Asymmetric Information and the Distribution of Trading Volume *

Matthijs Lof† Jos van Bommel‡

January 27, 2017

Abstract

We propose the coefficient of variation (ratio of standard deviation to mean) of trading volume, VCV, as a new and easily computable measure of information asymmetry in security markets. We derive from a simple microstructure model that VCV is an increasing function of the proportion of informed trade. Simulations confirm this result under more general assumptions. Empirically, we find that VCV is highly correlated to extant measures of asymmetric information in the cross-section of US stocks. Moreover, VCV sharply decreases after earnings announcements resolve information asymmetries.

Keywords: VCV, Trading volume, Informed trading.

JEL classification: D82, G12, G14

*We thank Jonathan Cohn, David Easley, Eliezer Fich, Carlos Forner Rodríguez, Pete Kyle, Deniz Okat, Conall O'Sullivan, Jillian Popadak, Valeri Sokolovski, Enrique Sentana, Christian Westheide and participants at the American Economic Association (Chicago), German Finance Association, Aalto University, University of Amsterdam, and University of Maastricht for constructive comments.

†Aalto University School of Business. E-mail: matthijs.lof@aalto.fi

‡Luxembourg School of Finance, University of Luxembourg. E-mail: jos.vanbommel@uni.lu

1 Introduction

In this paper, we show that the proportion of informed trade affects the distribution of trading volume, within the context of the seminal market micro structure model of Kyle (1985). We consider a market where liquidity seekers submit orders to competitive liquidity providers (market makers). These market makers match buy and sell orders and absorb the order imbalance, conditional on which they set the clearing price. Liquidity seekers are either be informed (insiders) or uninformed (noise traders). Uninformed liquidity seekers place uncorrelated (*i.i.d.*) orders, while informed orders are perfectly correlated. Uninformed orders are therefore mostly matched to each other, while informed orders generate order imbalances that need to be accommodated by market makers. This observation leads to simple expressions for the first two moments of the trading volume as functions of the proportion of informed trade. Specifically, we show that the coefficient of variation (the ratio of the standard deviation to the mean) of trading volume increases monotonically in the proportion of informed trade. We propose the volume coefficient of variation (VCV) as a novel measure of the proportion of information trade. VCV is very easy to compute and only requires observed trading volume.

The intuition of our measure is that the distribution of trading volume depends on the correlation of individual orders. If all liquidity seekers are uninformed and place uncorrelated orders, these orders are mostly netted out against each other, and net order flow is relatively low compared to the observed trading volume. Trading volume follows in this case a Normal-like, thin-tailed, slightly skewed distribution. In the presence of informed (and correlated) liquidity seeking demand, the net orderflow that is matched by market makers takes up a higher share of the total trading volume. These correlated order flows generate volume outcomes over a more dispersed distribution. The coefficient of variation adequately captures the transformation of the volume distribution as the number of informed traders increases. In addition to the analytical results, we conduct a simulation study and provide empirical evidence in support of our main result: VCV increases in the proportion of informed trade.[1]

Adverse selection and asymmetric information in security markets have been widely studied since Bagehot (1971) identified it as the key determinant of market illiquidity. Copeland and Galai (1983), Kyle (1985, 1989), Glosten and Milgrom (1985), Karpoff (1986), Easley and O'Hara (1992), Admati and Pfleider (1989), Foster and Viswanathan (1994), and many others, have increased our understanding of the strategic behavior of asymmetrically informed traders and their effect on security markets. There has been no shortage of subsequent papers that aim to measure information asymmetries in security markets. Stoll (1978), Huang and Stoll (1997), and Madhavan *et al.* (1997) provide methods to extract the adverse selection component from bid ask spreads, and other microstructure data.

Easley, Kiefer, O'Hara and Paperman (1996) develop a measure for the probability of informed trading, the well-known PIN measure. They use the model of Glosten and Milgrom (1985) to estimate the proportion of informed traders from the dynamics of the signed order process. The PIN meaure has been widely used to study information asymmetries in security markets.[2]

Both PIN and VCV are are expected to increase in the order imbalances generated by informed demand. The PIN measure is estimated from individual orders that are classified as buys or sells, such that the order imbalance is actually observable. VCV, on the other hand, is estimated using aggregate volume data, from which the order imbalance is implied. Computing VCV therefore does not require intraday transaction-level data. All empirical results in this paper are based on daily volumes obtained from CRSP. Johnson

[1]To the best of our knowledge, we are the first to relate the coefficient of variation of trading volume to asymmetric information. Chordia et al. (2002) use the coefficient of variation of trading volume when examining the relation between stock returns and the variability of trading volume, without relating this measure to asymmetric information.

[2]Easley *et al.* (1997a and 1997b) analyze the information content around trade lags and trade size. Applications of PIN include, among others, the impact of analyst coverage on informational content (Easley *et al.*, 1998), stock splits (Easley *et al.*, 2001), dealer vs. auction markets (Heidle and Huang, 2002), trader anonymity (Gramming *et al.*, 2001), information disclosure (Vega, 2006; Brown and Hillegeist, 2007), corporate investments and M&A (Chen *et al.*, 2007; Aktas *et al.*, 2007), and ownership structure (Brockman and Yan, 2009), and the January effect (Kang, 2010).

and So (2017) recently propose the multimarket information asymmetry (MIA) measure, which is based on the relative trading volume in option and equity markets, following the assumption that informed traders are more likely to trade in options than uninformed traders. Although MIA, like VCV, is a very simple measure to compute, it does require access to option trading volume in addition to equity trading volume.

The PIN measure has been subject to debate, among others by Duarte and Young (2009), who argue that the (unadjusted) PIN is not only measuring informed trade, but also general illiquidity.[3] Duarte and Young 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 compare VCV to the various measures estimated by Duarte and Young (2009) and find that VCV is strongly related to PIN, but even more so to Adjusted PIN, while the relation to PSOS is not robust, thereby confirming that VCV is a measure of informed trading, rather than general illiquidity.

Characteristics of institutional ownership are also indicative of asymmetric information. Boone and White (2015) find that institutional ownership is associated with improved disclosure of information, and therefore lower information asymmetry. In accordance with this result, we find using 13F filings that firms with a large number of institutional shareholders (i.e. high breadth of ownership) have a low VCV on average. We also look specifically 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 funds in the institution's portfolio

[3]Other papers in the debate on the validity of the PIN include Easley et al. (2010) and Akay et al. (2012). Other critiques focus on the trade classification (Boehmer *et al.* 2007) and the estimation robustness, particularly in high-turnover stocks (Lin and Ke, 2011; Yan and Zhang, 2012). In response to these latter critiques, and the advent of high frequency trading, Easley et al. (2012a) develop the volume synchronized PIN, or VPIN. This estimator captures not only information asymmetry but also order flow toxicity, the risk of unbalanced orderflows. The VPIN is computed over volume-time rather than clock time and used bulk-classification of trades to decrease the computational burden. More recently, Bongaerts *et al.* (2016) propose the XPIN, a measure that looks at the relative product of price impact and order imbalance. Bongaerts et al. (2016) develop a model in which informed traders take into account their price impact and when buying (selling) rebalance their portfolios with offsetting trades of which the aggregate price impact.

(Fich *et al.*, 2015), and *Dedicated* investors, defined as institutional investors that are predominantly making long-term investments by picking a selective set of stocks (Bushee and Noe, 2000; and Bushee, 2001). We find that the VCV is lower for firms with a relatively large number of monitoring or dedicated (i.e. informed) investors.

The crux of our analysis in Section 2 is a Kyle (1985) model with informed and uninformed liquidity seekers and price-setting market makers. Instead of focusing on the price, we analyze volume. We introduce a simple expression for the stochastic trading volume, and derive the first and second moments as a function of the number of market participants, their trading activity and the proportion of informed trade. We show that both the expected value and the standard deviation of volume increase linearly in the proportion of informed trade, but that the standard deviation does so at a higher rate, so that the coefficient of variation of the trading volume is a natural estimator of the proportion of informed trade. Our measure is powerful because for a large number of market participants, the VCV *only* depends on the proportion of informed trade. We demonstrate that the VCV increases continuously in the proportion of informed trade and find a parsimonious expression of the relationship if the number of liquidity seekers goes to infinity.

To further analyze the relation between the proportion of informed trading and the volume distribution, and to gain insight in the small sample properties of the VCV, we conduct a Monte Carlo analysis in Section 3. We find that the generated volume distributions change markedly as a function of the proportion of informed trade, in that with more informed (correlated) orders, the volume distribution becomes more dispersed. We confirm that the simulated VCV increases in the proportion of informed trade in a continuous fashion, and that it is virtually independent of the number of market participants. These findings continue to hold in small samples and after relaxing the assumptions of our model.

For our empirical analysis in Section 4, we compute annual firm-level observations of VCV from daily volumes of all NYSE and AMEX stocks from 1982 until 2014, obtained from CRSP. Our cross-sectional analysis shows that the VCV is significantly correlated with various PIN measures, in particular with adjusted PIN by Duarte and Young (2009), and with patterns in institutional ownership.

Section 5 documents patterns in VCV around earnings announcements. It has been widely recognized that over earnings announcement windows information asymmetries are resolved. We expect that the proportion of uninformed trading decreases closely to the earnings announcement as information asymmetries build up, and discourage uninformed traders to trade around such events (See Milgrom and Stokey, 1982; Black, 1986; Wang, 1986; and Chae, 2005). After the announcement, the playing field is levelled as the information prior held privately by the insiders is now publicly disclosed, making the market more attractive for the uninformed traders. Our empirical investigations bear this out. From a comprehensive sample of more than 40,000 (quarterly) earnings announcements of U.S. firms we find a significant increase in VCV closely before announcements and a subsequent drop to levels below those seen before earnings announcements. Moreover, this pattern in VCV is strongest when the earnings announcement is considered to contain surprising information, i.e.; when the resolution of information asymmetry between insiders and outsiders is most significant.

2 Theory

To develop our measure of informed trade, we postulate three kinds or traders in the market: (i) informed liquidity seekers, (ii) uninformed liquidity seekers, and (iii) competitive market makers. We assume that there are M individual liquidity seekers, of which a proportion η is informed. Both M and ηM are integers. The M individual traders submit orders to the market where buy orders are matched to sell orders and the residual unmatched orders (henceforth referred to as 'order imbalance' or 'net order flow') are taken up by the market makers who set the price. The model thus closely resembles that of Kyle

(1985), with the only difference that we consider the individual orders of the liquidity seekers, and analyze total trading volume in addition to the net order flow.

To be precise, we denote the individual demands of liquidity seekers by $\tilde{y}_i$, for which positive values indicate buy orders and negative values indicate sell orders. The order imbalance (net order flow) is the sum of orders, which is taken up by the market maker: $\sum_M \tilde{y}_i$. This imbalance is often not observable because it is difficult to empirically distinguish liquidity seekers from liquidity suppliers. This is particularly the case when only aggregate volume data is available rather than transaction-level trading data. Total volume is easier to measure and in many markets readily observable. Total trading volume can be written as:

$$\tilde{V} = \tfrac{1}{2}\left(\sum_M |\tilde{y}_i| + \left|\sum_M \tilde{y}_i\right|\right) \tag{1}$$

The term inside brackets is the "double-counted transaction volume", counting both buys and sells. This double-counted volume includes the trades amongst liquidity seekers, as well as the trades between the market makers and unmatched liquidity seekers.

As an example, consider five liquidity seekers whose demands are -1, 2, 2, -2, 1. The net order flow is two, which means that the market makers 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 market makers. The double-counted volume is thus ten, and the commonly recorded single-counted volume is half this number.

We let the demand of every liquidity seeker, whether informed or uninformed, to be Normally distributed with zero mean and standard deviation σ:[4]

$$\tilde{y}_i \sim N\left(0, \sigma^2\right) \tag{2}$$

[4]The assumption that individual uninformed and informed demands are of the same order of magnitude is innocuous. Having $\frac{\eta}{k}M$ informed traders with demand distributed as $N\left(0, k\sigma^2\right)$ is equivalent to our model. For this reason we refer to η as the proportion of informed *trading* rather than informed *traders* in the market.

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

$$\sum_M \tilde{y}_i \sim N\left(0, \sigma^2\left(\eta^2 M^2 + (1-\eta) M\right)\right), \tag{3}$$

where the variance of net order flow is a fuction of η, due to the different correlation of informed and uninformed demand. When most liquidity seekers are uninformed, their uncorrelated demands can be mostly matched between each other and net order flow is low. When most traders are informed, their correlated demands lead to large imbalances. As a result, the variance of net order flow 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[5] we find:

$$\begin{aligned} E\left(\tilde{V}\right) &= \tfrac{1}{2}\left(E\left(\textstyle\sum_M |\tilde{y}_i|\right) + E\left(\left|\textstyle\sum_M \tilde{y}_i\right|\right)\right) \\ &= \tfrac{1}{2}\left(\sigma\sqrt{\tfrac{2}{\pi}}M + \sigma\sqrt{\tfrac{2}{\pi}}\sqrt{\eta^2 M^2 + (1-\eta)M}\right) \\ &= \tfrac{\sigma M}{\sqrt{2\pi}}\left(1 + \sqrt{\eta^2 + (1-\eta)M^{-1}}\right). \end{aligned} \tag{4}$$

We see that, for large M, expected trading volume is proportional to $(1+\eta)\sigma M$, meaning that trading volume increases in η. The variance of the components of trading volume are given by:

$$Var\left(\sum_M |\tilde{y}_i|\right) = M\sigma^2\left(1 - \frac{\pi}{2}\right) \tag{5}$$

and

$$Var\left(\left|\sum_M \tilde{y}_i\right|\right) = \left(\eta^2 M^2 + (1-\eta)M\right)\sigma^2\left(1 - \frac{\pi}{2}\right) \tag{6}$$

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

For large M, it can be shown that the the variance of total trading volume is proportional to $M^2\sigma^2\eta^2$ (see appendix). The ratio of the standard deviation to the mean (coefficient of variation) of trading volume therefore increases in η.

<u>Proposition 1</u>

Consider a market where M liquidity seeking traders submit Normally distributed market orders with mean zero and standard deviation σ and where the net order flow is absorbed by liquidity suppliers (market makers). If a proportion η of the traders are informed:

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

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

$$\lim_{M\to\infty}\frac{\sigma_V}{\mu_V} = \sqrt{2\pi - 4}\frac{\eta}{\eta + 1} \tag{7}$$

<u>Proposition 2</u>

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

$$VCV \equiv \frac{\hat{\sigma}_V}{\hat{\mu}_V} \tag{8}$$

is a consistent estimator of $\frac{\sigma_V}{\mu_V}$. The Volume coefficient of Variation (VCV) is therefore a measure of informed trade as its expected value increases monotonically in η.

It is clear from the analysis above that Eq. (7) implies a direct estimator of the proportion of informed trade:

$$\hat{\eta} \equiv \frac{\hat{\sigma}_V}{\hat{\mu}_V\sqrt{2\pi - 4} - \hat{\sigma}_V.} \tag{9}$$

However, as our simulation results in the next section bear out, $\hat{\eta}$ is a consistent estimator when demand is Normally distributed and M is large, but behaves poorly in small samples or when we relax the assumptions of the model, primarily because the denominator can be close to zero or turn even negative. VCV, on the other hand, is found to be increasing in η under very general conditions, including non-Normality and heterogeneous samples. For this reason, we propose VCV, as opposed to $\hat{\eta}$, as our measure of informed trade.

3 Simulation

In this section we analyze the distribution of trading volume generated by our model, for different values of M (number of liquidity seekers) and η (fraction of informed liquidity seekers). To do this, we draw $1 + (1 - \eta)M$ random observations from the standard normal distribution $N(0, 1)$ to simulate the individual demands. The first observation is multiplied by ηM, and represents the aggregate informed demand. The remaining observations represent the individual uninformed demands. As in our theoretical analysis, the fraction of informed trade can be interpreted either as ηM informed traders that each place orders of similar magnitude as the uniformed traders, or as a single uninformed trader who places an order that corresponds in magnitude to ηM uninformed traders. We compute the observed trading value volume $\tilde{V}$ that follows from Eq. (1). For each (M, η) pair we generate a sample of T volume ($\tilde{V}$) observations.

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

In Table 1, we report the average Volume Coefficient of Variation (VCV) and $\widehat{\eta}$ from $R = 1,000,0000$ repetitions of simulating a sample of T trading sessions with M traders,

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

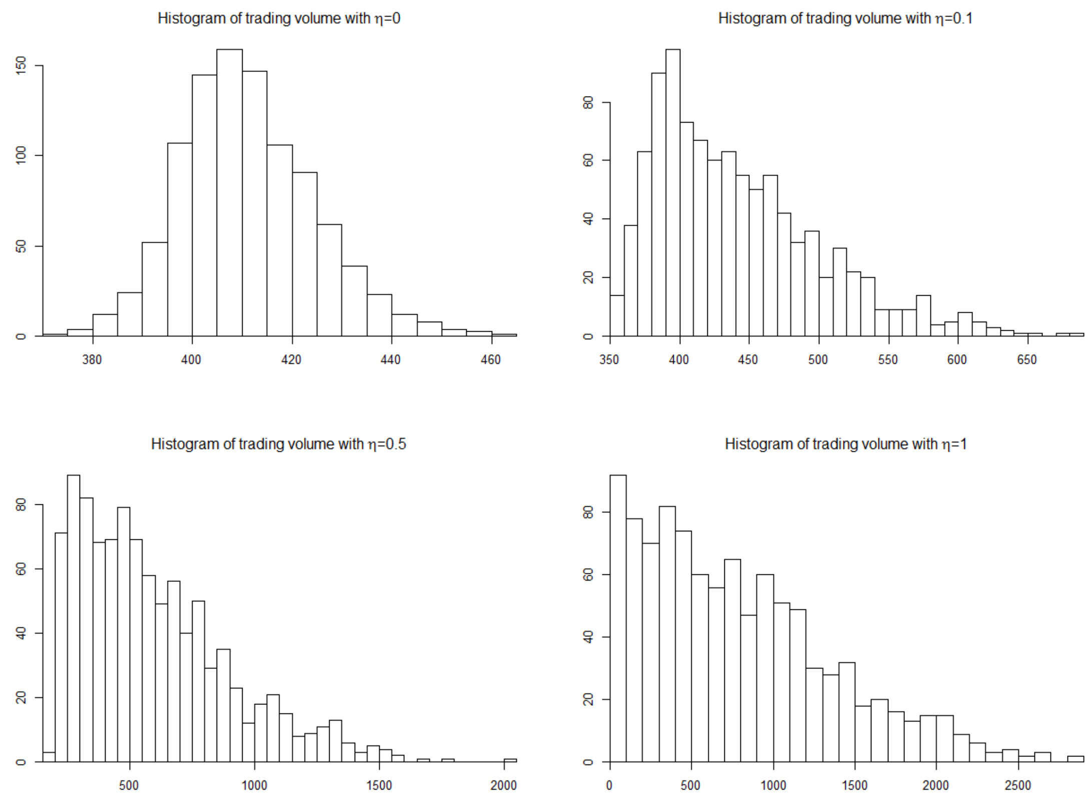

Figure 1: Histogram of T=1,000 volume realizations simulated from the model outlined in Section 2, for various values of the proportion of informed trading η. The number liquidity seekers (M) is 1,000 and the standard deviation of their demand is fixed at $\sigma = 1$.

for different values of η, T and M. In addition to averages of VCV and $\widehat{\eta}$ across the R repetitions, the table reports the simulated 90% confidence intervals. Panel A shows that with samples of $T = 100$ trading sessions, both VCV and $\widehat{\eta}$ increase monotonically in the true proportion of informed trade (η). This is even the case for markets with only a small number of liquidity seeking traders M. As Figure 2 shows, the VCV only deviates substantially from its theoretical value (Eq.(7)) when both M and η are low. Nevertheless, even for $M = 10$, VCV is strictly increasing in η. Also the estimator $\widehat{\eta}$ in our simulations traces the true value of η closely, in particular when either M or η are not too low.

The insensitivity to M, which is due to the relative weight of the informed volume, of which the shape is independent of M, is encouraging as it implies that there is little concern for confounding a high η with a low M. This insensitivity to M is also desirable from

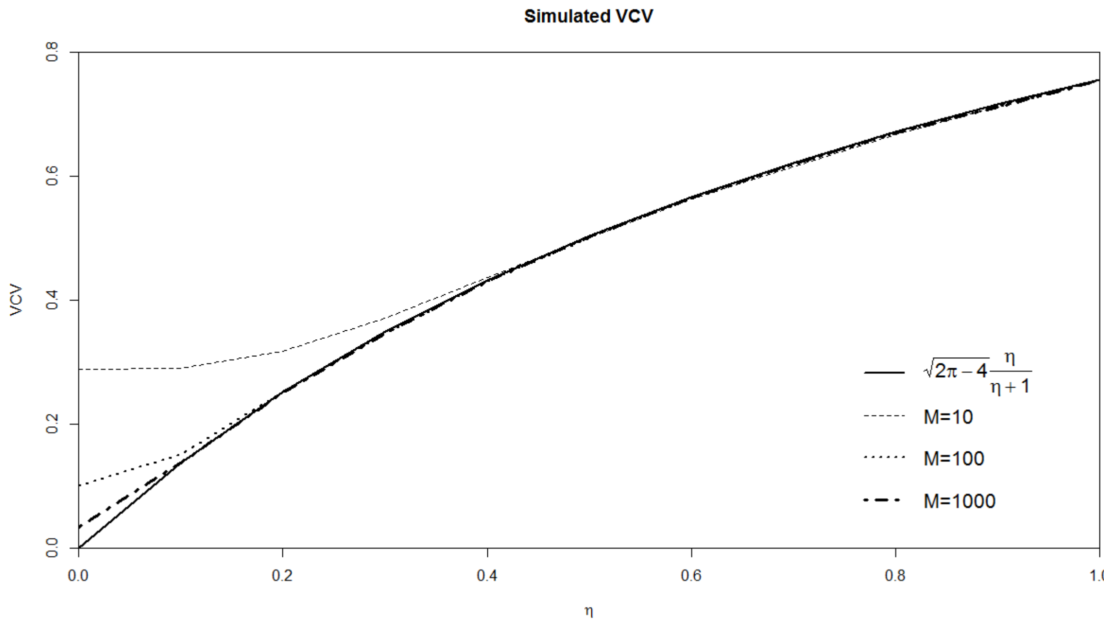

Figure 2: Average VCV obtained from $R = 1,000,000$ replications of $T = 100$ volume realizations simulated from the model outlined in Section 2, for various values of the proportion of informed trading η and number of liquidity seekers M.

an empirical perspective, because the number of order submitters in markets is typically unknown. Although some data-providers identify individual transactions, this number is different from the number of order-submitters because orders may be broken up.

Panel B of Table 1 shows the simulation results for smaller simulated samples of $T = 10$ trading sessions. We still find the average VCV and $\widehat{\eta}$ to increase monotonically in η, although the 90% confidence intervals indicate that VCV and in particular $\widehat{\eta}$ are less precisely estimated, which is not surprising given that these measure rely on means and standard deviations from samples of only $T = 10$ volume observations.

Next, we relax the assumptions of our model to investigate the robustness of VCV and $\widehat{\eta}$ as measures of asymmetric information. First, we repeat our simulation while relaxing the assumption of normally distributed demand and allow for leptokurtic and skewed demand distributions. In Table 2 we report the results where liquidity demand follows a uniform distribution (Panel A), a t-distribution (Panel B) and a skew-normal distribution (Panel C), for different values of η, while keeping $M = 1,000$ and $T = 100$ fixed.

Relaxing normality does not change the main result of our analysis: VCV and $\widehat{\eta}$ are still strictly increasing in η. However, the estimates of VCV have clearly more narrow confidence intervals than $\widehat{\eta}$. Moreover, with non-Normal demand, the estimates $\widehat{\eta}$ are no longer unbiased estimates of the true value of η.

We recognize that in practice the proportion of informed trade η is not necessarily constant across observations, and that we are typically interested in measuring the *average* proportion of informed trade, over either a time series or a cross section of observations. To gauge the precision of our measures in this context, we repeat the simulation analysis where we allow the proportion of informed trad η be random across observations. Panel A of Table 3 gives the results for the case where the number of uniformed liquidity seekers is fixed at 1,000, while the the number of informed liquidity seekers in each of the $T = 1,000$ trading sessions is randomly picked from a discrete Uniform distribution. We adjust the support of the uniform distribution to create variation in the average proportion of informed trade ($E[\eta]$). The results indicate again that both VCV and $\widehat{\eta}$ are increasing in the average proportion of informed trade, although $\widehat{\eta}$ diverges dramatically from the true average η and the precision of $\widehat{\eta}$ is significantly lower than that the of VCV.

In panel B, the number of uniformed liquidity seekers is again fixed at 1,000, while the amount of informed demand is binomially distributed. We choose the number of informed traders, but these traders participate in only one out of five trading sessions. That is, the number of informed traders in each trading session is zero with probability $\frac{4}{5}$. To create variation in the average proportion of informed trade, we adjust the potential number of informed traders. In this setting, $\widehat{\eta}$ clearly does not perform well as a measure of informed trading. The estimates of $\widehat{\eta}$ have wide confidence intervals, are not monotonically increasing in η, and are not bounded by 0 and 1. This occurs because the denominator in Eq. (9) can easily take on small positive or even negative numbers, which makes the estimator very imprecise. VCV, on the other hand, continues to be monotonically increasing in η while its confidence intervals remain fairly narrow.

Overall, the simulation results in this section demonstrate that VCV is very robust as a measure of asymmetric information. The basic results that the coefficient of variation of trading volume is monotonically increasing in the proportion of informed trade holds under general conditions and in small samples. In the remainder of this paper, we therefore focus on the VCV as our measure of informed trade, and investigate its properties using real empirical data.

Table 1: Simulation results

Panel A: T=100 observations												
M	η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
10	VCV	0.290	0.289	0.317	0.37	0.434	0.500	0.561	0.614	0.667	0.714	0.755
		[0.26,0.32]	[0.26,0.33]	[0.28,0.36]	[0.32,0.41]	[0.38,0.49]	[0.44,0.56]	[0.5,0.63]	[0.55,0.69]	[0.59,0.74]	[0.63,0.8]	[0.67,0.85]
10	$\widehat{\eta}$	0.238	0.237	0.266	0.325	0.404	0.497	0.593	0.689	0.795	0.905	1.009
		[0.2,0.27]	[0.2,0.28]	[0.23,0.31]	[0.27,0.38]	[0.34,0.48]	[0.41,0.59]	[0.49,0.71]	[0.57,0.83]	[0.64,0.97]	[0.72,1.13]	[0.8,1.28]
100	VCV	0.101	0.152	0.253	0.348	0.43	0.501	0.563	0.620	0.669	0.715	0.754
		[0.09,0.11]	[0.13,0.17]	[0.22,0.28]	[0.31,0.39]	[0.38,0.48]	[0.45,0.56]	[0.5,0.63]	[0.55,0.7]	[0.59,0.75]	[0.63,0.8]	[0.67,0.84]
100	$\widehat{\eta}$	0.072	0.112	0.201	0.300	0.398	0.498	0.596	0.699	0.800	0.907	1.008
		[0.06,0.08]	[0.1,0.13]	[0.17,0.23]	[0.26,0.35]	[0.34,0.46]	[0.42,0.59]	[0.5,0.71]	[0.57,0.86]	[0.64,0.98]	[0.71,1.14]	[0.79,1.27]
1000	VCV	0.033	0.137	0.251	0.347	0.43	0.502	0.566	0.618	0.669	0.717	0.756
		[0.03,0.04]	[0.12,0.15]	[0.22,0.28]	[0.31,0.39]	[0.38,0.48]	[0.45,0.56]	[0.5,0.63]	[0.55,0.69]	[0.59,0.75]	[0.64,0.81]	[0.67,0.85]
1000	$\widehat{\eta}$	0.022	0.100	0.199	0.298	0.399	0.499	0.601	0.697	0.800	0.910	1.012
		[0.02,0.03]	[0.09,0.11]	[0.17,0.23]	[0.26,0.35]	[0.34,0.46]	[0.42,0.59]	[0.5,0.72]	[0.56,0.85]	[0.65,0.98]	[0.73,1.15]	[0.8,1.3]

Panel B: T=10 observations												
M	η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
10	VCV	0.278	0.278	0.309	0.357	0.419	0.481	0.548	0.601	0.645	0.695	0.743
		[0.17,0.39]	[0.18,0.39]	[0.19,0.44]	[0.23,0.52]	[0.26,0.59]	[0.31,0.68]	[0.35,0.77]	[0.4,0.84]	[0.42,0.89]	[0.46,0.96]	[0.46,1.08]
10	$\widehat{\eta}$	0.229	0.229	0.262	0.317	0.396	0.486	0.598	0.707	0.799	0.943	1.205
		[0.13,0.35]	[0.13,0.35]	[0.14,0.41]	[0.18,0.53]	[0.21,0.63]	[0.25,0.82]	[0.3,1.03]	[0.36,1.26]	[0.39,1.42]	[0.44,1.73]	[0.44,2.48]
100	VCV	0.099	0.146	0.242	0.335	0.409	0.48	0.546	0.603	0.646	0.708	0.728
		[0.06,0.14]	[0.09,0.21]	[0.15,0.35]	[0.21,0.47]	[0.26,0.58]	[0.31,0.67]	[0.35,0.77]	[0.41,0.85]	[0.43,0.88]	[0.46,0.98]	[0.47,1.03]
100	$\widehat{\eta}$	0.070	0.108	0.194	0.290	0.382	0.483	0.596	0.717	0.797	0.988	1.065
		[0.04,0.1]	[0.06,0.16]	[0.11,0.3]	[0.17,0.45]	[0.21,0.62]	[0.26,0.8]	[0.3,1.03]	[0.37,1.28]	[0.4,1.38]	[0.43,1.84]	[0.45,2.14]
1000	VCV	0.032	0.131	0.24	0.333	0.417	0.489	0.544	0.604	0.648	0.703	0.736
		[0.02,0.05]	[0.08,0.19]	[0.15,0.34]	[0.22,0.48]	[0.27,0.58]	[0.32,0.69]	[0.36,0.76]	[0.4,0.86]	[0.41,0.9]	[0.45,0.99]	[0.49,1.04]
1000	$\widehat{\eta}$	0.022	0.096	0.192	0.289	0.392	0.497	0.594	0.714	0.819	1.034	1.087
		[0.01,0.03]	[0.05,0.14]	[0.11,0.29]	[0.17,0.46]	[0.22,0.62]	[0.27,0.84]	[0.31,1]	[0.37,1.33]	[0.37,1.49]	[0.43,1.88]	[0.47,2.24]

Notes: Average VCV and $\widehat{\eta}$ obtained from $R = 1,000,000$ replications of T volume realizations simulated from the model outlined in Section 2, for various values of the proportion of informed trading η and number of liquidity seekers M. In Panel A, the number of volume observations in each replication is $T = 100$. In panel B, $T = 10$. The table reports the average VCV and $\widehat{\eta}$ from R replications, and their 5^{th} and 95^{th} percentile (90% confidence interval) in square brackets.

Table 2: Simulation results: Non-Gaussian demand distribution

Panel A: Uniform distribution

η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
VCV	0.028	0.106	0.193	0.266	0.330	0.386	0.433	0.478	0.516	0.548	0.578
	[0.02,0.03]	[0.1,0.11]	[0.18,0.21]	[0.24,0.29]	[0.3,0.36]	[0.35,0.43]	[0.39,0.48]	[0.43,0.53]	[0.46,0.57]	[0.48,0.61]	[0.51,0.66]
$\widehat{\eta}$	0.019	0.076	0.146	0.214	0.279	0.344	0.402	0.464	0.520	0.571	0.623
	[0.02,0.02]	[0.07,0.08]	[0.13,0.16]	[0.19,0.24]	[0.24,0.31]	[0.3,0.39]	[0.34,0.46]	[0.39,0.54]	[0.44,0.6]	[0.47,0.68]	[0.5,0.77]

Panel B: T distribution

η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
VCV	0.041	0.176	0.319	0.444	0.562	0.635	0.731	0.790	0.862	0.919	0.965
	[0.04,0.05]	[0.14,0.24]	[0.25,0.43]	[0.35,0.59]	[0.44,0.76]	[0.51,0.82]	[0.58,0.95]	[0.64,1.01]	[0.7,1.14]	[0.73,1.2]	[0.78,1.21]
$\widehat{\eta}$	0.028	0.133	0.272	0.426	0.649	0.743	0.991	1.233	1.525	1.553	1.010
	[0.02,0.03]	[0.1,0.19]	[0.2,0.4]	[0.3,0.63]	[0.41,1]	[0.51,1.17]	[0.62,1.64]	[0.73,1.99]	[0.84,2.84]	[0.93,3.26]	[1.04,3.67]

Panel A: Skew-normal distribution

η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
VCV	0.024	0.079	0.153	0.229	0.303	0.379	0.455	0.532	0.608	0.680	0.755
	[0.02,0.03]	[0.07,0.09]	[0.13,0.17]	[0.2,0.26]	[0.27,0.34]	[0.34,0.42]	[0.4,0.51]	[0.47,0.59]	[0.54,0.68]	[0.61,0.77]	[0.66,0.85]
$\widehat{\eta}$	0.016	0.055	0.113	0.179	0.251	0.335	0.432	0.545	0.678	0.825	1.01
	[0.01,0.02]	[0.05,0.06]	[0.1,0.13]	[0.15,0.21]	[0.22,0.29]	[0.28,0.39]	[0.36,0.5]	[0.46,0.64]	[0.56,0.82]	[0.67,1.03]	[0.78,1.29]

Notes: Average VCV and $\widehat{\eta}$ obtained from $R = 1,000,000$ replications of $T = 100$ volume observations, simulated from a model with $M = 1000$ liquidity seekers. Different from Table 1, liquidity demand is not normally distributed. In Panel A, demand is uniformly distributed over the support $[-1, 1]$. In Panel B, demand is t-distributed with 4 degrees of freedom (t_4). In Panel C, demand is Skew-Normal distributed with shape parameter 10, indicating positive skew ($SN(0, 1, 10)$). The table reports the average VCV and $\widehat{\eta}$ from R replications, and their 5^{th} and 95^{th} percentile (90% confidence interval) in square brackets, for different values of η.

Table 3: Simulation results: Random proportion of informed traders

				Panel A: Uniform distribution					
Uninformed	1000	1000	1000	1000	1000	1000	1000	1000	1000
Informed	U[0,0]	U[0,200]	U[0,500]	U[0,800]	U[0,1300]	U[0,2000]	U[0,3000]	U[0,5000]	U[0,8000]
$E[\eta]$	0	0.09	0.2	0.29	0.39	0.5	0.6	0.71	0.8
VCV	0.033	0.171	0.346	0.459	0.588	0.694	0.779	0.868	0.925
	[0.03,0.04]	[0.14,0.2]	[0.29,0.4]	[0.39,0.53]	[0.51,0.67]	[0.6,0.79]	[0.68,0.88]	[0.76,0.99]	[0.81,1.05]
$\widehat{\eta}$	0.022	0.128	0.298	0.439	0.642	0.859	1.078	1.381	1.624
	[0.02,0.03]	[0.1,0.15]	[0.24,0.36]	[0.35,0.53]	[0.5,0.81]	[0.66,1.1]	[0.82,1.41]	[1.02,1.91]	[1.14,2.27]
				Panel B: Binomial distribution					
Uninformed	1000	1000	1000	1000	1000	1000	1000	1000	1000
Informed	B($\frac{1}{5}$,0)	B($\frac{1}{5}$,500)	B($\frac{1}{5}$,1250)	B($\frac{1}{5}$,2000)	B($\frac{1}{5}$,3250)	B($\frac{1}{5}$,5000)	B($\frac{1}{5}$,7500)	B($\frac{1}{5}$,12500)	B($\frac{1}{5}$,20000)
$E[\eta]$	0	0.09	0.2	0.29	0.39	0.5	0.6	0.71	0.8
VCV	0.033	0.412	0.828	1.118	1.437	1.703	1.913	2.162	2.330
	[0.03,0.04]	[0.31,0.51]	[0.66,0.99]	[0.92,1.3]	[1.23,1.65]	[1.48,1.98]	[1.65,2.23]	[1.84,2.56]	[1.95,2.82]
$\widehat{\eta}$	0.023	0.380	1.264	3.263	-3.477	-8.137	-6.562	-3.637	-3.068
	[0.02,0.03]	[0.26,0.51]	[0.77,1.9]	[1.55,6.12]	[-55.32,69.83]	[-48.7,34.94]	[-11.63,-3.1]	[-5.58,-2.44]	[-4.46,-2.16]

Notes: Average VCV and $\widehat{\eta}$ obtained from $R = 1,000,000$ replications of $T = 100$ volume observations. Different from Table 1, the number of uninformed liquidity seekers is kept constant at $1,000$, while the number of informed liquidity seekers is varying randomly across observations. In Panel A, the number of informed liquidity seekers follows a discrete uniform distribution over the support $[0, X]$. In Panel B, informed demand binomially distributed such that the number of informed traders in each trading session is with probability $\frac{4}{5}$ equal to zero and with probability $\frac{1}{5}$ equal to X. The table reports the average VCV and $\widehat{\eta}$ from R replications, and their 5^{th} and 95^{th} percentile (90% confidence interval) in square brackets, for different values of X, which determines the average proportion of informed trade $E[\eta]$.

4 The Cross-Section of VCV

After having established from theoretical and numerical analysis a positive monotonic relation between VCV and the proportion of informed traders, we now turn to empirical results. In this section, we describe cross-sectional variation in VCV for US stocks, while in the next section we study the time-series behavior of VCV. We compute annual Volume Coefficients of Variation (VCV) for US stocks and compare these figures to alternative measures of informed trading and illiquidity. We obtain daily trading volumes (number of shares traded multiplied by the closing price) from the CRSP daily stock file for all common stock listed on NYSE and AMEX over the period 1982-2014. We exclude firms listed on NASDAQ from our sample, in order to avoid biased caused by the differences in market structure. We disregard the most infrequently traded stocks by only considering stocks that had positive trading volume during at least 200 days in the previous calendar year.

Annual firm-level observations of VCV are computed by simply dividing the annual standard deviation of daily trading volumes by the annual mean of daily trading volumes. We compute VCV using three different measures of trading volume: Trading volume in USD (number of shares traded multipled by the closing price), Volume *Shares* (defined as daily volume of a stock as a percentage of total market volume on the same day) and daily *turnover* (defined as $\frac{shares\ traded_{i,t}}{shares\ outstanding_{i,t}}$).

Table 4 shows summary statistics for these three measures of VCV, while Table 5 reports the correlations between the measures. The sample means and other statistics of the three VCV measures are very close to each other. Table 4 also shows that VCV varies considerably both in the cross section and over time. As Table 5 shows, the three different measures of VCV are highly correlated with each other. In the remainder of this paper, our measure of informed trading VCV is defined as the annual coefficient of variation of daily volume shares ($\text{VCV}_{\%}$). Similar results are obtained when using any of the other VCV measures.

Table 4: VCV Summary Statistics

	N	T	mean	sd	sd(CS)	sd(TS)	min	max	median	ρ
VCV$_{USD}$	5939	33	1.17	0.68	0.66	0.47	0.24	12.98	1.03	0.18
VCV$_{\%}$	5939	33	1.15	0.70	0.67	0.48	0.16	14.00	1.02	0.19
VCV$_{TO}$	5939	33	1.14	0.66	0.63	0.44	0.23	10.18	1.01	0.19

Notes: 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 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 N; the number of distinct stocks in the sample, T; the number of time-series observations (years), mean, standard deviation, sd (CS), the time-series average of annual cross-sectional standard deviations, sd (TS), the cross-sectional average of stock-specific time-series standard deviations, min, max, median and 1st order autocorrelation (ρ). Sample: 1982-2014. Source: CRSP.

Table 5: VCV and other firm characteristics

	VCV$_{USD}$	VCV$_{\%}$	VCV$_{TO}$	Size	Illiq	Turnover	Coverage
VCV$_{USD}$		0.98	0.97	-0.55	0.60	-0.20	-0.44
VCV$_{\%}$	0.98		0.95	-0.55	0.60	-0.20	-0.45
VCV$_{TO}$	0.97	0.96		-0.52	0.57	-0.22	-0.43
Size	-0.66	-0.66	-0.63		-0.95	0.26	0.75
Illiq	0.69	0.69	0.66	-0.96		-0.41	-0.74
Turnover	-0.22	-0.22	-0.23	0.29	-0.46		0.25
Coverage	-0.56	-0.57	-0.55	0.80	-0.83	0.38	

Notes: This table reports the correlation between annual firm-level observations of VCV$_{USD}$, VCV$_{\%}$ and VCV$_{TO}$ (See Table 4 for definitions) and other annual firm-level characteristics. The upper diagonal entries show the correlations for levels. The lower diagonal entries show the average within-year rank correlations. *Size* is the log of market capitalization at the last trading day of June. *ILLIQ* is the log of the annual average of the daily ratio $\frac{|R_{i,t}|}{Volume\ (USD)_{i,t}}$ (Amihud, 2002). *Turnover* is the annual average of daily trading volume as a percentage of market capitalization. *Coverage* refers to analyst coverage and is equal to *log(1+number of analysts)*. Sample: 1982-2014. Source: CRSP and IBES.

Table 5 also shows the correlation between VCV and other firm characteristics: Size, Turnover, and Amihud (2002) Illiquidity. Size and Illiquidity are computed from CRSP data. Size is defined as the log of market capitalization at the last trading day of June. Amihud (2002) Illiquidity is defined as the the log of the annual average of the daily ratio $\frac{|R_{i,t}|}{Volume\ (USD)_{i,t}}$. VCV is negatively correlated to Size and Turnover and positively correlated to Illiquidity. These results are consistent with our proposition that VCV measures informed trading, since information asymmetry is likely to be predominant in smaller stocks and asymmetric information reduces liquidity. Table 5 also shows the correlation between VCV and analyst coverage, defined as the logarithm of one plus the number of distinct analysts covering a stock in a given year (Source: I/B/E/S). Analyst coverage is likely to reduce information asymmetries, which is reaffirmed by the negative correlation with VCV.

In the following two subsections, we compare VCV to other firm-level indicators of asymmetric information: The Probability of Informed Trading (PIN) and the characteristics of institutional ownership.

4.1 VCV and PIN

Table 6 shows the correlations between VCV and various annual PIN measure for US stocks. We make use of the various annual PIN measures kindly made publicly available by the authors of previous studies. These measures include the PIN measures estimated by Easley et al. (2010 – PIN_{EHO}); Brown, Hillegeist and Lo (2004 – PIN_{BHL}); Brown and Hillegeist (2007 – PIN_{BH}); and Duarte and Young (2006 – PIN_{DY}). [7]

Table 6 shows that our VCV measure is positively correlated to all these PIN related measures, suggesting that VCV is, like PIN, indicative of informed trading. The correlation between VCV and PIN is of similar magnitude as the correlations between the various PIN measures. Compared to these PIN measures, however, our VCV measure is far easier to compute and does not require intraday data on the order process.

Duarte and Young (2006) argue that PIN measures not only informed trading but also other illiquidity effects. They therefore decompose PIN into *Adjusted PIN*, which is proposed as a cleaner measure of asymmetric information; and *PSOS* (probability of symmetric order-flow shock), which is a measure of illiquidity unrelated to asymmetric information. These additional variables are included in Table 6. Both Adjusted PIN are PSOS are

[7] Annual firm-level observations of PIN_{DY} are made available by Jefferson Duarte (`http://www.owlnet.rice.edu/~jd10/`). Annual firm-level observations of PIN_{EHO} are made available by Søren Hvidkjær (`https://sites.google.com/site/hvidkjaer/data`). Annual firm-level observations of PIN_{BH} and PIN_{BHL} are made available by Stephen Brown (`http://scholar.rhsmith.umd.edu/sbrown/pin-data`)

Table 6: VCV and PIN

	VCV	PIN_{EHO}	PIN_{BHL}	PIN_{BH}	PIN_{DY}	Adj. PIN	PSOS
VCV		0.39	0.40	0.52	0.41	0.40	0.32
PIN_{EHO}	0.51		0.50	0.67	0.76	0.60	0.55
PIN_{BHL}	0.51	0.62		0.68	0.51	0.44	0.32
PIN_{BH}	0.64	0.67	0.72		0.67	0.69	0.42
PIN_{DY}	0.55	0.85	0.62	0.70		0.67	0.73
Adjusted PIN	0.51	0.63	0.55	0.70	0.70		0.34
PSOS	0.44	0.62	0.43	0.45	0.70	0.38	

Notes: This table reports the correlation between the annual firm-level coefficients of variation of daily volume shares (VCV) and various annual firm-level PIN measures (Probability of Informed Trading). The upper diagonal entries show the correlations for levels. The lower diagonal entries show the average within-year rank correlations. PIN_{EHO} is estimated by Easley, Hvidkjaer, and O'Hara (2010). PIN_{BHL} is estimated by Brown, Hillegeist and Lo (2004). PIN_{BH} is estimated by Brown and Hillegeist (2007). PIN_{DY}, Adjusted PIN, and the illiquidity measure PSOS are estimated by Duarte and Young (2009). Sources: CRSP and cited authors' websites.

positively correlated with VCV.

In table 7, we examine the correlation between VCV and the three measures by Duarte and Young (2006) in a regression context. The regression results indicate that VCV is mostly associated with adjusted PIN and is not significantly related to PSOS, thereby supporting our claim that VCV, like adjusted PIN, measures asymmetric information rather than general illiquidity.

4.2 VCV and institutional ownership

In this section we study the relation between VCV and various indicators of institutional ownership that we obtain from 13F filings. Table 8 reports the result from regressing VCV on various institutional ownership characteristics. These characteristics include institutional holdings (defined as the percentage of shares of a firm held by institutional investors at the end of the year) and breadth of ownership (defined as the number of institutional investors holding shares in the firm, as a percentage of the total number of institutional investors at the end of each year – Chen et al., 2002). Boone and White (2015) find that institutional ownership leads to higher transparency and therefore lower infor-

Table 7: VCV and Adjusted PIN

	VCV		
	(1)	(2)	(3)
PIN_{DY}	0.165		−0.031
	(0.123)		(0.096)
Adjusted PIN	0.895***	0.978***	1.001***
	(0.082)	(0.098)	(0.096)
PSOS	−0.156**	−0.104**	
	(0.064)	(0.050)	
Observations	39971	39971	39971
Adjusted R^2	0.324	0.324	0.323
Year fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes: This table shows the results from regressing annual firm-level coefficients of variation of daily volume 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 year fixed effects and control variables: Size (log of market capitalization), Amihud illiquidity and turnover. T-statistics based on two-way clustered standard errors at the year and firm level are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level. Source: CRSP and the website of Jefferson Duarte (`http://www.owlnet.rice.edu/~jd10/`)

mation asymmetry. The first column of Table 8 shows indeed that VCV has a significantly negative association with breadth of ownership. Consistent with Boone and White (2015) VCV is lower (implying lower information asymmetry) for firms that have high breadth of ownership.

In addition, we consider two measures that are specifically designed to identify informed investors: Monitoring investors and Dedicated investors. Following Fich et al (2015), we define an institutional investor X to be a monitor for firm Y if firm Y belongs to the top 10% of holdings in the portfolio of institution X. 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', meaning they make relatively long-term investment in a select

Table 8: VCV and institutional ownership

	VCV			
	(1)	(2)	(3)	(4)
Holdings (%)	-0.0001	-0.0002	-0.0001	-0.0002
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Breadth	-0.533^{***}	-1.013^{***}	-0.571^{***}	-1.027^{***}
	(0.156)	(0.111)	(0.152)	(0.110)
Monitors		0.901^{***}		0.861^{***}
		(0.152)		(0.143)
Dedicated			0.301^{***}	0.297^{***}
			(0.084)	(0.086)
Observations	47220	47220	45730	45730
Adjusted R^2	0.418	0.421	0.414	0.417
Year fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: This table shows the results from regressing annual firm-level coefficients of variation of daily volume shares (VCV) on various measures of institutional ownership. *Holdings%* is the fraction of shares of the firm held by institutional investors at the end of the year; *Breadth* is the percentage of all institutional investors that hold shares of the firm (Chen *et al.*, 2002); $Monitors$ is the fraction of institutional investors in each firm for which the firm is in the top 10% of the institution's holdings (Fich *et al.*, 2015); and *Dedicated* is the fraction of institutional investors in each firm that are classified as 'Dedicated' investors by Bushee and Noe (2000). All regressions include year fixed effects and control variables: Size (log of market capitalization), Amihud illiquidity and turnover. T-statistics based on two-way clustered standard errors at the year and firm level are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level. Sources: CRSP, 13F and the website of Brian Bushee `http://acct.wharton.upenn.edu/faculty/bushee/`

group of firms, as opposed to 'quasi-indexing' investors and 'transient' investors.[8]

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

The relation between patterns in institutional ownership and VCV reported in Table 8 strongly reaffirms that VCV is a measure of informed trade. Consider firm A which is owned by only few institutional investors, and for several of those few institutions holding shares in firm A, this investment is a large fraction of their portfolio. That is, *Breadth* is low, but *Monitors* and *Dedicated* are expected to be high. Ownership of firm A is therefore relatively concentrated in the hands of a small number of presumably well informed investors. When trading firm A, asymmetric information should be a significant concern, as it is likely that the counter party is one of these better informed investors. On the other hand, consider firm B that is widely held among institutional investors, each of which holds only a relatively small share of the firm. That is, *Breadth* is high, while *Monitors* and *Dedicated* are expected to be low. For firm B the risk of asymmetric information should be therefore lower, which is in accordance with the results reported in Table 8.

5 VCV around earnings announcements

In this section we document the pattern of VCV around earnings announcements. It is widely recognized that earnings announcement resolve information asymmetries. In this section we show, consistent with this view, that the VCV increases prior to announcements and declines afterwards, suggesting that uninformed traders delay their trades until information asymmetries are resolved after the announcement.

[8]This classification of institutional investors in 13F is made available on the website of Brian Bushee `http://acct.wharton.upenn.edu/faculty/bushee/`.

We obtain quarterly earnings announcement dates from COMPUSTAT for all NYSE and AMEX listed US firms. We compute the VCV over 10 day windows before and after the announcement, while skipping the 5 days closely around the announcement. That is, the before window includes days -12:-3 and the after window includes days 3:12, where day 0 is the announcement date.[9] Table 9 shows summary statistics of the VCV in each 10 day window. The table reports that the median VCV in the before window is higher than in the after window. Also the average difference is negative and statistically significant. The table further shows that the number of firms for which the VCV decreases following the announcement is around 52%. Although this number is only slightly higher than 50% (which it would be if earnings announcements have no effect on information asymmetry), the difference is statistically significant: A reduction in VCV around earnings announcements is significantly more common than an increase in VCV . For each earnings announcement date, we also obtain a *placebo date*, which is a date within 100 trading days before or after the actual announcement, randomly drawn from a uniform distribution. Around these placebo dates, the fraction of firms that see a reduction in VCV is approximately 50%.

The second and third column of Table 9 show the same statistics for surprising and non-surprising announcement separately. We choose a market-based definition of surprise: An announcement is classified as surprising when the absolute value of the cumulative abnormal return over days -1, 0 and 1 around the announcement exceeds its median: $|CAR_{-1:+1}| > median(|CAR_{-1:+1}|)$. Abnormal returns are defined as the stock's return minus the market return on the same day. As expected, the patterns in VCV around announcement dates are strongest when the announcement is surprising. That is, when the announcement is surprising, the announcement is more informative and is more effective in reducing information asymmetry, which is reflected by stronger variation of the VCV.

[9]We also consider other windows including 5 day and 20 day windows. Results are are qualitatively similar to 10 day windows and available upon request.

Table 9: VCV around Earnings Announcement

	All	Surprising	Non-surprising
Median VCV before announcement	0.526	0.537	0.515
Median VCV after announcement	0.515	0.523	0.508
Mean change	-0.010	-0.014	-0.005
T-statistic (H0: No change)	-11.545	-12.217	-4.311
% VCV before >VCV after	0.517	0.520	0.514
T-statistic (H0: 50%)	15.105	12.701	8.698
Observations	197,975	97,221	100,754
% VCV before >VCV after	0.500	0.500	0.501
T-statistic (H0: 50%)	0.310	-0.154	0.584

Notes: This table shows nonparametric statistics on the pattern of VCV computed over 10 day windows before and after quarterly earnings announcements, excluding the 5 days window closely around the announcement (i.e. the before window includes days -12:-3 and the after window includes days 3:12, where day 0 is the announcement date.) The first two rows show the median VCV during the windows before and after the announcement. The third and fourth row show the average change in VCV following the announcement, and a t-statistic for the hypothesis that the change is zero. The final rows show the number of firms for which the VCV decreases following the announcement and a t-statistic for the hypothesis that this percentage is 50%. The final rows show these percentages for randomly chosen placebo dates within 100 trading days before or after the actual announcement. Column 2 and 3 differentiate surprising and non-surprising announcements. An earnings announcement is considered surprising when the absolute cumulative announcement window return exceeds the median absolute return around earnings announcements ($|CAR_{-1:+1}| > median(|CAR_{-1:+1}|)$).

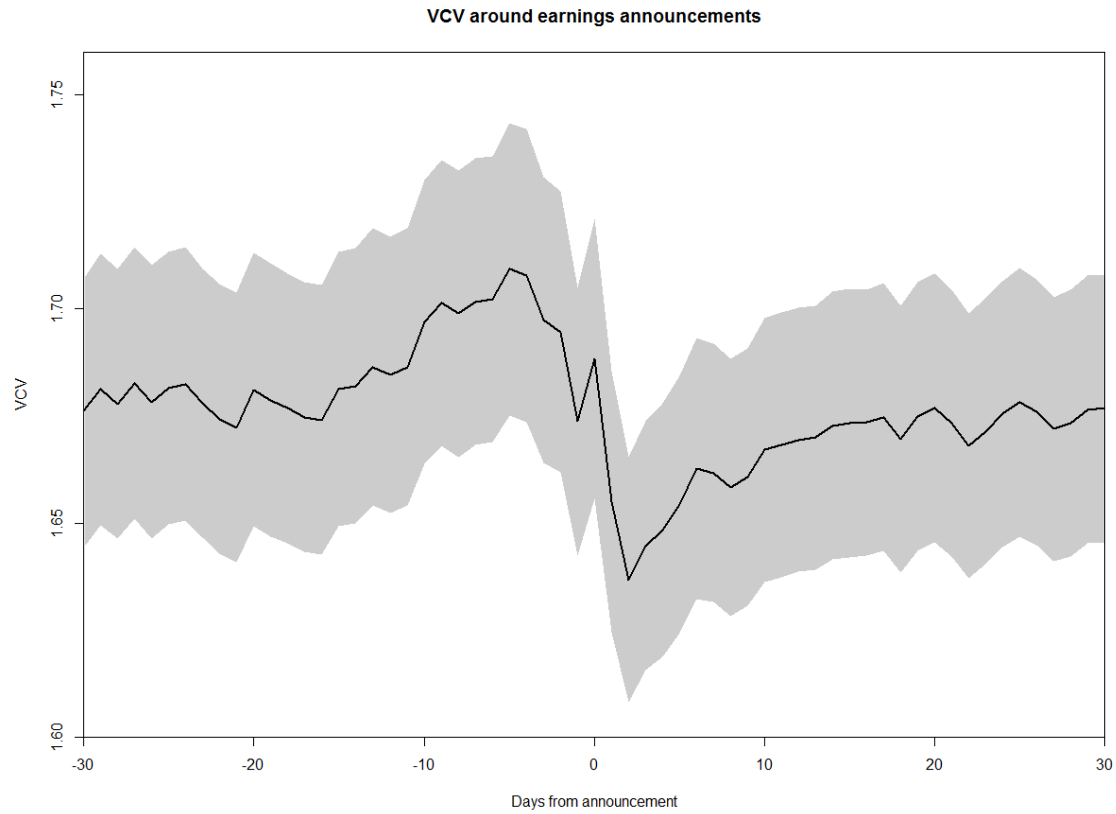

Figure 3: This figure shows the evolution of the daily cross-sectional VCV on days -30 to 30 around quarterly earnings announcements. That is, at day d, the sample includes the stock's trading volume on day d days after the announcement, for all quarterly announcements in COMPUSTAT. The black line shows the coefficient of variation for this sample. Grey areas show 95% confidence intervals derived following Albrecher *et al.* (2010).

Computing the VCV over 10 day windows is not optimal. First of all, the small sample causes the estimates to be noisy, which explains the small (although significant) changes in Table 9. More worrying is that Table 9 is showing that the distribution of trading volume changes around the announcement, hence there is no a priori reason to assume that the distribution does not change *within* the 10 day windows before and after the announcement. For this reason, we also compute the so-called *cross-sectional VCV* for each day around the announcement. We calculate the coefficient of variation at day d after the announcement, using the daily trading volumes at day d for all 197,975 earnings announcements in our sample (all volumes are as before volume shares, i.e. volumes as a percentage as total trading volume on that calendar date). This cross sectional VCV is then computed for all days d over the interval -30 days before the announcement to +30 days after the announcement. The black line in Figure 3 shows the pattern of VCV over this interval, while the shaded areas indicate 95% confidence bounds. Clearly, the VCV increases when the announcement date is approaching, as uninformed investors are delaying their trading activity. When information asymmetries are resolved at the announcement date, the VCV sharply declines and stays relatively low for around 10 trading days (i.e. around 2 weeks). After 30 trading days, the VCV is roughly equal to the VCV 30 trading days prior to the announcement. Johnson and So (2017) document a very similar pattern in their Multimarket Information Asymmetry (MIA) measure, calculated from the relative trading volume of options and stocks. Like VCV, MIA increases in the days before earnings announcements, and rapidly declines around the announcement.

6 Conclusion

In this paper we derive from the Kyle (1985) model that the distribution of trading volume diverges from a normal distribution in the presence of informed trading. Specifically, we show that the Volume Coefficient of Variation (VCV) increases in the proportion on informed trading. We therefore propose VCV as a measure of adverse selection. Monte Carlo simulations show indeed that VCV increases in the proportion of informed liquidity seekers.

Our empirical results indicate that stocks for which daily trading volume has a high coefficient of variation, also tend to have other characteristics that are typically associated with asymmetric information (e.g.: high PINs, low mutual fund ownership, low analyst coverage, high illiquidity) and vice versa.

Around quarterly earnings announcements we find, as expected, that informed trading increases shortly before the announcement and rapidly decreases after the announcement. The post-earnings announcement drift (PEAD) and momentum effect are lower when our VCV measure is high, suggesting that PEAD and momentum are mostly driven by noise-traders.

VCV is an appealing measure of information asymmetry because of its simplicity: It is calculated by simply dividing the sample standard deviation of daily trading volumes over the sample mean. The measure is applicable in both cross-sections and time-series. Unlike other measures of information asymmetry, estimating VCV requires only total daily trading volumes, as opposed to intraday order-level data.

References

Admati, A.R. and Pfleiderer, P. (1989), Divide and conquer: A theory of intraday and day-of-the-week effects, *Review of Financial Studies*, 2, 189-223.

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

Akay, O., Cyree, H.B., Griffiths, M.D., and Winters, D.B., (2012) What does PIN identify? Evidence from the T-bill market, *Journal of Financial Markets*, 15, 29-46.

Aktas, N., de Bodt, E. Declerck, F., and van Oppens, H (2007) The PIN anomaly around M&A announcements, *Journal of Financial Markets*, 10, 169-191.

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

Bagehot, W. (1971), The only game in town, *Financial Analyst Journal*, 27, 14-22.

Black, F. (1986). Noise. The journal of finance, 41, 528-543.

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

Bongaerts, D., Rösch, D., and Van Dijk, M. A. (2016). Cross-Sectional Identification of Informed Trading. Working paper.

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

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

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

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

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

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

Chae, J. (2005). Trading volume, information asymmetry, and timing information. *The journal of finance*, 60, 413-442.

Chen, Q., Goldstein, I., Jiang, W. (2007) Price informativeness and investment sensitivity to stock price, *Review of Financial Studies*, 20, 619-650.

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

Chordia, Tarun, Avanidhar Subrahmanyam, and V. Ravi Anshuman (2001). Trading activity and expected stock returns. *Journal of Financial Economics* 59, no. 1 : 3-32.

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

Copeland, T.E., Galai D. (1983) Information effects and the bid-ask spread, *Journal of Finance*, 38, 1457-1469.

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

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

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

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

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

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

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

Easley, D., and O'Hara, M. (1992), Adverse selection and large trading volume: the implication for market efficiency, *Journal of Financial and Quantitative Analysis*, 27, 185-208.

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

Easley, D., O'Hara, M., and Saar, G. (2001), How stock splits affect trading: a microstructure approach, *Journal of Financial and Quantitative Analysis*, 36, 25-51.

Fich, Eliezer M., Jarrad Harford, and Anh L. Tran (2015). Motivated monitors: The importance of institutional investors? portfolio weights. *Journal of Financial Economics* 118, no. 1: 21-48.

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

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

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

Heidle, H., and Huang, R. (2002) Information-based trading in dealer and auction markets: an analysis of exchange listings, *Journal of Financial and Quantitative Analysis*, 37, 391-424.

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

Kang, M. (2010) Probability of information-based trading and the January effect, *Journal of Banking and Finance*, 34, 2985-2994.

Karpoff, J.M. (1986), A theory of trading volume, *Journal of Finance*, 41, 1069-1087.

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

Kyle, A. S. (1989). Informed speculation with imperfect competition. *The Review of Economic Studies*, 56(3), 317-355.

Lin, H.W., and Ke, W.C. (2011), A computing bias in estimating the probability of informed trading. *Journal of Financial Markets*, 14, 625-640

Madhavan, A., Richardson, M. and Roomans M. (1997), Why do security prices change? A transaction-level analysis of NYSE stocks, *The Review of Financial Studies*, 10, 4, 1035-1064.

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

Stoll, H.R. (1978), The supply of dealer services in securities markets, *Journal of Finance*, 33, 1133-1151.

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

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

Yan, Y. and Zhang, S. (2012), An improved estimation method and empirical properties of the probability of informed trading, *Journal of Banking and Finance*, 36, 454-467.

Appendix A: Variance of trading volume

Define $\tilde{Y}_{MM} = |\sum_M \tilde{y}_i|$ as the part of double-counted volume traded by market makers (order imbalance), $\tilde{Y}_I = \sum_{1\ldots\eta M} |\tilde{y}_i|$ as the part traded by informed investors and $\tilde{Y}_U = \sum_{\eta M+1\ldots M} |\tilde{y}_i|$ as the part traded by uninformed investors. Then Eq. (1) can be rewritten as:

$$\tilde{V} = \tfrac{1}{2}\left(\tilde{Y}_I + \tilde{Y}_U + \tilde{Y}_{MM}\right). \tag{10}$$

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

$$\begin{aligned} Var\left(2\tilde{V}\right) &= Var\left(\tilde{Y}_I\right) + Var\left(\tilde{Y}_U\right) + Var\left(\tilde{Y}_{MM}\right) \\ &\quad +2Cov\left(\tilde{Y}_I, \tilde{Y}_U\right) + 2Cov\left(\tilde{Y}_I, \tilde{Y}_{MM}\right) + 2Cov\left(\tilde{Y}_U, \tilde{Y}_{MM}\right) \end{aligned} \tag{11}$$

We can assume that $Cor\left(\tilde{Y}_I, \tilde{Y}_U\right) = 0$, because the demands of informed and uninformed liquidity seekers are independent. Moreover, when $M \to \infty$ and $\eta > 0$, the order imbalance consists mainly of orders submitted by informed liquidity seekers. The orders of uninformed traders tend to cancel each other out because of the *i.i.d* property. In other words, in the limit the market makers trade exclusively to ofset the imbalance from informed traders. Therefore, $Cor\left(\tilde{Y}_U, \tilde{Y}_{MM}\right) \to 0$ and $Cor\left(\tilde{Y}_I, \tilde{Y}_{MM}\right) \to 1$. Substituting these into the variance gives:

$$Var\left(2\tilde{V}\right) \to Var\left(\tilde{Y}_I\right) + Var\left(\tilde{Y}_U\right) + Var\left(\tilde{Y}_{MM}\right) + 2s.d.\left(\tilde{Y}_I\right) s.d.\left(\tilde{Y}_{MM}\right), \tag{12}$$

which, using the properties of the Half Normal distribution, results in:

$$
\begin{aligned}
Var\left(2\tilde{V}\right) &\rightarrow \eta^2 M^2 \sigma^2 \left(1-\tfrac{2}{\pi}\right) + (1-\eta) M \sigma^2 \left(1-\tfrac{2}{\pi}\right) + \left(\eta^2 M^2 + (1-\eta) M\right) \sigma^2 \left(1-\tfrac{2}{\pi}\right) \\
&\quad +2\sqrt{\eta^2 M^2 \sigma^2 \left(1-\tfrac{2}{\pi}\right)} * \sqrt{\left(\eta^2 M^2 + (1-\eta) M\right) \sigma^2 \left(1-\tfrac{2}{\pi}\right)} \\
&\rightarrow 2M^2 \sigma^2 \left(1-\tfrac{2}{\pi}\right) \left(\eta^2 + (1-\eta) M^{-1} + \eta \sqrt{\eta^2 + (1-\eta) M^{-1}}\right) \\
&\rightarrow 4M^2 \sigma^2 \left(1-\tfrac{2}{\pi}\right) \eta^2
\end{aligned}
\tag{13}
$$

where the last step follows from $M^{-1} \rightarrow 0$ for large M. The standard deviation of trading volume is thus equal to $s.d.\left(\tilde{V}\right) = M\sigma\eta\sqrt{1-\frac{2}{\pi}}$, from which Proposition 1 is easily derived.