i.d.*, Normally distributed with mean zero and standard deviation σ_u. Due to risk neutrality and zero-interest, the ex-ante expected liquidation value $E[v]$ equals the previous clearing price, p_0.

The $n + m$ individual liquidity seekers submit orders to the market where buy orders are matched to sell orders and the order imbalance $\sum_n u_i + \sum_m x_j$, is taken up by liquidity providers who set the price at the expected value given the order imbalance. Our model is thus a modified Kyle (1985) model, in which multiple imperfectly informed insiders compete. Similar models have been analyzed by Holden and Subrahmaniam (1992), Foster and Viswanathan (1994, 1996) and others.

We use the terminology and symbols of Kyle (1985), and look for the linear equilibrium in which the informed traders choose their trade as a linear function of their signal and the last traded price p_0:

$$x_j = \beta_j(s_j - p_0), \tag{A.1}$$

and the competitive market makers use the following linear pricing function:

$$p = p_0 + \lambda \left(\sum_n u_i + \sum_m x_j \right). \tag{A.2}$$

In equilibrium β_j and λ are determined jointly: the βs follow from the profit optimization problem of the informed traders, who take λ, s_j, p_0, and the other parameters (n, m, σ_u, σ_v,σ_ε) as given, and (Kyle's) λ is determined by the liquidity providers who set the price at the expected value given the observed order imbalance and knowledge on the trading aggressiveness of the informed investors (i.e. the βs).

The m profit maximizing informed investors each solve:

$$\max_{x_j} x_j(E[v|s_j] - p_0 - \lambda(x_j + E[\sum_n u_i + \sum_{m-1} x_{-j}|s_j]), \tag{A.3}$$

where x_{-j} denotes the orders of the informed traders other than j. The first order condition is $x_j^*(s_j) = \frac{s_j - p_0}{2\lambda} - \frac{m-1}{2} E[x_{-j}|s_j]$. Since the signal's noise components ε_j are *i.i.d.*, the final term, $E[x_{-j}|s_j]$, equals $\beta_{-j}(s_j - p_0)$, where β_{-j} is the trading aggressiveness for all traders except j. Hence, all traders set their demand following $x_j^*(s_j) = (s_j - p_0)\left(\frac{1}{2\lambda} - \frac{m-1}{2}\beta_{-j}\right)$, so that in equilibrium we have $\beta_j = \beta_{-j} = \beta = \frac{1}{\lambda(m+1)}$.

Simultaneously, the market makers set the equilibrium price at the expected value of v, conditional on the order imbalance $\sum_n u_i + \sum_m x_j$. From the projection theorem, we know that

$$E[v|\sum_n u_i, \sum_m x_j; p_0, \beta, \sigma_v, \sigma_u, n, m] = p_0 + \frac{m\beta\left(\sigma_v^2 + \sigma_\varepsilon^2\right)}{n\sigma_u^2 + m^2\beta^2\left(\sigma_v^2 + \sigma_\varepsilon^2\right)} \left(\sum_n u_i + \sum_m x_j \right), \tag{A.4}$$

implying that $\lambda = \frac{m\beta\left(\sigma_v^2 + \sigma_\varepsilon^2\right)}{n\sigma_u^2 + m^2\beta^2(\sigma_v^2 + \sigma_\varepsilon^2)}$. Combining these two results, we find that in equilibrium:

$$\beta = \frac{\sqrt{n}\sigma_u}{\sqrt{m\left(\sigma_v^2 + \sigma_\varepsilon^2\right)}}; \quad \lambda = \frac{\sqrt{m\left(\sigma_v^2 + \sigma_\varepsilon^2\right)}}{(m+1)\sqrt{n}\sigma_u}. \tag{A.5}$$

We now express trading volume as a function of the model's parameters and observe that total trading volume can now be written, similar to Eq.(1) in the paper, as:

$$V = \frac{1}{2} \left(\sum_n |u_i| + \sum_m |x_j| + \left| \sum_n u_i + \sum_m x_j \right| \right). \tag{A.6}$$

The demands u_i and x_j both follow a Normal distribution around zero. We find from (A.1) and (A.5) that the variance of informed demand (σ_x^2) is only dependent on the variance of uninformed demand (σ_u^2) and the ratio of uninformed to informed investors:

$$\sigma_x^2 = \beta^2 \left(\sigma_v^2 + \sigma_\varepsilon^2\right) = \frac{n}{m}\sigma_u^2. \tag{A.7}$$

The intuition of (A.7) is that the trading aggressiveness of each informed investor increases in the number of uninformed investors, and decreases in the number of informed investors. The distribution of trading volume thus depends only on n, m, and σ_u and is, unlike the distribution of prices and returns, independent of σ_v and σ_ε. Given that all components of (A.6) follow (correlated) Normal distributions, we can find the first two moments of total trading volume from integration. We find that both the mean and the standard deviation of volume are linear in σ_u. In particular we have:

<u>Proposition A.1</u>

Consider a market where n uninformed liquidity seeking traders submit Normally distributed market orders with mean zero and variance σ_u^2; m informed liquidity seeking traders, who receive noisy signals on the asset's liquidation value, submit Normally distributed market orders with mean zero and variance $\beta^2\left(\sigma_v^2 + \sigma_\varepsilon^2\right)$; and the order imbalance is absorbed by competitive liquidity suppliers. In equilibrium:

i. *The expected value of trading volume is given by:*

$$E[V] = \frac{\sigma_u}{\sqrt{2\pi}}(n + \sqrt{nm} + \sqrt{n(m+1)}). \tag{A.8}$$

ii. *The variance of trading volume is given by:*

$$\begin{aligned} Var(V) \quad = \quad & 2n\sigma_u^2 \textstyle\int_0^\infty x^2 (m\Phi(\sqrt{m}x) + \Phi(\frac{x}{\sqrt{mn+n-1}}))\phi(x)dx \\ & + \sigma_u^2 \frac{n\sqrt{m}(1-(m+1)^{\frac{3}{2}}) + (mn+n-1)^{\frac{3}{2}} - (mn+n)^{\frac{3}{2}} - n(m+1)^2}{\pi(m+1)}, \end{aligned} \tag{A.9}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the probability density function and cumulative density function of the Standard Normal distribution.

iii. *The coefficient of variation of trading volume $\frac{\sigma_V}{\mu_V}$ is a function of n and m only. For a given number of uninformed traders n, $\frac{\sigma_V}{\mu_V}$ increases in the number of informed traders m. For a given m, $\frac{\sigma_V}{\mu_V}$ decreases in n.*

Proof: See Internet Appendix A.2.2.

Panel A of Figure A.1 graphically depicts the relationship between $\frac{\sigma_V}{\mu_V}$ and the number of uninformed and informed traders. For the case where $n = m = 1$, $\frac{\sigma_V}{\mu_V}$ is equal to $\frac{\sqrt{3\pi-4\sqrt{2}}}{2+\sqrt{2}}$. There is no closed form solution for all other finite (n, m) combinations. The figure shows that for a given n, $\frac{\sigma_V}{\mu_V}$ is a concave increasing function of the number of informed investors. As m goes to infinity, $\frac{\sigma_V}{\mu_V}$ approaches $\frac{1}{2}\sqrt{2\pi - 4} \approx 0.756$, which is the coefficient of variation of the Half-Normal distribution and the coefficient of variation implied by our earlier model (Proposition 1) with $\eta = 1$. Additionally, for a given number of informed traders m, $\frac{\sigma_V}{\mu_V}$ is decreasing in the number of uninformed traders n. When n is very large relative to m, $\frac{\sigma_V}{\mu_V}$ approaches zero, as in Proposition 1 with $\eta = 0$. Panel A also shows that $\frac{\sigma_V}{\mu_V}$ decreases in $\frac{n}{m}$ along any diagonal with constant $(n + m)$.

Panel B of Figure A.1 shows how the price elasticity of net order flow (Kyle's λ) varies with n and m. It is interesting to see that λ shows a very different pattern than $\frac{\sigma_V}{\mu_V}$. Kyle's λ is not a strictly increasing function of the proportion of informed traders (for a given number of traders): moving along diagonals with constant $n + m$, we see that λ is a non-monotonic convex function of $\frac{n}{m}$. The intuition for this pattern is that an increase in the number of informed traders increases the total informed order flow (and price informativeness), thereby lowering the price elasticity. As the number of informed traders increases further, they become less aggressive (i.e. β declines), reducing λ.

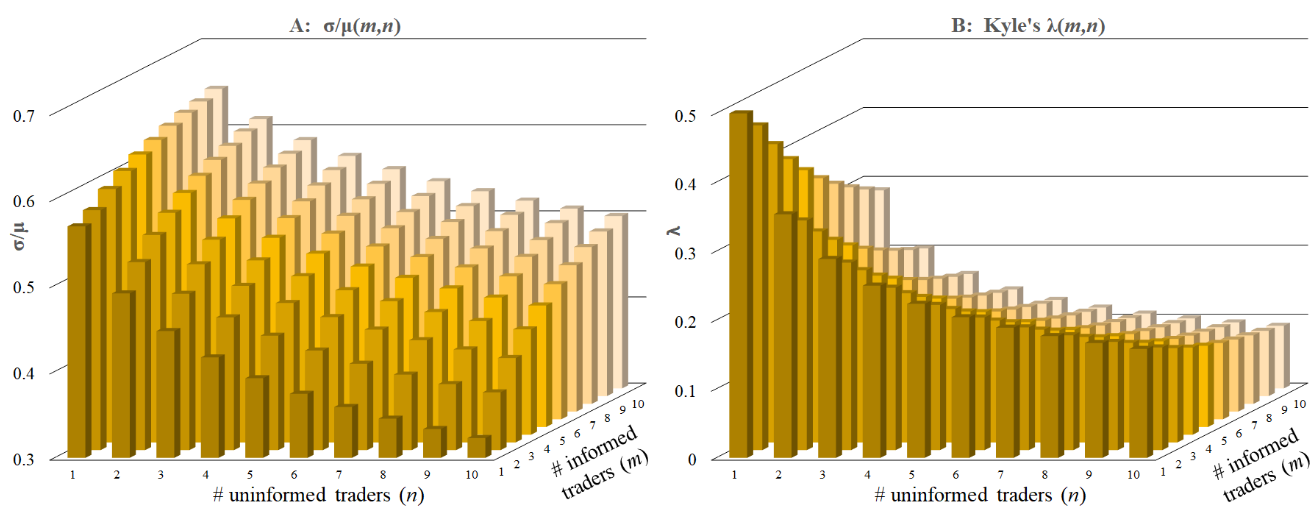

Figure A.1: Coefficient of variation of trading volume $\frac{\sigma_V}{\mu_V}$ (Panel A) and Kyle's λ (Panel B) as a function of n uninformed and m informed traders. Kyle's λ (Eq. A.5) is divided by $\sqrt{\frac{\sigma_v^2+\sigma_\varepsilon^2}{\sigma_u^2}}$ to make it invariant to σ_ε, σ_v, and σ_u.

The model outlined in this subsection is one specific example of how the proportion of informed trade η can be endogenously determined in equilibrium. In this example, η is a function of the numbers of uninformed and informed traders only. We can see that the assumption of equal trading intensity for all traders in the prior subsection is for convenience only, and that we can define η and M as:

$$\eta = \frac{\sigma_x m}{\sigma_x m + \sigma_u n}; \qquad M = n + \frac{\sigma_x}{\sigma_u} m. \tag{A.10}$$

That is, M is a measure of total *trading activity*, in which the number of individual traders are weighted by their trading intensities, while η is the proportion of informed *trade*, rather than the proportion of informed *traders*.

If informed traders are risk neutral and receive signals with *i.i.d.* noise terms, we find from Eq.(A.7) that $\eta = \frac{\sqrt{nm}}{\sqrt{nm}+n}$ and $M = n + \sqrt{nm}$. Further enriching the model with risk aversion, long lived information, or strategic uninformed trading (as in Admati and Pfleiderer, 1988, or Foster and Viswathan, 1990) will change the above expressions for the equilibrium η and M, but will not change Proposition 1, as there will always be a proportion of informed trade, and an equivalent number of market participants. Indeed, VCV will be a consistent estimator of $\frac{\sigma_V}{\mu_V}$ and a useful measure to answer the empirical question of how information asymmetry depends on (or causes) variation in discretionary informed or uninformed trading, or on trading or reporting circumstances.

A.2 Variance of trading volume

A.2.1 Benchmark model (section 2 of the paper)

Define $Y_{MM} = |\sum_M y_i|$ as the part of double-counted volume traded by liquidity providers (the order imbalance), $Y_I = \sum_{1 \ldots \eta M} |y_i|$ as the part traded by informed liquidity seekers and $Y_U = \sum_{\eta M+1 \ldots M} |y_i|$ as the part traded by uninformed liquidity seekers. Then Eq.(1) in the paper can be rewritten as:

$$V = \tfrac{1}{2} \left(Y_I + Y_U + Y_{MM} \right). \tag{A.11}$$

The variance of double-counted trading volume is given by:

$$\begin{aligned} Var\left(2V\right) &= Var\left(Y_I\right) + Var\left(Y_U\right) + Var\left(Y_{MM}\right) \\ &\quad +2Cov\left(Y_I, Y_U\right) + 2Cov\left(Y_I, Y_{MM}\right) + 2Cov\left(Y_U, Y_{MM}\right). \end{aligned} \tag{A.12}$$

Using the properties of the Half Normal distribution, we find that:

$$\begin{aligned} Var\left(Y_I\right) &= \eta^2 M^2 \sigma^2 \left(1-\tfrac{2}{\pi}\right) \\ Var\left(Y_U\right) &= (1-\eta) M \sigma^2 \left(1-\tfrac{2}{\pi}\right) \\ Var\left(Y_{MM}\right) &= \left(\eta^2 M^2 + (1-\eta)M\right) \sigma^2 \left(1-\tfrac{2}{\pi}\right). \end{aligned} \tag{A.13}$$

$Cov\left(Y_I, Y_U\right) = 0$, because the demands of informed and uninformed liquidity seekers are independent. Moreover, when M is large and $\eta > 0$, the order imbalance consists mainly of orders submitted by informed liquidity seekers. The orders of uninformed traders tend to net out against each other because of the *i.i.d* property. This implies that in the limit ($M \to \infty$), the liquidity suppliers trade exclusively to offset the imbalance from informed seekers. Therefore, $\lim_{M\to\infty} Cor\left(Y_U, Y_{MM}\right) = 0$ and $\lim_{M\to\infty} Cor\left(Y_I, Y_{MM}\right) = 1$. Given these correlations, Eq.(A.12) implies that when $M \to \infty$:

$$Var\left(\tfrac{2V}{M}\right) = Var\left(\tfrac{Y_I}{M}\right) + Var\left(\tfrac{Y_U}{M}\right) + Var\left(\tfrac{Y_{MM}}{M}\right) + 2\sqrt{Var\left(\tfrac{Y_I}{M}\right) Var\left(\tfrac{Y_{MM}}{M}\right)}, \tag{A.14}$$

which, given the variances in Eq.(A.13), results in:

$$\begin{aligned} Var\left(\tfrac{2V}{M}\right) &= \eta^2\sigma^2\left(1-\tfrac{2}{\pi}\right) + (1-\eta)M^{-1}\sigma^2\left(1-\tfrac{2}{\pi}\right) + \left(\eta^2 + (1-\eta)M^{-1}\right)\sigma^2\left(1-\tfrac{2}{\pi}\right) \\ &\quad + 2\sqrt{\eta^2\sigma^2\left(1-\tfrac{2}{\pi}\right)}\sqrt{\left(\eta^2 + (1-\eta)M^{-1}\right)\sigma^2\left(1-\tfrac{2}{\pi}\right)} \\ &= 2\sigma^2\left(1-\tfrac{2}{\pi}\right)\left(\eta^2 + (1-\eta)M^{-1} + \eta\sqrt{\eta^2 + (1-\eta)M^{-1}}\right), \\ &= 4\sigma^2\left(1-\tfrac{2}{\pi}\right)\eta^2, \end{aligned} \tag{A.15}$$

where the last step follows from $M^{-1} \to 0$ for large M. The standard deviation of trading volume divided by M thus equals to $\sigma\eta\sqrt{1-\frac{2}{\pi}}$, from which Proposition 1 is easily derived:

$$\lim_{M\to\infty} \frac{s.d.(V)}{E\left[V\right]} = \lim_{M\to\infty} \frac{s.d.(V/M)}{E\left[V/M\right]} = \sqrt{2\pi - 4}\frac{\eta}{\eta + 1}. \tag{A.16}$$

A.2.2 Proof of Proposition A.1

The expected value of trading volume is found by applying the properties of the Half-Normal distribution, given that u_i, x_j and $\left(\sum_n u_i + \sum_m x_j\right)$ all follow a Normal distribution around zero. To evaluate the variance of the trading volume we use the following lemma:

Lemma: If r and s are two i.i.d. random variables from the Standard Normal distribution, and α *is a positive scalar,* *we have:*

$$Cov(|r|,|r+\alpha s|) = 4\int_0^\infty r^2\Phi(\frac{r}{\alpha})\phi(r)dr + \frac{2\alpha^3 - 2(\alpha^2+1)^{\frac{3}{2}}}{(\alpha^2+1)\pi} - 1, \tag{A.17}$$

Where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the probability density function and cumulative density function of the Standard Normal distribution.

Proof:

$$\begin{aligned} Cov(|r|,|r+\alpha s|) &= E[|r||r+\alpha s|] - E[|r|]E[|r+\alpha s|] \\ &= E[|r||r+\alpha s|] - \tfrac{2}{\pi}\sqrt{1+\alpha^2}. \end{aligned} \tag{A.18}$$

We evaluate the first term by integration:

$$\begin{aligned} E[|r||r+\alpha s|] &= \textstyle\iint |r(r+\alpha s)| d\phi(r) d\phi(s) \\ &= \textstyle\int_{-\infty}^{0} (\int_{-\infty}^{-\frac{r}{\alpha}} (r^2+\alpha rs)\phi(s)ds + \int_{-\frac{r}{\alpha}}^{\infty} (-r^2-\alpha rs)\phi(s)ds)dr \\ &\quad \textstyle + \int_0^\infty (\int_{-\infty}^{-\frac{r}{\alpha}} (-r^2-\alpha rs)\phi(s)ds + \int_{-\frac{r}{\alpha}}^{\infty} (r^2+\alpha rs)\phi(s)ds)\phi(r)dr \\ &= \textstyle\int_{-\infty}^{0} (r^2\Phi(-\frac{r}{\alpha}) - \frac{\alpha r}{\sqrt{2\pi}}e^{-\frac{r^2}{2\alpha^2}} - r^2(1-\Phi(-\frac{r}{\alpha})) - \frac{\alpha r}{\sqrt{2\pi}}e^{-\frac{r^2}{2\alpha^2}})\phi(r)dr \\ &\quad \textstyle + \int_0^\infty (-x^2\Phi(-\frac{r}{\alpha}) + \frac{\alpha r}{\sqrt{2\pi}}e^{-\frac{r^2}{2\alpha^2}} + r^2(1-\Phi(-\frac{r}{\alpha})) + \frac{\alpha r}{\sqrt{2\pi}}e^{-\frac{r^2}{2\alpha^2}})\phi(r)dr \\ &= \textstyle 2\int_{-\infty}^{0} r^2\Phi(-\frac{r}{\alpha})\phi(r)dr - \int_{-\infty}^{0} r^2\phi(r)dr - \frac{2\alpha}{2\pi}\int_{-\infty}^{0} re^{-\frac{r^2}{2\alpha^2}-\frac{r^2}{2}}dr \\ &\quad \textstyle -2\int_0^\infty r^2\Phi(-\frac{r}{\alpha})\phi(r)dr + \int_0^\infty r^2\phi(r)dr + \frac{2\alpha}{2\pi}\int_0^\infty re^{-\frac{r^2}{2\alpha^2}-\frac{r^2}{2}}dr \\ &= \textstyle 2\int_0^\infty r^2\Phi(\frac{r}{\alpha})\phi(r)dr - \frac{1}{2} + \frac{\alpha^3}{(\alpha^2+1)\pi} \\ &\quad \textstyle -2\int_0^\infty r^2(1-\Phi(\frac{r}{\alpha}))\phi(r)dr + \frac{1}{2} + \frac{\alpha^3}{(\alpha^2+1)\pi} \\ &= \textstyle 4\int_0^\infty r^2\Phi(\frac{r}{\alpha})\phi(r)dr + \frac{2\alpha^3}{(\alpha^2+1)\pi} - 2\int_0^\infty r^2\phi(r)dr \\ &= \textstyle 4\int_0^\infty r^2\Phi(\frac{r}{\alpha})\phi(r)dr + \frac{2\alpha^3}{(\alpha^2+1)\pi} - 1. \end{aligned} \tag{A.19}$$

Substitute (A.19) into (A.18) to obtain the lemma (A.17).

To evaluate the variance of the trading volume we use:

$$\begin{aligned} Var(2V) &= \textstyle\sum_n var(|u_i|) + var(|\sum_m x_j|) + var(|z|) \\ &\quad \textstyle +2Cov(|\sum_m x_j|,|z|) + 2\sum_n Cov(|u_i|,|z|), \end{aligned} \tag{A.20}$$

where $z = \sum_m x_i + \sum_n u_i$. Note that we can consider the total informed demand, $x \equiv \sum_m x_j = m\beta v = \sqrt{mn}\frac{\sigma_u}{\sigma_v}v$ as a single random variable. Using again the properties of the Half-Normal distribution, we find that the variance terms are:

$$\begin{aligned}\sum_n Var(|u_i|) + Var(|x|) + Var(|z|) &= n\sigma_u^2(1-\tfrac{2}{\pi}) + mn\sigma_u^2(1-\tfrac{2}{\pi}) \\ &\quad +(1+m)n\sigma_u^2(1-\tfrac{2}{\pi}) \\ &= 2(m+1)n\sigma_u^2(1-\tfrac{2}{\pi}).\end{aligned} \tag{A.21}$$

The first covariance term in (A.20) can be evaluated as follows:

$$Cov(|\sum_m x_j|, |z|) = Cov(|x|, |x+u|), \tag{A.22}$$

where $x \equiv \sum_m x_j \sim N(0, mn\sigma_u^2)$ and $u \equiv \sum_n u_i \sim N(0, n\sigma_u^2)$, and x and u are independent. From the Lemma, it follows that:

$$Cov(|x|, |x+u|) = mn\sigma_u^2(4\int_0^\infty x^2\Phi(\sqrt{m}x)\phi(x)dx + \frac{2(1-(m+1)^{\frac{3}{2}})}{\pi\sqrt{m}(m+1)} - 1). \tag{A.23}$$

The final covariance terms in (A.20) are all identical, and can be evaluated as:

$$Cov(|u_i|, |z|) = Cov(|u_i|, |u_i + z_{-i}|), \tag{A.24}$$

where $u_i \sim N(0, \sigma_u^2)$ and $z_{-i} \equiv \sum_{n|i} u_j + x \sim N(0, (n-1+mn)\sigma_u^2)$ are two *i.i.d.* Normally distributed random variables. From the Lemma, it then follows that:

$$\begin{aligned}Cov(|u_i|, |z|) &= \sigma_u^2(4\int_0^\infty x^2\Phi(\tfrac{x}{\sqrt{mn+n-1}})\phi(x)dx \\ &\quad + \tfrac{2(mn+n-1)^{\frac{3}{2}} - 2(mn+n)^{\frac{3}{2}}}{\pi(mn+n)} - 1).\end{aligned} \tag{A.25}$$

Combining the variance terms (A.21), and the covariance terms (A.23) and (A.25) gives, after re-arranging, the variance of trading volume ($Var(V)$) as given in Proposition A.1.

B Detailed simulation results

In this section, we report detailed results for the Monte Carlo simulation exercise reported in Section 3 of the paper. To demonstrate the robustness of VCV, we also present simulation results for various modifications of our benchmark model, including Non-Gaussian demand, autocorrelated demand, stochastic liquidity demand and provision, heterogeneous information, and endogenous trading intensity. Table B.1 lists the various model specifications considered. Detailed simulation results are reported on the following pages. The various specifications considered in this appendix confirm the robustness of the result that VCV is strictly increasing in the proportion of informed trade η. The alternative measure $\widehat{\eta}$ (Eq. 8 in the paper) performs however rather poorly in small samples and in various modifications of the benchmark model.

Table B.1: Simulation exercise: Summary table

Table	Assumptions	Notes
B.2	Benchmark model: $T = 100$. Gaussian demand. Constant number of traders (M), trading intensity (σ), and proportion of informed traders (η).	
B.3	Small sample ($T = 10$)	Demonstrates that VCV can be estimated with smaller samples (e.g. monthly windows).
B.4	Non-Gaussian demand	VCV robust to skewness and excess kurtosis in trading volume (e.g. caused by outliers).
B.5	Dynamic / persistent demand	VCV robust to autocorrelation in trading volume.
B.6	Random proportion of informed trade (η)	VCV increases in *average* proportion of informed trade when applied to a sample of observations with different proportions of informed trade.
B.7	Random trading intensity (σ)	VCV robust to variation in trading volume unrelated to private information.
B.8	Random number of traders (M)	VCV robust to variation in trading volume unrelated to private information.
B.9	Random liquidity supply	VCV robust to variation in market maker capacity.
B.10	Heterogeneous information	VCV robust to different groups of informed traders trading on distinct signals.
B.11	Trading intensity of informed investors proportional to uninformed order flow	VCV robust to endogenous proportion of informed trade.

B.1 Benchmark simulations

Tables B.2 and B.3 provide detailed simulation results for the benchmark model (Selected results of these simulations are reported in Table 1 in the paper).

Table B.2: Simulation results: Benchmark model

This table reports the sample average, standard deviation, and selected percentiles of VCV (Panel A) and $\widehat{\eta}$ (Panel B), obtained from $R = 1,000,000$ replications of $T = 100$ volume realizations simulated from the model outlined in Section 2 in the paper, for various values of the proportion of informed trade η and number of liquidity seekers M.

					Panel A: VCV						
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
					$M = 10$						
Mean	0.28	0.28	0.31	0.36	0.42	0.48	0.54	0.60	0.65	0.70	0.74
s.d.	0.07	0.07	0.07	0.09	0.10	0.11	0.12	0.14	0.15	0.16	0.17
$q_{0.05}$	0.17	0.17	0.19	0.22	0.27	0.31	0.35	0.39	0.43	0.46	0.48
Median	0.28	0.28	0.3	0.35	0.41	0.47	0.53	0.59	0.64	0.69	0.73
$q_{0.95}$	0.40	0.40	0.44	0.51	0.59	0.67	0.76	0.83	0.91	0.98	1.05
					$M = 100$						
Mean	0.10	0.15	0.25	0.35	0.43	0.50	0.56	0.62	0.67	0.71	0.75
s.d.	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.04	0.05	0.05	0.06
$q_{0.05}$	0.09	0.13	0.22	0.31	0.38	0.45	0.50	0.55	0.59	0.63	0.67
Median	0.10	0.15	0.25	0.35	0.43	0.50	0.56	0.62	0.67	0.71	0.75
$q_{0.95}$	0.11	0.17	0.28	0.39	0.48	0.56	0.63	0.69	0.75	0.80	0.85
					$M = 1000$						
Mean	0.03	0.14	0.25	0.35	0.43	0.50	0.56	0.62	0.67	0.71	0.75
s.d.	0.00	0.01	0.02	0.02	0.03	0.03	0.04	0.04	0.05	0.05	0.06
$q_{0.05}$	0.03	0.12	0.22	0.31	0.38	0.45	0.50	0.55	0.59	0.63	0.67
Median	0.03	0.14	0.25	0.35	0.43	0.50	0.56	0.62	0.67	0.71	0.75
$q_{0.95}$	0.04	0.16	0.28	0.39	0.48	0.56	0.63	0.69	0.75	0.80	0.85
					Panel B: $\widehat{\eta}$						
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
					$M = 10$						
Mean	0.24	0.24	0.27	0.33	0.41	0.50	0.59	0.69	0.80	0.90	1.01
s.d.	0.02	0.02	0.03	0.03	0.04	0.05	0.07	0.08	0.10	0.12	0.15
$q_{0.05}$	0.20	0.20	0.23	0.28	0.34	0.41	0.49	0.57	0.64	0.72	0.79
Median	0.24	0.24	0.27	0.32	0.40	0.49	0.59	0.69	0.79	0.89	0.99
$q_{0.95}$	0.27	0.28	0.31	0.38	0.48	0.59	0.71	0.84	0.98	1.12	1.28
					$M = 100$						
Mean	0.07	0.11	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.01
s.d.	0.01	0.01	0.02	0.03	0.04	0.05	0.07	0.08	0.10	0.13	0.15
$q_{0.05}$	0.06	0.10	0.17	0.25	0.34	0.42	0.5	0.57	0.65	0.72	0.79
Median	0.07	0.11	0.20	0.30	0.40	0.49	0.59	0.69	0.79	0.89	0.99
$q_{0.95}$	0.08	0.13	0.23	0.35	0.46	0.59	0.71	0.85	0.98	1.13	1.28
					$M = 1000$						
Mean	0.02	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.01
s.d.	0.00	0.01	0.02	0.03	0.04	0.05	0.07	0.08	0.10	0.13	0.15
$q_{0.05}$	0.02	0.09	0.17	0.26	0.34	0.42	0.50	0.57	0.65	0.72	0.79
Median	0.02	0.10	0.20	0.30	0.40	0.50	0.59	0.69	0.79	0.89	0.99
$q_{0.95}$	0.03	0.11	0.23	0.35	0.47	0.59	0.72	0.85	0.98	1.13	1.28

Table B.3: Simulation results: Small samples ($T = 10$)

This table reports the sample average, standard deviation, and selected percentiles of VCV (Panel A) and $\widehat{\eta}$ (Panel B), obtained from $R = 1,000,000$ replications of $T = 10$ volume realizations simulated from the model outlined in Section 2 in the paper, for various values of the proportion of informed trade η and number of liquidity seekers M.

η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Panel A: VCV											
$M = 10$											
Mean	0.28	0.28	0.31	0.36	0.42	0.48	0.54	0.60	0.65	0.70	0.74
s.d.	0.07	0.07	0.07	0.09	0.10	0.11	0.12	0.14	0.15	0.16	0.17
$q_{0.05}$	0.17	0.17	0.19	0.22	0.27	0.31	0.35	0.39	0.43	0.46	0.48
Median	0.28	0.28	0.30	0.35	0.41	0.47	0.53	0.59	0.64	0.69	0.73
$q_{0.95}$	0.40	0.40	0.44	0.51	0.59	0.67	0.76	0.83	0.91	0.98	1.05
$M = 100$											
Mean	0.1	0.15	0.24	0.33	0.41	0.48	0.55	0.60	0.65	0.70	0.74
s.d.	0.02	0.04	0.06	0.08	0.09	0.11	0.12	0.14	0.15	0.16	0.17
$q_{0.05}$	0.06	0.09	0.15	0.21	0.27	0.32	0.36	0.40	0.43	0.46	0.48
Median	0.10	0.14	0.24	0.33	0.41	0.48	0.54	0.59	0.64	0.69	0.73
$q_{0.95}$	0.14	0.21	0.34	0.47	0.57	0.67	0.76	0.84	0.91	0.98	1.04
$M = 1000$											
Mean	0.03	0.13	0.24	0.33	0.41	0.48	0.55	0.60	0.65	0.70	0.74
s.d.	0.01	0.03	0.06	0.08	0.09	0.11	0.12	0.13	0.15	0.16	0.17
$q_{0.05}$	0.02	0.08	0.15	0.21	0.27	0.32	0.36	0.40	0.43	0.46	0.48
Median	0.03	0.13	0.24	0.33	0.41	0.48	0.54	0.60	0.64	0.69	0.73
$q_{0.95}$	0.05	0.19	0.34	0.47	0.57	0.67	0.76	0.84	0.91	0.98	1.05
Panel B: $\widehat{\eta}$											
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$M = 10$											
Mean	0.23	0.23	0.26	0.32	0.40	0.49	0.59	0.70	0.81	0.94	1.27
s.d.	0.07	0.07	0.08	0.10	0.14	0.18	0.23	0.30	4.12	3.00	50.50
$q_{0.05}$	0.13	0.13	0.15	0.17	0.21	0.26	0.31	0.35	0.39	0.43	0.46
Median	0.22	0.22	0.25	0.30	0.38	0.46	0.55	0.64	0.73	0.83	0.93
$q_{0.95}$	0.36	0.36	0.41	0.51	0.64	0.81	1.00	1.23	1.50	1.84	2.25
$M = 100$											
Mean	0.07	0.11	0.19	0.29	0.39	0.49	0.59	0.71	0.83	1.06	1.11
s.d.	0.02	0.03	0.06	0.09	0.12	0.17	0.23	0.31	0.46	32.19	3.05
$q_{0.05}$	0.04	0.06	0.11	0.16	0.22	0.27	0.31	0.35	0.39	0.43	0.47
Median	0.07	0.10	0.19	0.28	0.37	0.46	0.55	0.65	0.74	0.83	0.93
$q_{0.95}$	0.10	0.16	0.30	0.45	0.61	0.8	1.00	1.24	1.51	1.85	2.23
$M = 1000$											
Mean	0.02	0.10	0.19	0.29	0.39	0.49	0.60	0.71	0.78	0.96	1.16
s.d.	0.01	0.03	0.06	0.09	0.13	0.17	0.23	0.31	14.28	1.64	24.67
$q_{0.05}$	0.01	0.06	0.11	0.17	0.22	0.27	0.31	0.36	0.40	0.43	0.46
Median	0.02	0.09	0.19	0.28	0.37	0.46	0.56	0.65	0.74	0.83	0.93
$q_{0.95}$	0.03	0.14	0.29	0.45	0.61	0.80	1.01	1.24	1.52	1.85	2.26

B.2 Non-Gaussian demand, dynamic demand, and random proportion of informed trade

In this section, we simulate from modified versions of the benchmark model in Section 1, by deviating from the assumptions that all M liquidity seekers submit *i.i.d.* Normally distributed orders and that a constant proportion η of the liquidity seekers is informed.

Table B.4 reports simulations for a model where we relax the assumption of Normally distributed demand and allow for leptokurtic and skewed demand distributions, to generate jumps in trading volume that are unrelated to informed trading. In the top panel, demand is Uniformly distributed over the support $[-1, 1]$. In the middle panel, demand is t-distributed with 4 degrees of freedom (t_4). In the bottom panel, demand is Skew-Normally distributed with shape parameter 10, indicating positive skew ($SN(0, 1, 10)$). Relaxing the assumption of Normality does not change the main result of our analysis: VCV and $\widehat{\eta}$ are still strictly increasing in η. However, the standard deviations are clearly smaller for VCV than for $\widehat{\eta}$. More importantly, the average $\widehat{\eta}$ is no longer closely following the true value of η, implying that, in the case of non-Gaussian demand, $\widehat{\eta}$ should not be interpreted as a direct estimator of the true value of η.

Table B.5 reports simulations for a model where we relax the assumption of *i.i.d* demand and instead allow for dynamic demand, generating persistent or autocorrelated volume. In the benchmark model in Section 2 of the paper, demand by trader i in period t ($y_{i,t}$) is *i.i.d.* distributed with mean zero and standard deviation one. Instead, we now specify demand as an AR(1) process with *i.i.d.* innovations: $y_{i,t} = \rho y_{i,t-1} + \varepsilon_{i,t}$, where $\varepsilon_{i,t}$ is *i.i.d.* normally distributed with mean zero and standard deviation one. We calibrate the AR(1) parameter at $\rho = 0.3$ and $\rho = 0.9$ to generate lower and higher autocorrelation in volume, as well as $\rho = 1.$ which causes demand and therefore to follow a non-stationary random walk process. (Note that the case $\rho = 0$ corresponds to the benchmark model.) In all three cases we continue to find that VCV increases monotonically in the proportion of informed trade. $\widehat{\eta}$ does so as well, but is (similar to Table B.4) diverging from the true value η when ρ is high.

In Table B.6 we relax the assumption of a constant proportion of informed trade. In the top panel, the number of uninformed traders is held fixed at 1000, while the number of uninformed investors is in each $T = 100$ of the trading sessions randomly drawn from a Uniform distribution. We adjust the support of the Uniform distribution to create variation in the average proportion of informed trade ($E[\eta]$). The results indicate again that VCV is strictly increasing in the *average* proportion of informed trade. this is important, because the proportion of informed trade η is not necessarily constant across observations, and we are typically interested in measuring the average proportion of informed trade. Note that the

Table B.4: Simulation results: Non-Gaussian demand distributions

This table reports the sample average, standard deviation, and selected percentiles of VCV (Panel A) and $\widehat{\eta}$ (Panel B), obtained from $R = 1,000,000$ replications of $T = 100$ volume observations, simulated from a model with $M = 1000$ liquidity seekers. Different from Table A1, liquidity demand is not Normally distributed. In the top panel, demand is Uniformly distributed over the support $[-1, 1]$. In the middle panel, demand is t-distributed with 4 degrees of freedom (t_4). In the bottom panel, demand is Skew-Normally distributed with shape parameter 10, indicating positive skew ($SN(0, 1, 10)$).

					Panel A: VCV						
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
					Uniform distribution						
Mean	0.03	0.11	0.19	0.27	0.33	0.38	0.43	0.48	0.51	0.55	0.58
s.d.	0.00	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.04	0.04
$q_{0.05}$	0.02	0.10	0.18	0.24	0.30	0.35	0.39	0.43	0.46	0.48	0.51
Median	0.03	0.11	0.19	0.27	0.33	0.39	0.43	0.48	0.51	0.55	0.58
$q_{0.95}$	0.03	0.11	0.21	0.29	0.36	0.42	0.48	0.53	0.57	0.61	0.65
					t-distribution						
Mean	0.04	0.18	0.32	0.45	0.55	0.64	0.72	0.80	0.86	0.92	0.97
s.d.	0.00	0.04	0.07	0.09	0.11	0.12	0.13	0.14	0.15	0.16	0.16
$q_{0.05}$	0.04	0.13	0.25	0.35	0.44	0.51	0.58	0.64	0.69	0.74	0.78
Median	0.04	0.17	0.31	0.43	0.53	0.62	0.70	0.77	0.83	0.89	0.94
$q_{0.95}$	0.05	0.24	0.43	0.6	0.73	0.85	0.94	1.03	1.11	1.18	1.24
					Skew-Normal distribution						
Mean	0.02	0.08	0.15	0.23	0.30	0.38	0.45	0.53	0.61	0.68	0.75
s.d.	0.00	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.04	0.05	0.06
$q_{0.05}$	0.02	0.07	0.13	0.20	0.27	0.34	0.40	0.47	0.54	0.60	0.67
Median	0.02	0.08	0.15	0.23	0.30	0.38	0.45	0.53	0.61	0.68	0.75
$q_{0.95}$	0.03	0.09	0.17	0.26	0.34	0.42	0.51	0.59	0.68	0.76	0.85
					Panel B: $\widehat{\eta}$						
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
					Uniform distribution						
Mean	0.02	0.08	0.15	0.21	0.28	0.34	0.40	0.46	0.52	0.57	0.62
s.d.	0.00	0.00	0.01	0.01	0.02	0.03	0.04	0.04	0.05	0.06	0.08
$q_{0.05}$	0.02	0.07	0.13	0.19	0.25	0.30	0.35	0.39	0.43	0.47	0.51
Median	0.02	0.08	0.15	0.21	0.28	0.34	0.40	0.46	0.51	0.57	0.62
$q_{0.95}$	0.02	0.08	0.16	0.24	0.31	0.39	0.46	0.54	0.61	0.68	0.75
					t-distribution						
Mean	0.03	0.13	0.28	0.43	0.60	0.78	1.08	1.25	1.23	1.24	1.60
s.d.	0.00	0.05	0.29	1.03	2.14	7.57	27.35	14.14	81.28	113.16	48.82
$q_{0.05}$	0.02	0.10	0.20	0.30	0.40	0.51	0.62	0.73	0.84	0.95	1.05
Median	0.03	0.13	0.26	0.4	0.54	0.70	0.86	1.03	1.22	1.41	1.61
$q_{0.95}$	0.03	0.19	0.4	0.65	0.92	1.26	1.64	2.11	2.63	3.27	3.97
					Skew-Normal distribution						
Mean	0.02	0.06	0.11	0.18	0.25	0.34	0.43	0.54	0.67	0.83	1.01
s.d.	0.00	0.00	0.01	0.02	0.02	0.03	0.04	0.06	0.08	0.11	0.15
$q_{0.05}$	0.01	0.05	0.10	0.15	0.22	0.29	0.37	0.45	0.55	0.67	0.79
Median	0.02	0.06	0.11	0.18	0.25	0.33	0.43	0.54	0.67	0.82	0.99
$q_{0.95}$	0.02	0.06	0.13	0.20	0.29	0.39	0.51	0.64	0.81	1.02	1.28

sample of $T = 100$ trading sessions can both refer to a time-series or a cross-section of observations, such that VCV indicates either the average proportion of informed trade for a single asset over time, or the

Table B.5: Simulation results: Dynamic demand

This table reports the sample average, standard deviation, and selected percentiles of VCV (Panel A) and $\widehat{\eta}$ (Panel B), obtained from $R = 1,000,000$ replications of $T = 100$ volume observations, simulated from a model with $M = 1000$ liquidity seekers. Different from Table A1, liquidity demand is not *i.i.d.* but persistent. Specifically, demand by trader i in period t is equal to $y_{i,t} = \rho y_{i,t-1} + \varepsilon_{i,t}$, where $\varepsilon_{i,t}$ is *i.i.d.* normally distributed with mean zero. The table reports results for $\rho \in (0.3, 0.9, 1)$. The benchmark model with *i.i.d.* demand (Table A.1) corresponds to $\rho = 0$.

					Panel A: VCV						
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
					$\rho = 0.3$						
Mean	0.03	0.14	0.25	0.35	0.43	0.50	0.56	0.62	0.67	0.71	0.75
s.d.	0.00	0.01	0.02	0.02	0.03	0.03	0.04	0.04	0.05	0.05	0.06
$q_{0.05}$	0.03	0.12	0.22	0.31	0.38	0.44	0.50	0.55	0.59	0.63	0.66
Median	0.03	0.14	0.25	0.35	0.43	0.50	0.56	0.62	0.67	0.71	0.75
$q_{0.95}$	0.04	0.16	0.28	0.39	0.48	0.56	0.63	0.69	0.75	0.80	0.85
					$\rho = 0.9$						
Mean	0.09	0.16	0.25	0.34	0.41	0.48	0.54	0.59	0.64	0.68	0.72
s.d.	0.00	0.02	0.04	0.05	0.05	0.06	0.06	0.07	0.08	0.08	0.09
$q_{0.05}$	0.08	0.12	0.19	0.27	0.34	0.40	0.45	0.49	0.53	0.56	0.59
Median	0.09	0.15	0.25	0.33	0.41	0.48	0.54	0.59	0.64	0.68	0.72
$q_{0.95}$	0.10	0.20	0.32	0.42	0.51	0.59	0.66	0.72	0.77	0.82	0.88
					$\rho = 1$						
Mean	0.35	0.36	0.37	0.40	0.43	0.46	0.49	0.53	0.57	0.60	0.63
s.d.	0.01	0.04	0.07	0.09	0.10	0.11	0.11	0.12	0.13	0.14	0.16
$q_{0.05}$	0.33	0.29	0.27	0.27	0.29	0.31	0.33	0.35	0.37	0.38	0.39
Median	0.35	0.35	0.37	0.39	0.41	0.45	0.48	0.51	0.56	0.6	0.63
$q_{0.95}$	0.37	0.43	0.49	0.55	0.60	0.65	0.7	0.74	0.79	0.84	0.9
					Panel B: $\widehat{\eta}$						
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
					$\rho = 0.3$						
Mean	0.02	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
s.d.	0.00	0.01	0.02	0.03	0.04	0.05	0.07	0.08	0.1	0.13	0.15
$q_{0.05}$	0.02	0.09	0.17	0.25	0.34	0.42	0.50	0.57	0.65	0.72	0.78
Median	0.02	0.10	0.20	0.3	0.4	0.49	0.59	0.69	0.79	0.89	0.99
$q_{0.95}$	0.03	0.12	0.23	0.35	0.46	0.59	0.72	0.85	0.99	1.12	1.28
					$\rho = 0.9$						
Mean	0.06	0.11	0.2	0.29	0.38	0.47	0.57	0.66	0.75	0.84	0.95
s.d.	0.00	0.02	0.04	0.05	0.07	0.09	0.11	0.13	0.16	0.19	0.25
$q_{0.05}$	0.06	0.09	0.15	0.22	0.29	0.36	0.42	0.48	0.54	0.59	0.64
Median	0.06	0.11	0.20	0.28	0.37	0.46	0.55	0.64	0.73	0.82	0.91
$q_{0.95}$	0.07	0.15	0.27	0.38	0.51	0.63	0.77	0.9	1.05	1.20	1.38
					$\rho = 1$						
Mean	0.30	0.31	0.33	0.36	0.4	0.45	0.50	0.56	0.63	0.71	0.79
s.d.	0.01	0.05	0.08	0.11	0.14	0.16	0.18	0.22	0.26	0.32	0.41
$q_{0.05}$	0.28	0.24	0.22	0.21	0.23	0.25	0.28	0.31	0.33	0.33	0.34
Median	0.30	0.30	0.32	0.34	0.38	0.42	0.46	0.51	0.58	0.65	0.71
$q_{0.95}$	0.32	0.39	0.48	0.58	0.66	0.76	0.85	0.97	1.10	1.25	1.47

average proportion of informed trade for a group of assets at a given point in time.

In the second panel of Table B.6 the number of informed liquidity seekers is in each trading session randomly drawn from a Bernoulli distribution. The number of active informed traders in each trading session is equal to X with probability $\frac{1}{5}$ and zero with probability $\frac{4}{5}$, such that the informed traders participate in only one out of five trading sessions on average. To create variation in the average proportion of informed trade, we adjust the potential number of informed traders X. This version of our model can be interpreted as a hybrid of our Kyle (1985)-type model in Section 2 and the PIN model by Easley et al. (1996), in which arrival of information is random, similar to the model by Back et al. (2018). The results confirm that VCV continues to be monotonically increasing in $E[\eta]$ while its standard deviations remain fairly low. In these settings, $\widehat{\eta}$ clearly does not perform well as a measure of informed trading. The simulated observations of $\widehat{\eta}$ are more dispersed than VCV. In the case of Bernoulli distributed informed trade, the averages of $\widehat{\eta}$ are not monotonically increasing in $E[\eta]$, and are not bounded by 0 and 1.

Table B.6: Simulation results: Random proportion of informed trade

This table reports the sample average, standard deviation, and selected percentiles of VCV (Panel A) and $\widehat{\eta}$ (Panel B), obtained from $R = 1,000,000$ replications of $T = 100$ volume observations. Different from Table A.1, the number of uninformed liquidity seekers is kept constant at $1,000$, while the number of informed liquidity seekers is varying randomly across observations. In the top panel, the number of informed liquidity seekers follows a discrete Uniform distribution over the support $[0, X]$. In the second panel, informed demand is Bernoulli distributed such that the number of active informed traders in each trading session is with probability $\frac{4}{5}$ equal to zero and with probability $\frac{1}{5}$ equal to X. we consider different values of X, which determine the average proportion of informed trade $E[\eta]$.

Panel A: VCV											
Uniform distribution: Informed investors $\sim U[0, X]$											
Uninformed	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
X	0.00	200	500	800	1300	2000	3000	5000	8000	20000	50000
$E[\eta]$	0.00	0.09	0.2	0.29	0.39	0.5	0.6	0.71	0.8	0.91	0.96
Mean	0.03	0.17	0.34	0.46	0.58	0.69	0.78	0.87	0.92	0.99	1.02
s.d.	0.00	0.02	0.03	0.04	0.05	0.06	0.06	0.07	0.08	0.08	0.08
$q_{0.05}$	0.03	0.14	0.29	0.39	0.50	0.6	0.68	0.75	0.81	0.86	0.89
Median	0.03	0.17	0.34	0.46	0.58	0.69	0.78	0.86	0.92	0.99	1.02
$q_{0.95}$	0.04	0.20	0.40	0.53	0.67	0.79	0.89	0.99	1.05	1.13	1.17
Bernoulli distribution: Informed investors $\sim B(1/5, X)$											
Uninformed	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
X	0.00	500	1250	2000	3250	5000	7500	12500	20000	50000	125000
$E[\eta]$	0.00	0.09	0.2	0.29	0.39	0.5	0.6	0.71	0.8	0.91	0.96
Mean	0.03	0.41	0.83	1.12	1.43	1.7	1.93	2.16	2.32	2.50	2.59
s.d.	0.00	0.06	0.10	0.11	0.13	0.15	0.18	0.22	0.26	0.31	0.34
$q_{0.05}$	0.03	0.31	0.66	0.93	1.23	1.47	1.67	1.84	1.94	2.06	2.11
Median	0.03	0.41	0.84	1.12	1.43	1.69	1.91	2.14	2.29	2.47	2.55
$q_{0.95}$	0.04	0.51	0.99	1.30	1.65	1.96	2.24	2.55	2.78	3.06	3.21
Panel B: $\widehat{\eta}$											
Uniform distribution: Informed investors $\sim U[0, X]$											
Uninformed	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
X	0.00	200	500	800	1300	2000	3000	5000	8000	20000	50000
$E[\eta]$	0.00	0.09	0.2	0.29	0.39	0.5	0.6	0.71	0.8	0.91	0.96
Mean	0.02	0.13	0.3	0.44	0.64	0.85	1.08	1.37	1.62	1.99	2.19
s.d.	0.00	0.02	0.04	0.06	0.09	0.14	0.19	0.28	0.38	0.58	1.01
$q_{0.05}$	0.02	0.10	0.24	0.35	0.50	0.66	0.81	1.00	1.14	1.33	1.43
Median	0.02	0.13	0.29	0.43	0.63	0.84	1.05	1.33	1.56	1.88	2.05
$q_{0.95}$	0.03	0.15	0.36	0.54	0.80	1.10	1.42	1.88	2.3	3.01	3.39
Bernoulli distribution: Informed investors $\sim B(1/5, X)$											
Uninformed	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
X	0.00	500	1250	2000	3250	5000	7500	12500	20000	50000	125000
$E[\eta]$	0.00	0.09	0.2	0.29	0.39	0.5	0.6	0.71	0.8	0.91	0.96
Mean	0.02	0.38	1.28	3.30	10.87	-13.8	-5.73	-3.66	-3.09	-2.69	-2.56
s.d.	0.00	0.08	0.35	113.51	2725.51	3861.71	49.41	1.65	0.80	0.60	0.55
$q_{0.05}$	0.02	0.25	0.78	1.58	-71.55	-46.10	-10.57	-5.61	-4.50	-3.77	-3.54
Median	0.02	0.38	1.24	2.85	8.00	-8.11	-4.74	-3.40	-2.93	-2.58	-2.46
$q_{0.95}$	0.03	0.51	1.90	6.14	78.05	28.04	-3.07	-2.46	-2.19	-1.97	-1.89

B.3 Stochastic liquidity

Next, we demonstrate the robustness of VCV to random variation in liquidity demand and liquidity provision. Tables B.7 and B.8 provide simulations for variations of the model in which liquidity demand (the number of market participants M and their trading intensities σ) are varying randomly (Uniformly distributed) across observations. These modifications of the benchmark model generate variation in trading volume across observations that is unrelated to informed trading. For example, when VCV is applied to a time-series of volumes of the same asset, there could be seasonal patterns in the level of trading volume. When VCV is applied to a cross-section of volumes, there are generally different levels of volume for different assets. These simulation results indicate that VCV continues to increase monotonically in the proportion of informed trade η after allowing for random variation in liquidity demand. $\widehat{\eta}$ increases in η as well, but is far more dispersed than VCV.

Table B.7: Simulation results: Random trading intensity σ

This table reports the sample average, standard deviation, and selected percentiles of VCV (Panel A) and $\widehat{\eta}$ (Panel B), obtained from $R = 1,000,000$ replications of $T = 100$ volume observations, simulated from a model with $M = 1000$ liquidity seekers. Different from Table A1, in which the trading intensity σ for both informed and uninformed investors is kept constant at unity, σ is now drawn randomly before each trading session from a Uniform distribution over the support [0.2,1.8].

Panel A: VCV											
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Mean	0.46	0.49	0.54	0.60	0.66	0.72	0.77	0.82	0.87	0.91	0.95
s.d.	0.03	0.03	0.04	0.04	0.05	0.06	0.06	0.07	0.07	0.07	0.08
$q_{0.05}$	0.42	0.44	0.48	0.53	0.58	0.63	0.68	0.72	0.75	0.79	0.82
Median	0.46	0.49	0.54	0.60	0.66	0.72	0.77	0.82	0.86	0.91	0.94
$q_{0.95}$	0.51	0.54	0.60	0.67	0.74	0.81	0.88	0.94	0.99	1.04	1.08
Panel B: $\widehat{\eta}$											
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Mean	0.44	0.48	0.55	0.66	0.78	0.92	1.06	1.21	1.37	1.55	1.74
s.d.	0.04	0.05	0.06	0.08	0.11	0.14	0.18	0.22	0.28	0.36	0.44
$q_{0.05}$	0.38	0.40	0.46	0.54	0.63	0.71	0.81	0.91	1.00	1.10	1.20
Median	0.44	0.47	0.55	0.65	0.77	0.90	1.04	1.18	1.33	1.50	1.66
$q_{0.95}$	0.51	0.56	0.66	0.80	0.97	1.17	1.38	1.63	1.89	2.20	2.53

In our benchmark model, we assume the existence of a deep-pocketed market maker that is always able to trade against the order imbalance. We next analyze the performance in the case of limited market maker capacity. Specifically, in Table B.9, we analyze the distribution of trading volume in the presence of stochastic liquidity provision: we assume that in each trading session liquidity providers do not absorb the order imbalance $\sum_M y_i$ with probability p. Total trading volume can thus be written as:

$$V = \begin{cases} \frac{1}{2}\left(\sum_M |y_i| + |\sum_M y_i|\right) & \text{with probability (1-}p\text{)} \\ \frac{1}{2}\left(\sum_M |y_i| - |\sum_M y_i|\right) & \text{with probability } p \end{cases} \tag{B.1}$$

That is, with probability (1-p), liquidity provision is high and the counting of volume is identical to the benchmark model in our paper (Eq.1 in the paper): double-counted volume includes the trades among

Table B.8: Simulation results: Random number of market participants M

This table reports the sample average, standard deviation, and selected percentiles of VCV (Panel A) and $\widehat{\eta}$ (Panel B), obtained from $R = 1,000,000$ replications of $T = 100$ volume observations, simulated from a model with M liquidity seekers. Different from Table A1, in which the number liquidity seekers is kept constant at $M = 1000$, M is now before each trading session randomly drawn from a Uniform distribution over the support [200,1800].

						Panel A: VCV					
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Mean	0.03	0.15	0.28	0.38	0.47	0.55	0.62	0.69	0.74	0.79	0.83
s.d.	0.00	0.01	0.02	0.03	0.04	0.04	0.05	0.05	0.06	0.06	0.07
$q_{0.05}$	0.03	0.13	0.24	0.33	0.41	0.48	0.55	0.60	0.65	0.69	0.73
Median	0.03	0.15	0.28	0.38	0.47	0.55	0.62	0.68	0.74	0.79	0.83
$q_{0.95}$	0.04	0.18	0.32	0.44	0.54	0.63	0.71	0.78	0.84	0.9	0.95
						Panel B: $\widehat{\eta}$					
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Mean	0.02	0.11	0.22	0.34	0.46	0.58	0.71	0.84	0.97	1.11	1.25
s.d.	0.00	0.01	0.02	0.04	0.06	0.07	0.10	0.13	0.16	0.19	0.24
$q_{0.05}$	0.02	0.09	0.19	0.28	0.38	0.47	0.57	0.66	0.75	0.84	0.93
Median	0.02	0.11	0.22	0.34	0.46	0.58	0.70	0.83	0.95	1.09	1.22
$q_{0.95}$	0.03	0.13	0.27	0.41	0.56	0.71	0.89	1.07	1.25	1.46	1.69

liquidity seekers, as well as the trades between the liquidity providers and unmatched liquidity seekers. With probability p, liquidity is low and the volume only includes liquidity seeking demands that can be matched among each other. As the results indicate, VCV is increasing in the proportion of informed trade for different levels of p. However, we also see that the level of VCV is different for different levels of market maker capacity. These levels differences primarily appear when the proportion of informed trade η is high. In particular, for high levels of η, VCV tends to be higher when liquidity provision is low (i.e. p is high). For lower levels of η, the impact of market maker capacity on VCV is much smaller.[1]

Overall, the results in Table B.9 show that VCV increases in the proportion of informed trade even in the presence of limited limited liquidity provision. Panel B shows that $\widehat{\eta}$ is not strictly positive and highly erratic.

[1]We revisit the impact of market maker capacity on VCV empirically below in Table C.17.

Table B.9: Simulation results: Stochastic liquidity provision

This table reports the sample average, standard deviation, and selected percentiles of VCV (Panel A) and $\widehat{\eta}$ (Panel B), obtained from $R = 1,000,000$ replications of $T = 100$ volume observations, simulated from a model with $M = 1,000$ liquidity seekers. Different from Table A1, in which the order imbalance is always absorbed by liquidity suppliers, liquidity supply is now random and follows a Bernoulli distribution: in each trading session, liquidity provision is high with probability (1-p) and the counting of volume is identical to the benchmark model (Eq.1 in the paper), while liquidity is low with probability p such that volume only includes liquidity seeking demands that can be matched among each other:

$$V = \begin{cases} \frac{1}{2}\left(\sum_M |y_i| + \left|\sum_M y_i\right|\right) & \text{with probability (1-}p\text{)} \\ \frac{1}{2}\left(\sum_M |y_i| - \left|\sum_M y_i\right|\right) & \text{with probability } p \end{cases}$$

The table reports results for $p \in (0.25, 0.5, 0.75, 1)$. The benchmark model (Table A.1) corresponds to $p = 0$.

						Panel A: VCV					
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
						$p = 0.25$					
Mean	0.04	0.15	0.29	0.41	0.52	0.63	0.72	0.81	0.90	0.97	1.05
s.d.	0.00	0.01	0.02	0.03	0.03	0.04	0.05	0.06	0.07	0.07	0.08
$q_{0.05}$	0.04	0.13	0.25	0.36	0.47	0.56	0.65	0.72	0.79	0.85	0.92
Median	0.04	0.15	0.28	0.41	0.52	0.62	0.72	0.81	0.89	0.97	1.04
$q_{0.95}$	0.05	0.17	0.32	0.45	0.58	0.69	0.80	0.91	1.01	1.10	1.19
						$p = 0.5$					
Mean	0.05	0.15	0.29	0.44	0.58	0.73	0.87	1.02	1.17	1.32	1.47
s.d.	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.09	0.11	0.14
$q_{0.05}$	0.04	0.13	0.25	0.38	0.51	0.65	0.78	0.91	1.03	1.15	1.26
Median	0.05	0.15	0.29	0.44	0.58	0.73	0.87	1.02	1.17	1.31	1.46
$q_{0.95}$	0.05	0.17	0.33	0.49	0.65	0.81	0.97	1.14	1.32	1.51	1.70
						$p = 0.75$					
Mean	0.04	0.13	0.26	0.40	0.57	0.76	0.97	1.22	1.52	1.87	2.32
s.d.	0.00	0.02	0.03	0.05	0.07	0.08	0.09	0.10	0.12	0.17	0.29
$q_{0.05}$	0.04	0.10	0.20	0.32	0.46	0.62	0.83	1.06	1.33	1.61	1.91
Median	0.04	0.13	0.26	0.40	0.57	0.76	0.97	1.22	1.51	1.86	2.29
$q_{0.95}$	0.05	0.16	0.31	0.48	0.67	0.88	1.11	1.38	1.73	2.17	2.84
						Panel B: $\widehat{\eta}$					
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
						$p = 0.25$					
Mean	0.03	0.11	0.23	0.37	0.53	0.71	0.92	1.17	1.49	1.88	2.37
s.d.	0.00	0.01	0.02	0.03	0.05	0.08	0.12	0.18	0.28	0.45	1.16
$q_{0.05}$	0.03	0.10	0.20	0.32	0.45	0.59	0.75	0.92	1.10	1.30	1.54
Median	0.03	0.11	0.23	0.37	0.53	0.70	0.91	1.16	1.45	1.80	2.22
$q_{0.95}$	0.03	0.13	0.27	0.43	0.62	0.85	1.14	1.50	2.01	2.68	3.72
						$p = 0.5$					
Mean	0.03	0.11	0.24	0.41	0.63	0.94	1.40	2.15	3.83	-4.48	-1.69
s.d.	0.00	0.01	0.03	0.04	0.07	0.12	0.23	0.50	2.47	898.31	2335.03
$q_{0.05}$	0.03	0.09	0.20	0.34	0.52	0.76	1.08	1.52	2.17	2.42	-76.92
Median	0.03	0.11	0.24	0.41	0.63	0.93	1.37	2.07	3.37	6.19	7.40
$q_{0.95}$	0.04	0.13	0.28	0.48	0.76	1.16	1.82	3.07	6.75	30.25	78.03
						$p = 0.75$					
Mean	0.03	0.09	0.2	0.36	0.61	1.02	1.88	5.77	-8.17	-3.09	-3.09
s.d.	0.00	0.01	0.03	0.06	0.11	0.21	0.52	62.16	1436.4	465.7	6.22
$q_{0.05}$	0.03	0.07	0.15	0.26	0.43	0.7	1.2	2.34	-108.11	-14.58	-4.80
Median	0.03	0.09	0.20	0.36	0.60	1.01	1.80	4.14	3.81	-5.25	-2.95
$q_{0.95}$	0.03	0.12	0.26	0.47	0.81	1.39	2.79	10.49	92.4	-3.22	-2.14

B.4 Heterogeneous information

Table B.10 provides simulation results for a version of our model in which there is heterogeneous information (or differences of opinion) among the informed investors. Instead of all ηM informed investors making the same order based on the same information, we divide the informed investors into two groups that trade on independent private signals and therefore make independent orders. That is, the first $\frac{1}{2}\eta M$ informed traders each submit identical orders and the second $\frac{1}{2}\eta M$ informed traders each submit identical orders that are uncorrelated to the orders by the first group. The demands of each of the $(1-\eta)\,M$ uninformed liquidity seekers remain uncorrelated as in the benchmark model in Section 2. Information heterogeneity does not affect our main result: both VCV and $\widehat{\eta}$ increase monotonically in η.

Table B.10: Simulation results: Heterogeneous information

This table reports the sample average, standard deviation, and selected percentiles of VCV (Panel A) and $\widehat{\eta}$ (Panel B), obtained from $R = 1,000,000$ replications of $T = 100$ volume observations, simulated from a model with $M = 1000$ liquidity seekers. Different from Table A1, the ηM informed traders are divided into two equal-sized groups that each make an independent order.

						Panel A: VCV					
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Mean	0.03	0.09	0.17	0.24	0.30	0.36	0.41	0.45	0.49	0.53	0.57
s.d.	0.00	0.01	0.01	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.04
$q_{0.05}$	0.03	0.08	0.15	0.21	0.27	0.31	0.36	0.40	0.44	0.47	0.50
Median	0.03	0.09	0.17	0.24	0.30	0.36	0.41	0.45	0.49	0.53	0.57
$q_{0.95}$	0.04	0.11	0.19	0.27	0.34	0.40	0.46	0.51	0.56	0.60	0.64
						Panel B: $\widehat{\eta}$					
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Mean	0.02	0.07	0.13	0.19	0.25	0.31	0.37	0.43	0.49	0.55	0.60
s.d.	0.00	0.01	0.01	0.02	0.02	0.03	0.04	0.05	0.05	0.06	0.07
$q_{0.05}$	0.02	0.06	0.11	0.16	0.21	0.26	0.31	0.36	0.40	0.45	0.49
Median	0.02	0.07	0.13	0.19	0.25	0.31	0.37	0.43	0.48	0.54	0.60
$q_{0.95}$	0.03	0.08	0.15	0.22	0.29	0.36	0.43	0.51	0.58	0.66	0.73

B.5 Endogenous informed trading intensity

Table B.11 provides results of simulations in which the informed trading intensity is endogenized, by making it explicitly dependent on the volume generated by the uninformed traders. The demands of the $(1-\eta)M$ uninformed investors are normally distributed with mean zero and standard deviation of one, as in our benchmark model, while the standard deviation of the ηM informed investors' demand is conditional on the demand by the uninformed investors. Specifically, the standard deviation of informed demand is increasing in the absolute order imbalance generated by the uninformed investors, such that informed investors attempt to 'hide' their orders in the uninformed order flow, as in Kyle (1985). The ηM informed investors observe in each trading session t the uninformed order flow $\sum_{\eta M+1 \ldots M} y_{i,t}$ and adjust their trading intensity accordingly:

$$\begin{aligned} y_{informed,t} &\sim N\left(0, \sigma_t^2\right) \\ \sigma_t &= \frac{\left|\sum_{i=\eta M+1 \ldots M} y_{i,t}\right|}{E\left[\left|\sum_{i=\eta M+1 \ldots M} y_{i,t}\right|\right]} = \frac{\left|\sum_{i=\eta M+1 \ldots M} y_{i,t}\right|}{\sqrt{2(1-\eta M)/\pi}} \end{aligned} \tag{B.2}$$

Hence, the unconditional expected value of the informed investor's trading intensity σ_t is unity, as for the uninformed investors, but the informed investors trade more (less) aggressively when the uninformed net order flow is relatively high (low). We obtain qualitatively similar results when the informed trading intensity is dependent on the cumulative size of the uninformed orders ($\sum_{\eta M+1 \ldots M} |y_{i,t}|$), instead of on the uninformed order imbalance. Both VCV and $\widehat{\eta}$ increase on average monotonically in η, while the standard deviations show that $\widehat{\eta}$ suffers from extreme outliers.

Table B.11: Simulation results: Endogenous informed trading

This table reports the sample average, standard deviation, and selected percentiles of VCV (Panel A) and $\widehat{\eta}$ (Panel B), obtained from $R = 1,000,000$ replications of $T = 100$ volume observations, simulated from a model with $M = 1000$ liquidity seekers. Different from Table A1, the ηM informed traders orders are normally distributed with a standard deviation that is in each trading session proportional to the absolute net order flow of the uninformed investors as in Eq. (1) of this appendix. The case $\eta = 1$ is omitted, because the trading intensity of informed investors in undefined in the absence of uninformed investors.

						Panel A: VCV					
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Mean	0.03	0.22	0.4	0.55	0.68	0.79	0.89	0.98	1.06	1.13	.
s.d.	0.00	0.03	0.05	0.07	0.08	0.09	0.10	0.11	0.12	0.12	.
$q_{0.05}$	0.03	0.17	0.32	0.45	0.56	0.66	0.75	0.82	0.89	0.95	.
Median	0.03	0.22	0.39	0.54	0.67	0.79	0.88	0.97	1.05	1.12	.
$q_{0.95}$	0.04	0.28	0.49	0.67	0.83	0.96	1.07	1.18	1.27	1.35	.
						Panel B: $\widehat{\eta}$					
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Mean	0.02	0.17	0.36	0.58	0.84	1.15	1.48	2.09	2.64	3.02	.
s.d.	0.00	0.03	0.07	0.12	0.21	0.37	17.6	10.96	29.27	193.43	.
$q_{0.05}$	0.02	0.13	0.26	0.42	0.59	0.77	0.98	1.2	1.44	1.68	.
Median	0.02	0.17	0.35	0.56	0.80	1.08	1.41	1.8	2.26	2.83	.
$q_{0.95}$	0.03	0.22	0.49	0.81	1.20	1.73	2.44	3.51	5.08	7.64	.

C Supplementary summary statistics and empirical results

In this section, we report summary statistics, subsample analyses, and additional empirical results to supplement the empirical results presented in Sections 4, 5 and 6 of the paper.

C.1 Volume and VCV summary statistics

Table C.1, reports summary statistics on the three definitions on trading volume used in the paper: V_{USD}, $V_{\%}$, and V_{TO} (See Eq. 19, 20 , and 21 in the paper, respectively)

Table C.1: Volume summary statistics

This table reports summary statistics of daily stock-level trading volumes, for three definitions of volume: daily dollar trading volume in US dollars (VCV$_{USD}$), daily volume market shares (daily dollar volume as a percentage of total market dollar volume – VCV$_{\%}$), and turnover (dollar volume as a fraction of market capitalization – VCV$_{TO}$). The table reports the total number of observations, the number of distinct stocks in the sample (N), the number of time-series observations/days (T), mean, standard deviation, s.d. (CS), the time-series average of daily cross-sectional standard deviations, s.d. (TS), the cross-sectional average of stock-specific time-series standard deviations, selected quantiles (q), and the cross-sectional average of stock-specific first-order autocorrelations (ρ). The bottom two rows report the time-series averages of within-day rank (Spearman) correlations between the different volume measures. Sample: 1980-2020.

	V_{USD} (in millions USD)	$V_{\%}$ (in percentage points)	V_{TO} (in percentage points)
Observations	$37,434,085$	$37,434,085$	$37,434,085$
N	$16,571$	$16,571$	$16,571$
T	$10,340$	$10,340$	$10,340$
Mean	18.946	0.026	0.539
s.d.	137.564	0.120	7.807
s.d. (CS)	71.961	0.093	1.157
s.d. (TS)	10.659	0.014	0.812
$q_{0.1}$	0.008	0.00003	0.021
$q_{0.25}$	0.047	0.0002	0.072
Median	0.442	0.002	0.221
$q_{0.75}$	4.687	0.011	0.568
$q_{0.9}$	28.656	0.052	1.174
Correlations			
$V_{\%}$	1		
V_{TO}	0.713	0.713	

Table C.2 provides subsample summary statistics of VCV, complementing the full-sample summary statistics in Table 2 of the paper. These subsamples include *(i)* NYSE/AMEX listed stocks, *(ii)* NASDAQ-listed stocks, *(iii)* observations prior to 2000 (1980-1999), and *(iv)* observations after 2000 (2000-2020).

Table C.2: VCV Summary Statistics

This table reproduces the results in Table 2 of the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX stocks and for subsamples of observations prior to 2000 (1980-1999) and post 2000 (2000-2020).

	NASDAQ			NYSE/AMEX		
	VCV_{USD}	$VCV_{\%}$	VCV_{TO}	VCV_{USD}	$VCV_{\%}$	VCV_{TO}
Observations	84,297	84,297	84,297	64,376	64,376	64,376
N	12,066	12,066	12,066	6,070	6,070	6,070
T	38	38	38	41	41	41
Mean	1.53	1.51	1.46	1.12	1.11	1.10
s.d.	0.90	0.91	0.83	0.69	0.71	0.67
s.d. (CS)	0.88	0.89	0.83	0.67	0.68	0.64
s.d. (TS)	0.59	0.59	0.55	0.45	0.46	0.44
$q_{0.1}$	0.70	0.66	0.68	0.50	0.45	0.50
$q_{0.25}$	0.99	0.96	0.94	0.65	0.62	0.65
Median	1.35	1.33	1.28	0.97	0.95	0.95
$q_{0.75}$	1.80	1.79	1.72	1.39	1.38	1.35
$q_{0.9}$	2.44	2.42	2.36	1.88	1.88	1.83
ρ	0.15	0.15	0.16	0.23	0.24	0.24
Correlations						
$VCV_{\%}$	0.98			0.98		
VCV_{TO}	0.96	0.95		0.98	0.96	

	Pre-2000			Post-2000		
	VCV_{USD}	$VCV_{\%}$	VCV_{TO}	VCV_{USD}	$VCV_{\%}$	VCV_{TO}
Observations	75,603	75,603	75,603	73,070	73,070	73,070
N	12,217	12,217	12,217	8,982	8,982	8,982
T	20	20	20	21	21	21
Mean	1.47	1.46	1.42	1.23	1.20	1.18
s.d.	0.71	0.73	0.67	0.95	0.95	0.87
s.d. (CS)	0.68	0.69	0.65	0.95	0.96	0.88
s.d. (TS)	0.49	0.49	0.46	0.55	0.55	0.52
$q_{0.1}$	0.79	0.77	0.78	0.50	0.45	0.50
$q_{0.25}$	1.03	1.02	1.00	0.64	0.60	0.63
Median	1.34	1.33	1.28	0.96	0.93	0.92
$q_{0.75}$	1.72	1.72	1.66	1.50	1.46	1.43
$q_{0.9}$	2.24	2.25	2.17	2.22	2.18	2.13
ρ	0.07	0.08	0.09	0.13	0.13	0.14
Correlations						
$VCV_{\%}$	0.98			0.99		
VCV_{TO}	0.97	0.96		0.97	0.96	

C.2 Monthly and Quarterly VCV

This section reports summary statistics of VCV computed over months and quarters. Similar to the annual VCV (Eq. 18 in the paper), the quarterly (monthly) VCV for a stock is obtained by dividing the standard deviation of daily trading volumes by its mean, estimated from all daily trading volumes within the quarter (month). We disregard infrequently traded stocks from the quarterly (monthly) sample, by only including firm-quarter (firm-month) observations with positive trading for volume least 45 (15) days during that quarter (month).

Table C.3: Summary Statistics: Quarterly and Monthly VCV

This table reports summary statistics of quarterly and monthly firm-level observations of the Volume Coefficient of Variation (VCV) of daily dollar trading volume in US dollars (VCV_{USD}), daily volume market shares (daily dollar volume as a percentage of total market dollar volume – $VCV_{\%}$), and turnover (dollar volume as a fraction of market capitalization – VCV_{TO}). The bottom two rows report the time-series averages of within-quarter and within-month rank (Spearman) correlations between the different VCV measures. Sample:1980-2020.

	Quarterly			Monthly		
	VCV_{USD}	$VCV_{\%}$	VCV_{TO}	VCV_{USD}	$VCV_{\%}$	VCV_{TO}
Observations	590,770	590,770	590,770	1,767,688	1,767,688	1,767,688
N	16,571	16,571	16,571	16,571	16,571	16,571
T	164	164	164	492	492	492
Mean	1.069	1.055	1.056	0.868	0.855	0.862
s.d.	0.617	0.627	0.607	0.482	0.491	0.480
s.d. (CS)	0.592	0.600	0.582	0.453	0.459	0.450
s.d. (TS)	0.466	0.468	0.459	0.393	0.393	0.390
$q_{0.1}$	0.474	0.437	0.472	0.377	0.346	0.376
$q_{0.25}$	0.649	0.623	0.642	0.521	0.499	0.519
Median	0.942	0.932	0.928	0.767	0.758	0.761
$q_{0.75}$	1.309	1.308	1.291	1.087	1.085	1.079
$q_{0.9}$	1.771	1.774	1.752	1.466	1.466	1.457
Correlations						
$VCV_{\%}$	0.976			0.969		
VCV_{TO}	0.987	0.965		0.994	0.963	

Table C.3 shows that, on average, the monthly VCV is higher than the quarterly VCV, which is in turn on average higher than the annual VCV. The reason for this pattern is that the annual standard deviation of daily volumes includes within-month and within-quarter variation, but also between-quarter and between-month seasonal variation. This is evident from Table C.4, showing that the annual average of monthly standard deviations (0.46) is much lower than the annual standard deviation (0.75). The annual average of the monthly means is however nearly equivalent to the annual mean. Because the levels of the monthly/quarterly/annual standard deviations are different, so are the levels of the monthly/quarterly/annual VCV. We should therefore not directly compare the levels of VCVs estimated over time windows of different length. Table C.5 shows that the annual averages of monthly/quarterly VCVs are highly correlated to the annual VCV.

Table C.4: Volume seasonality

The first column reports the averages of firm-month observations of VCV (See Table C.3) by calendar month. The second and third column report the averages of the monthly firm-month observations of the mean ($\widehat{\mu_V}$) and standard deviation ($\widehat{\sigma_V}$) of daily volumes (See Eq. 18 in the paper). Means, standard deviations, and VCV are computed from daily turnover (V_{TO}, Eq. 21). The bottom rows of the table report the average of the monthly figures, and the corresponding annual estimates of VCV, $\widehat{\mu_V}$, and $\widehat{\sigma_V}$. Sample:1980-2020.

Calendar month	VCV	$\widehat{\mu_V}$	$\widehat{\sigma_V}$
January	0.836	0.553	0.422
February	0.835	0.551	0.429
March	0.860	0.559	0.450
April	0.849	0.531	0.423
May	0.856	0.530	0.417
June	0.924	0.550	0.483
July	0.866	0.512	0.421
August	0.876	0.573	0.656
September	0.856	0.565	0.522
October	0.880	0.535	0.437
November	0.863	0.531	0.438
December	0.847	0.531	0.451
Average	0.862	0.543	0.463
Annual VCV,$\widehat{\mu_V}$,$\widehat{\sigma_V}$	1.302	0.540	0.753

Table C.5: Correlations between Annual, Quarterly, and Monthly VCV

We calculate firm-year averages of the firm-month and firm-quarter observations of $VCV_{\%}$, VCV_{USD} and VCV_{TO} (See Table C.3 and compare these to the firm-year observations of the annual VCV (See Table 2 in the paper). The table reports the time-series averages of within-year rank (Spearman) correlations between the monthly, quarterly and annual VCV measures. Sample:1980-2020.

Panel A: $VCV_{\%}$		
	Annual	Quarterly
Quarterly	0.943	
Monthly	0.883	0.960
Panel B: VCV_{USD}		
	Annual	Quarterly
Quarterly	0.937	
Monthly	0.875	0.957
Panel C: VCV_{TO}		
	Annual	Quarterly
Quarterly	0.950	
Monthly	0.889	0.959

C.3 VCV and firm characteristics

Table C.6 reports summary statistics for the firm-level characteristics considered in Section 4.1 of the paper. Table C.7 reports subsample correlations between VCV and the firm characteristics (full-sample correlations are reported in Table 3 of the paper)

Table C.6: Summary statistics of firm characteristics

This table reports summary statistics on the variables used in Table 3 of the paper. Size, Illiquidity and Coverage are measured in logs. The sample period for the Bid-ask spread and Coverage are 1983-2020 and 1990-2018, respectively. For other variables the sample is 1980-2020.

	Size	BM	Age	Vol.	TO	Illiq	Bid-Ask	Roll	Cov.
Observations	148,673	145,697	148,673	148,673	148,673	145,697	123,818	148,673	74,004
N	16,571	16,426	16,571	16,571	16,571	16,472	14,297	16,571	10,289
T	41	41	41	41	41	41	38	41	29
Mean	12.45	0.66	15.89	3.46	0.54	−16.63	2.23	0.97	2.06
s.d.	2.07	0.55	15.68	2.29	4.33	3.33	3.21	3.00	0.81
s.d. (CS)	1.93	0.49	15.54	2.01	1.38	2.93	2.35	2.68	0.80
s.d. (TS)	0.68	0.34	3.77	1.29	0.38	1.30	1.45	2.28	0.33
$q_{0.1}$	9.87	0.17	2	1.45	0.08	−21.15	0.06	−2.05	0.69
$q_{0.25}$	10.93	0.31	4	1.96	0.14	−19.21	0.21	−0.99	1.39
Median	12.31	0.54	11	2.88	0.29	−16.55	1.22	0.84	2.08
$q_{0.75}$	13.83	0.85	22	4.29	0.63	−14.00	3.02	2.26	2.64
$q_{0.9}$	15.21	1.26	36	6.11	1.16	−12.22	5.60	4.30	3.18

Table C.7: VCV and other firm characteristics

Notes: This table reproduces the correlations between VCV and other firm characteristics, reported in Table 3 of the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX stocks and for subsamples of observations prior to 2000 (1980-1999) and Post 2000 (2000-2020).

	NASDAQ	NYSE/AMEX	Pre-2000	Post-2000
	VCV	VCV	VCV	VCV
Size	−0.56	−0.67	−0.54	−0.77
BM ratio	0.20	0.19	0.17	0.17
Age	−0.14	−0.31	−0.26	−0.33
Volatility	0.32	0.38	0.29	0.52
Turnover	−0.28	−0.21	−0.29	−0.27
Illiquidity	0.64	0.68	0.62	0.76
Bid-Ask Spread	0.62	0.64	0.51	0.73
Roll's measure	0.29	0.08	0.28	0.23
Coverage	−0.48	−0.55	−0.59	−0.55

C.4 VCV and return reversals

Tables C.8 and C.9 report subsample results for the analyses in Section 4.2 of the paper, examining the relation between VCV, illiquidity, and return reversals.

Table C.8: VCV and the Bid-Ask Spread

This table reproduces the results in Table 4 of the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX stocks and for subsamples of observations prior to 2000 (1980-1999) and Post 2000 (2000-2020).

	Roll: Low	2	3	High	High-Low
NASDAQ					
Bid-Ask: Low	1.023	0.960	0.869	1.020	−0.002
2	1.203	1.207	1.207	1.100	−0.103**
3	1.407	1.348	1.520	1.447	0.040*
High	2.164	1.728	1.855	2.106	−0.058*
High-Low	1.141***	0.767***	0.985***	1.085***	
NYSE/AMEX					
Bid-Ask: Low	0.827	0.690	0.591	0.691	−0.136
2	1.017	0.938	0.807	0.772	−0.246***
3	1.239	1.254	1.114	1.009	−0.230**
High	1.822	1.693	1.675	1.889	0.067*
High-Low	0.994***	1.003***	1.084***	1.198***	
Pre-2000					
Bid-Ask: Low	1.084	1.043	1.013	1.367	0.284*
2	1.359	1.400	1.492	1.716	0.357**
3	1.607	1.531	1.699	1.730	0.123*
High	1.918	1.706	1.775	1.923	0.005
High-Low	0.834***	0.663***	0.762***	0.555**	
Post-2000					
Bid-Ask: Low	0.703	0.581	0.630	0.810	0.10**8
2	0.892	0.770	0.777	0.880	−0.012
3	1.221	1.158	1.094	1.153	−0.068**
High	2.064	1.705	1.824	2.225	0.161**
High-Low	1.361***	1.124***	1.195***	1.415***	

Table C.9: VCV and monthly reversals: Subsamples

This table reproduces the results in Table 5 of the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX stocks and for subsamples of observations prior to 2000 (January 1980 - December 1999) and Post 2000 (January 2000 - December 2020).

	Illiq: Low	2	3	High	High-Low
NASDAQ					
VCV: Low	−0.022	−0.034	−0.063	−0.137	−0.115***
2	−0.007	−0.008	−0.050	−0.117	−0.110***
3	0.001	−0.010	−0.015	−0.067	−0.068***
High	−0.003	−0.015	−0.009	−0.045	−0.042***
High-Low	0.019***	0.018***	0.055***	0.092***	
NYSE/AMEX					
VCV: Low	−0.031	−0.023	−0.054	−0.165	−0.134***
2	−0.023	−0.011	−0.052	−0.100	−0.076***
3	−0.018	−0.027	−0.041	−0.061	−0.044***
High	−0.026	−0.016	−0.008	−0.057	−0.030***
High-Low	0.004	0.007	0.046***	0.108***	
Pre-2000					
VCV: Low	−0.018	−0.035	−0.063	−0.134	−0.117***
2	0.010	−0.004	−0.048	−0.105	−0.115***
3	0.004	−0.014	−0.022	−0.070	−0.074***
High	−0.013	−0.009	−0.011	−0.053	−0.039***
High-Low	0.005	0.026***	0.052***	0.082***	
Post-2000					
VCV: Low	−0.029	−0.020	−0.036	−0.215	−0.186***
2	−0.018	−0.017	−0.032	−0.090	−0.073***
3	0.002	−0.016	−0.016	−0.055	−0.057***
High	−0.006	−0.016	−0.004	−0.041	−0.036***
High-Low	0.023***	0.004	0.033***	0.173***	

We consider weekly return reversals as another robustness check on the relation between VCV and reversals. We compute weekly return autocorrelations for each firm within each year in our sample. We then double-sort stocks within each year into quartiles based on annual estimates of VCV and Amihud's (2002) Illiquidity. Table C.10 reports the average weekly return autocorrelation, for each of these 16 groups. These weekly reversals are generally stronger than the monthly reversals reported in Table 5 in the paper and Table C.9 above. Nevertheless, we find a similar pattern as in Table 5, with return reversals increasing in Amihud Illiquidity and decreasing in VCV, consistent with the notion that VCV measures informed trade.

Table C.10: VCV and weekly reversals

This table reports the sample average of weekly return autocorrelations for 16 groups of stocks double-sorted within each year on Amihud (2002) *Illiquidity* and *VCV*. The final row and column report the difference in average weekly return autocorrelations between high and low quartiles, with significant differences at the 10%, 5%, and 1% level indicated by *, **, and ***. Source: CRSP.

	Illiq: Low	2	3	High	High-Low
FULL SAMPLE					
VCV: Low	−0.062	−0.065	−0.091	−0.112	−0.051**
2	−0.048	−0.047	−0.067	−0.106	−0.058***
3	−0.041	−0.039	−0.050	−0.096	−0.054***
High	−0.041	−0.027	−0.042	−0.100	−0.059***
High-Low	0.021*	0.039***	0.049***	0.013	
NASDAQ					
VCV: Low	−0.064	−0.070	−0.095	−0.103	−0.039
2	−0.050	−0.051	−0.073	−0.105	−0.054***
3	−0.035	−0.045	−0.052	−0.099	−0.065***
High	−0.033	−0.029	−0.047	−0.104	−0.071***
High-Low	0.031	0.041***	0.048***	−0.001	
NYSE/AMEX					
VCV: Low	−0.064	−0.070	−0.095	−0.103	−0.039**
1	−0.061	−0.061	−0.083	−0.137	−0.076***
2	−0.046	−0.041	−0.052	−0.109	−0.063***
3	−0.044	−0.032	−0.043	−0.082	−0.037***
High	−0.042	−0.024	−0.031	−0.080	−0.038
High-Low	0.019*	0.037***	0.052***	0.057	
Pre-2000					
VCV: Low	−0.060	−0.060	−0.083	−0.112	−0.053**
2	−0.039	−0.044	−0.062	−0.105	−0.066***
3	−0.036	−0.034	−0.049	−0.100	−0.063***
High	−0.040	−0.022	−0.037	−0.101	−0.062***
High-Low	0.020*	0.038***	0.046***	0.011	
Post-2000					
VCV: Low	−0.063	−0.068	−0.099	−0.119	−0.056
2	−0.055	−0.048	−0.068	−0.097	−0.042**
3	−0.046	−0.044	−0.049	−0.083	−0.037***
High	−0.043	−0.038	−0.048	−0.097	−0.054**
High-Low	0.020	0.030**	0.050***	0.022	

C.5 VCV and other measures of information asymmetry

Table C.11 reports summary statistics for the information asymmetry measures considered in Section 4.3 of the paper. Table C.12 reports subsample correlations between VCV and the other information asymmetry measures (full-sample correlations are reported in Table 6 of the paper). Table C.12 reports subsample regressions on the relation between VCV and the PIN (full-sample regressions are reported in Table 7 of the paper).

Table C.11: Summary statistics of information asymmetry measures

This table reports summary statistics on the information asymmetry measures used in Table 6 of the paper.

	PIN_{BHL}	PIN_{BH}	PIN_{EHO}	PIN_{DY}	Adj.PIN	PSOS	MIA	C2
Observations	76,410	76,870	32,526	37,918	37,918	37,918	25,461	122,553
N	11,316	11,334	4,365	4,634	4,634	4,634	3,877	13,746
T	18	18	19	22	22	22	23	40
Mean	0.19	0.21	0.19	0.21	0.17	0.28	0.41	0.02
s.d.	0.10	0.11	0.07	0.09	0.08	0.15	0.10	0.11
s.d. (CS)	0.10	0.10	0.06	0.09	0.07	0.15	0.10	0.11
s.d. (TS)	0.07	0.06	0.04	0.06	0.05	0.10	0.06	0.09
$q_{0.1}$	0.05	0.09	0.11	0.11	0.09	0.13	0.27	−0.12
$q_{0.25}$	0.11	0.13	0.15	0.14	0.12	0.17	0.34	−0.05
Median	0.20	0.19	0.19	0.19	0.16	0.23	0.42	0.02
$q_{0.75}$	0.26	0.27	0.23	0.24	0.21	0.33	0.47	0.08
$q_{0.9}$	0.31	0.36	0.28	0.33	0.28	0.50	0.52	0.14

Table C.12: VCV and other information asymmetry measures

This table reproduces the correlations between VCV and other information asymmetry measures, reported in Table 6 of the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX, and for subsamples of observations prior to 2000 (1980-1999) and post 2000 (2000-2020).

	NASDAQ VCV	NYSE/AMEX VCV	Pre-2000 VCV	Post-2000 VCV
PIN_{BHL}	0.43	0.52	0.39	0.62
PIN_{BH}	0.52	0.65	0.46	0.69
PIN_{EHO}	0.39	0.53	0.51	0.71
PIN_{DY}	0.40	0.57	0.52	0.73
Adjusted PIN	0.40	0.52	0.47	0.71
PSOS	0.29	0.46	0.43	0.57
MIA	0.22	0.31	0.10	0.28
C2	0.10	0.05	0.18	0.05

Table C.13: VCV and Adjusted PIN

This table reproduces the main regression in Table 7 of the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX, and for subsamples of observations prior to 2000 (1980-1999) and post 2000 (2000-2020).

	NASDAQ	NYSE/AMEX	Pre-2000	Post-2000
	VCV	VCV	VCV	VCV
PIN_{DY}	0.311	0.468***	0.321**	1.316***
	(0.265)	(0.144)	(0.120)	(0.214)
Adjusted PIN	0.756**	0.961***	0.912***	0.835**
	(0.278)	(0.160)	(0.171)	(0.213)
PSOS	−0.244*	0.060	0.052	0.088
	(0.130)	(0.058)	(0.064)	(0.073)
Observations	1,728	36,190	29,677	8,241
Adjusted R^2	0.342	0.356	0.319	0.498
Fixed effects	Yes	Yes	Yes	Yes

C.6 VCV and institutional ownership

Table C.14 reports summary statistics for the firm-quarter variables on institutional ownership considered in Section 4.4 of the paper. Table C.15 reports subsample regressions on the relation between quarterly VCV and institutional ownership (full-sample regressions are reported in Table 8 of the paper). To demonstrate robustness, Table C.16 reproduces Table 8 of the paper using firm-year observations of annual VCV and end-of-year (Quarter 4) observations of the institutional ownership variables.

Table C.14: Summary statistics of institutional ownership characteristics

This table reports summary statistics on the measures of institutional ownership used in Table 8 of the paper. Sample: 1980Q1-2018Q4.

	Holdings	Breadth	Monitors	Dedicated
Observations	346,858	346,858	346,858	346,679
N	12,176	12,176	12,176	12,169
T	116	116	116	116
Mean	49.20	5.25	3.54	6.27
s.d.	30.64	7.21	6.02	9.61
s.d. (CS)	27.72	7.21	6.00	5.10
s.d. (TS)	10.14	0.91	1.66	5.11
$q_{0.1}$	7.94	0.46	0	0
$q_{0.25}$	22.48	1.12	0	0
Median	48.90	2.95	1.43	1.68
$q_{0.75}$	74.43	6.15	4.58	9.73
$q_{0.9}$	90.01	12.39	9.59	19.05

Table C.15: VCV and institutional ownership: Subsamples

This table reproduces the main regression in Table 8 of the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX, and for subsamples of observations prior to 2000 (1980Q1-1999Q4) and post 2000 (2000Q1-2018Q4).

	NASDAQ	NYSE/AMEX	Pre-2000	Post-2000
	VCV	VCV	VCV	VCV
Holdings	0.001	-0.0001	0.002^{**}	-0.0001
	(0.001)	(0.0003)	(0.001)	(0.0003)
Breadth	-1.364^{**}	-1.254^{***}	-0.983^{***}	-1.168^{**}
	(0.573)	(0.348)	(0.269)	(0.434)
Monitors	0.640^{***}	0.645^{*}	0.050	0.645
	(0.192)	(0.337)	(0.103)	(0.382)
Dedicated	0.377^{***}	0.791^{**}	0.235^{***}	0.959^{***}
	(0.069)	(0.265)	(0.044)	(0.174)
Observations	201,265	145,414	126,120	220,559
Adjusted R^2	0.336	0.423	0.324	0.415
Fixed effects	Yes	Yes	Yes	Yes

Table C.16: VCV and institutional ownership: Annual data

This table reports the results from regressing annual firm-level coefficients of variation of daily volume market shares (VCV) on various measures of institutional ownership. *Holdings* is the percentage of shares of the firm held by institutional investors at the end of the year; *Breadth* is the percentage of all institutional investors that hold shares of the firm (Chen et al., 2002); $Monitors$ is the fraction of institutional investors in each firm for which the firm is in the top 10% of the institution's holdings (Fich et al., 2015); and *Dedicated* is the fraction of institutional investors in each firm that are classified as 'Dedicated' investors by Bushee and Noe (2000). All regressions include fixed effects for each year, industry, size decile, book-to-market decile and illiquidity decile. Two-way clustered standard errors, clustered at the year and industry level, are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5%, and 1% level. Sources: CRSP, 13F and the website of Brian Bushee `http://acct.wharton.upenn.edu/faculty/bushee/`. Sample: 1990-2018.

FULL SAMPLE	VCV			
	(1)	(2)	(3)	(4)
Holdings	-0.001	-0.001	-0.001	-0.001^{*}
	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Breadth	-1.098^{***}	-1.741^{***}	-1.067^{***}	-1.704^{***}
	(0.067)	(0.108)	(0.067)	(0.099)
Monitors		1.061^{***}		1.050^{***}
		(0.142)		(0.135)
Dedicated			0.487^{***}	0.481^{***}
			(0.134)	(0.130)
Observations	88,584	88,584	88,538	88,538
Adjusted R^2	0.408	0.410	0.410	0.411
Fixed effects	Yes	Yes	Yes	Yes
SUBSAMPLES	NASDAQ	NYSE/AMEX	Pre-2000	Post-2000
	VCV	VCV	VCV	VCV
Holdings	-0.001	-0.001^{*}	0.002^{***}	-0.001^{*}
	(0.001)	(0.0003)	(0.001)	(0.0005)
Breadth	-2.000^{***}	-1.637^{***}	-1.268^{***}	-1.787^{***}
	(0.297)	(0.138)	(0.086)	(0.124)
Monitors	1.343^{***}	0.969^{***}	0.347^{**}	1.259^{***}
	(0.213)	(0.150)	(0.123)	(0.136)
Dedicated	0.432^{***}	0.916^{***}	0.169^{***}	1.416^{***}
	(0.135)	(0.163)	(0.024)	(0.323)
Observations	51,656	36,882	32,541	55,997
Adjusted R^2	0.341	0.463	0.340	0.452
Fixed effects	Yes	Yes	Yes	Yes

C.7 Time series patterns of VCV

The solid black line in Figure C.1 panel A shows the cross-sectional equal-weighted average of the firm-level $VCV_{\%}$, for each year of our sample analyzed in Section 4 (1980-2020). The black dashed line and gray line show the annual averages of $VCV_{\$}$ and VCV_{TO}, respectively.

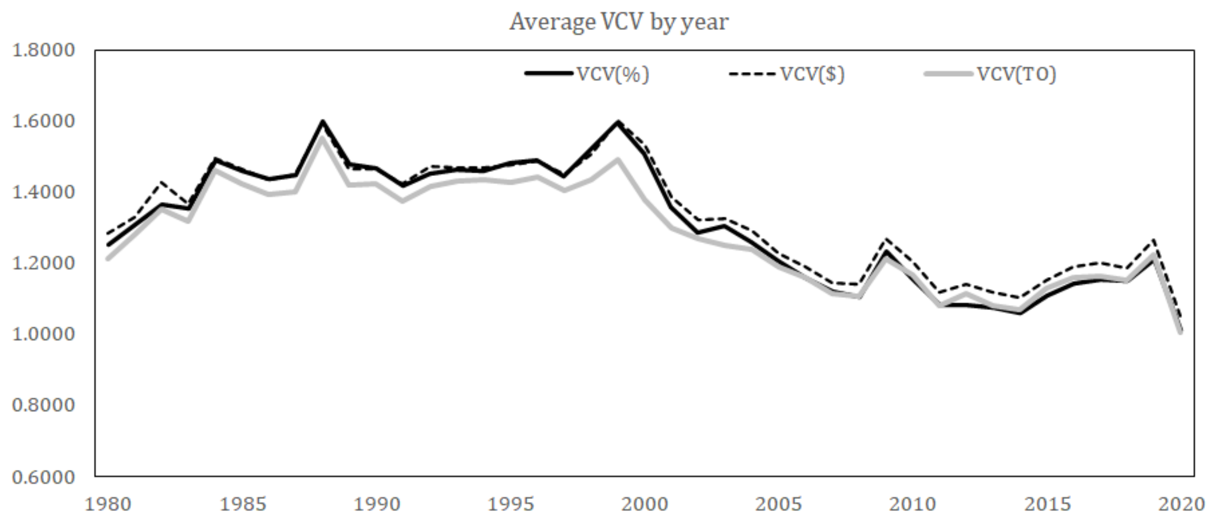

Figure C.1: This figure report the the annual cross-sectional average of annual firm-level VCVs, calculated from volume market shares ($VCV_{\%}$, black solid line), US dollar volume ($VCV_{\$}$, dashed line) and turnover (VCV_{TO}, gray line), over the period 1980-2020.

Table C.17 shows the results from regressing time-series changes in the quarterly average of firm-quarter observations of $VCV_{\%}$ (See Table C.3) on the percentage changes of three different indicators of liquidity supply: $Repo$, $Noise$ and TED. $Repo$ is the net repo rate, proposed by Adrian and Shin (2010) and also used by Johnson and So (2018) as an indicator of market maker leverage. Net repo is defined as the difference between end-of-quarter repo assets and liabilities on the aggregated balance sheets of Security Brokers and Dealers.[2] $Noise$ is the estimated standard error of yields on all traded Treasury bonds in deviation of the zero-coupon yield curve. Hu, Pan, and Wang (2013) argue that this measure is indicative of the aggregate supply of arbitrage capital.[3] TED is the end-of-quarter TED spread: the spread between 3-Month LIBOR and the 3-Month Treasury Bill yield (source: FRED), which Frazzini and Pedersen (2014) use as a measure of funding liquidity.

The regressions in Table C.17 are in differences instead of levels, because of the persistence of both VCV and the liquidity measures. We also include quarter-of-year dummies to control for seasonal variation in VCV (See Table C.4). The coefficients on all three liquidity measures are insignificant. Our

[2]Items FL662151003.Q and FL662151003.Q on the Financial Accounts of the United States - Z.1 (`https://www.federalreserve.gov/releases/z1/`)

[3]Daily observations of the noise measure are provided by Jun Pan: `http://en.saif.sjtu.edu.cn/junpan/`. We use the end-of-quarter observations.

simulation results in Table B.9 show that level differences in VCV can arise for different levels of market maker capacity, especially when the proportion of informed trade is high. The regression results in C.17 however suggest that observed aggregate variation in market maker capacity, proxied by $Repo$, $Noise$, and TED, does not have a significant impact on the average VCV.

Table C.17: VCV and Liquidity Supply

This table reports the result from regressing time-series changes in the average quarterly $VCV_{\%}$ on percentage changes Repo, Noise, and TED (See variable definitions above). I(Qi) is a dummy variable indicating the i^{th} quarter of the year. Newey-West standard errors are reported in parentheses. *, ** and *** indicate statistical significance at the 10%, 5%, and 1% level. Sample: 1990Q1-2016Q4

	ΔVCV			
	(1)	(2)	(3)	(4)
ΔRepo	0.057			0.059
	(0.039)			(0.037)
ΔNoise		-0.004		0.004
		(0.013)		(0.021)
ΔTED			-0.014	-0.015
			(0.012)	(0.018)
I(Q1)	-0.028***	-0.027***	-0.028***	-0.030***
	(0.007)	(0.007)	(0.007)	(0.007)
I(Q2)	0.039***	0.038***	0.039***	0.040***
	(0.014)	(0.013)	(0.014)	(0.012)
I(Q3)	-0.031	-0.030	-0.027	-0.029*
	(0.019)	(0.022)	(0.021)	(0.015)
I(Q4)	0.002	0.003	0.0002	-0.001
	(0.011)	(0.012)	(0.010)	(0.010)
Observations	107	107	107	107
Adjusted R^2	0.212	0.193	0.201	0.205

C.8 Brokerage closures

Table C.18 reports a differences-in-differences regression similar to Table 9 in the paper, with the difference that the control group does not include the full sample, but a subset of observations matched by firm size and analyst coverage.

Table C.18: Brokerage closures: Matched control group

This table reproduces the results in Table 9 of the paper, with the difference that the control group contains a selective set of matched observations. The treatment sample consist of 1,764 stocks. For each firm in the treatment group, we select three control firms matched by firm size and analyst coverage in the calendar year prior to the event. For all 7,056 stocks, we compute VCV over the months [-14,-3], and over the months [3,14], with the brokerage closure occurring in month 0.

	Full sample	Coverage < 10	Coverage < 5
	VCV	VCV	VCV
After × Treated	0.028**	0.055***	0.099**
	(0.014)	(0.021)	(0.048)
After	−0.068***	−0.085***	−0.094***
	(0.012)	(0.014)	(0.017)
Treated	−0.013	−0.033	−0.017
	(0.009)	(0.023)	(0.039)
Observations	14,112	4,232	1,320
Adjusted R^2	0.401	0.395	0.434
Fixed effects	Yes	Yes	Yes

C.9 Earnings announcements

Figure C.2 displays the path of the cross-sectional VCV for subsamples of the data. Figure 4 in the paper shows the full-sample plot.

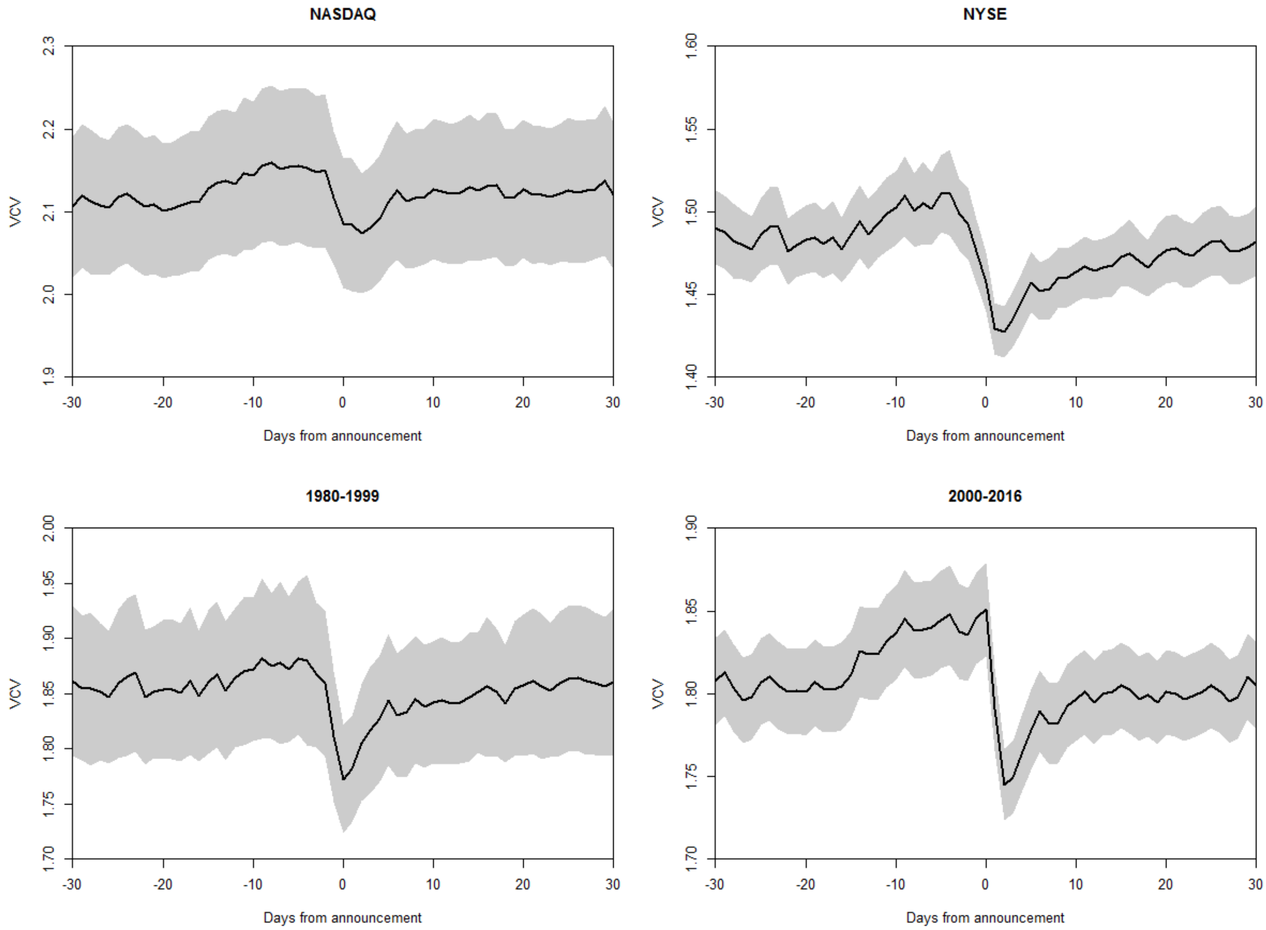

Figure C.2: This figure reproduces Figure 4 in the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX, and for subsamples of observations prior to 2000 (1980-1999) and post 2000 (2000-2016).

In addition to the cross-sectional VCV, we also consider changes in VCV from short time-series around the announcement dates. For each quarterly earnings announcement, we compute the coefficient of variation of trading volume shares ($V_{\%}$) using volumes over the days -11 to -2 before the announcement day and over the days +2 to +11 after the announcement date. In Table C.19, we report the means and medians for the VCV before and after, across all N quarterly earnings announcements. The results indicate that the firm-level VCV on average decreases around the earnings announcement date, consistent with the path of the cross-sectional VCV in Figure 4 of the paper. The difference is small but highly significant. Also the median VCV is higher before the announcement. Table C.19 reports these results on the full sample and various subsamples.

Table C.19: 10-day VCV before and after Earnings announcements

This table reports the means and medians of firm-level VCVs before (days -11:-2) and after (days +2:+11) quarterly earnings announcements, computed from 10-day time-series of daily volume shares. The table includes full-sample results as well as results for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX, and for subsamples of observations prior to 2000 (1980-1999) and post 2000 (2000-2016). The final columns show the results for surprising and non-surprising announcements, defined by the absolute Standardized Unexpected Earnings (SUE - See Figure 5 in the paper for details).

	Full sample	NASDAQ	NYSE	Pre-2000	Post-2000	Low |SUE|	High |SUE|
Mean before	2.666	2.692	2.637	2.810	2.549	2.599	2.650
Mean after	2.649	2.675	2.621	2.794	2.532	2.582	2.629
Mean difference	-0.017	-0.018	-0.016	-0.016	-0.017	-0.017	-0.022
t-statistic	-21.435	-15.831	-14.454	-11.700	-19.822	-9.865	-11.217
Median before	2.562	2.590	2.532	2.721	2.460	2.499	2.545
Median after	2.549	2.577	2.520	2.707	2.449	2.483	2.529
Median difference	-0.009	-0.010	-0.009	-0.011	-0.008	-0.011	-0.011
N	270933	141848	129085	121083	149850	40083	42523

References

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

Adrian, T., and Shin, H. S. (2010). Liquidity and leverage. *Journal of Financial Intermediation*, 19(3), 418-437.

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

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

Foster, F. D., and Viswanathan, S. (1990). A theory of the interday variations in volume, variance, and trading costs in securities markets. The Review of Financial Studies, 3(4), 593-624.

Foster, F. D., and Viswanathan, S. (1994). Strategic trading with asymmetrically informed traders and long-lived information. *Journal of Financial and Quantitative Analysis*, 29, 499-518.

Foster, F. D., and Viswanathan, S. (1996). Strategic trading when agents forecast the forecasts of others. *Journal of Finance*, 51, 1437-1478.

Frazzini, A., and Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, 111(1), 1-25.

Holden, C. W., and Subrahmanyam, A. (1992). Long-lived private information and imperfect competition. *Journal of Finance*, 47, 247-270.

Hu, G. X., Pan, J., and Wang, J. (2013). Noise as information for illiquidity. *The Journal of Finance*, 68(6), 2341-2382.

Johnson, T. L., and So, E. C. (2018). Asymmetric trading costs prior to earnings announcements: Implications for price discovery and returns. Journal of Accounting Research, 56(1), 217-263.

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