Bank of Finland Research Discussion Papers
1 • 2018

Matthijs Lof – Jos van Bommel

Asymmetric information and the distribution of trading volume

Bank of Finland
Research

Bank of Finland Research Discussion Papers
Editor-in-Chief Esa Jokivuolle

Bank of Finland Research Discussion Paper 1/2018
22 January 2018

Matthijs Lof – Jos van Bommel
Asymmetric information and the distribution of trading volume

ISBN 978-952-323-206-8, online
ISSN 1456-6184, online

Bank of Finland
Research Unit

PO Box 160
FIN-00101 Helsinki

Phone: +358 9 1831

Email: research@bof.fi
Website: www.bof.fi/en/research

Asymmetric Information and the Distribution of Trading Volume *

Matthijs Lof† Jos van Bommel‡

January 17, 2018

Abstract

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

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

*We thank Anna Bayona, Alex Boulatov, Jonathan Cohn, David Easley, Eliezer Fich, Carlos Forner Rodríguez, Thierry Foucault, Bruce Grundy, Sam Hartzmark, Dirk Jenter, Pete Kyle, Anna Obizhaeva, Deniz Okat, Conall O'Sullivan, Paolo Pasquariello, Jillian Popadak, Jorg Prokop, Francesco Sangiorgi, Valeri Sokolovski, Enrique Sentana, Erik Theissen, Hugo van Buggenum, Christian Westheide and participants at the American Economic Association (Chicago), the EBC network, the French Finance Association, the German Finance Association, the SAFE Market Microstructure Conference, the Spanish Finance Association, Aalto University, the Bank of England, the Bank of Finland, Erasmus University Rotterdam, the Higher School of Economics, University of Amsterdam, University of Liverpool, University of Maastricht, and University of Turku 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 analyze the distribution of trading volume in security markets, and investigate how it depends on the proportion of informed trade. We consider a market where liquidity seekers submit orders to competitive liquidity providers (market makers), who absorb the order imbalance and set the clearing price, as in Kyle (1985). Liquidity seekers are either informed or uninformed. Uninformed liquidity seekers place uncorrelated 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 absorbed by market makers. We derive simple expressions for the first two moments of the distribution of total 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. A major advantage of VCV is its computational simplicity: computing VCV requires only observations of trading volume, as opposed to quotes, prices, or signed order flow.

The intuition behind 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, their orders are mostly netted out against each other, and net order flow is relatively low compared to the observed trading volume. In this case, trading volume follows a slightly skewed Normal-like distribution. As more market participants are informed, liquidity demand becomes more correlated, so that the expected net order flow takes up a higher share of total trading volume, resulting in a more dispersed and skewed distribution of trading volume. We show that the coefficient of variation adequately captures the mapping from the proportion of informed trade to the volume distribution. 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]

[1]To the best of our knowledge, we are the first to relate the coefficient of variation of trading volume

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.

Easley et al. (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 measure has been widely used to study information asymmetries in security markets.[2]

Both PIN and VCV are expected to increase in the order imbalance generated by informed demand. The PIN measure is estimated from transaction-level data and requires trades to be classified as either buy- or sell-initiated. Also other information asymmetry measures, such as order flow volatility (Chordia et al., 2017) 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 due to high-frequency trading (Johnson and So, 2017; Easley et al., 2012; Boehmer et al., 2007). Computing VCV does not require signed transaction-level data. Instead, VCV is estimated from aggregate volume data.

Duarte and Young (2009) argue that (unadjusted) PIN is not only measuring informed

to asymmetric information. Chordia et al. (2001) use the coefficient of variation of trading volume when examining the relation between stock returns and the variability of trading volume, without relating this measure to asymmetric information.

[2]Easley et al. (1997a and 1997b) analyze the information content around trade lags and trade size. Applications of PIN include, among others, the pricing of information asymmetry (Easley et al., 2002), 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).

trade, but also general illiquidity unrelated to information asymmetry.[3] 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 compare VCV to various PIN measures and find that VCV is strongly related to PIN, but even more so to Adjusted PIN, while the relationship to PSOS is weak. This corroborates that VCV is a measure of informed trading, rather than general illiquidity.

Johnson and So (2017) propose the multimarket information asymmetry (MIA) measure, which is based on the relative daily trading volumes in options and stocks, following the premise that informed investors are more likely than uninformed investors to trade in options. Although MIA, like VCV, is a simple measure to compute, it requires access to option trading volume in addition to equity trading volume. Llorente et al. (2002) propose the C2-measure that captures the relation between daily trading volume and return persistence, as a proxy for information asymmetry. C2 deduces the proportion of informed trade from the premise that returns generated by informed trade are likely to be persistent, while uninformed trade leads to return reversals. We find that our VCV measure is positively correlated with both the MIA measure of Johnson and So (2017) and the C2 measure of Llorente et al. (2002).

Recent studies find that institutional ownership is associated with improved disclosure of information (Boone and White, 2015) and more informative market prices (Bai et al., 2016). Consistent with these results, we find from 13F filings that firms with more institutional shareholders (i.e. high breadth of ownership) have on average lower VCVs. 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 (Fich et al.,

[3]Other papers in the debate on the validity of PIN include Easley et al. (2010), Akay et al. (2012), and Duarte et al. (2017). Other studies focus on 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. (2012) develop the volume synchronized PIN, or VPIN. This estimator captures not only information asymmetry but also order flow toxicity, i.e.: the risk of unbalanced orderflows.

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 or dedicated (i.e. informed) investors.

The crux of our theoretical 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 observed total 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 demonstrate 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. The coefficient of variation of trading volume is therefore a natural measure of the proportion of informed trade, as it increases monotonically in the proportion of informed trade, while it is asymptotically independent of the number of market participants and their trading intensity.

To further analyze the relation between informed trading and the distribution of volume, and to gain insight into the small sample properties of VCV, we conduct a Monte Carlo analysis in Section 3. We start by simulating our benchmark model and find, as predicted, that the generated volume distribution changes markedly as a function of the proportion of informed trade. With more informed (correlated) orders, the volume distribution becomes more skewed and dispersed. Our simulations confirm that VCV strictly increases in the proportion of informed trade, even when the sample size or the number of market participants is low.

In our benchmark model, volume jumps only occur due to informed trading: in the absence of informed investors, the volatility of volume is very low. To demonstrate that VCV also works as an indicator of informed trading outside of this specific environment, we repeat the simulation analysis for several variations of our benchmark model. These variations include non-Gaussian demand, random arrival of information, heterogeneity among the informed investors, and random variation in the number of market participants and their trading intensity. We continue to find a strong positive and monotonic relationship between VCV and the proportion of informed trade in all of these alternative specifications.

We recognize that the proportion of informed traders and their trading intensities are endogenous and dependent on the strength of the informational advantage and on the trading activity of the uninformed traders, as in Kyle (1985) and Admati and Pfleiderer (1988). Collin-Dufresne and Fos (2015) provide empirical evidence of such strategic timing by informed investors. In addition, the proportion of informed trade may depend on the informed traders' risk aversion, as in Subrahmanyam (1991), or on the time horizon of their informational advantage and the competition with other informed traders, as in Holden and Subrahmanyam (1992) and Foster and Viswanathan (1996). In this paper, we take the econometrician's point of view, meaning that we are agnostic about about the number of informed traders, their preferences or time horizons, and are only interested in the resulting proportion of liquidity seeking demand that comes from informed market participants. In our simulation exercise in Section 3, we also show that VCV is indicative of informed trading regardless of whether informed trading intensity is exogenously given, or specified endogenously as a function of uninformed trading activity.

The sampling frequency of volume is not explicitly specified in our model. All empirical results in this paper are based on daily trading volumes, but VCV can also be computed from intraday, weekly or monthly volumes. The VCV of a specific asset can be obtained from a time-series of volumes, and is used to compare information asymmetries across assets. Alternatively, VCV can be computed from a cross-section of trading volumes to evaluate information asymmetry over calendar time or event time.

For our empirical analysis in Section 4, we compute annual firm-level observations of VCV from daily volumes of all NYSE, AMEX and NASDAQ stocks from 1980 until 2016, obtained from CRSP. We use three distinct volume measures: *(i)* trading volume in US dollars, *(ii)* turnover, and *(iii)* volume market shares (volume in US dollars as a fraction of total market dollar volume). All three measures provide highly similar results, confirming that VCV is robust to aggregate-level variation in trading volume that is unrelated to idiosyncratic firm-level information. Our cross-sectional analysis shows that VCV correlates as expected with various firm-level characteristics: firms that are smaller and younger have higher VCVs, as do stocks that see lower turnover, higher return volatility, and wider bid-ask spreads. In addition, VCV is significantly correlated with other indicators of asymmetric information, in particular with Adjusted PIN by Duarte and Young (2009), and with patterns in institutional ownership. As further evidence of VCV measuring informed trading, we show that, controlling for illiquidity, return reversals are weaker for high-VCV stocks, consistent with informed trading being predictive of future price changes. As we report in our internet appendix, all results are qualitatively similar in subsamples of NYSE/AMEX and NASDAQ stocks, as well as in pre- and post-2000 periods, validating the robustness of VCV as a measure of information asymmetry in different market environments.

Section 5 documents time-series patterns of VCV around information events. First, we exploit exogenous terminations in analyst coverage due to brokerage closures, similar to Kelly and Ljungqvist (2012) and Derrien and Kecskes (2013). We find that the VCV of affected firms significantly increases in the year following such disruptions to the information environment. We expect the impact of coverage terminations to be more severe for firms that already have low analyst coverage prior to the brokerage closure. Our results confirm this hypothesis: the increase in VCV following closure-induced coverage terminations is much larger for stocks with low analyst coverage. Next, we analyze the cross-sectional VCV computed from the cross section of trading volumes around earnings announcements. It has been widely recognized that information asymmetries are resolved around earnings announcements. We expect that the proportion of uninformed trading is low close 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, 1994; and Chae, 2005). After the announcement, the playing field is levelled and the market is more attractive for uninformed traders. Our empirical investigations bear this out. From a comprehensive sample of over 300,000 quarterly earnings announcements of U.S. firms, we find VCV to be relatively high prior to announcements, while VCV is significantly lower in the days following the announcement.

2 Theory

To analyze the distribution of trading volume, we postulate three kinds of traders: *(i)* informed liquidity seekers, *(ii)* uninformed liquidity seekers, and *(iii)* competitive liquidity providers (market makers). We assume that there are M individual liquidity seekers, of which a proportion η is informed. Both M and ηM are integers. During independent trading sessions, the M individual liquidity seekers submit orders to the market where buy orders are matched to sell orders and the residual orders (henceforth referred to as 'order imbalance' or 'net order flow') are taken up by the liquidity providers who set the price. The model thus closely resembles that of Kyle (1985), with the difference that we consider the individual orders of the liquidity seekers, and our focus is on analyzing total trading volume.

To be precise, we denote the individual demands of liquidity seekers $\tilde{y}_i$, for which positive values indicate buy orders and negative values indicate sell orders. The order imbalance (net order flow) is the sum of all orders, $\sum_M \tilde{y}_i$, which is taken up by the liquidity provider. This imbalance is typically not publicly observable. Total trading volume can then be written as:

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

The term inside brackets is the "double-counted transaction volume", counting both buys and sells, of (i) the liquidity seekers (the first term) and (ii) 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.[4]

[4]This expression for trading volume is similar to that in Admati and Pfleiderer (1988) and Grundy and

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 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.

We let the demand of every liquidity seeker, whether informed or uninformed, be Normally distributed with zero mean and standard deviation σ, which we denote trading intensity.[5]

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

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 net order flow follows a Normal distribution around zero as in Kyle (1985):

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

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

We now derive the first two moments of the observable total trading volume (Eq.(1)) as

McNichols (1989).

[5]The assumption that individual uninformed and informed demands are of the same order of magnitude is without loss of generality. 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 *trade* rather than informed *traders* in the market.

a function of η. Using the properties of the Half Normal distribution we find:[6]

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

As the number of market participants M increases, expected trading volume per capita converges to a strictly increasing function of the proportion of informed trade η:

$$\lim_{M\to\infty} E\left[\frac{V}{M}\right] = \frac{\sigma}{\sqrt{2\pi}}\left(1+\eta\right). \tag{5}$$

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

$$\lim_{M\to\infty} Var\left(\frac{V}{M}\right) = \sigma^2\left(1-\frac{2}{\pi}\right)\eta^2. \tag{6}$$

For large M, the expected value of trading volume thus increases linearly in $1+\eta$ (Eq.(5)), while its standard deviation increases linearly in η (Eq.(6)). The ratio of the standard deviation to the mean (the coefficient of variation) of trading volume is therefore strictly increasing in η. Moreover, for large M, the coefficient of variation is independent of both the number of market participants M and their trading intensity σ.

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

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 net order flow 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:*

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

where μ_V and σ_V denote the population mean and standard deviation of trading volume V.

Corollary

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

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

is a consistent estimator of $\frac{\sigma_V}{\mu_V}$. The Volume Coefficient of Variation (VCV) is a measure of informed trade. $E[VCV]$ increases monotonically in η.

The above analysis also shows that a direct estimator of the proportion of informed trade is implied from Eq.(7):

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

However, as our simulation results in the next section show, $\hat{\eta}$ is a consistent estimator of η only when demand is Normally distributed, M is large, and η is constant across observations. We find that $\hat{\eta}$ behaves poorly in small samples or when we relax the assumptions of the model, primarily because its denominator can be close to zero or turn negative. We find that VCV, on the other hand, is monotonically increasing in η under general conditions, including non-Normality, time-varying proportions of informed trade and endogenous informed trade. For this reason, we propose VCV, as opposed to $\hat{\eta}$, as our measure of informed trade.

3 Simulations

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 (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. The aggregate informed demand can be interpreted either as ηM informed traders who each place orders of similar magnitude as the uninformed traders, or as a single informed trader who places orders of size $\frac{\eta}{1-\eta}$ times the average magnitude of orders submitted by the $(1 - \eta) M$ uninformed traders. We compute the observed trading value volume V that follows from Eq.(1). For each (M, η) pair, we generate a sample of T volume (V) observations, from which we compute VCV.

Figure 1 displays four histograms of simulated volumes with $M = 1,000$ liquidity seekers, for different values of η. The sample size is $T = 1,000$ trading sessions. The simulation confirms the analysis in the previous section: In case of no informed traders ($\eta = 0$), the volume distribution follows a slightly skewed bell-curve, while in the presence of informed traders volume is higher in level and far more dispersed. The simulated VCVs for the four panels are 0.03, 0.14, 0.48 and 0.77, respectively. [7]

[7] The slightly skewed bell-curved volume distribution for $\eta = 0$ converges (as $M \to \infty$) to the distribution

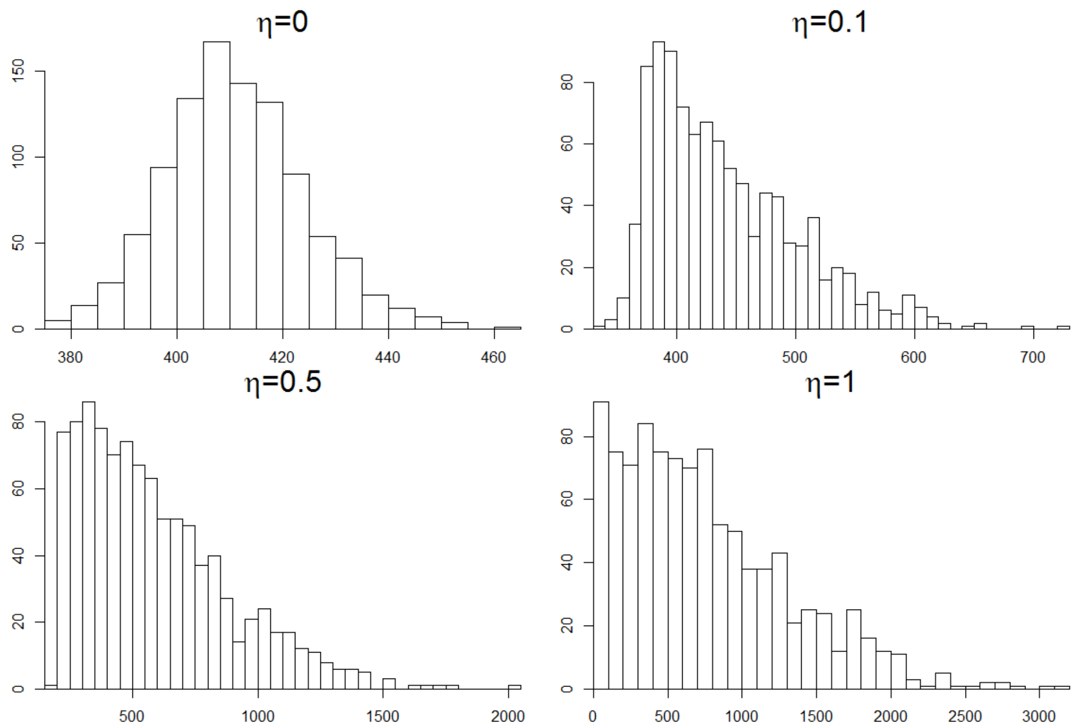

Figure 1: Histogram of T=1,000 volume realizations simulated from the model outlined in Section 2, for various values of the proportion of informed trading η. The number of liquidity seekers (M) is 1,000 and the trading intensity (σ) is fixed at unity.

Figure 2 reports the average Volume Coefficient of Variation (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 Figure 2 shows, the average VCV only deviates substantially from its theoretical value (Eq.(7)) 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 order submitters (M) in markets is typically unknown.

In Table 1, Panel A, we report the average VCV as plotted in Figure 2 for selected values of η, as well as the standard deviation of VCV to evaluate its precision. In addition to VCV, we also report these statistics on simulated values of $\widehat{\eta}$ (Eq.(9)). Both VCV and $\widehat{\eta}$ increase monotonically in the true proportion of informed trade (η). This is even the case

of the maximum of two Normally distributed random variables, which was first described by Clark (1961).

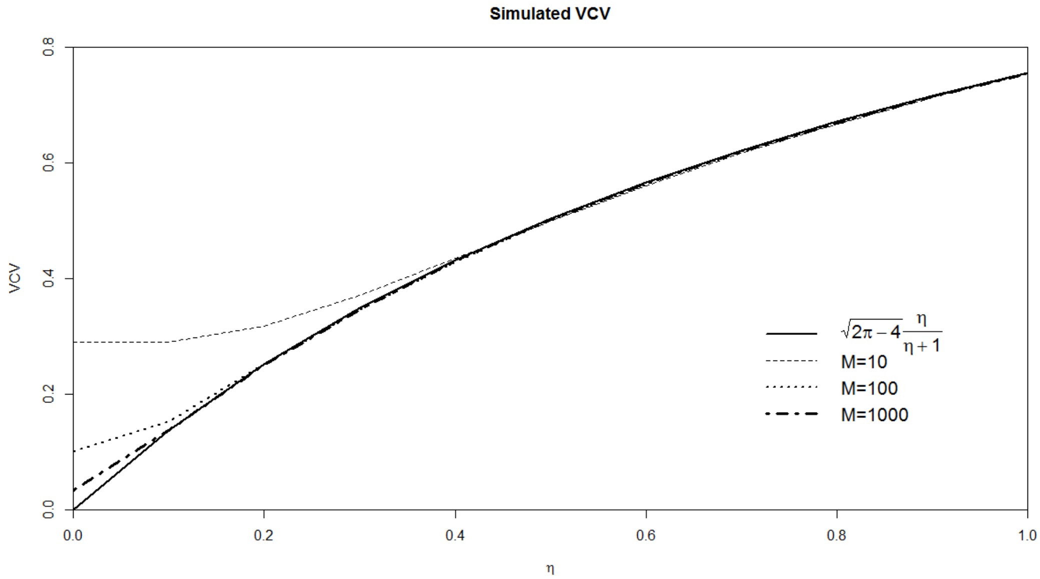

Figure 2: Average VCV obtained from $R = 1,000,000$ replications of $T = 100$ volume realizations simulated from the model outlined in Section 2, for various values of the proportion of informed trading η and number of liquidity seekers M.

for markets with only a small number of liquidity seeking traders M. Also, the estimator $\widehat{\eta}$ in our simulations traces the true value of η closely, in particular when either M or η are not too low. Panel B of Table 1 reports simulation results for smaller simulated samples, of only $T = 10$ trading sessions. We still find the average VCV and $\widehat{\eta}$ to increase monotonically in η, although VCV, and in particular $\widehat{\eta}$, are less precisely estimated.

Next, we investigate the robustness of VCV as a measure of asymmetric information by simulating trading volumes from various modified versions of our benchmark model. First, we repeat our simulation while relaxing the assumption of Normally distributed demand (Eq.(2)) and allow for leptokurtic and skewed demand distributions, to generate outliers in trading volume that are unrelated to informed trading. In Table 2, Panel A, we report the average VCV and $\widehat{\eta}$ from R simulations in which liquidity demand follows a leptokurtic t-distribution with 4 degrees of freedom, or a Skew-Normal distribution with shape parameter 10 (indicating positive skewness), for selected values of η, while keeping $M = 1,000$ and $T = 100$ fixed. We find that relaxing the assumption of Normality does not change the main result of our analysis: VCV and $\widehat{\eta}$ are still strictly increasing in η. However, the standard deviations are

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					$\widehat{\eta}$		
			$M = 10$					$M = 10$		
Avg	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
			$M = 100$					$M = 100$		
Avg	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
			$M = 1000$					$M = 1000$		
Avg	0.03	0.25	0.50	0.67	0.75	0.02	0.20	0.50	0.80	1.01
s.d.	0.00	0.02	0.03	0.05	0.06	0.00	0.02	0.05	0.10	0.15
Panel B: $T = 10$										
η	0	0.2	0.5	0.8	1	0	0.2	0.5	0.8	1
			VCV					$\widehat{\eta}$		
			$M = 10$					$M = 10$		
Avg	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
			$M = 100$					$M = 100$		
Avg	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
			$M = 1000$					$M = 1000$		
Avg	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

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

clearly smaller for VCV than for $\widehat{\eta}$. More importantly, the average $\widehat{\eta}$ is no longer closely following the true value of η, implying that, in the case of non-Gaussian demand, $\widehat{\eta}$ should not be interpreted as a direct estimator of the true value of η. We obtain qualitatively similar results when simulating demand from a Uniform distribution, or from t- and Skew-normal distributions with different degrees-of-freedom and shape parameters.

In practice, the proportion of informed trade η is not necessarily constant across observations, and we are typically interested in measuring the *average* proportion of informed trade, 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 while allowing the proportion of informed trade η to be random across observations. This version of our model can be interpreted as a hybrid of our Kyle (1985)-type model in Section 2 and the PIN model by Easley et al. (1996), in which arrival of information is random, similar to Back et al. (2017). Panel B of 2 reports simulation results for the case where the number of uninformed liquidity seekers is fixed at 1,000, while the number of informed liquidity seekers is in each of the $T = 100$ trading sessions randomly drawn from a Binomial distribution. The number of active informed traders is in each trading session equal to X with probability $\frac{1}{5}$ and zero with probability $\frac{4}{5}$, such that the informed traders participate in only one out of five trading sessions on average. To create variation in the average proportion of informed trade, we adjust the potential number of informed traders X. In this setting, $\widehat{\eta}$ clearly does not perform well as a measure of informed trading. The simulated observations of $\widehat{\eta}$ are widely dispersed, while their averages are not monotonically increasing in $E[\eta]$, and are not bounded by 0 and 1. This poor performance of $\widehat{\eta}$ occurs because the denominator in Eq.(9) can easily take on small or negative numbers, which makes the estimator highly erratic. VCV, on the other hand, continues to be monotonically increasing in $E[\eta]$ while its standard deviations remain fairly low. These results are robust to various alternative distributions for the number of active informed investors.

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

$$\begin{aligned} \tilde{y}_{informed,t} &\sim N\left(0, \sigma_t^2\right) \\ \sigma_t &= \frac{\left|\sum_{i=\eta M+1 \ldots M} \tilde{y}_{i,t}\right|}{E\left[\left|\sum_{i=\eta M+1 \ldots M} \tilde{y}_{i,t}\right|\right]} = \frac{\left|\sum_{i=\eta M+1 \ldots M} \tilde{y}_{i,t}\right|}{\sqrt{2(1-\eta M)/\pi}} \end{aligned} \tag{10}$$

Hence, the unconditional expected value of the informed investor's trading intensity σ_t is unity, as for the uninformed investors, but the informed investors trade more (less) aggressively when the uninformed net order flow is relatively high (low). Panel C of Table 2 reports the average simulated VCV and $\widehat{\eta}$ from this model. The case $\eta = 1$ is omitted, because the trading intensity of informed investors in undefined in the absence of uninformed investors. As in our benchmark model, we find a positive and monotonic relation between η and VCV. Also $\widehat{\eta}$ is increasing in η, although it is estimates are highly dispersed. We obtain qualitatively similar results when the informed trading intensity is dependent on the cumulative size of the uninformed orders ($\sum_{\eta M+1 \ldots M} |\tilde{y}_{i,t}|$), instead of on the uninformed order imbalance.

Finally, we allow for heterogeneous information, or differences of opinion, in our model. Instead of all ηM informed investors making the same order based on the same information, we divide the informed investors into two groups that trade on independent private signals and therefore make independent orders. That is, the first $\frac{1}{2}\eta M$ informed traders each submit identical orders and the second $\frac{1}{2}\eta M$ informed traders each submit identical orders that are independent of the orders by the first group. The demands of each of the $(1-\eta)M$ uninformed liquidity seekers remain uncorrelated as in Section 2. Table 2 Panel D reports the average simulated VCV and $\widehat{\eta}$ from this model. Both the average VCV and $\widehat{\eta}$ are strictly increasing in η. VCV has clearly lower standard deviations than $\widehat{\eta}$, while the averages of $\widehat{\eta}$ again diverge from the true value of η.

Overall, the simulation results in this section demonstrate that VCV is highly robust as a measure of asymmetric information. The basic result that the coefficient of variation of

Table 2: Simulation results: Robustness

Panel A: Non-Gaussian demand distributions

η	0	0.2	0.5	0.8	1	0	0.2	0.5	0.8	1
			VCV					$\widehat{\eta}$		
			t-distribution					*t-distribution*		
Avg	0.04	0.32	0.64	0.86	0.97	0.03	0.28	0.78	1.23	1.60
s.d.	0.00	0.07	0.12	0.15	0.16	0.00	0.29	7.57	81.28	48.82
			Skew-Normal distribution					*Skew-Normal distribution*		
Avg	0.02	0.15	0.38	0.61	0.75	0.02	0.11	0.34	0.67	1.01
s.d.	0.00	0.01	0.03	0.04	0.06	0.00	0.01	0.03	0.08	0.15

Panel B: Random proportion of informed trade [*Informed investors* $\sim B(1/5, X)$]

X	0	1250	5000	20000	125000	0	1250	5000	20000	125000
$E[\eta]$	0	0.2	0.5	0.8	0.96	0	0.2	0.5	0.8	0.96
			VCV					$\widehat{\eta}$		
Avg	0.03	0.83	1.70	2.32	2.59	0.02	1.28	-13.8	-3.09	-2.56
s.d.	0.00	0.10	0.15	0.26	0.34	0.00	0.35	3861.71	0.80	0.55

Panel C: Endogenous informed trade

η	0	0.2	0.5	0.8	1	0	0.2	0.5	0.8	1
			VCV					$\widehat{\eta}$		
Avg	0.03	0.40	0.79	1.06	.	0.02	0.36	1.15	2.64	.
s.d.	0.00	0.05	0.09	0.12	.	0.00	0.07	0.37	29.27	.

Panel D: Heterogeneous information

η	0	0.2	0.5	0.8	1	0	0.2	0.5	0.8	1
			VCV					$\widehat{\eta}$		
Avg	0.03	0.17	0.36	0.49	0.57	0.02	0.13	0.31	0.49	0.60
s.d.	0.00	0.01	0.03	0.04	0.04	0.00	0.01	0.03	0.05	0.07

Notes: Average and standard deviation of VCV (left) and $\widehat{\eta}$ (right) obtained from $R = 1,000,000$ replicated samples of $T = 100$ volume realizations, simulated from various generalizations of the model outlined in Section 2, with $M = 1000$ liquidity seekers, for various values of the proportion of informed trading η. In Panel A, demand is t-distributed with 4 degrees of freedom (t_4), or Skew-Normally distributed with shape parameter 10, indicating positive skew ($SN(0, 1, 10)$). In Panel B, the number of uninformed liquidity seekers is kept constant at $1,000$, while the number of informed liquidity seekers is varying randomly across observations and follows a binomial distribution such that the number of active informed traders in each trading session is with probability $\frac{4}{5}$ equal to zero and with probability $\frac{1}{5}$ equal to X. The table reports the average VCV and $\widehat{\eta}$ for different values of X, which determines the average proportion of informed trade $E[\eta]$. In Panel C, the trading intensity of the ηM informed investors is increasing in the realized uninformed absolute net order flow, as in Eq.(10). In Panel D, the ηM informed traders are divided into two equal-sized groups that each make an independent order. Detailed simulation results and results based on additional model-variations are reported in Internet Appendix Tables A.3-6.

trading volume is monotonically increasing in the proportion of informed trade holds under rather general conditions and in small samples. Additional simulation results, including simulations with random variation in the number of market participants and their trading intensities across observations, are presented in Internet Appendix A. Also these results reaffirm the robustness of VCV. In the remainder of this paper, we therefore focus on VCV as our measure of informed trade, and investigate its properties using real empirical data.

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 trade, we now turn to the data to analyze the empirical properties of VCV. In this section, we describe the cross-sectional variation in VCV for US stocks, while we study the time-series behavior of VCV in the next section. We compute annual Volume Coefficients of Variation (VCV) for US stocks and compare these figures with other firm-level characteristics, including indicators of informed trade and illiquidity. We obtain daily trading volumes from the CRSP daily stock file for all common stocks listed on NYSE, AMEX and NASDAQ over the period 1980-2016. We disregard the most infrequently traded stocks by only including firm-year observations for stocks with positive trading volume in at least 200 days during that year.[8]

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 firm i in year τ is defined as:

$$VCV_{i,\tau} = \frac{\hat{\sigma}_{V(i,t\in\tau)}}{\hat{\mu}_{V(i,t\in\tau)}}, \tag{11}$$

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

where $\hat{\sigma}_{V(i,t\in\tau)}$ is the sample standard deviation of all daily trading volumes of firm i, $V_{i,t}$, in year τ and $\hat{\mu}_{V(i,t\in\tau)}$ is the sample average of of all daily trading volumes of firm i in year τ. We compute VCV using three different measures of trading volume: Trading volume in US dollars:

$$V_{USD,i,t} = shares\ traded_{i,t} \times closing\ price_{i,t}, \tag{12}$$

Volume *market shares*, 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:

$$V_{\%,i,t} = \frac{V_{USD,i,t}}{\sum_i V_{USD,i,t}}, \tag{13}$$

and daily *turnover*, to control for differences in market capitalization:

$$V_{TO,i,t} = \frac{shares\ traded_{i,t}}{shares\ outstanding_{i,t}}. \tag{14}$$

Table 3 reports summary statistics for these three measures of VCV. The sample averages, as well as other statistics, are highly similar for the three distinct VCV measures. The bottom rows of Table 3 show that the three different measures of VCV are highly correlated. The similarity between the three VCV measures offers support for the theoretical analysis in 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.(7)). 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 are virtually equivalent. The results in Table 3 support this premise. In the remainder of this paper, our measure of informed trading VCV is defined as the annual coefficient of variation of daily volume market shares ($\text{VCV}_{\%}$). Highly similar results are obtained when using any of the other volume definitions. In Internet Appendix Table B.1, we report these summary

Table 3: VCV Summary Statistics

	VCV_{USD}	$VCV_{\%}$	VCV_{TO}
Observations	129,199	129,199	129,199
N	15,628	15,628	15,628
T	37	37	37
Mean	1.295	1.275	1.248
s.d.	0.618	0.624	0.588
s.d. (CS)	0.584	0.586	0.559
s.d. (TS)	0.439	0.440	0.418
$q_{0.1}$	0.614	0.577	0.604
$q_{0.25}$	0.849	0.823	0.826
median	1.197	1.182	1.148
$q_{0.75}$	1.582	1.572	1.519
$q_{0.9}$	2.079	2.063	2.007
ρ	0.173	0.179	0.182
Correlations			
$VCV_{\%}$	0.980		
VCV_{TO}	0.965	0.953	

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 market shares (daily dollar volume as a percentage of total market dollar volume – $VCV_{\%}$), and turnover (dollar volume as a fraction of market capitalization – VCV_{TO}). The table reports the total number of observations, the number of distinct stocks in the sample (N), the number of time-series observations/years (T), mean, standard deviation, s.d. (CS), the time-series average of annual cross-sectional standard deviations, s.d. (TS), the cross-sectional average of stock-specific time-series standard deviations, selected quantiles (q), and the cross-sectional average of stock-specific first order autocorrelations (ρ). The bottom two rows report the time-series averages of within-year rank (Spearman) correlations between the different VCV measures. Sample: 1980-2016.

statistics for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX, and for subsamples of observations prior to 2000 (1980-1999) and post 2000 (2000-2016), showing that the three measures of VCV behave fairly similar across these subsamples.

4.1 VCV and other firm characteristics

Table 4 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, return

Table 4: VCV and other firm characteristics

	VCV	Size	BM	Age	Vol.	Turn.	Illiq	B-A	Roll
Size	−0.60								
BM ratio	0.15	−0.21							
Age	−0.27	0.32	0.13						
Volatility	0.36	−0.58	−0.05	−0.39					
Turnover	−0.28	0.33	−0.21	−0.02	0.18				
Illiquidity	0.65	−0.94	0.22	−0.32	0.54	−0.53			
Bid-Ask spread	0.58	−0.86	0.24	−0.23	0.58	−0.42	0.91		
Roll's measure	0.22	−0.34	0.19	−0.04	0.21	−0.30	0.41	0.51	
Coverage	−0.54	0.77	−0.24	0.16	−0.29	0.48	−0.80	−0.69	−0.29

Notes: This table reports the correlations between annual firm-level observations of VCV (obtained from daily volume market shares) and other annual firm-level characteristics. Each entry reports the time-series average of within-year rank (Spearman) correlations. *Size* is the log of market capitalization at the last trading day of June. *BM ratio* is the ratio of the book value to the market value of equity. *Age* is the number of years since the firm's first appearance in CRSP. *Volatility* is the annual standard deviation of daily returns. *Turnover* is the annual average of daily trading volume as a percentage of market capitalization. *Illiquidity* is the log of the annual average of the daily ratio $\frac{|R_{i,t}|}{V_{USD,i,t}}$ (Amihud, 2002). *Bid-Ask spread* is the annual average of daily bid-ask spreads $\frac{ask_{i,t}-bid_{i,t}}{price_{i,t}}$. *Roll's measure* is the square root of the negative of the daily return autocovariance $\sqrt{-Cov\left(R_{i,t},R_{i,t-1}\right)}$. *Coverage* refers to the number of distinct analysts covering a stock in a given year Sample: 1980-2016. Source: CRSP, COMPUSTAT, and IBES.

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

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

these variables and subsample analyses are provided in Internet Appendix Tables B.2 and B.3.

As can be seen from Table 4, VCV is negatively correlated to size and turnover and positively correlated to return volatility, Amihud illiquidity and the bid-ask spread. 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. The negative correlation with firm age suggest that information asymmetry is lower for more mature firms. Analyst coverage is likely to reduce information asymmetries, which is reaffirmed by the negative correlation with VCV. In Section 5, we study the impact of exogenous reductions in analyst coverage due to brokerage closures.

4.2 Return reversals

The correlation between VCV and the bid-ask spread reported in Table 4, is clearly higher than the correlation between VCV and Roll's (1984) estimate of the bid-ask spread. This result is expected, as it is well known from Huang and Stoll (1997) and others, that Roll's measure underestimates the bid-ask spread in the presence of information asymmetries, since price changes due to informed trading are likely to be predictive of future price changes, thereby pushing up return autocorrelations. To further evaluate the relationship between VCV and the bid-ask spread, we double-sort stocks within each year into quartiles based on the bid-ask spread and on Roll's measure. Table 5 shows the average VCV for each of these sixteen groups of firms. We find that VCV is monotonically increasing in the bid-ask spread but not in Roll's measure, which is consistent with the downward bias of Roll's measure in the presence of information asymmetry. Stocks with high information asymmetry are expected to have a relatively high bid-ask spread but a relatively low value of Roll's measure. We see from Table 5 that these stocks are precisely the stocks with a high VCV.

To further study the relation between VCV and return autocorrelation, we consider weekly reversals. It is well known that returns on individual stocks, in particular illiquid stocks, ex-

Table 5: VCV and the Bid-Ask spread

	Roll: Low	2	3	High	High-Low
Bid-Ask: Low	0.962	0.831	0.767	0.946	−0.016
2	1.144	1.097	1.093	0.990	*-0.154*
3	1.325	1.286	1.407	1.325	−0.000
High	1.730	1.619	1.672	1.783	*0.053*
High-Low	**0.768**	**0.788**	**0.905**	**0.837**	0.069

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

hibit significant short-term reversals (e.g. Jegadeesh, 1990). We compute weekly return autocorrelations for each firm within each year in our sample. We then double-sort stocks within each year into quartiles based on Amihud's (2002) Illiquidity and VCV. In Table 6, we report the average weekly return autocorrelation, for each of these 16 groups. Across all 16 groups, we find weekly return reversals (i.e. negative autocorrelation). These reversals are stronger for the more illiquid stocks. However, within each liquidity quartile, we find that reversals are decreasing in VCV. The final column of Table 6 shows that return autocorrelation is lower for High VCV stocks than for Low VCV stocks. This result implies, similar to Table 5, that short-term reversals are in general more profound for illiquid stocks, but these reversals are weaker when the illiquidity is associated with information asymmetry.

The results reported in this section are qualitatively similar across exchanges and in different time periods. Subsample analyses are reported in Internet Appendix Tables B.4 and B.5. These results are also consistent with existing research. Johnson and So (2017), Bongaerts et al. (2016), Hameed et al. (2008), and Llorente et al., (2002) use various measures to show that asymmetric information is associated with weaker short-term reversals. In the next subsection, we have a closer look at the empirical relation between VCV and existing measures of asymmetric information.

Table 6: VCV and weekly reversals

	Illiq: Low	2	3	High	High-Low
VCV: Low	−0.058	−0.065	−0.087	−0.102	**-0.044**
2	−0.047	−0.046	−0.065	−0.094	**-0.046**
3	−0.039	−0.039	−0.049	−0.084	**-0.045**
High	−0.040	−0.028	−0.040	−0.081	**-0.041**
High-Low	*0.018*	**0.037**	**0.047**	0.021	0.003

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

4.3 VCV and other measures of asymmetric information

In this subsection, we compare VCV with various incumbent measures of asymmetric information. These measures include the probability of informed trade (PIN), the Multimarket Information Asymmetry (MIA) measure, and C2. PIN is estimated by fitting a structural microstructure model to signed transaction data (Easley et al., 1996). MIA is based on relative trading volume in options and stocks, under the assumption that informed traders are more likely to trade in options (Johnson and So, 2017). C2 measures the relation between daily volume and return persistence, based on the premise that price changes due to informed trading are predictive of future price changes (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 (2017) and 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}).[10] We compute annual firm-level observations of MIA as

[10] Annual firm-level observations of PIN_{DY}, PIN_{EHO}, PIN_{BH} and PIN_{BHL} are made available by Jefferson Duarte (`http://www.owlnet.rice.edu/~jd10/`), Søren Hvidkjær (`https://sites.google.com/site/hvidkjaer/data`) and Stephen Brown (`http://scholar.rhsmith.umd.edu/sbrown/pin-data`), respectively. Daily firm-level observations of MIA are made available by Travis Johnson (`http://travislakejohnson.com/data.html`). Summary statistics of the measures employed in this section, as well as subsample analyses, are provided in Internet Appendix Tables B.6-B.8.

the annual average of the available daily observations for each firm. We derive annual firm-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 7: VCV and other information asymmetry measures

	VCV	PIN_{BHL}	PIN_{BH}	PIN_{EHO}	PIN_{DY}	Adj.PIN	PSOS	MIA
PIN_{BHL}	0.52							
PIN_{BH}	0.57	0.74						
PIN_{EHO}	0.48	0.61	0.65					
PIN_{DY}	0.52	0.64	0.68	0.85				
Adjusted PIN	0.48	0.57	0.70	0.61	0.71			
PSOS	0.43	0.46	0.43	0.62	0.70	0.38		
MIA	0.16	0.32	0.38	−0.01	0.11	0.19	0.004	
C2	0.10	0.13	0.11	0.02	0.02	0.03	0.04	0.03

Notes: This table reports the correlation between the annual firm-level coefficients of variation of daily volume market shares (VCV) and various annual firm-level information asymmetry measures. Each entry reports the time-series average of within-year rank (Spearman) correlations. PIN_{BHL} is estimated by Brown, Hillegeist and Lo (2004). PIN_{BH} is estimated by Brown and Hillegeist (2007). PIN_{EHO} is estimated by Easley, Hvidkjaer, and O'Hara (2010). PIN_{DY}, Adjusted PIN, and the illiquidity measure PSOS are estimated by Duarte and Young (2009). MIA is the annual average of firm-day level observations estimated by Jonhson and So (2017). C2 is estimated following Llorente et al. (2002). Sources: CRSP and cited authors' websites.

Table 7 shows the correlations between VCV and various annual firm-level information asymmetry measures. Our VCV measure is positively correlated to all PIN measures. 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. The correlations between VCV and the MIA and C2 measures are substantially lower, although still positive.

Duarte and Young (2009) argue that PIN does not only measure informed trading, but also other illiquidity effects. They therefore decompose PIN into *Adjusted PIN*, which is proposed as a cleaner measure of asymmetric information; and *PSOS* (probability of symmetric order-flow shock), which is a measure of illiquidity unrelated to asymmetric information. These additional variables are included in Table 7. Both Adjusted PIN are PSOS are positively correlated with VCV.

Table 8: VCV and Adjusted PIN

	VCV			
	(1)	(2)	(3)	(4)
PIN_{DY}	0.281^{***}		0.661^{***}	0.286^{***}
	(0.105)		(0.141)	(0.092)
Adjusted PIN	0.687^{***}	0.832^{***}		0.684^{***}
	(0.111)	(0.141)		(0.113)
PSOS	0.004	0.091^{**}	-0.125^{***}	
	(0.038)	(0.036)	(0.041)	
Observations	35,498	35,498	35,498	35,498
Adjusted R^2	0.365	0.365	0.362	0.365
Fixed effects	Yes	Yes	Yes	Yes

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

In Table 8, 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, Fama-French 48 industry fixed effects, and decile fixed effects for size, book-to-market and Amihud illiquidity deciles.[11] The regression results indicate that VCV is mostly associated with adjusted PIN, while VCV is not positively related to PSOS, thereby supporting our claim that VCV, like adjusted PIN, is indicative of asymmetric information rather than general illiquidity.

[11] 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.

4.4 VCV and institutional ownership

In this subsection, we study the relationship between VCV and various indicators of institutional ownership that we obtain from 13F filings recorded in the Spectrum database. Table 9 reports the results 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 reported in the Spectrum 13F database at the end of each year – Chen et al., 2002). Boone and White (2015) find that institutional ownership leads to an improvement in disclosure practices and therefore lower information asymmetry. The first column of Table 9 shows indeed that VCV has a significantly negative association with breadth of ownership. 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 to be a monitor for a certain firm if that firm belongs to the top 10% of holdings in the institution's portfolio. These monitoring investors are likely to be better informed about the firm than non-monitoring investors. Dedicated investors are those institutional investors that Bushee and Noe (2000) and Bushee (2001) classify as 'dedicated'. They are characterized by large, stable holdings in a small number of firms , as opposed to 'quasi-indexing' investors and 'transient' investors.[12]

The variable *Monitors* in Table 9 is the percentage of institutional investors in each firm that can be defined as monitoring investors. The variable *Dedicated* in Table 9 is the percentage of institutional investors in each firm that are classified as dedicated investors. Columns 2–4 of Table 9 show that these variables are both significantly positively associated

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

Table 9: VCV and institutional ownership

	VCV			
	(1)	(2)	(3)	(4)
Holdings	0.0001	−0.0001	0.0001	−0.0001
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Breadth	−1.558***	−1.890***	−1.518***	−1.819***
	(0.118)	(0.141)	(0.120)	(0.139)
Monitors		0.858***		0.815***
		(0.135)		(0.133)
Dedicated			0.270***	0.271***
			(0.067)	(0.065)
Observations	78,628	78,628	72,875	72,875
Adjusted R^2	0.455	0.456	0.451	0.453
Fixed effects	Yes	Yes	Yes	Yes

Notes: This table shows the results from regressing annual firm-level coefficients of variation of daily volume market shares (VCV) on various measures of institutional ownership. *Holdings* is the 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 fixed effects for each year, industry, size decile, book-to-market decile and illiquidity decile. Two-way clustered standard errors, clustered at the year and industry level, are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level. Sources: CRSP, 13F and the website of Brian Bushee `http://acct.wharton.upenn.edu/faculty/bushee/`

with VCV, consistent with our proposition that VCV measures informed trade.

The relationship between patterns in institutional ownership and VCV reported in Table 9 reaffirms that VCV is a measure of asymmetric information. Suppose that a firm is held by only a small number of institutional investors, and this investment makes up a relatively large fraction of the portfolio for several of those institutional shareholders (i.e. *Breadth* is low, while *Monitors* and *Dedicated* are expected to be 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 firm, information asymmetry should be a significant concern, as it is likely that the counter party 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 expected to be low), the risk of asymmetric information should be lower, which is in accordance with the results reported in Table 9. Summary statistics of the measures employed in this section, as well as subsample analyses, are provided in Internet Appendix Tables B.9-B.10.

5 VCV around information events

After having analyzed cross-sectional variation in VCV, we now study the behavior of VCV over time. The solid black line in Figure 3 shows the cross-sectional average of the firm-level VCVs (derived from volume market shares), for each year in our sample analyzed in Section 4. The gray bars show the cross-sectional average of the firm-level mean and standard deviation of volume market shares. The declining trend in VCV post-2000 suggests that asymmetric information has reduced over this period, which is consistent with recent studies (e.g. Duarte at al., 2008; Lambert et al., 2012; Horton et al., 2013) that find improved market transparency as the result of regulatory changes, such as the enactment by the SEC in 2000 of Regulation Fair Disclosure (Reg FD) and the adoption of International Financial Reporting Standards (IFRS). In Internet Appendix Figure B1, we reproduce this graph for firm-level VCVs computed from dollar volume and turnover, showing once again that VCV is highly similar for the three volume measures.

In the remainder of this section, we study the behavior of VCV around information events. First, we exploit a natural experiment to identify exogenous changes in information asymmetry: brokerage closures. Various recent studies (e.g. Kelly and Ljungqvist, 2012; Derrien and Kecskes, 2013) consider reductions in analyst coverage as a result of brokerage closures as exogenous shocks to the information environment of stocks. Consistent with the hypothesis that information asymmetry increases following such reductions in analyst coverage, we document an increase in VCV.

Next, we analyze VCV around quarterly earnings announcements and find that VCV is relatively high shortly before, and significantly lower after announcements.

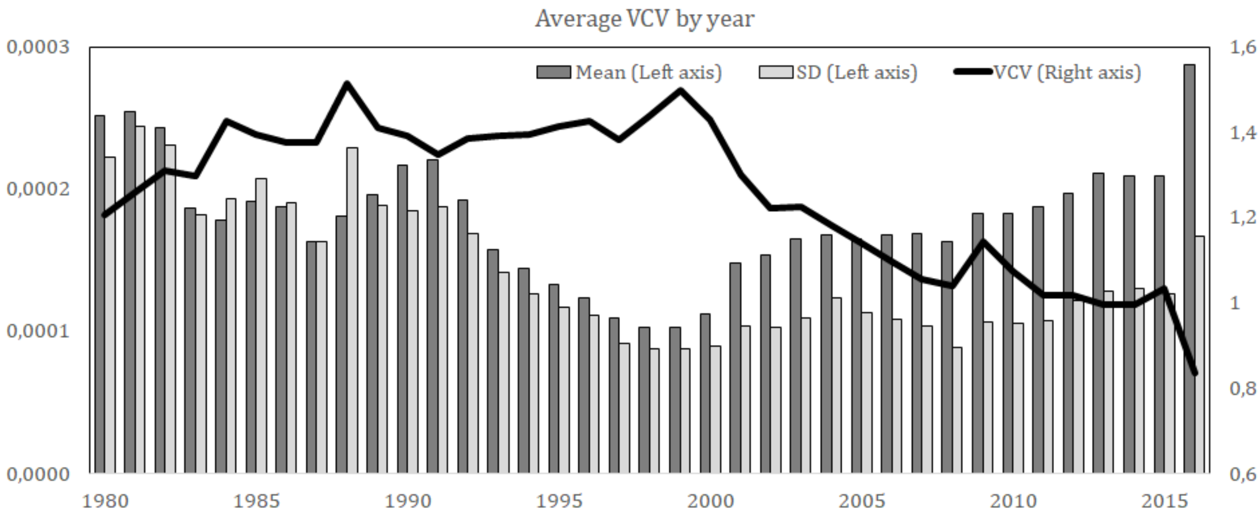

Figure 3: The black line shows the annual cross-sectional average of annual firm-level VCVs, calculated from volume market shares (Eq.(13)) over the period 1980-2016. The bars show the annual cross-sectional average of annual firm-level means and standard deviations of volume market shares.

5.1 VCV around brokerage closures

Kelly and Ljungqvist (2012) find that information asymmetry increases following terminations in analyst coverage that are caused by exogenous closures or acquisitions of brokerage firms. For the 22 brokerage closures between April 2000 and January 2008 listed in Appendix A of Kelly and Ljungqvist (2012), we identify in the IBES database a treatment sample of a total of 1,764 observations of firms that experience reductions in analyst coverage due to one of these closures.

We perform a simple difference-in-differences regression, to compare the VCV of treated firms (i.e. firms that experience closure-induced coverage terminations) to non-treated firms (the control group), before and after the brokerage closure. For each brokerage closure, our control group includes all non-treated firms in our sample analyzed in Section 4, for which analyst coverage in the calendar year prior to the brokerage closure is strictly positive. The

VCV before closure is defined as the coefficient of variation of daily volume market shares over a 12-month period before the closure, while the VCV after closure is calculated over a 12-month period after the closure. Following Derrien and Kecskes (2013), we impose a three-month gap 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 of the two dummy variables.

The results of the difference-in-differences regression are reported in the first column of Table 10. The coefficient of "After" is negative, which reflects that VCV is on average decreasing over time, as can be seen from Figure 3. The "Treated" coefficient indicates that there is a minor difference between the VCV of treated and control firms, prior to the event. Finally, the interaction term coefficient is significantly positive, meaning that the VCV of firms that face analyst reductions as a result of brokerage closures increases relative to the VCV of control firms that are not exposed to the brokerage closures.[13]

The second and third column of Table 10 show that this result is stronger when we restrict the sample to firms with low 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 a 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 10 provide strong evidence for our proposition that VCV measures information asymmetry.

[13] In Internet Appendix Tables B.11-B12, we report results for VCV computed over a 6-month period, and for a regression with a smaller control sample matched on firm size and analyst coverage, and find qualitatively similar results.

Table 10: Brokerage closures

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

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

5.2 VCV around earnings announcements

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

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

$$VCV_{XS,d} = \frac{\hat{\sigma}_{V(i,t=t_i+d)}}{\hat{\mu}_{V(i,t=t_i+d)}}, \tag{15}$$

where $\hat{\sigma}_{V(i,t=t_i+d))}$ is the sample standard deviation of all firms' daily trading volume on day d after the firm-specific announcement date t_i, and $\hat{\mu}_{V(i,t=t_i+d)}$ is the sample average of all firms' daily trading volume 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.[14]

This cross-sectional VCV is computed for all days d over the interval from -30 days before the announcement to +30 days after the announcement. The black line in Figure 4 shows the pattern of VCV over this interval, while the shaded areas indicate 95% confidence bounds, computed from the asymptotic distribution of sample coefficients of variation as derived by Albrecher et al. (2010). Figure 4 clearly shows that VCV is higher in the weeks prior to the announcement, which could be due to uninformed investors delaying their trading activity when 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. In internet Appendix Figure B.1, we reproduce Figure 4 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 pattern of VCV around earnings announcements is consistent with the behavior of alternative information asymmetry measures. Johnson and So (2017) document 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. Also Chordia et al. (2016) find

[14]The results are similar when $VCV_{XS,d}$ is computed from daily turnovers or dollar volumes.

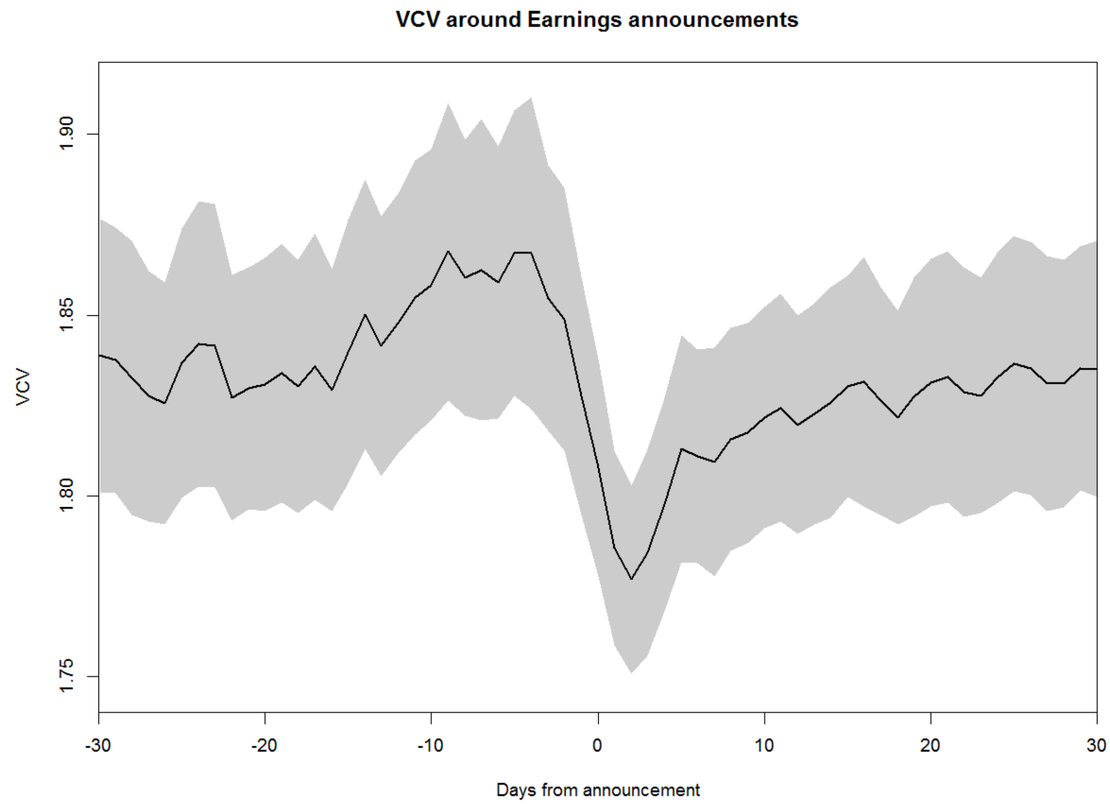

Figure 4: The black line shows the evolution of the daily cross-sectional VCV_{XS} around quarterly earnings announcements. The full sample includes all daily trading volumes over 61 days windows (day -30:30) around $N = 339,257$ quarterly announcements (sources: CRSP and COMPUSTAT). The reported VCV at d days after the announcement is estimated from the subsample of each stock's trading volume market shares at date d after each firm's announcement. The gray shaded areas indicate 95% confidence intervals: $VCV_{XS,d} \pm 1.96 \times S.E.(VCV_{XS,d})$. Standard errors ($S.E.$) are derived following Albrecher et al. (2010).

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) and Duarte et al. (2017) find that PIN is in fact lower prior to earnings announcements and higher afterwards. Duarte et al. (2017) explain this puzzling result by showing that the PIN measure mis-identifies asymmetric information when applied on a daily frequency, and instead simply indicates abnormal turnover. Easley et al. (2008), on the other hand, estimate a generalized PIN model in which the arrival rate of information is time-varying and find that PIN is high (low) before (after) earnings announcements, resembling the pattern in Figure 4.

6 Conclusion

In this paper, we use the Kyle (1985) model to demonstrate that the distribution of trading volume depends on the proportion of informed and thus correlated liquidity seeking demand. Specifically, we show that the Volume Coefficient of Variation (VCV) increases in the proportion on informed trade. We therefore propose 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 also tend to have other characteristics that are typically associated with asymmetric information (e.g.: high PIN, low breadth of institutional ownership, low analyst coverage, high illiquidity) and vice versa. Consistent with the hypothesis that informed trade is predictive of future price changes, we find that short-term return reversals are weaker for high VCV stocks, confirming that VCV is not just a measure of illiquidity. Our finding that VCV significantly increases following exogenous reductions in analyst overage due to brokerage closures, provides further evidence that VCV captures information asymmetry.

We introduce the cross-sectional VCV, which can be applied to evaluate information asymmetry in event time, e.g. following 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.

VCV is an appealing proxy for information asymmetry because of its simplicity: Computing VCV, by dividing the sample standard deviation of daily trading volumes over the sample mean, is very straightforward. Unlike alternative measures of information asymmetry, estimating VCV requires only total trading volumes, as opposed to intraday transaction-level data. The measure is therefore applicable to any security for which trading volume is observable and can be implemented both in cross-sections and in time-series. The potential applications of our measure are numerous. For example, VCV can be used as a control variable in empirical corporate finance research when there is need to control for information asymmetry, as a sorting characteristic in empirical asset pricing when studying the pricing effects of asymmetric information, or as the dependent variable of interest to compare information asymmetry across firms, countries, asset classes, or over time.

Appendix A: Variance of trading volume

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

$$V = \tfrac{1}{2}\left(\tilde{Y}_I + \tilde{Y}_U + \tilde{Y}_{MM}\right). \tag{16}$$

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

$$\begin{aligned} Var\,(2V) &= Var\left(\tilde{Y}_I\right) + Var\left(\tilde{Y}_U\right) + Var\left(\tilde{Y}_{MM}\right) \\ &\quad +2Cov\left(\tilde{Y}_I, \tilde{Y}_U\right) + 2Cov\left(\tilde{Y}_I, \tilde{Y}_{MM}\right) + 2Cov\left(\tilde{Y}_U, \tilde{Y}_{MM}\right) \end{aligned} \tag{17}$$

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

$$\begin{aligned} Var\left(\tilde{Y}_I\right) &= \eta^2 M^2 \sigma^2 \left(1 - \tfrac{2}{\pi}\right) \\ Var\left(\tilde{Y}_U\right) &= (1-\eta) M \sigma^2 \left(1 - \tfrac{2}{\pi}\right) \\ Var\left(\tilde{Y}_{MM}\right) &= \left(\eta^2 M^2 + (1-\eta) M\right) \sigma^2 \left(1 - \tfrac{2}{\pi}\right). \end{aligned} \tag{18}$$

$Cov\left(\tilde{Y}_I, \tilde{Y}_U\right) = 0$, because the demands of informed and uninformed liquidity seekers are independent. Moreover, when M is large and $\eta > 0$, the order imbalance consists mainly of orders submitted by informed liquidity seekers. The orders of uninformed traders tend to net out against each other because of the $i.i.d$ property. This implies that in the limit ($M \to \infty$), the liquidity suppliers trade exclusively to offset the imbalance from informed seekers. Therefore, $\lim_{M\to\infty} Cor\left(\tilde{Y}_U, \tilde{Y}_{MM}\right) = 0$ and $\lim_{M\to\infty} Cor\left(\tilde{Y}_I, \tilde{Y}_{MM}\right) = 1$. Given these correlations, Eq.(17) implies that when $M \to \infty$:

$$Var\left(\tfrac{2V}{M}\right) = Var\left(\tfrac{\tilde{Y}_I}{M}\right) + Var\left(\tfrac{\tilde{Y}_U}{M}\right) + Var\left(\tfrac{\tilde{Y}_{MM}}{M}\right) + 2\sqrt{Var\left(\tfrac{\tilde{Y}_I}{M}\right) Var\left(\tfrac{\tilde{Y}_{MM}}{M}\right)}, \tag{19}$$

which, given the variances in Eq.(18), results in:

$$\begin{aligned} Var\left(\tfrac{2V}{M}\right) &= \eta^2\sigma^2\left(1-\tfrac{2}{\pi}\right) + (1-\eta)M^{-1}\sigma^2\left(1-\tfrac{2}{\pi}\right) + \left(\eta^2+(1-\eta)M^{-1}\right)\sigma^2\left(1-\tfrac{2}{\pi}\right) \\ &\quad +2\sqrt{\eta^2\sigma^2\left(1-\tfrac{2}{\pi}\right)}\sqrt{\left(\eta^2+(1-\eta)M^{-1}\right)\sigma^2\left(1-\tfrac{2}{\pi}\right)} \\ &= 2\sigma^2\left(1-\tfrac{2}{\pi}\right)\left(\eta^2+(1-\eta)M^{-1}+\eta\sqrt{\eta^2+(1-\eta)M^{-1}}\right) \\ &= 4\sigma^2\left(1-\tfrac{2}{\pi}\right)\eta^2, \end{aligned} \tag{20}$$

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

$$\lim_{M\to\infty}\frac{s.d.(V)}{E\left[V\right]} = \lim_{M\to\infty}\frac{s.d.(V/M)}{E\left[V/M\right]} = \sqrt{2\pi-4}\frac{\eta}{\eta+1}. \tag{21}$$

References

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

Admati, A. 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.

Back, K., Crotty, K. and Li, T. (2017). Identifying information asymmetry in securities markets. *Review of Financial Studies*, forthcoming.

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

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

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

Black, F. (1986). Noise. *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 White, J. T. (2015). The effect of institutional ownership on firm transparency and information production. *Journal of Financial Economics*, 117 , 508-533.

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

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

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

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

Bushee, B. J., 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. *Journal of Finance*, 60, 413-442.

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

Chordia, T., Hu, J., Subrahmanyam, A., and Tong, Q. (2017). Order flow volatility and equity costs of capital. *Management Science*, forthcoming.

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

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

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

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

Duarte, J., Han, X., Harford, J., and Young, L. (2008). Information asymmetry, information dissemination and the effect of regulation FD on the cost of capital. *Journal of Financial Economics*, 87(1), 24-44.

Duarte, J., Hu, E. and Young, L. (2017). Does the PIN model mis-identify private information and if so, what are our alternatives? Working paper.

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

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

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

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

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

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

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

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.

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

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

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.

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

Horton, J., Serafeim, G., and Serafeim, I. (2013). Does mandatory IFRS adoption improve the information environment? *Contemporary Accounting Research*, 30(1), 388-423.

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

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of*

Finance, 45(3), 881-898.

Johnson, T. L., and So, E. C. (2017). A Simple Multimarket Measure of Information Asymmetry. *Management Science*, forthcoming.

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.

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

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.

Lambert, R. A., Leuz, C., and Verrecchia, R. E. (2011). Information asymmetry, information precision, and the cost of capital. *Review of Finance*, 16(1), 1-29.

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

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

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

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

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

Subrahmanyam, A. (1991). Risk aversion, market liquidity, and the price efficiency. *Review of Financial Studies*, 4, 417-441.

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.

Asymmetric Information and the Distribution of Trading Volume

Internet Appendix

Matthijs Lof Jos van Bommel

January 17, 2018

A Detailed simulation results

In this section, we report detailed and supplementary results for the Monte Carlo simulation exercise in Section 3 of the paper. Tables A.1 and A.2 provide detailed results of the simulation of the benchmark model (Table 1 in the paper). Table A.3 provides detailed results of the simulations with non-Gaussian demand distributions (Table 2 Panel A), as well as results of simulating a Uniform demand distribution. Table A.4 provides detailed results of the simulations with Binomially distributed proportion of informed trade (Table 2 Panel B), as well as results of simulating a Uniformly distributed proportion of informed trade. Table A.5 provides detailed results of the simulations with endogenous informed demand (Table 2 Panel C). Table A.6 provides detailed results of the simulations with heterogeneous information among the informed investors (Table 2 Panel D). Finally, Table A.7 and A.8 provide additional simulations in which the number of market participants M and their trading intensities σ are varying (following a Uniform distribution) across observations.

Table A.1: Simulation results: Benchmark model

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

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

Table A.2: Simulation results: Small samples ($T = 10$)

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

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

Table A.3: Simulation results: Non-Gaussian demand distributions

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

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

Table A.4: Simulation results: Random proportion of informed trade

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

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

Table A.5: Simulation results: Endogenous informed trading

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

Notes: This table reports the sample average, standard deviation, and selected percentiles of VCV (Panel A) and $\widehat{\eta}$ (Panel B), obtained from $R = 1,000,000$ replications of $T = 100$ volume observations, simulated from a model with $M = 1000$ liquidity seekers. Different from Table A1, the ηM informed traders orders are normally distributed with a standard deviation that is in each trading session proportional to the absolute net order flow of the uninformed investors as in Eq. (10) of the paper.

Table A.6: Simulation results: Heterogeneous information

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

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

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

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

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

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

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

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

B Supplementary summary statistics and empirical results

In this section, we report summary statistics, subsample analyses, and additional empirical results to supplement the empirical results presented in Sections 4 and 5 of the paper. Tables B.1, B.3, B.4. B.5, B.7, B.8, B.10, and Figure B.2 provide subsample analyses of the results reported in the paper. These subsamples include *(i)* NYSE/AMEX listed stocks, *(ii)* NASDAQ-listed stocks, *(iii)* observations prior to 2000 (1980-1999), and *(iv)* observations after 2000 (2000-2016). Tables B.2, B.6 and B.9 provide summary statistics on several firm-level variables used in the paper. Tables B.11 and B.12 report results from the difference in differences regression around brokerage closures (Table 10 in the paper), using VCV computed over a period of 6 months, and using a smaller control group. Figure B.1 plots average VCV over time (as Figure 3 in the paper), using volumes in US dollars and turnover, instead of volume market shares.

Table B.1: VCV Summary Statistics

	NASDAQ			NYSE/AMEX		
	VCV_{USD}	$VCV_{\%}$	VCV_{TO}	VCV_{USD}	$VCV_{\%}$	VCV_{TO}
Observations	73,304	73,304	73,304	55,895	55,895	55,895
N	11,314	11,314	11,314	5,793	5,793	5,793
T	34	34	34	37	37	37
Mean	1.43	1.41	1.37	1.12	1.10	1.09
s.d.	0.63	0.63	0.60	0.56	0.57	0.54
s.d.(CS)	0.60	0.60	0.58	0.52	0.53	0.50
s.d.(TS)	0.46	0.45	0.43	0.39	0.39	0.37
$q_{0.1}$	0.74	0.70	0.71	0.54	0.50	0.54
$q_{0.25}$	1.00	0.98	0.96	0.70	0.68	0.69
Median	1.32	1.31	1.26	1.01	1.00	0.99
$q_{0.75}$	1.71	1.70	1.64	1.38	1.37	1.34
$q_{0.9}$	2.23	2.20	2.15	1.82	1.82	1.77
ρ	0.13	0.13	0.14	0.20	0.22	0.21
Correlations						
$VCV_{\%}$	0.98			0.98		
VCV_{TO}	0.95	0.94		0.97	0.96	

	Pre-2000			Post-2000		
	VCV_{USD}	$VCV_{\%}$	VCV_{TO}	VCV_{USD}	$VCV_{\%}$	VCV_{TO}
Observations	71,146	71,146	71,146	58,053	58,053	58,053
N	11,978	11,978	11,978	8,111	8,111	8,111
T	20	20	20	17	17	17
Mean	1.40	1.40	1.35	1.16	1.13	1.12
s.d.	0.55	0.55	0.52	0.67	0.67	0.63
s.d.(CS)	0.52	0.53	0.50	0.66	0.65	0.62
s.d.(TS)	0.40	0.40	0.38	0.43	0.43	0.41
$q_{0.1}$	0.82	0.80	0.80	0.52	0.48	0.52
$q_{0.25}$	1.04	1.02	1.00	0.67	0.63	0.65
Median	1.31	1.30	1.25	0.98	0.95	0.94
$q_{0.75}$	1.65	1.65	1.59	1.46	1.43	1.39
$q_{0.9}$	2.09	2.10	2.04	2.05	2.01	1.95
ρ	0.09	0.09	0.09	0.10	0.11	0.11
Correlations						
$VCV_{\%}$	0.98			0.99		
VCV_{TO}	0.96	0.95		0.97	0.96	

Notes: This table reproduces the results in Table 3 of the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX stocks and for subsamples of observations prior to 2000 (1980-1999) and post 2000 (2000-2016).

Table B.2: Summary statistics of firm-characteristics

	Size	BM	Age	Vol.	TO	Illiq	Bid-Ask	Roll	Cov.
Observations	129,199	129,199	129,199	129,199	129,199	129,199	106,051	129,199	66,005
N	15,628	15,628	15,628	15,628	15,628	15,628	13,421	15,628	9,770
T	37	37	37	37	37	37	34	37	26
Mean	12.34	0.67	15.10	3.41	0.48	−16.46	2.22	0.87	2.02
s.d.	1.90	0.55	14.74	2.01	0.68	3.22	2.71	2.76	0.79
s.d.(CS)	1.76	0.48	14.59	1.72	0.52	2.85	2.01	2.46	0.78
s.d.(TS)	0.65	0.34	3.69	1.16	0.27	1.29	1.33	2.15	0.33
$q_{0.1}$	9.93	0.18	2	1.48	0.08	−20.83	0.08	−2.09	0.69
$q_{0.25}$	10.93	0.32	4	2.00	0.14	−18.93	0.27	−1.05	1.39
Median	12.25	0.55	11	2.90	0.28	−16.38	1.36	0.83	2.08
$q_{0.75}$	13.66	0.85	21	4.26	0.58	−13.93	3.12	2.22	2.64
$q_{0.9}$	14.90	1.26	34	5.99	1.09	−12.21	5.53	4.15	3.09

Notes: This table reports summary statistics on the variables used in Table 4 of the paper. Size, Illiquidity and Coverage are measured in logs.

Table B.3: VCV and other firm characteristics

	NASDAQ VCV	NYSE/AMEX VCV	Pre-2000 VCV	Post-2000 VCV
Size	−0.50	−0.62	−0.49	−0.74
BM ratio	0.18	0.17	0.14	0.16
Age	−0.13	−0.27	−0.24	−0.31
Volatility	0.26	0.33	0.27	0.46
Turnover	−0.27	−0.23	−0.24	−0.33
Illiquidity	0.59	0.65	0.58	0.74
Bid-Ask sPread	0.57	0.59	0.47	0.70
Roll's measure	0.25	0.06	0.25	0.19
Coverage	−0.47	−0.52	−0.56	−0.53

Notes: This table reproduces the correlations between VCV and other firm characteristics, reported in Table 4 of the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX stocks and for subsamples of observations prior to 2000 (1980-1999) and Post 2000 (2000-2016).

Table B.4: VCV and the Bid-Ask Spread

	Roll: Low	2	3	High	High-Low
NASDAQ					
Bid-Ask: Low	1.034	0.990	0.904	1.031	−0.003
2	1.217	1.227	1.223	1.098	*-0.119*
3	1.392	1.343	1.469	1.391	−0.001
High	1.799	1.666	1.721	1.799	0.000
High-Low	**0.765**	**0.677**	**0.817**	**0.768**	0.003
NYSE/AMEX					
Bid-Ask: Low	0.871	0.743	0.621	0.695	−0.176
2	1.041	0.958	0.831	0.758	**-0.283**
3	1.217	1.216	1.106	0.966	*-0.251*
High	1.624	1.563	1.545	1.665	0.040
High-Low	**0.753**	**0.821**	**0.924**	**0.969**	0.003
Pre-2000					
Bid-Ask: Low	1.094	1.059	1.033	1.323	0.229
2	1.338	1.360	1.433	1.621	*0.283*
3	1.520	1.461	1.601	1.634	*0.114*
High	1.720	1.559	1.648	1.738	0.017
High-Low	**0.626**	*0.500*	**0.614**	**0.614**	-0.212
Post-2000					
Bid-Ask: Low	0.734	0.620	0.656	0.832	0.099
2	0.905	0.799	0.796	0.868	−0.038
3	1.204	1.148	1.082	1.099	*-0.105*
High	1.727	1.638	1.688	1.825	*0.098*
High-Low	**0.993**	*1.018*	**1.032**	**0.993**	-0.000

Notes: This table reproduces the results in Table 5 of the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX stocks and for subsamples of observations prior to 2000 (1980-1999) and Post 2000 (2000-2016).

Table B.5: VCV and weekly reversals

	Illiq: Low	2	3	High	High-Low
NASDAQ					
VCV: Low	−0.063	−0.071	−0.092	−0.095	−0.032
2	−0.051	−0.051	−0.072	−0.094	**-0.043**
3	−0.031	−0.045	−0.052	−0.087	**-0.055**
High	−0.035	−0.032	−0.044	−0.083	*-0.048*
High-Low	0.028	**0.039**	**0.048**	0.012	-0.016
NYSE/AMEX					
VCV: Low	−0.057	−0.060	−0.079	−0.121	*-0.064*
2	−0.045	−0.040	−0.049	−0.092	**-0.046**
3	−0.043	−0.033	−0.043	−0.075	**-0.032**
High	−0.042	−0.025	−0.031	−0.073	*-0.031*
High-Low	0.015	**0.036**	**0.048**	0.048	0.033
Pre-2000					
VCV: Low	−0.055	−0.060	−0.078	−0.101	**-0.047**
2	−0.038	−0.043	−0.059	−0.091	**-0.053**
3	−0.033	−0.035	−0.047	−0.083	**-0.050**
High	−0.040	−0.024	−0.035	−0.077	**-0.037**
High-Low	0.015	**0.036**	**0.043**	0.024	0.010
Post-2000					
VCV: Low	−0.062	−0.068	−0.098	−0.119	−0.057
2	−0.055	−0.048	−0.069	−0.097	**-0.042**
3	−0.045	−0.045	−0.049	−0.082	**-0.037**
High	−0.041	−0.039	−0.046	−0.083	*-0.043*
High-Low	0.021	**0.029**	**0.029**	0.035	0.014

Notes: This table reproduces the results in Table 6 of the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX, and for subsamples of observations prior to 2000 (1980-1999) and post 2000 (2000-2016).

Table B.6: Summary statistics of information asymmetry measures

	PIN_{BHL}	PIN_{BH}	PIN_{EHO}	PIN_{DY}	Adj.PIN	PSOS	MIA	C2
Observations	72,290	72,295	30,529	35,498	35,498	35,498	19,611	120,706
N	11,103	11,119	4,319	4,585	4,585	4,585	3,452	15,118
T	18	18	19	22	22	22	20	33
Mean	0.19	0.21	0.20	0.21	0.18	0.28	0.42	0.02
s.d.	0.10	0.10	0.06	0.09	0.07	0.15	0.10	0.10
s.d.(CS)	0.10	0.10	0.06	0.09	0.07	0.15	0.10	0.10
s.d.(TS)	0.07	0.06	0.04	0.06	0.05	0.10	0.07	0.08
$q_{0.1}$	0.05	0.10	0.12	0.11	0.10	0.14	0.29	−0.11
$q_{0.25}$	0.11	0.13	0.15	0.14	0.13	0.17	0.36	−0.04
Median	0.20	0.19	0.19	0.19	0.16	0.23	0.42	0.02
$q_{0.75}$	0.26	0.27	0.24	0.24	0.21	0.33	0.48	0.08
$q_{0.9}$	0.31	0.36	0.28	0.33	0.28	0.50	0.53	0.13

Notes: This table reports summary statistics on the information asymmetry measures used in Table 7 of the paper.

Table B.7: VCV and other information asymmetry measures

	NASDAQ VCV	NYSE/AMEX VCV	Pre-2000 VCV	Post-2000 VCV
PIN_{BHL}	0.43	0.49	0.38	0.61
PIN_{BH}	0.49	0.61	0.42	0.67
PIN_{EHO}	0.36	0.48	0.46	0.68
PIN_{DY}	0.35	0.52	0.46	0.70
Adjusted PIN	0.37	0.47	0.42	0.67
PSOS	0.26	0.43	0.39	0.54
MIA	0.18	0.20	0.01	0.21
C2	0.06	0.06	0.14	0.06

Notes: This table reproduces the correlations between VCV and other information asymmetry measures, reported in Table 7 of the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX, and for subsamples of observations prior to 2000 (1980-1999) and post 2000 (2000-2016).

Table B.8: VCV and Adjusted PIN

	NASDAQ VCV	NYSE/AMEX VCV	Pre-2000 VCV	Post-2000 VCV
PIN_{DY}	0.200	0.290**	0.154*	0.948***
	(0.272)	(0.121)	(0.089)	(0.265)
Adjusted PIN	0.757***	0.682***	0.632***	0.825***
	(0.285)	(0.112)	(0.118)	(0.203)
PSOS	−0.202	0.019	0.025	0.034
	(0.141)	(0.037)	(0.044)	(0.069)
Observations	1,693	33,805	27,832	7,666
Adjusted R^2	0.350	0.364	0.325	0.487
fixed effects	Yes	Yes	Yes	Yes

Notes: This table reproduces the main regression in Table 8 of the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX, and for subsamples of observations prior to 2000 (1980-1999) and post 2000 (2000-2016).

Table B.9: Summary statistics of institutional ownership characteristics

	Holdings	Breadth	Monitors	Dedicated
Observations	78,628	78,628	78,628	72,875
N	11,553	11,553	11,553	10,957
T	26	26	26	24
Mean	47.15	4.56	3.15	6.78
s.d.	29.61	5.34	4.83	9.93
s.d.(CS)	27.20	5.34	4.79	5.34
s.d.(TS)	11.68	1.11	1.93	5.90
$q_{0.1}$	7.66	0.45	0	0
$q_{0.25}$	21.79	1.06	0	0
Median	46.65	2.83	1.39	1.92
$q_{0.75}$	70.94	5.80	4.50	11.11
$q_{0.9}$	86.20	11.03	8.74	20

Notes: This table reports summary statistics on the measures of institutional ownership used in Table 9 of the paper.

Table B.10: VCV and institutional ownership

	NASDAQ	NYSE/AMEX	Pre-2000	Post-2000
	VCV	VCV	VCV	VCV
Holdings	0.0003	−0.0002	0.002***	−0.0004
	(0.001)	(0.0003)	(0.001)	(0.0004)
Breadth	−2.756***	−1.791***	−1.547***	−2.026***
	(0.475)	(0.125)	(0.122)	(0.192)
Monitors	0.864***	0.863***	0.218**	1.162***
	(0.132)	(0.176)	(0.088)	(0.126)
Dedicated	0.249***	0.491***	0.199***	0.695***
	(0.063)	(0.135)	(0.031)	(0.165)
Observations	42,835	30,040	31,320	41,555
Adjusted R^2	0.380	0.483	0.371	0.505
fixed effects	Yes	Yes	Yes	Yes

Notes: This table reproduces the main regression in Table 9 of the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX, and for subsamples of observations prior to 2000 (1980-1999) and post 2000 (2000-2016).

Table B.11: Brokerage closures: 6 month VCV

	Full sample	Coverage < 10	Coverage < 5
	VCV	VCV	VCV
After	−0.037***	−0.047***	−0.053***
	(0.009)	(0.005)	(0.007)
Treated	−0.003	−0.014	−0.025
	(0.008)	(0.012)	(0.028)
After × Treated	0.011*	0.036***	0.114***
	(0.006)	(0.011)	(0.019)
Observations	66,850	46,952	27,760
Adjusted R^2	0.401	0.395	0.434
Fixed effects	Yes	Yes	Yes

Notes: This table reproduces the results in Table 10 of the paper, with the difference that the VCV is measured over a period of 6 months. That is, the VCV before closure is measured over months -8:-3, while the VCV after closure is measured over the period +3:+8, with the brokerage closure occuring in month 0.

Table B.12: Brokerage closures: Matched control group

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

Notes: This table reproduces the results in Table 10 of the paper, with the difference that the control group contains a selective set of matched observations. The treatment sample consist of 1,764 stocks. For each firm in the treatment group, we select three control matched by firm size and analyst coverage in the calendar year prior to the event. For all 7,056 stocks, we compute VCV over the months -14:-3, and over the months 3:14, with the brokerage closure occurring in month 0.

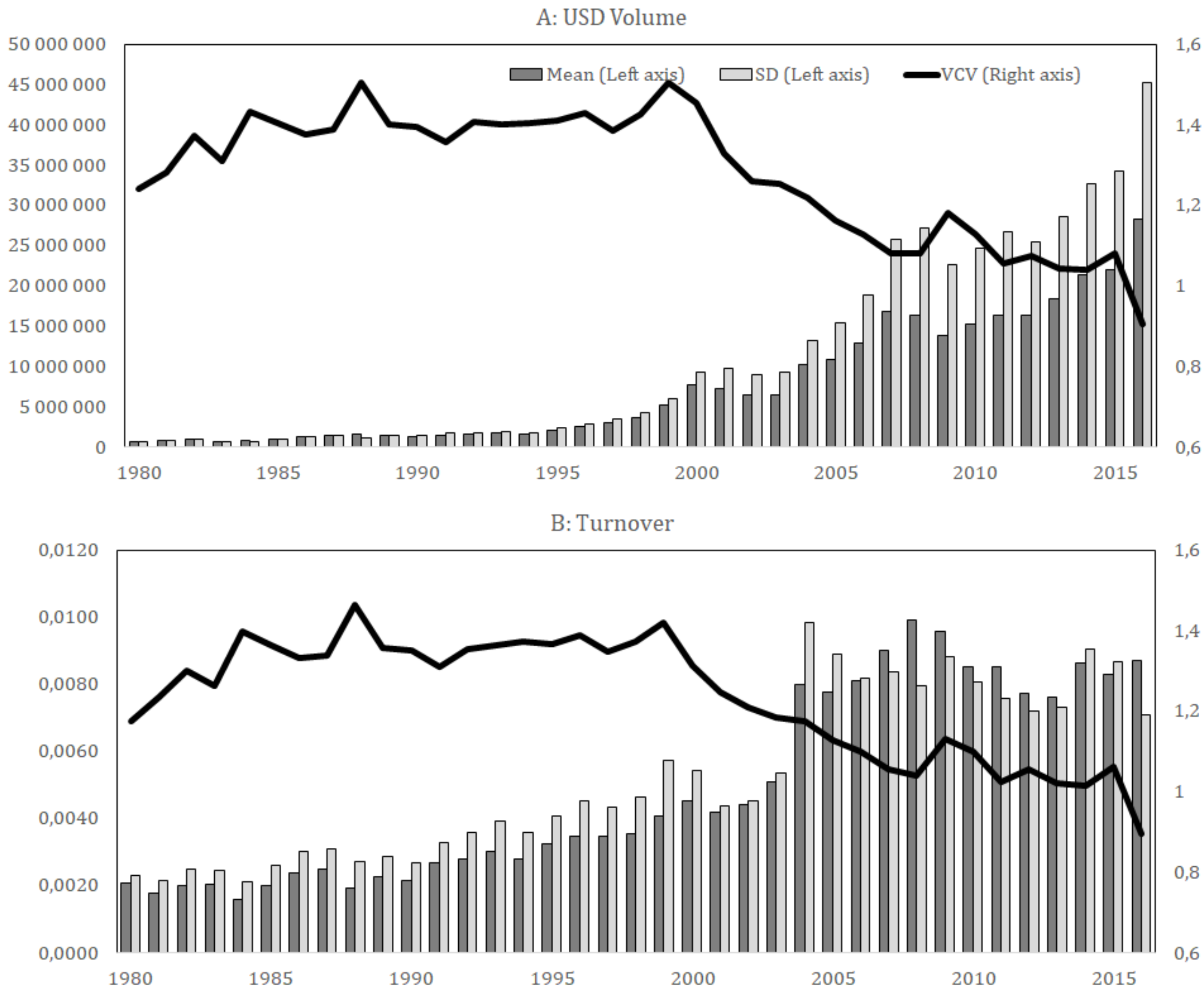

Figure B.1: This figure reproduces Figure 3 in the paper, with the exception that the annual cross-sectional average of firm-level estimates of the mean, standard deviation, and coefficient of variation (VCV) of volume are not based on volume market shares, but on US dollar volume (panel A) and turnover (panel B). Although the means and standard deviations (gray bars) are very different for the three volume measures, VCV (black line) is highly similar.

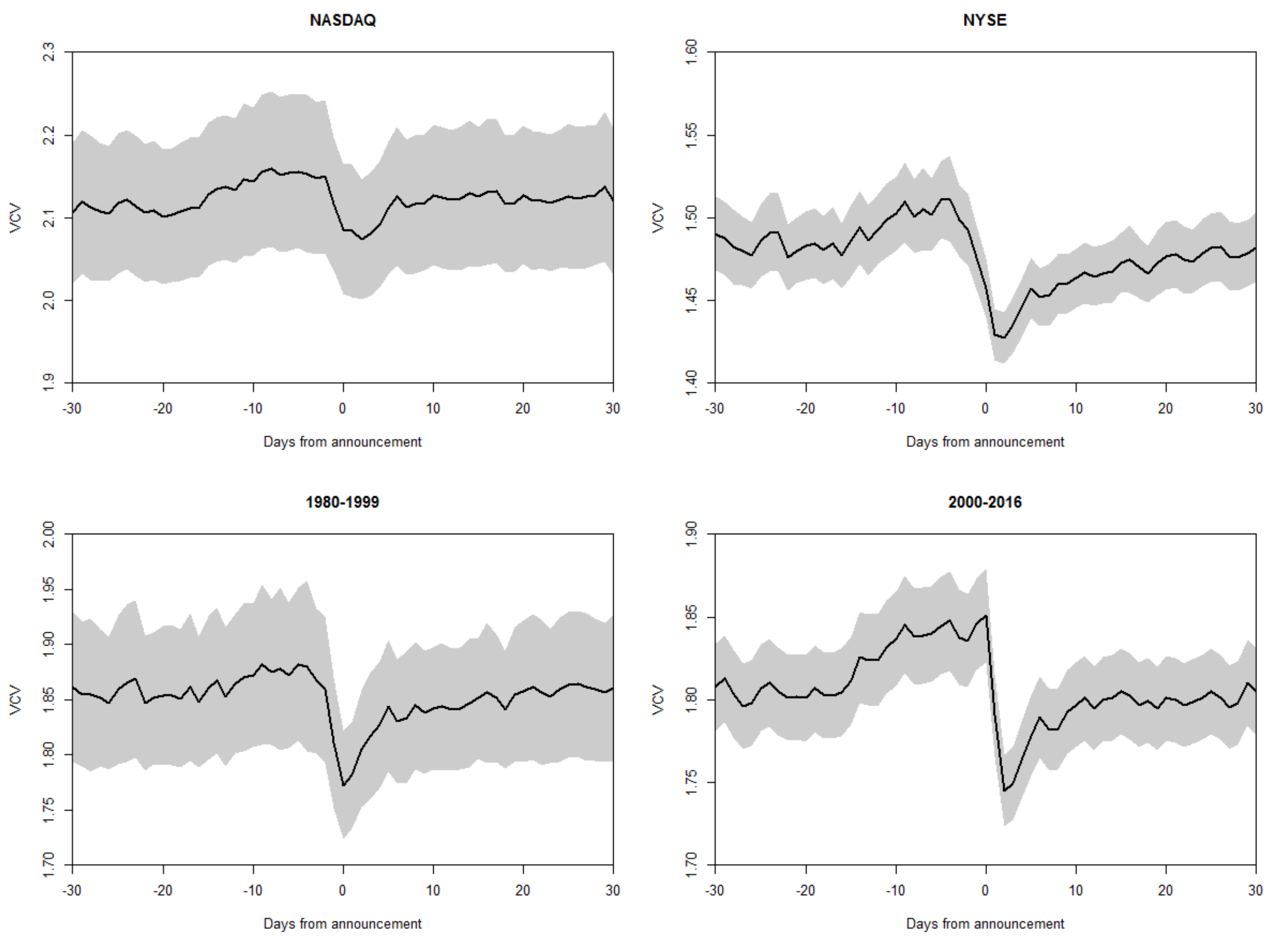

Figure B.2: This figure reproduces Figure 4 in the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX, and for subsamples of observations prior to 2000 (1980-1999) and post 2000 (2000-2016).

Bank of Finland Research Discussion Papers 2018

ISSN 1456-6184, online

1/2018 Matthijs Lof – Jos van Bommel
Asymmetric information and the distribution of trading volume
ISBN 978-952-323-206-8, online