Lof, Matthijs; van Bommel, Jos

Asymmetric information and the distribution of trading volume

Published in:
Journal of Corporate Finance

DOI:
10.1016/j.jcorpfin.2023.102464

Published: 01/10/2023

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Lof, M., & van Bommel, J. (2023). Asymmetric information and the distribution of trading volume. *Journal of Corporate Finance*, *82*, Article 102464. https://doi.org/10.1016/j.jcorpfin.2023.102464

Journal of Corporate Finance 82 (2023) 102464

Contents lists available at ScienceDirect

Journal of Corporate Finance

journal homepage: www.elsevier.com/locate/jcorpfin

Asymmetric information and the distribution of trading volume[☆]

Matthijs Lof[a,*], Jos van Bommel[b]

[a] Aalto University School of Business, P.O. BOX 21210, 00076 Aalto, Finland
[b] University of Luxembourg, 6, rue Richard Coudenhove-Kalergi L-1359, Luxembourg

A R T I C L E I N F O

Editor: E Lyandres

JEL classification:
D82
G12
G14

Keywords:
VCV
Trading volume
Informed trading
Information asymmetry

A B S T R A C T

We propose the Volume Coefficient of Variation (VCV), the ratio of the standard deviation to the mean of trading volume, as a new and simple 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 obtain VCV from daily observations of trading volume and provide extensive evidence supporting the hypothesis that VCV indicates information asymmetry, by studying return reversals, institutional ownership, and 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.

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 use a canonical microstructure model to show that the coefficient of variation of trading volume increases monotonically in the proportion of informed traders in the market. We present simulation and empirical analysis providing evidence that the Volume Coefficient of Variation (VCV) is a robust and powerful measure of information asymmetry. VCV is easy to compute and only requires observable trading volume. VCV can be computed from both time-series and cross-sections of trading volume, to analyze information asymmetry across firms and over time.[1]

Using a market microstructure model based on Kyle (1985), we 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

☆ We thank an anonymous referee, Anna Bayona, Alex Boulatov, Darwin Choi, Jaewon Choi, Tarun Chordia, Jonathan Cohn, David Easley, Eliezer Fich, Carlos Forner Rodríguez, Thierry Foucault, Bruce Grundy, Alex Gümbel, 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 and Nikolas Breikopf for help with the EDGAR data.

* Corresponding author.
E-mail addresses: matthijs.lof@aalto.fi (M. Lof), jos.vanbommel@uni.lu (J. van Bommel).

[1] To the best of our knowledge, we are the first to relate the coefficient of variation of trading volume to asymmetric information. Subrahmanyam and Anshuman (2001) consider 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.

https://doi.org/10.1016/j.jcorpfin.2023.102464
Received 19 December 2022; Received in revised form 10 June 2023; Accepted 21 July 2023
Available online 26 July 2023

does so at a higher rate. The coefficient of variation of trading volume 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 financial economics literature. Within the context of capital markets, information asymmetry affects liquidity, asset prices, and financing and investment decisions. Since information asymmetry is not directly observable, numerous empirical proxies have been proposed in the literature. Early papers use the bid–ask spread or other illiquidity measures to proxy for information asymmetry (e.g. Amihud and Mendelson, 1986; Welker, 1995; Healy 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). We find that, unlike VCV, illiquidity measures such as bid–ask spreads and Amihud (2002) Illiquidity predict return reversals, which are a consequence of uninformed rather than informed order flow, and should therefore be avoided as proxies of asymmetric information. Easley et al. (1996, 1997a,b) 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), Jurkatis (2022)). 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, while Yang et al. (2020) suggest abnormal idiosyncratic volatility (AIV) prior to earnings announcements as an indicator of information asymmetry. 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 even from relatively small samples. In addition to robustness to sample size, we examine various modifications of our benchmark 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 simulated VCVs 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 1962 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 as a firm-level measure is not sensitive to aggregate market-level variation in trading volume. We compare VCV to annual firm-level characteristics and other information asymmetry measures. We find that VCV is higher for smaller and younger firms with lower analyst coverage, 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 and bid–ask spreads, 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

[2] The PIN measure has been widely used to study information asymmetry and its relation to numerous other characteristics, including analyst coverage (Easley et al., 1998; Dang et al., 2021), CEO myopia (Antia et al., 2010), disclosure quality (Vega, 2006; Brown and Hillegeist, 2007; Dumitrescu and Zakriya, 2022), dividend policy (De Cesari and Huang-Meier, 2015) earnings surprises (Brown et al., 2009), executives' social media activity (Feng and Johansson, 2019), insider trading (Lee et al., 2014), institutional ownership (Brockman and Yan, 2009; Boone and White, 2015), labor investment efficiency (Ben-Nasr and Alshwer, 2016), pricing of information asymmetry (Easley et al., 2002, 2010; Mohanram and Rajgopal, 2009; Hwang et al., 2013), and reputation incentives (Sila et al., 2017).

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), Bushman et al. (2017), Chen and Lin (2017), To et al. (2018), and Luong and Qiu (2021). Exogenous terminations in analyst coverage disrupt the information environment of firms 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), 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., 2022).

VCV is a very flexible measure, and can be computed for volume observations from different time intervals. 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 also examine subsamples of NYSE/AMEX and NASDAQ stocks, as well as time-series subsamples, 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, return reversals, 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 8-K filings. 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. We refer to σ as *trading intensity*. A proportion η of the M liquidity seekers is informed, with ηM being an integer. 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:

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

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]

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):

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

[3] This expression for trading volume is also used by Admati and Pfleiderer (1988) and Grundy and McNichols (1989). 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 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]:

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

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:

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

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 $\left(|\sum_M y_i|\right)$. The variances and covariances of these three components are derived in Internet Appendix A. For large M, we find that the variance 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.

Proposition 1. *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. *Consider the sample mean ($\hat{\mu}_V$) and sample standard deviation ($\hat{\sigma}_V$) as consistent estimators of the population mean (μ_V) and population standard deviation (σ_V) of trading volume. Then, by the continuous mapping theorem,*

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

is a consistent estimator of $\frac{\sigma_V}{\mu_V}$. It thus follows from Proposition 1 that E[VCV] increases monotonically in η such that the Volume Coefficient of Variation (VCV) is a measure of informed trade.

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. In subsequent sections, we present simulations and empirical analyses in the next sections that strongly support this result. Our simulation analysis furthermore show that VCV increases in the *average* proportion of informed trade, when we allow η, M and σ to vary across observations

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. Since its denominator can become very small or negative, $\hat{\eta}$ behaves poorly in small samples or alternative model specifications. 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 consider informed and uninformed traders having different trading intensities, and define η and M as:

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

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

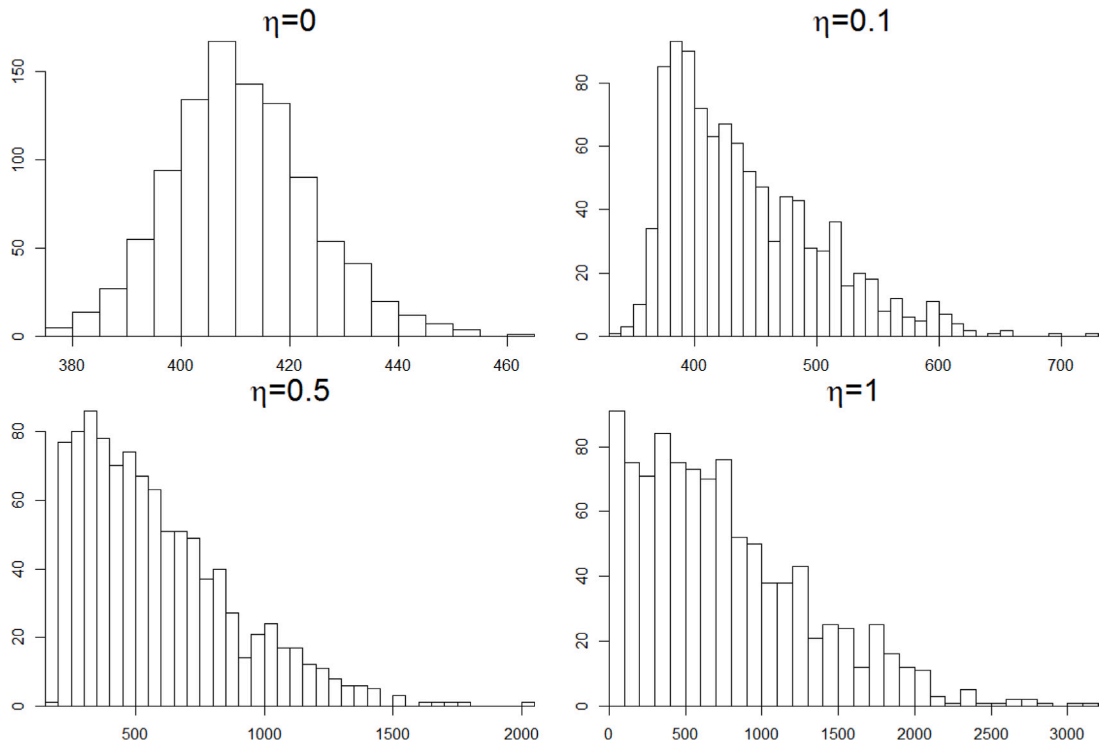

Fig. 1. Histogram of T = 1000 volume realizations simulated from the model outlined in Section 2, for various values of the proportion of informed trading η. The number of liquidity seekers (M) is 1000 and the trading intensity (σ) is fixed at unity.

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, η is the proportion of informed *trade*, rather than the proportion of informed *traders*. M is a measure of the number of intensity-weighted traders, with the uninformed trading intensity as numeraire.

It is important to notice that our model is silent about the *motivations* of trade of both informed and uninformed investors, but merely establishes the statistical relationship between VCV and the proportion of informed trade. In Internet Appendix A.2, 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. We find (similar to Holden and Subrahmanyam, 1992; Foster and Viswanathan, 1994, 1996 and others) that informed traders increase their trading intensity in the presence of competition, but that the relationship between the *equilibrium* proportion of informed trade and the coefficient of variation of trading volume is equivalent to Proposition 1 with η and M defined as in (9).

Further enriching the model with risk aversion, long lived information, or strategic uninformed trading (as in Foster and Viswanathan, 1990 or Admati and Pfleiderer, 1988) will not change Proposition 1, as there will always be an equilibrium proportion of informed trade η and a weighted number of liquidity seekers M. 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.

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.

Fig. 1 displays four histograms of simulated volumes with $M = 1000$ liquidity seekers, for different values of η. The sample size is $T = 1000$ trading sessions. The simulation confirms the analysis in the previous section: in the case of no informed traders ($\eta = 0$),

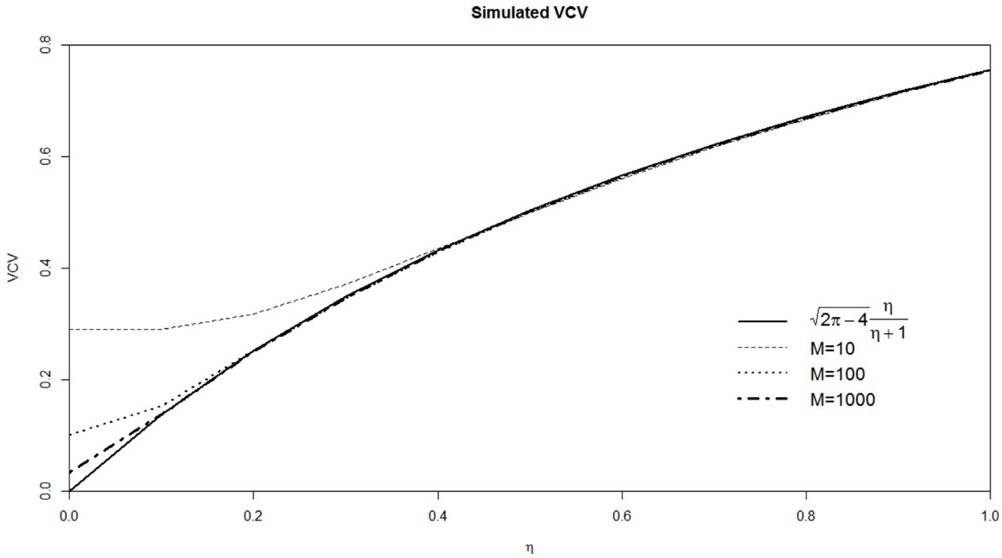

Fig. 2. Average VCV obtained from $R = 1,000,000$ replications of $T = 100$ volume realizations simulated from the model outlined in Section 2, for various values of the proportion of informed trading η and number of liquidity seekers M.

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.[5]

Fig. 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.

In Table 1, Panel A, we report the average VCV as plotted in Fig. 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 $\hat{\eta}$ (Eq. (8)). Both VCV and $\hat{\eta}$ 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 $\hat{\eta}$ 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 results for smaller simulated samples, of $T = 10$ trading sessions. We still find the average VCV and $\hat{\eta}$ 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 $\hat{\eta}$, 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 main 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

[5] 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.

Panel A: $T = 100$										
η	0	0.2	0.5	0.8	1	0	0.2	0.5	0.8	1
			VCV					$\hat{\eta}$		
			$M = 10$					$M = 10$		
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					$\hat{\eta}$		
			$M = 10$					$M = 10$		
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

This table reports the average and standard deviation of VCV (left) and $\hat{\eta}$ (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.

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 $\hat{\eta}$ clearly does not perform well as a measure of informed trading beyond the benchmark model. The simulated observations of $\hat{\eta}$ 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 $\hat{\eta}$ 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 over the period January 1962–December 2020. Our sample consists of firm-year observations for common stocks listed on NYSE, AMEX, and NASDAQ. Inclusion in the sample requires that, for a given calendar year, the stock has a strictly positive mean and standard deviation of daily trading volume in CRSP along with a recorded book value in COMPUSTAT.[6]

[6] For NASDAQ listed firms, we adjust trading volume prior to 2004 following Gao and Ritter (2010): 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.

Table 2
VCV summary statistics.

	VCV_{USD}	$VCV_{\%}$	VCV_{TO}
Observations	205,863	205,863	205,863
N	19,435	19,435	19,435
T	59	59	59
Mean	1.511	1.493	1.460
s.d.	1.003	1.019	0.967
s.d. (CS)	0.894	0.903	0.860
s.d. (TS)	0.638	0.643	0.611
$q_{0.1}$	0.617	0.580	0.609
$q_{0.25}$	0.887	0.859	0.860
Median	1.281	1.263	1.228
$q_{0.75}$	1.814	1.804	1.744
$q_{0.9}$	2.582	2.581	2.509
ρ	0.238	0.246	0.254
Correlations			
$VCV_{\%}$	0.982		
VCV_{TO}	0.977	0.966	

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: 1962–2020.

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:

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

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 (Kumar and Lee, 2006) and common liquidity shocks (Admati and Pfleiderer, 1988; Brogaard et al., 2018), 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 the other volume definitions.[7]

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

[7] 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 (1962–1999) and post 2000 (2000–2020), showing that the three measures of VCV behave fairly similar across these subsamples.

Table 3
VCV and other firm characteristics.
Sources: CRSP, COMPUSTAT, and IBES.

	VCV	Size	BM	Age	Vol.	Turn.	Illiq	B-A	Roll
Size	−0.64								
BM	0.21	−0.31							
Age	−0.27	0.34	0.11						
Volatility	0.34	−0.58	0.02	−0.36					
Turnover	−0.28	0.24	−0.21	−0.03	0.25				
Illiquidity	0.69	−0.94	0.32	−0.35	0.53	−0.44			
Bid–Ask spread	0.66	−0.87	0.30	−0.22	0.58	−0.48	0.92		
Roll's measure	0.23	−0.33	0.23	−0.03	0.22	−0.28	0.39	0.53	
Coverage	−0.55	0.77	−0.23	0.22	−0.29	0.46	−0.79	−0.69	−0.28

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 closing bid–ask spreads $\frac{ask_{i,t}-bid_{i,t}}{\frac{1}{2}(ask_{i,t}+bid_{i,t})}$ (daily closing bid and ask prices are available in CRSP from 1982). *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 (available in IBES starting from 1981).

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 counterparts, 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 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 its midpoint $\frac{ask_{i,t}-bid_{i,t}}{\frac{1}{2}(ask_{i,t}+bid_{i,t})}$, following Chung and Zhang (2014). Roll's (1984) measure is the square root of the negative of the daily return autocovariance $\sqrt{-Cov\left(R_{i,t}, R_{i,t-1}\right)}$.[8] 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 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. VCV and illiquidity

We can see from Table 3 that VCV is strongly correlated with firm-level liquidity indicators: Amihud (2002) Illiquidity, bid–ask spread, size, and to a lesser extent volatility, turnover, and Roll's measure. To illustrate that VCV differs from these liquidity characteristics and proxies specifically for information asymmetry, we analyze 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)), which are generally considered a sign of uninformed order flow. As argued by Huang and Stoll (1997), price changes due to informed trading are less likely to be reversed by the bid–ask bounce, and are thus characterized by weaker reversals.

[8] 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 4
VCV and monthly reversals.
Source: CRSP. Sample: 1962–2020 (Panel A) and 1982–2020 (Panel B).

A	*ILLIQ*: Low	2	3	High	High-Low
VCV: Low	−0.022	−0.026	−0.057	−0.128	−0.106***
2	−0.008	−0.011	−0.05	−0.117	−0.109***
3	−0.006	−0.02	−0.022	−0.071	−0.065***
High	−0.013	−0.014	−0.011	−0.053	−0.041***
High-Low	0.009**	0.012***	0.047***	0.074***	
B	*Bid–Ask*: Low	2	3	High	High-Low
VCV: Low	−0.022	−0.025	−0.058	−0.145	−0.123***
2	−0.013	−0.017	−0.032	−0.123	−0.110***
3	−0.003	−0.017	−0.023	−0.073	−0.070***
High	−0.013	−0.011	−0.013	−0.049	−0.036***
High-Low	0.009**	0.014***	0.044***	0.097***	

Panel A 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 monthly autocorrelation between high and low quartiles, with significant differences at the 10%, 5%, and 1% level indicated by *, **, and ***. In Panel B, stocks are double sorted by monthly estimate of VCV and the Bid–Ask spread.

Liquidity proxies in the literature are broadly separated into measures of price impact and proxies of the bid–ask spread (see, e.g. Goyenko et al. (2009), Fong et al. (2017), Le and Gregoriou (2020)). Amihud's (2002) illiquidity ratio of absolute daily return to volume is the most popular measure of price impact in the finance literature. In a recent survey, Le and Gregoriou (2020) find that compared to other proxies of the bid–ask spread, the daily closing bid ask-spread proposed by Chung and Zhang (2014) provides a close approximation of the intraday (TAQ) spread. Both Amihud Illiquidity and the daily closing bid–ask spread are low-frequency proxies that are (similar to VCV) based on daily (CRSP) data and are therefore easily accessible to researchers and cover longer time periods than intraday measures based on TAQ data. Our sample of monthly averages of Amihud Illiquidity covers the same sample as VCV (1962–2020) while the monthly average closing bid–ask spread is available from 1982.

We show in this subsection that these two indicators of liquidity (Amihud Illiquidity and the closing bid–ask spread) predict stronger short-term reversals, suggesting they are indicative of uninformed rather than informed order flow. VCV is on the other hand associated with weaker reversals, consistent with the hypothesis that VCV indeed measures informed trade.[9]

To study the relation between VCV, illiquidity and reversals, 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: $cor(R_{i,t}, R_{i,t+1})$. Panel A of Table 4 reports the average of these monthly return autocorrelations, for each of the 16 groups. Across all groups, we find return reversals (negative autocorrelation). These reversals are clearly stronger for the more illiquid stocks, as shown in the last column. However, within each liquidity quartile, we find that reversals are decreasing in VCV: The final row of Table 4 shows that return autocorrelation is on average significantly higher (i.e. reversals are weaker) for High VCV stocks than for Low VCV stocks. In Panel B of Table 4, we repeat this analysis by double-sorting stocks on VCV and the bid–ask spread. The results are similar to Panel A, with short-term reversals significantly increasing in the bid–ask spread, while decreasing in VCV. Internet Appendix Section C.4 reports various additional test to demonstrate the robustness of this result.[10]

Since short-term reversals are a consequence of uninformed order flow, these results provide evidence that Amihud's illiquidity measure and the bid–ask spread are increasing in uninformed order imbalance, and therefore not effective as proxies for informed trading. VCV, on the other hand, is not a proxy of general illiquidity but specifically captures information asymmetry. In the next subsection, we have a closer look at the empirical relation between VCV and existing measures of asymmetric information.

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.

[9] Also Llorente et al. (2002), Hameed et al. (2008), Odders-White and Ready (2008), Bongaerts et al. (2016), and Johnson and So (2018) use various asymmetric information measures to show that asymmetric information is associated with weaker short-term reversals, while Wang (2021) provides evidence of insider trades predicting future returns.

[10] Tables C.8 and C.9 reports subsample results. In Table C.10, we also analyze other liquidity proxies besides Amihud Illiquidity and the bid–ask spread, including size, firm age, volatility, analyst coverage, and institutional ownership. In Table C.11 we furthermore analyze weekly instead of monthly reversals.

Table 5
VCV and other information asymmetry measures.
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}	0.65	0.74						
PIN_{EHO}	0.57	0.58	0.68					
PIN_{DY}	0.61	0.61	0.70	0.86				
Adjusted PIN	0.56	0.55	0.71	0.65	0.74			
PSOS	0.51	0.44	0.48	0.65	0.74	0.44		
MIA	0.25	0.36	0.43	0.11	0.22	0.31	0.04	
C2	0.10	0.14	0.14	0.02	0.03	0.03	0.04	0.01

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 et al. (2004). PIN_{BH} is estimated by Brown and Hillegeist (2007). PIN_{EHO} is estimated by Easley et al. (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).

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 et al. (2004) – PIN_{BHL}); Brown and Hillegeist (2007) – PIN_{BH}); and Duarte and Young (2006 – PIN_{DY}).[11] 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 (detrended) turnover, while controlling for daily lagged returns (see Llorente et al. (2002, for details).

Table 5 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 1962–2020, while the available PIN measures are not available prior to 1983 and post 2010. As discussed in the introduction, the classification into buy and sell orders has been increasingly problematic in the time of continuous trading.[12]

Duarte and Young (2009) argue that PIN does not only capture informed trading, but also other illiquidity effects. They therefore propose two alternative variables: *Adjusted PIN*, which is proposed as a cleaner measure of asymmetric information; and *PSOS* (probability of symmetric order-flow shock), which is considered a measure of illiquidity unrelated to asymmetric information. These additional variables are included in Table 5. Both Adjusted PIN and PSOS are positively correlated with VCV.

In Table 6, 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.[13] The regression results indicate that VCV is significantly associated with both PIN and adjusted PIN, while there is no robust relation between VCV and PSOS, thereby supporting the reversal results in the previous subsection that VCV is indicative of asymmetric information rather than general illiquidity.

4.4. VCV and institutional ownership

Several studies find that institutional ownership leads to an improvement in disclosure practices and therefore lower information asymmetry (e.g. Bushee and Noe (2000), Boone and White (2015), Bai et al. (2016), Cheng et al. (2020)). In this subsection, we study the relationship between VCV and various indicators of institutional ownership that we obtain from quarterly 13F filings

[11] 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 C.5.

[12] The time-series samples and number of distinct stocks N in our sample with coverage of VCV, PIN_{BH}, PIN_{DY} and MIA are: 1962–2020, N = 19,435 (VCV); 1993–2010, N = 12,374 (PIN_{BH}); 1993–2010, N = 12,346 (PIN_{BHL}); 1983–2001, N = 4550 (PIN_{EHO}); 1983–2004, N = 4826 (PIN_{DY}); 1996–2018, N = 3910 (MIA). See Internet Appendix Table C.12 for details.

[13] 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 (See also, e.g., Gassen et al. (2020), who apply fixed effects to control for nonlinear effects of illiquidity). We show in Table C.15 of the Internet Appendix that the main results are robust to the direct inclusion of size, book-to-market and illiquidity as controls, instead of fixed effects.

Table 6
VCV and Adjusted PIN.
Sources: CRSP and the website of Jefferson Duarte (http://www.owlnet.rice.edu/~jd10/). Sample 1983–2004.

	VCV			
	(1)	(2)	(3)	(4)
PIN_{DY}	1.200***		1.673***	1.104***
	(0.135)		(0.153)	(0.129)
Adjusted PIN	0.834***	1.592***		0.899***
	(0.117)	(0.154)		(0.109)
PSOS	−0.088	0.318***	−0.280***	
	(0.059)	(0.066)	(0.052)	
Observations	40,968	40,968	40,968	40,968
Adjusted R^2	0.401	0.396	0.400	0.403
Fixed effects	Yes	Yes	Yes	Yes

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.

Table 7
VCV and institutional ownership.
Sources: CRSP, Refinitiv (13F) and Brian Bushee http://acct.wharton.upenn.edu/faculty/bushee/. Sample period: 1980Q1-2020Q4.

	VCV				
	(1)	(2)	(3)	(4)	(5)
Holdings	0.047	0.068	0.063	0.063	0.057
	(0.048)	(0.053)	(0.050)	(0.053)	(0.050)
Breadth		−0.921***	−1.224***	−0.964***	−1.302***
		(0.183)	(0.236)	(0.208)	(0.258)
Monitors			0.611***		0.664***
			(0.180)		(0.180)
Dedicated				0.255***	0.250***
				(0.074)	(0.072)
Observations	508,072	508,072	508,072	502,721	502,721
Adjusted R^2	0.420	0.422	0.423	0.424	0.425
Fixed effects	Yes	Yes	Yes	Yes	Yes

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.

recorded in the Refinitiv database on WRDS. We consider institutional holdings (*Holdings*), which is defined as the percentage of a firm's shares held by institutional investors at the end of each quarter. The association between *Holdings* and information asymmetry is not straightforward. For instance, a firm could have a high value of *Holdings*, even if ownership is highly concentrated with only a small number of institutional investors possessing a significant portion of the firm's shares. We hypothesize a more robust link between information asymmetry and the number of distinct institutional investors in a specific firm. This is captured by the variable breadth of ownership (*Breadth*), which is 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 13F database at the end of each quarter (Chen et al., 2002). Grullon et al. (2004) and Cao and Wu (2022) find that a high institutional ownership base (i.e. high breadth of ownership) is associated with higher firm visibility and investor recognition.

The first two columns of Table 7 report the results from regressing quarterly estimates of VCV on end-of-quarter characteristics on *Holdings* and *Breadth*. The results show indeed that VCV has a significantly negative association with *Breadth*, while the relation with *Holdings* is insignificant. VCV is thus on average lower (implying lower information asymmetry) for firms that have high breadth of ownership.

In addition, we consider two variables 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 and are generally considered better informed than 'quasi-indexing' and 'transient' investors, (Cheng et al., 2020).[14]

The variable *Monitors* in Table 7 is the percentage of institutional investors in each firm that are defined as monitoring investors. The variable *Dedicated* in Table 7 is the percentage of institutional investors in each firm that are classified as dedicated investors. Columns 3–5 of Table 7 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 7 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 7.[15]

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), Chen and Lin (2017), To et al. (2018), Luong and Qiu (2021)) 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 1764 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 4.1, 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 8. 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 negative, which reflects that VCV is on average decreasing over time.[16] The *Treated* coefficient indicates that there is a minor difference between the VCV of treated and control firms, prior to the event.[17]

The second and third column of Table 8 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 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 8 provide strong evidence for our proposition that VCV measures information asymmetry.

[14] 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 Refinitiv database is made available on the website of Brian Bushee http://acct.wharton.upenn.edu/faculty/bushee/.

[15] 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. We also reproduce the regressions in Table 7 using an unscaled version of Breadth, finding qualitatively similar results. Finally, we include the proportion of Quasi-indexing and Transient investors to the regression, in addition to Dedicated. We find that VCV decreases in the proportion of Quasi-indexers, consistent with Hillegeist and Weng (2021), who find that find high quasi-indexer ownership causes less insider trading.

[16] 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. Collver (2007), Chen et al. (2010), Petacchi (2015), Beaver et al. (2018), Pawlewicz (2018)).

[17] Internet Appendix Section C.8 provides results for a smaller 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.

Table 8
Brokerage closures.

	Full sample VCV	Coverage ≤ 10 VCV	Coverage ≤ 5 VCV
After ×Treated	0.042***	0.060***	0.081**
	(0.010)	(0.006)	(0.031)
After	−0.030**	−0.041**	−0.047**
	(0.012)	(0.014)	(0.016)
Treated	−0.047**	−0.039*	−0.044
	(0.019)	(0.018)	(0.026)
Observations	98,696	69,354	45,146
Adjusted R^2	0.428	0.363	0.320
Fixed effects	Yes	Yes	Yes

This table reports the results from difference-in-differences regressions around brokerage closure-induced terminations of analyst coverage. The treatment sample consist of 1889 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 47,459 observations. For all 49,348 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 98,696 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 10 (5) or less 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.

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 $N = 596{,}881$ quarterly earnings announcement dates from COMPUSTAT, from a total of 17,994 distinct NYSE, AMEX, and NASDAQ listed US firms over the period 1971–2020. 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[18]:

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

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.

The black line in Fig. 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). Fig. 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.

Table 9 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 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 9 shows, the differences between VCV are statistically significant up to ten 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.

[18] 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.

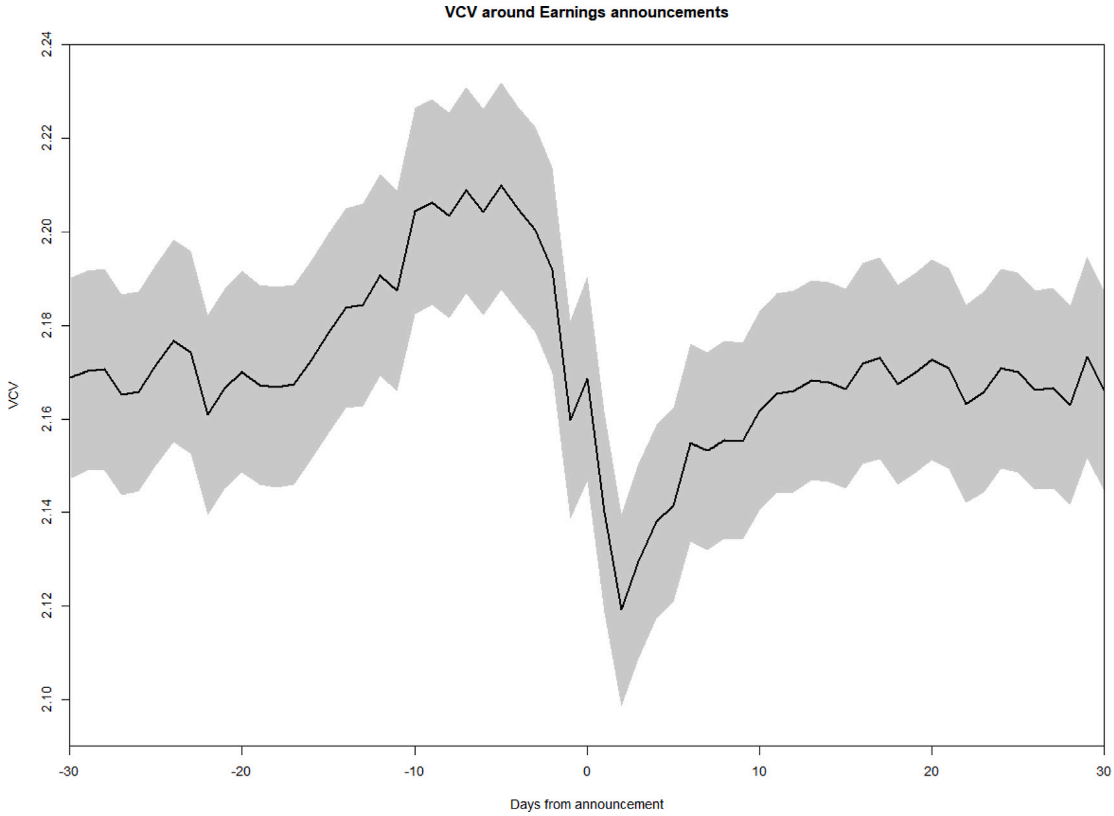

Fig. 3. The black line shows the evolution of the daily cross-sectional VCV_{XS} around quarterly earnings announcements. The full sample includes all daily trading volumes over a 61 day event-window (day −30:30) around $N = 596,881$ quarterly announcements (sources: CRSP and COMPUSTAT). The reported 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).

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 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 Fig. 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 the absolute value of SUE.

Fig. 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 Fig. 3. This is because both the surprising and non-surprising subsamples are restricted to those firms for which analyst expectations are available. For completeness, Fig. 4 also displays the cross-sectional VCV around the announcements for which no analyst forecasts are reported in IBES, using the same time series-sample for which SUE is available (1983–2020). 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 8. Overall, the breakdown in Fig. 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

Table 9
Volume around earnings announcements.

d	$\hat{\mu}_d \times 1000$			$\hat{\sigma}_d \times 1000$			$VCV_{XS,d}$		
	Before	After	Diff	Before	After	Diff	Before	After	Diff
0	0.996	0.996	0.000	2.159	2.159	0.000	2.169	2.169	0.000
1	0.988	1.014	0.026***	2.134	2.170	0.036***	2.160	2.140	−0.020*
2	0.951	1.018	0.067***	2.085	2.157	0.072***	2.192	2.119	−0.073***
3	0.941	1.003	0.062***	2.070	2.136	0.065***	2.200	2.130	−0.071***
4	0.932	0.992	0.061***	2.054	2.122	0.067***	2.205	2.138	−0.067***
5	0.934	0.985	0.052***	2.063	2.110	0.047***	2.210	2.141	−0.068***
6	0.934	0.978	0.045***	2.058	2.108	0.050***	2.204	2.155	−0.049***
7	0.937	0.975	0.038***	2.070	2.100	0.029***	2.209	2.153	−0.056***
8	0.934	0.978	0.044***	2.058	2.108	0.049***	2.204	2.156	−0.048**
9	0.936	0.973	0.037***	2.065	2.098	0.032***	2.206	2.155	−0.051***
10	0.947	0.969	0.022***	2.087	2.094	0.007**	2.204	2.162	−0.042***
11	0.947	0.966	0.019***	2.072	2.091	0.020***	2.187	2.166	−0.021
12	0.950	0.968	0.017***	2.081	2.095	0.014***	2.191	2.166	−0.025
13	0.950	0.966	0.016***	2.076	2.094	0.018***	2.185	2.168	−0.017
14	0.952	0.961	0.009**	2.079	2.083	0.004	2.184	2.168	−0.016
15	0.959	0.961	0.002	2.089	2.081	−0.007***	2.179	2.166	−0.012
16	0.961	0.962	0.001	2.088	2.090	0.002	2.172	2.172	−0.000
17	0.959	0.964	0.005	2.079	2.095	0.016***	2.167	2.173	0.006
18	0.958	0.965	0.007*	2.076	2.092	0.017***	2.166	2.167	0.001
19	0.960	0.963	0.004	2.080	2.090	0.010***	2.167	2.169	0.003
20	0.961	0.960	−0.001	2.086	2.087	0.001	2.170	2.173	0.003
21	0.966	0.962	−0.004	2.093	2.088	−0.005*	2.167	2.171	0.004
22	0.961	0.967	0.006*	2.076	2.092	0.016***	2.161	2.163	0.002
23	0.964	0.966	0.003	2.095	2.092	−0.002	2.174	2.165	−0.009
24	0.962	0.964	0.001	2.095	2.092	−0.003	2.177	2.171	−0.006
25	0.962	0.962	0.000	2.089	2.088	−0.001	2.171	2.170	−0.001
26	0.965	0.962	−0.004	2.091	2.083	−0.007***	2.166	2.166	0.000
27	0.961	0.964	0.003	2.080	2.089	0.008***	2.165	2.167	0.002
28	0.964	0.965	0.001	2.093	2.088	−0.006**	2.171	2.163	−0.008
29	0.966	0.963	−0.003	2.097	2.093	−0.004	2.170	2.173	0.003
30	0.967	0.963	−0.004	2.098	2.086	−0.011***	2.169	2.166	−0.003

This table reports the cross-sectional mean $\hat{\mu}_d$, standard deviation $\hat{\sigma}_d$ (both multiplied by 1000), and coefficient of variation $VCV_{CS,d}$ of all firms' daily trading volume shares on day d before and after $N = 596,881$ 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.

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 Fig. 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 reaffirms findings by Beaver et al. (2018) and Pawlewicz (2018), who document 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 $N = 379,060$ filing events between 2005 and 2020. 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 Fig. 5 shows, similar to Fig. 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 requirement for companies to file form 8-K at most four days after the event that triggered the filing. As Ben-Rephael et al. (2022) 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 Figs. 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.

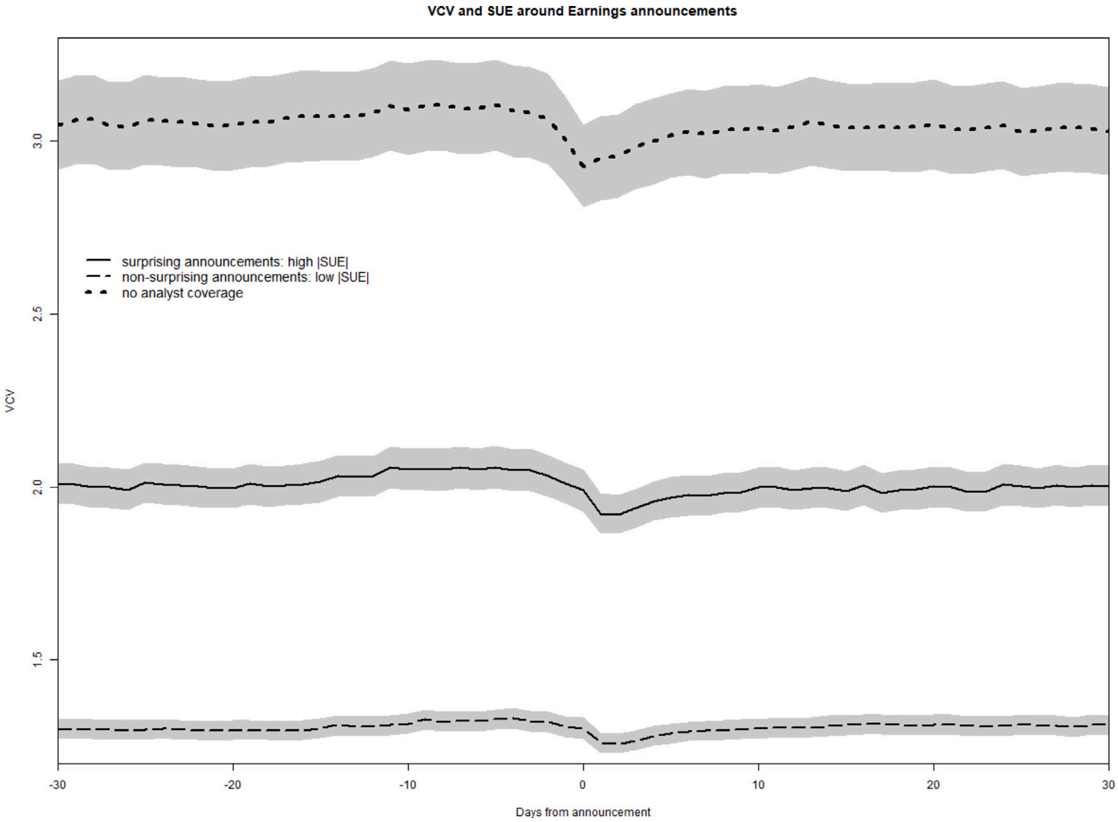

Fig. 4. This figure shows the evolution of the daily cross-sectional VCV_{XS} over a 61 day event-window (day −30:30) around quarterly earnings announcements. The sample is divided into surprising announcements (solid line, $N = 88,243$), non-surprising announcements (dashed line, $N = 88,594$), and earnings announcement for which no analyst forecasts are reported (dotted line, $N = 259,216$). 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 Fig. 3 for details.

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 of informed trade. We therefore propose the sample coefficient of variation of trading volume, 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. This finding confirms that, unlike general illiquidity proxies such as bid–ask spreads and Amihud Illiquidity, VCV specifically increases in informed rather than uninformed order flow. 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. VCV is therefore applicable to any security for which trading volume is observable, including stocks, bonds, asset-backed securities, credit-default swaps, options and other derivatives. For example, Ghosh et al. (2020) apply VCV to study information asymmetry in the

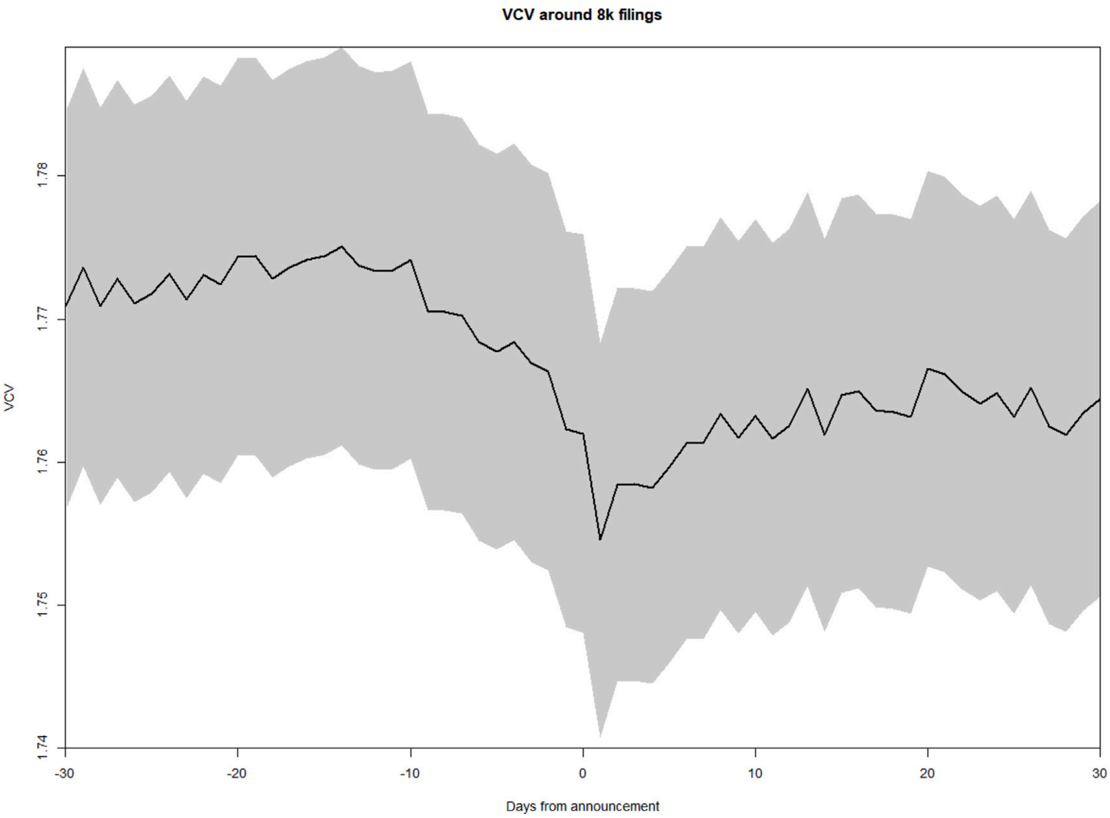

Fig. 5. This figure shows the evolution of the daily cross-sectional VCV_{XS} around SEC form 8-K filings. The full sample includes all daily trading volumes over a 61 day event-window (day −30:30) around $N = 379,060$ filings (sources: CRSP and SEC EDGAR). The reported VCV at d days after the filing 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).

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 firms or over time.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jcorpfin.2023.102464.

References

Admati, A., Pfleiderer, P., 1988. A theory of intraday patterns: volume and price variability. Rev. Financ. Stud. 1, 3–40.
Akbas, F., 2016. The calm before the storm. J. Finance 71 (1), 225–266.
Albrecher, H., Ladoucette, S., Teugels, J., 2010. Asymptotics of the sample coefficient of variation and the sample dispersion. J. Statist. Plann. Inference 140 (2), 358–368.
Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. J. Financial Mark. 5, 31–56.

Amihud, Y., Mendelson, H., 1986. Asset pricing and the bid-ask spread. Journal of financial Economics 17 (2), 223–249.
Antia, M., Pantzalis, C., Park, J.C., 2010. CEO decision horizon and firm performance: An empirical investigation. J. Corp. Finance 16 (3), 288–301.
Back, K., Crotty, K., Li, T., 2018. Identifying information asymmetry in securities markets. Rev. Financ. Stud. 31, 2277–2325.
Bai, J., Philippon, T., Savov, A., 2016. Have financial markets become more informative? J. Financ. Econ. 122 (3), 625–654.
Ball, R., Brown, P., 1968. An empirical evaluation of accounting income numbers. J. Account. Res. 15, 9–178.
Beaver, W.H., 1968. The information content of annual earnings announcements. J. Account. Res. 6, 7–92.
Beaver, W.H., McNichols, M.F., Wang, Z.Z., 2018. The information content of earnings announcements: new insights from intertemporal and cross-sectional behavior. Rev. Account. Stud. 23 (1), 95–135.
Ben-Nasr, H., Alshwer, A.A., 2016. Does stock price informativeness affect labor investment efficiency? J. Corp. Finance 38, 249–271.
Ben-Rephael, A., Da, Z., Easton, P.D., Israelsen, R.D., 2022. Who pays attention to SEC form 8-K? Account. Rev. 97 (5), 59–88.
Benos, E., Jochec, M., 2007. Testing the PIN Variable. Working paper.
Black, F., 1986. Noise. J. Finance 41, 528–543.
Boehmer, E., Gramming, J., Theissen, E., 2007. Estimating the probability of informed trading – does trade misclassification matter? J. Financial Mark. 10, 26–47.
Bongaerts, D., Rösch, D., Van Dijk, M.A., 2016. Cross-sectional Identification of Informed Trading. Working paper.
Boone, A., White, J.T., 2015. The effect of institutional ownership on firm transparency and information production. J. Financ. Econ. 117, 508–533.
Brockman, P., Yan, X.S., 2009. Block ownership and firm-specific information. J. Bank. Financ. 33, 308–316.
Brogaard, J., Carrion, A., Moyaert, T., Riordan, R., Shkilko, A., Sokolov, K., 2018. High frequency trading and extreme price movements. J. Financ. Econ. 128 (2), 253–265.
Brown, S., Hillegeist, S.A., 2007. How disclosure quality affects the level of information asymmetry. Rev. Account. Stud. 12 (2–3), 443–477.
Brown, S., Hillegeist, S.A., Lo, K., 2004. Conference calls and information asymmetry. J. Account. Econ. 37, 343–366.
Brown, S., Hillegeist, S.A., Lo, K., 2009. The effect of earnings surprises on information asymmetry. J. Account. Econ. 47 (3), 208–225.
Bushee, B.J., 2001. Do institutional investors prefer near-term earnings over long-run value? Contemp. Account. Res. 18 (2), 207–246.
Bushee, B.J., Leuz, C.an., 2005. Economic consequences of SEC disclosure regulation: Evidence from the OTC bulletin board. J. Account. Econ. 39 (2), 233–264.
Bushee, B.J., Noe, C.F., 2000. Corporate disclosure practices, institutional investors, and stock return volatility. J. Account. Res. 171–202.
Bushman, R.M., Williams, C.D., Wittenberg-Moerman, R., 2017. The informational role of the media in private lending. J. Account. Res. 55 (1), 115–152.
Cao, Z., Wu, W., 2022. Ownership breadth: Investor recognition or short-sale constraints? Finance Res. Lett. 47, 102847.
Chae, J., 2005. Trading volume, information asymmetry, and timing information. J. Finance 60, 413–442.
Chen, Z., Dhaliwal, D.S., Xie, H., 2010. Regulation fair disclosure and the cost of equity capital. Rev. Account. Stud. 15 (1), 106–144.
Chen, J., Hong, H., Stein, J.C., 2002. Breadth of ownership and stock returns. J. Financ. Econ. 66, 171–205.
Chen, T., Lin, C., 2017. Does information asymmetry affect corporate tax aggressiveness? J. Financ. Quant. Anal. 52 (5), 2053–2081.
Cheng, H., Huang, D., Luo, Y., 2020. Corporate disclosure quality and institutional investors' holdings during market downturns. J. Corp. Finance 60, 101523.
Chordia, T., Hu, J., Subrahmanyam, A., Tong, Q., 2019. Order flow volatility and equity costs of capital. Manage. Sci. 65, 1520–1551.
Chung, K.H., Zhang, H., 2014. A simple approximation of intraday spreads using daily data. J. Financial Mark. 17, 94–120.
Clark, Charles E., 1961. The greatest of a finite set of random variables. Oper. Res. 9 (2), 145–162.
Collver, C.D., 2007. Is there less informed trading after regulation fair disclosure? J. Corp. Finance 13 (2–3), 270–281.
Dang, C., Foerster, S., Li, Z.F., Tang, Z., 2021. Analyst talent, information, and insider trading. J. Corp. Finance 67, 101803.
De Cesari, A., Huang-Meier, W., 2015. Dividend changes and stock price informativeness. J. Corp. Finance 35, 1–17.
Derrien, F., Kecskes, A., 2013. The real effects of financial shocks: evidence from exogenous changes in analyst coverage. J. Finance 68 (4), 1407–1440.
Duarte, J., Hu, E., Young, L., 2020. A comparison of some structural models of private information arrival? J. Financ. Econ. 135 (3), 795–815.
Duarte, J., Young, L., 2009. Why is PIN priced? J. Financ. Econ. 91, 119–138.
Dumitrescu, A., Zakriya, M., 2022. Governance, information flow, and stock returns. J. Corp. Finance 72, 102168.
Easley, D., Engle, R.F., O'Hara, M., Wu, L., 2008. Time-varying arrival rates of informed and uninformed trades. J. Financ. Econom. 6 (2), 171–207.
Easley, D., Hvidkjaer, S., O'Hara, M., 2002. Is information risk a determinant of asset returns? J. Finance 57 (5), 2185–2221.
Easley, D., Hvidkjaer, S., O'Hara, M., 2010. Factoring information into returns. J. Financial Quant. Anal. 45, 293–309.
Easley, D., Kiefer, N.M., O'Hara, M., 1997a. One day in the life of a very common stock. Rev. Financ. Stud. 10, 805–835.
Easley, D., Kiefer, N.M., O'Hara, M., 1997b. The information content of the trading process. J. Empir. Financ. 4, 159–186.
Easley, D., Kiefer, M., Paperman, J.B., 1996. Liquidity, information, and infrequently traded stocks. J. Finance 51, 1405–1436.
Easley, D., O'Hara, M., Paperman, J.B., 1998. Financial analysts and information based trade. J. Financial Mark. 1, 175–201.
Easley, D., Lopez de Prado, M., O'Hara, M., 2012. Flow toxicity and liquidity in a high frequency world. Rev. Financ. Stud. 25, 1457–1493.
Feng, X., Johansson, A.C., 2019. Top executives on social media and information in the capital market: Evidence from China. J. Corp. Finance 58, 824–857.
Fich, E.M., Harford, J., Tran, A.L., 2015. Motivated monitors: The importance of institutional investors? portfolio weights. J. Financ. Econ. 118, 21–48.
Fong, K.Y., Holden, C.W., Trzcinka, C.A., 2017. What are the best liquidity proxies for global research? Rev. Finance 21 (4), 1355–1401.
Foster, F.D., Viswanathan, S., 1990. A theory of the interday variations in volume, variance, and trading costs in securities markets. Rev. Financ. Stud. 3 (4), 593–624.
Foster, F.D., Viswanathan, S., 1994. Strategic trading with asymmetrically informed traders and long-lived information. J. Financ. Quant. Anal. 29, 499–518.
Foster, F.D., Viswanathan, S., 1996. Strategic trading when agents forecast the forecasts of others. J. Finance 51, 1437–1478.
Gao, X., Ritter, J.R., 2010. The marketing of seasoned equity offerings. Journal of Financial Economics 97 (1), 33–52.
Gassen, J., Skaife, H.A., Veenman, D., 2020. Illiquidity and the measurement of stock price synchronicity. Contemp. Account. Res. 37 (1), 419–456.
George, T.J., Kaul, G., Nimalendran, M., 1994. Trading volume and transaction costs in specialist markets. J. Finance 49 (4), 1489–1505.
Ghosh, C., Liang, M., Petrova, M.T., 2020. The effect of fair value method adoption: Evidence from real estate firms in the EU. J. Real Estate Finance Econ. 60 (1), 205–237.
Goyenko, R.Y., Holden, C.W., Trzcinka, C.A., 2009. Do liquidity measures measure liquidity? J. Financial Econ. 92 (2), 153–181.
Grullon, G., Kanatas, G., Weston, J.P., 2004. Advertising, breadth of ownership, and liquidity. Rev. Financ. Stud. 17 (2), 439–461.
Grundy, B.D., McNichols, M., 1989. Trade and the revelation of information through prices and direct disclosure. Rev. Financial Stud. 2 (4), 495–526.
Hameed, A., Hong, D., Warachka, M., 2008. Momentum and Informed Trading. Working paper.
Healy, P.M., Hutton, A.P., Palepu, K.G., 1999. Stock performance and intermediation changes surrounding sustained increases in disclosure. Contemp. Account. Res. 16 (3), 485–520.
Heflin, F.L., Shaw, K.W., Wild, J.J., 2005. Disclosure policy and market liquidity: Impact of depth quotes and order sizes. Contemp. Account. Res. 22 (4), 829–865.
Hillegeist, S.A., Weng, L., 2021. Quasi-indexer ownership and insider trading: Evidence from russell index reconstitutions. Contemp. Account. Res. 38, 2192–2223.
Holden, C.W., Subrahmanyam, A., 1992. Long-lived private information and imperfect competition. J. Finance 47, 247–270.
Huang, R., Stoll, H.R., 1997. The components of the bid–ask spread: A general approach. Rev. Financ. Stud. 19, 995–1033.
Hwang, L.S., Lee, W.J., Lim, S.Y., Park, K.H., 2013. Does information risk affect the implied cost of equity capital? An analysis of PIN and adjusted PIN. J. Account. Econ. 55 (2–3), 148–167.

Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. J. Finance 45 (3), 881–898.
Johnson, T.L., So, E.C., 2018. A simple multimarket measure of information asymmetry. Manage. Sci. 64 (3), 1055–1080.
Jurkatis, S., 2022. Inferring trade directions in fast markets. J. Financial Mark. 58, 100635.
Kelly, B., Ljungqvist, A., 2012. Testing asymmetric-information asset pricing models. Rev. Financ. Stud. 25 (5), 1366–1413.
Kumar, A., Lee, C.M., 2006. Retail investor sentiment and return comovements. J. Finance 61 (5), 2451–2486.
Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53 (6), 1315–1335.
Le, H., Gregoriou, A., 2020. How do you capture liquidity? A review of the literature on low-frequency stock liquidity. J. Econ. Surv. 34 (5), 1170–1186.
Lee, I., Lemmon, M., Li, Y., Sequeira, J.M., 2014. Do voluntary corporate restrictions on insider trading eliminate informed insider trading? J. Corp. Finance 29, 158–178.
Lee, C., Ready, M., 1991. Inferring trade direction from intradaily data. J. Finance 46, 733–746.
Lerman, A., Livnat, J., 2010. The new form 8-K disclosures. Rev. Account. Stud. 15 (4), 752–778.
Li, K.K., You, H., 2015. What is the value of sell-side analysts? Evidence from coverage initiations and terminations. J. Account. Econ. 60 (2–3), 141–160.
Livnat, J., Mendenhall, R.R., 2006. Comparing the post-earnings announcement drift for surprises calculated from analyst and time series forecasts. J. Account. Res. 44 (1), 177–205.
Llorente, G., Michaely, R., Saar, G., Wang, J., 2002. Dynamic volume-return relation of individual stocks. Rev. Financial Stud. 15 (4), 1005–1047.
Luong, T.S., Qiu, B., 2021. Does it pay to be socially connected with wall street brokerages? Evidence from cost of equity. J. Corp. Finance 68, 101939.
Milgrom, P., Stokey, N., 1982. Information, trade and common knowledge. J. Econom. Theory 26, 17–27.
Mohanram, P., Rajgopal, S., 2009. Is PIN priced risk? J. Account. Econ. 47 (3), 226–243.
Odders-White, E.R., Ready, M.J., 2008. The probability and magnitude of information events. J. Financ. Econ. 87 (1), 227–248.
Pawlewicz, R., 2018. The effect of regulation on the timeliness and informational role of earnings announcements. Contemp. Account. Res. 35 (4), 1675–1701.
Petacchi, R., 2015. Information asymmetry and capital structure: Evidence from regulation FD. J. Account. Econ. 59 (2–3), 143–162.
Roll, R., 1984. A simple implicit measure of the effective bid–ask spread in an efficient market. J. Finance 39 (4), 1127–1139.
Sila, V., Gonzalez, A., Hagendorff, J., 2017. Independent director reputation incentives and stock price informativeness. J. Corp. Finance 47, 219–235.
Subrahmanyam, A.Chordia.T., Anshuman, R.V., 2001. Trading activity and expected stock returns. J. Financ. Econ. 59, 3–32.
To, T.Y., Navone, M., Wu, E., 2018. Analyst coverage and the quality of corporate investment decisions. J. Corp. Finance 51, 164–181.
Vega, C., 2006. Stock price reaction to public and private information. J. Financ. Econ. 82, 103–133.
Wang, J., 1994. A model of competitive stock trading volume. J. Polit. Econ. 102, 127–168.
Wang, Y., 2021. How are institutions informed? Proactive trading, information flows, and stock selection strategies. Contemp. Account. Res. 38, 1849–1887.
Welker, M., 1995. Disclosure policy, information asymmetry, and liquidity in equity markets. Contemp. Account. Res. 11 (2), 801–827.
Weller, B.M., 2018. Does algorithmic trading reduce information acquisition? Rev. Financ. Stud. 31 (6), 2184–2226.
Yang, Y.C., Zhang, B., Zhang, C., 2020. Is information risk priced? Evidence from abnormal idiosyncratic volatility. J. Financ. Econ. 135 (2), 528–554.