Asymmetric Information and the Distribution of Trading Volume *

Matthijs Lof† Jos van Bommel‡

May 29, 2019

Abstract

We propose the Volume Coefficient of Variation (VCV), the ratio of the standard deviation to the mean of trading volume, as a new and easily computable measure of information asymmetry in security markets. We use a microstructure model to demonstrate that VCV is strictly increasing in the proportion of informed trade. Empirically, we find that 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, Darwin Choi, Tarun Chordia, 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, David Reiffen, Dagfinn Rime, Francesco Sangiorgi, Andriy Shkilko, Valeri Sokolovski, Enrique Sentana, Erik Theissen, Ian Tonks, Hugo van Buggenum, Michael Weber, Christian Westheide and participants at the ABFER, CEPR and CUHK Annual Symposium (Hong Kong), the American Economic Association (Chicago), the Eastern Finance Association (Philadelphia), the EBC network, the French Finance Association, the 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, Hong Kong Baptist University, University of Amsterdam, University of Liverpool, University of Maastricht, University of Toulouse, and University of Turku for constructive comments.

†Aalto University School of Business. E-mail: matthijs.lof@aalto.fi
‡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, 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 highly correlated. Uninformed orders are therefore mostly matched to each other, while informed orders generate order imbalances that trigger market maker intervention and thus additional trading volume. 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 and intuitive measure of information asymmetry. In addition to the analytical results, we conduct an extensive simulation exercise and provide ample empirical evidence in support of our main main hypothesis that VCV increases in the proportion of informed trade. VCV is easy to compute and requires only observations of trading volume, as opposed to quotes, prices, or signed order flow. VCV can be computed from time-series, to investigate cross-sectional differences in information asymmetry, or cross-sections of trading volumes, to analyze information asymmetry over (event-) time.[1]

The intuition behind our measure is that the distribution of trading volume depends on the correlation of individual orders. When liquidity seekers are uninformed and have uncorrelated liquidity needs, most orders will be netted out against each other, so that the order imbalance is relatively low compared to the observed trading volume. In this case, trading volume follows a Normal-like distribution. If more liquidity seekers become informed, order

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

imbalances become more prominent, leading to a more dispersed and more skewed distribution of recorded trading volume. We show that the coefficient of variation adequately captures the mapping from the proportion of informed and hence correlated trade to the volume distribution.

Our analytical expression of VCV in terms of the proportion of informed trade holds regardless of the mechanism by which the proportion of informed trading is determined. To illustrate this, we consider a model in which multiple informed traders choose their demand strategically, so that the proportion of informed trade is endogenously determined in equilibrium, and find a positive and monotonic relation between VCV and the proportion of informed trade. This analysis also demonstrates that the relation between VCV and price impact is weak and not monotonic. The reason is that when the number of informed traders increases, they each trade less aggressively to mitigate price impact. Also in our empirical analysis, we find that VCV is clearly distinct from measures of price impact or general illiquidity, such as Amihud (2002) illiquidity and Roll's (1984) measure.

While our theory is based on the Kyle (1985) auction setting, the intuition that correlated orders will lead to larger trading volumes than uncorrelated orders, and thus to a more dispersed distribution, also pertains in continuous markets: liquidity providers will accumulate more inventory (positive or negative) when confronted with correlated liquidity seeking demand. As they find themselves on the wrong side of the market, they will increase their efforts to unwind these inventories, resulting in excessive sequential 'hot potato' transactions and 'intermediation chains' (see e.g. Jovanovic and Menkveld, 2016; Glode and Opp, 2016; Bayraktar and Munk, 2017; and Rosu, 2019).[2]

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 oth-

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

ers, 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.[3] Both PIN and VCV are expected to increase in the order imbalance generated by correlated 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 (e.g. Boehmer et al., 2007; Easley et al., 2012; Johnson and So, 2018). Computing VCV does not require signed transaction-level data. Instead, VCV is estimated from volume data only, either from time series or cross sections.

Duarte and Young (2009) argue that (unadjusted) PIN is not only measuring informed trade, but also general illiquidity unrelated to information asymmetry.[4] 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

[3]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 (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), earnings surprises (Brown et al., 2009), corporate investments and M&A (Chen et al., 2007; Aktas et al., 2007), ownership structure (Brockman and Yan, 2009), and the January effect (Kang, 2010).

[4]Other papers in the debate on the validity of PIN include Easley et al. (2010), Akay et al. (2012), Back et al. (2018), and Duarte et al. (2018). 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 order flows.

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.

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 assertion that returns generated by informed trade are likely to be persistent, while uninformed trade leads to return reversals. In a recent paper, Johnson and So (2018) 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. We find that our VCV measure is positively correlated with both the MIA measure of Johnson and So (2018) 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 the institution's portfolio (Fich et al., 2015), and *dedicated* investors, defined as institutional investors that predominantly make long-term investments in a selective set of stocks (Bushee and Noe, 2000; Bushee, 2001). We find that, controlling for breadth of ownership, VCV is higher for firms with monitoring and/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 prices and order flows, we analyze total trading 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 intensity, 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 VCV 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 total trading 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. 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. To verify the robustness of VCV as a measure of information asymmetry, we consider various modifications of our benchmark model, allowing for random variation in the proportion of informed trade, stochastic liquidity, endogenous proportion of informed trade, and heterogeneous information. For all these model specifications, our simulation results show a strictly positive relation between VCV and the realized proportion of informed trade, confirming that VCV detects information asymmetry under very general conditions.

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 dollars, *(ii)* turnover, and *(iii)* volume market shares (dollar-volume as a fraction of total market dollar-volume). These three measures of VCV turn out to be virtually identical, implying that VCV is not sensitive to aggregate market-level variation in trading volume. This is important, since it is well known that other factors besides idiosyncratic firm-level information can drive variation in trading activity, such as sentiment (Kumar and Lee, 2006), or common liquidity shocks (Admati and Pfleiderer, 1988; Brogaard et al., 2018).

We find 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 (Duarte and Young, 2009), and with patterns in institutional ownership. As further evidence of VCV measuring informed trading, we show that, controlling for Amihud (2002) illiquidity, return reversals are weaker for high-VCV stocks, consistent with informed trading being predictive of future price changes.

Section 5 documents patterns of VCV around information events. First, we exploit exogenous terminations in analyst coverage due to brokerage closures, similar to Kelly and Ljungqvist (2012), Derrien and Kecskes (2013), Li and You (2015), and Chen and Lin (2017). 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.

Finally, we analyze VCV in event-time computed from cross sections of trading volumes around earnings announcements. It has been widely recognized that information asymmetries are resolved around these events. We find that VCV is relatively high prior to announcements, and drops significantly in the days following the announcement. This suggests that information asymmetries build up and discourage uninformed traders to trade just before earnings announcements (See Milgrom and Stokey, 1982; Black, 1986; Wang, 1994; and Chae, 2005), while the market is more attractive for uninformed traders after these information events.

2 Theory

To analyze the distribution of trading volume, we first present a simple model in which we postulate M individual liquidity seekers, who each submit Normally distributed orders with mean zero and standard deviation σ, and where competitive liquidity providers (market makers) absorb the order imbalance. Proportion η of the M liquidity seekers is informed, with ηM being an integer. We refer to σ as *trading intensity*. The assumption that informed and uninformed traders have equal trading intensity is for convenience only and without loss of generality: the trading volume distribution would be identical if the there were be $k\eta M$ informed traders who each trade with intensity $\frac{\sigma}{k}$. For now, we assume η to be exogenous. In the next subsection, we endogenize η by allowing informed investors to choose their trading intensity strategically.

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

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

The term inside brackets is the "double-counted transaction volume", counting both buys and sells, of (*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.[5]

As an example, consider five liquidity seekers whose demands are -1, 2, 2, -2, 1. The order imbalance is two, meaning that the liquidity providers end up selling two units. The observed trading volume is five: we have three units sold by liquidity seekers, five units

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

bought by liquidity seekers and two units sold by liquidity providers. The double-counted volume is thus ten, and the commonly recorded single-counted volume is half this number.

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

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

The variance of the order imbalance is a nonlinear function of η, due to the different correlations of informed and uninformed demand. When most liquidity seekers are uninformed, their orders will be mostly matched to each other and the order imbalance is low. When most traders are informed, their correlated demands can lead to large imbalances. As a result, the standard deviation of the unobservable order imbalance 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 a function of η. Using the properties of the Half Normal distribution we find:[6]

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

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

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

To analyze the variance of the observed trading volume, we consider each of the three components of the double-counted volume that can be attributed to *(i)* informed liquidity seekers $\left(\sum_{1\ldots\eta M}|y_i|\right)$, *(ii)* uninformed liquidity seekers $\left(\sum_{\eta M+1\ldots M}|y_i|\right)$, and *(iii)* liquidity

[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 $Var\left(|x|\right) = \sigma^2\left(1-\frac{2}{\pi}\right)$.

providers ($|\sum_M y_i|$). The variances and covariances of these three components are derived in Appendix A. For large M, we find that the variance of of the per capita trading volume increases in η^2:

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

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

<ins>Proposition 1</ins>

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

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

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

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

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

<ins>Corollary</ins>

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

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

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

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

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

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

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

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

2.1 VCV in equilibrium

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

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

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

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

$$x_j = \beta_j(s_j - p_0), \tag{9}$$

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

$$p = p_0 + \lambda \left(\sum_n u_i + \sum_m x_j \right). \tag{10}$$

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

The m profit maximizing informed investors each solve:

$$\max_{x_j} x_j(E[v|s_j] - p_0 - \lambda(x_j + E[\sum_n u_i + \sum_{m-1} x_{-j}|s_j]), \tag{11}$$

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

Simultaneously, the market makers set the equilibrium price at the expected value of v, conditional on the order imbalance $\sum_n u_i + \sum_m x_j$. From the projection theorem, we know that $E[v|\sum_n u_i, \sum_m x_j; p_0, \beta, \sigma_v, \sigma_u, n, m] = p_0 + \frac{m\beta\left(\sigma_v^2+\sigma_\varepsilon^2\right)}{n\sigma_u^2+m^2\beta^2(\sigma_v^2+\sigma_\varepsilon^2)}\left(\sum_n u_i + \sum_m x_j\right)$, implying that $\lambda = \frac{m\beta\left(\sigma_v^2+\sigma_\varepsilon^2\right)}{n\sigma_u^2+m^2\beta^2(\sigma_v^2+\sigma_\varepsilon^2)}$. Combining these two results, we find that in equilibrium:

$$\beta = \frac{\sqrt{n}\sigma_u}{\sqrt{m\left(\sigma_v^2+\sigma_\varepsilon^2\right)}}; \quad \lambda = \frac{\sqrt{m\left(\sigma_v^2+\sigma_\varepsilon^2\right)}}{(m+1)\sqrt{n}\sigma_u}. \tag{12}$$

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

$$V = \frac{1}{2}\left(\sum_n |u_i| + \sum_m |x_j| + \left|\sum_n u_i + \sum_m x_j\right|\right). \tag{13}$$

The demands u_i and x_j both follow a Normal distribution around zero. We find from (9) and (12) that the variance of informed demand (σ_x^2) is only dependent on the variance of uninformed demand (σ_u^2) and the ratio of uninformed to informed investors:

$$\sigma_x^2 = \beta^2\left(\sigma_v^2+\sigma_\varepsilon^2\right) = \frac{n}{m}\sigma_u^2. \tag{14}$$

The intuition of (14) is that the trading aggressiveness of each informed investor increases in the number of uninformed investors, and decreases in the number of other informed investors. The distribution of trading volume thus depends only on n, m, and σ_u and is, unlike the distribution of prices and returns, independent of σ_v and σ_ε. Given that all components of (13) follow (correlated) Normal distributions, we can find the first two moments of total trading volume from integration. We find that both the mean and the standard deviation of volume are linear in σ_u. In particular we have:

Proposition 2

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

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

$$E[V]=\frac{\sigma_u}{\sqrt{2\pi}}(n+\sqrt{nm}+\sqrt{n(m+1)}). \tag{15}$$

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

$$\begin{aligned} Var(V) \;=\; & 2n\sigma_u^2\int_0^\infty x^2(m\Phi(\sqrt{m}x)+\Phi(\frac{x}{\sqrt{mn+n-1}}))\phi(x)dx \\ & +\sigma_u^2\frac{n\sqrt{m}(1-(m+1)^{\frac{3}{2}})+(mn+n-1)^{\frac{3}{2}}-(mn+n)^{\frac{3}{2}}-n(m+1)^2}{\pi(m+1)}, \end{aligned} \tag{16}$$

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

iii. *The coefficient of variation of trading volume is a function of n and m only. For a given number of uninformed traders n, the volume coefficient of variation increases in the number of informed traders m. For a given m, the volume coefficient of variation decreases in n.*

Proof: See Appendix A.

Panel A of Figure 1 graphically depicts the relationship between the equilibrium VCV and the number of uninformed and informed traders. For the case where $n=m=1$ the VCV is equal to $\frac{\sqrt{3\pi-4\sqrt{2}}}{2+\sqrt{2}}$. There is no closed form solution for all other finite (n,m) combinations. The figure shows that for a given n, the equilibrium VCV is a concave increasing function of the number of informed investors. As m goes to infinity, VCV approaches $\frac{1}{2}\sqrt{2\pi-4}\approx 0.756$,

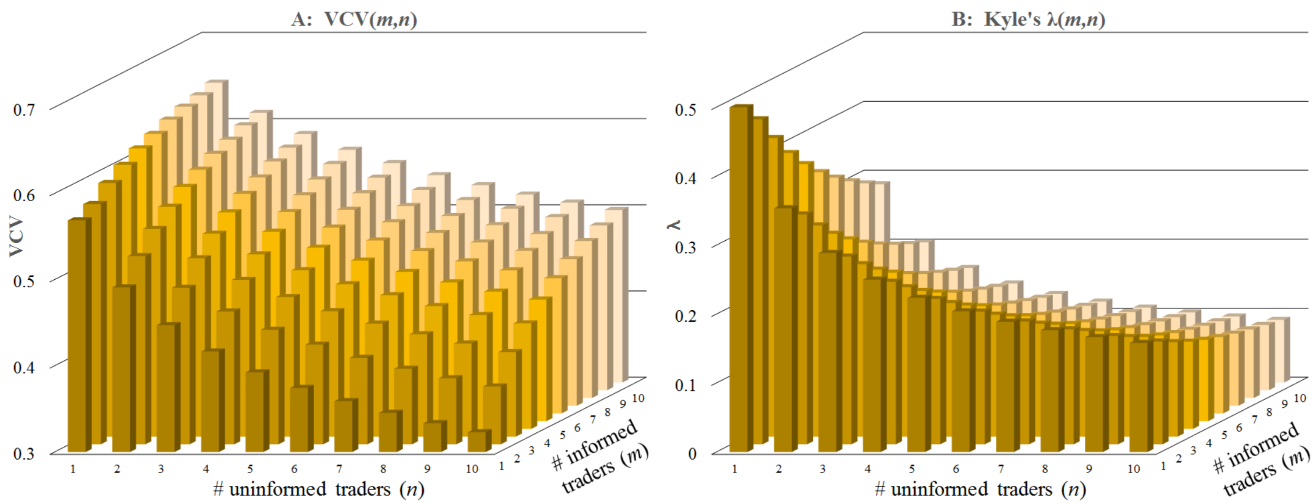

Figure 1: Equilibrium VCV (Panel A) and Kyle's λ (Panel B) as a function of n uninformed and m informed traders. Kyle's λ (Eq. 12) is divided by $\sqrt{\frac{\sigma_v^2+\sigma_\varepsilon^2}{\sigma_u^2}}$ to make it invariant to σ_ε, σ_v, and σ_u.

which is the coefficient of variation of the Half-Normal distribution and the VCV implied by our earlier model (Proposition 1) with $\eta = 1$. Additionally, for a given number of informed traders m, VCV is decreasing in the number of uninformed traders n. When n is very large relative to m, VCV approaches zero, as in Proposition 1 with $\eta = 0$. Panel A also shows that VCV decreases in $\frac{n}{m}$ along any diagonal with constant $(n + m)$.

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

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

η and M as:

$$\eta = \frac{\sigma_x m}{\sigma_x m + \sigma_u n}; \quad M = n + \frac{\sigma_x}{\sigma_u} m. \tag{17}$$

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

If informed traders are risk neutral and receive signals with *i.i.d.* noise terms, we find from Eq.(14) that $\eta = \frac{\sqrt{nm}}{\sqrt{nm}+n}$ and $M = n + \sqrt{nm}$. Further enriching the model with risk aversion, or long lived information will change the above expressions for the equilibrium η and M, but will not change Proposition 1, as there will always be a proportion of informed trade, and an equivalent number of market participants.

3 Simulations

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

Figure 2 displays four histograms of simulated volumes with $M = 1,000$ liquidity seekers, for different values of η. The sample size is $T = 1,000$ trading sessions. The simulation confirms the analysis in the previous section: in the case of no informed traders ($\eta = 0$), the volume distribution follows a slightly skewed bell-curve, while in the presence of informed traders volume is higher in level and far more dispersed. The simulated VCVs for the four

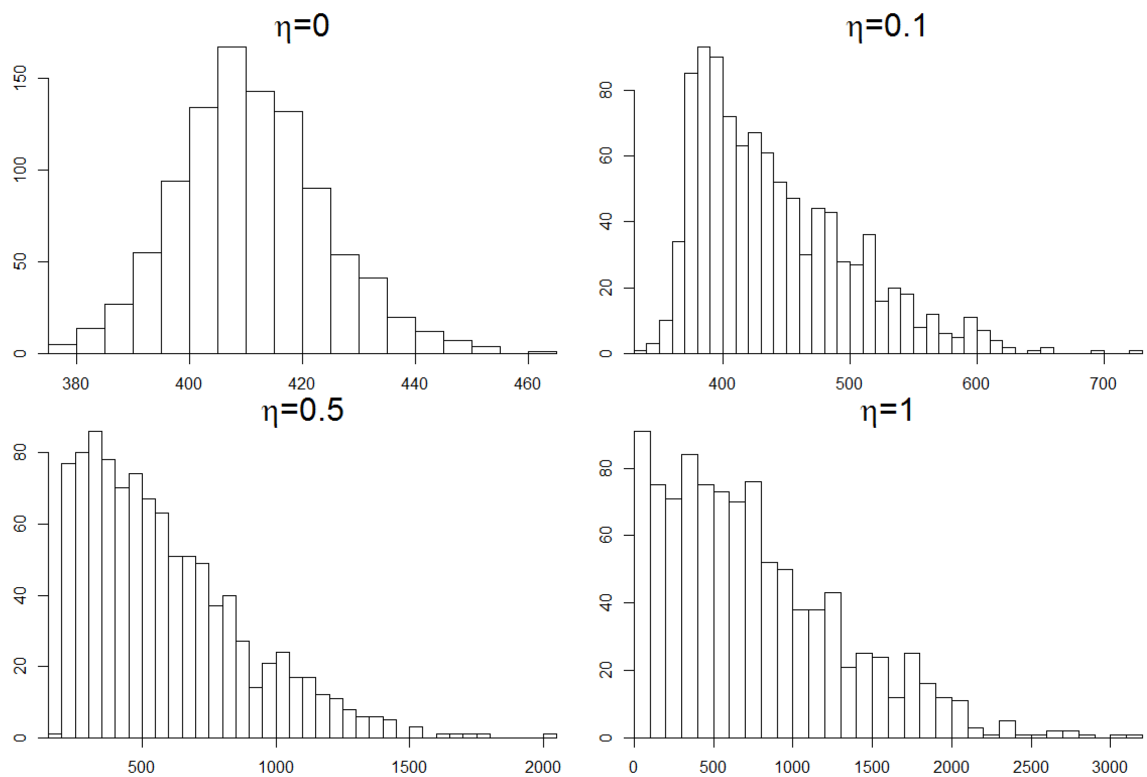

Figure 2: Histogram of T=1,000 volume realizations simulated from the model outlined in Section 2, for various values of the proportion of informed trading η. The number of liquidity seekers (M) is 1,000 and the trading intensity (σ) is fixed at unity.

panels are 0.03, 0.14, 0.48 and 0.77, respectively.[7]

Figure 3 reports the average VCV from $R = 1,000,000$ repetitions of simulating a sample of $T = 100$ trading sessions with M traders, for different values of η and M. As we can see, the average VCV only deviates substantially from its theoretical value (Eq.6) when both M and η are low. Nevertheless, even for small M, the average VCV is strictly increasing in η. The insensitivity to M is encouraging as it implies that there is little concern for confounding a high η with a low M. The insensitivity to M is also desirable from an empirical perspective, because the number of traders (M) in markets is typically unknown.

In Table 1, Panel A, we report the average VCV as plotted in Figure 3 for selected values of η, as well as the standard deviations to evaluate VCV's precision. In addition to VCV, we also report these statistics on simulated values of $\widehat{\eta}$ (Eq.8). Both VCV and $\widehat{\eta}$ increase

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

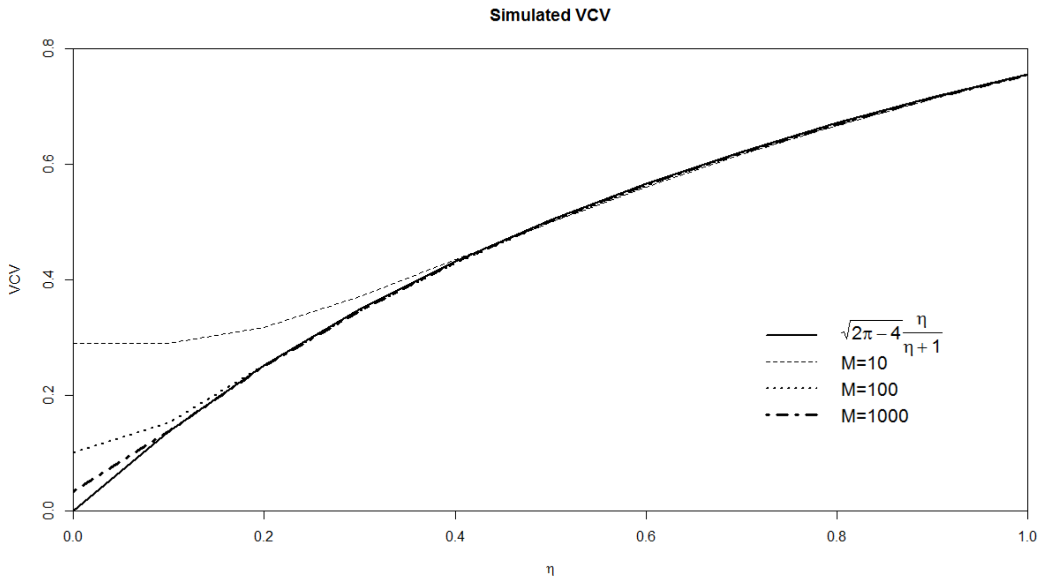

Figure 3: Average VCV obtained from $R = 1,000,000$ replications of $T = 100$ volume realizations simulated from the model outlined in Section 2, for various values of the proportion of informed trading η and number of liquidity seekers M.

monotonically in the true proportion of informed trade (η). This is even the case for markets with low trading activity M. Also, the estimator $\widehat{\eta}$ in our simulations traces the true value of η closely, in particular when either M or η are not too low. Panel B of Table 1 reports simulation results for smaller simulated samples, of $T = 10$ trading sessions. We still find the average VCV and $\widehat{\eta}$ to increase monotonically in η, although VCV, and more so $\widehat{\eta}$, 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 and allow for leptokurtic and skewed demand distributions, to generate jumps 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

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

				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

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,

Table 2: Simulation results - Robustness

This table reports the average and standard deviation of VCV (left) and $\widehat{\eta}$ (right) obtained from $R = 1,000,000$ replicated samples of $T = 100$ 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 Bernoulli 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 $E[\eta]$.

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

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 the model by Back et al. (2018). Panel B of Table 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 Bernoulli distribution. The number of active informed traders in each trading session is equal to X with probability $\frac{1}{5}$ and zero with probability $\frac{4}{5}$, such that the informed traders participate in only one out of five trading sessions on average. To create variation in the average proportion of informed trade, we adjust the potential number of informed traders X. 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.(8) 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.

Overall, the simulation results in this section demonstrate the robustness of VCV as a measure of asymmetric information. The basic result that the coefficient of variation of trading volume is monotonically increasing in the proportion of informed trade holds under general conditions and in small samples. Additional simulation results are presented in Internet Appendix A. These simulations are based on further variations of the basic model including random variation in liquidity demand (the number of liquidity seekers and their trading intensities) across observations, stochastic liquidity supply, heterogeneity among the informed investors, and endogenous informed demand. These supplementary results provide further evidence for the robustness of VCV as measure of information asymmetry: for all model specifications considered, VCV is strictly increasing in the proportion of informed trade. 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 analytical and numerical analysis a positive monotonic relation between VCV and the proportion of informed trade, we now turn to the data to analyze the empirical properties of VCV. In this section, we describe cross-sectional variation in VCV for US stocks, while we study time-series behavior 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 stock i in year τ is defined as:

$$VCV_{i,\tau} = \frac{\hat{\sigma}_{V(i,t\in\tau)}}{\hat{\mu}_{V(i,t\in\tau)}}, \tag{18}$$

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

$$V_{USD,i,t} = shares\ traded_{i,t} \times closing\ price_{i,t}, \tag{19}$$

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, such as macro-level sentiment and liquidity shocks:

$$V_{\%,i,t} = \frac{V_{USD,i,t}}{\sum_i V_{USD,i,t}}, \tag{20}$$

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

$$V_{TO,i,t} = \frac{shares\ traded_{i,t}}{shares\ outstanding_{i,t}}. \tag{21}$$

[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.

Table 3: VCV Summary Statistics

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

	VCV_{USD}	$VCV_{\%}$	VCV_{TO}
Observations	137,522	137,522	137,522
N	15,918	15,918	15,918
T	37	37	37
Mean	1.362	1.343	1.310
s.d.	0.798	0.811	0.748
s.d. (CS)	0.762	0.773	0.720
s.d. (TS)	0.536	0.540	0.506
$q_{0.1}$	0.594	0.556	0.586
$q_{0.25}$	0.840	0.814	0.817
Median	1.215	1.200	1.164
$q_{0.75}$	1.650	1.641	1.583
$q_{0.9}$	2.224	2.213	2.146
ρ	0.172	0.178	0.189
Correlations			
$VCV_{\%}$	0.983		
VCV_{TO}	0.970	0.959	

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 strong similarity between the three VCV measures offers support for the theoretical analysis of Section 2: although trading intensity (σ) and participation (M) are determinants of the level and variance of volume, VCV is independent of both σ and M (Eq.(6)). Market-wide variation in the number of market participants and their trading intensity should therefore have little impact, so that VCV derived from dollar volume, volume market shares, or turnover, are virtually equivalent. The results in Table 3 support this premise. In the remainder of this section, our measure of informed trading VCV is defined as the annual coefficient of variation of daily volume market shares ($VCV_{\%}$), which controls for market-wide variation in volume that is unrelated to firm-specific information. Highly similar results are obtained when using any of the other volume definitions. In the Internet Appendix Table B.1, we report the VCV summary statistics for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX, and for subsamples of observations prior to 2000 (1980-1999) and post 2000 (2000-2016), 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 is defined as the log of market capitalization on the last trading day of June. Return volatility is the annual standard deviation of daily returns. Amihud (2002) illiquidity is defined as the the log of the annual average of the daily ratio $\frac{|R_{i,t}|}{V_{USD,i,t}}$. The bid-ask spread is the annual average of daily closing bid-ask spreads as a percentage of the closing price $\frac{ask_{i,t}-bid_{i,t}}{price_{i,t}}$. Roll's (1984) measure is the square root of the negative of the daily return autocovariance $\sqrt{-Cov\left(R_{i,t}, R_{i,t-1}\right)}$.[9] The book-to-market ratio is the ratio of the book value of equity at the fiscal year end, obtained from COMPUSTAT, to the market value of equity at the end of the same calendar year. Firm age is proxied by the number of years passed since the firm appeared for the first time in the CRSP database. Analyst coverage is defined as the number of distinct analysts covering a stock in a given year (Source: IBES). Summary statistics of these variables and subsample analyses are provided in Internet Appendix 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 is a measure of informed trading, since informa-

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

Table 4: VCV and other firm characteristics

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

	VCV	Size	BM	Age	Vol.	Turn.	Illiq	B-A	Roll
Size	−0.64								
BM ratio	0.17	−0.22							
Age	−0.29	0.34	0.11						
Volatility	0.38	−0.61	−0.04	−0.40					
Turnover	−0.31	0.33	−0.21	−0.02	0.16				
Illiquidity	0.68	−0.94	0.22	−0.33	0.55	−0.52			
Bid-Ask spread	0.62	−0.87	0.25	−0.24	0.61	−0.42	0.91		
Roll's measure	0.25	−0.36	0.19	−0.05	0.24	−0.30	0.41	0.52	
Coverage	−0.58	0.79	−0.25	0.19	−0.32	0.47	−0.81	−0.71	−0.29

tion asymmetry is likely to be stronger 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 asymmetry, which is consistent with the negative correlation with VCV. In Section 5, we study the impact of exogenous reductions in analyst coverage due to brokerage closures and find that reductions in analyst coverage are associated with an increase in VCV.

4.2 Return reversals

The correlation between VCV and the bid-ask spread reported in Table 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 less likely to be reversed by the bid-ask bounce, and are

Table 5: VCV and the Bid-Ask spread

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

	Roll: Low	2	3	High	High-Low
Bid-Ask: Low	0.949	0.803	0.748	0.949	−0.001
2	1.157	1.114	1.114	1.018	−0.139**
3	1.384	1.350	1.488	1.404	0.020
High	1.949	1.717	1.822	2.015	0.067*
High-Low	0.999***	0.914***	1.074***	1.067***	0.067

characterized by less negative autocorrelations. To further evaluate the relationship between VCV and the bid-ask spread, we double-sort stocks within each year into quartiles based on the bid-ask spread and on Roll's measure. Table 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 return reversals. It is well known that returns on individual stocks, in particular illiquid stocks, exhibit significant short-term reversals (e.g. Jegadeesh, 1990). We 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 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 6: VCV and weekly reversals

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

	Illiq: Low	2	3	High	High-Low
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

Table 5, that short-term reversals are in general more profound for illiquid stocks, but that these reversals are weaker when the illiquidity is associated with information asymmetry.

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: Llorente et al. (2002), Hameed et al. (2008), Odders-White and Ready (2008), Bongaerts et al. (2016), and Johnson and So (2018) use various measures to show that asymmetric information is associated with weaker short-term reversals. In the next subsection, we have a closer look at the empirical relation between VCV and existing measures of asymmetric information.

4.3 VCV and other measures of asymmetric information

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

Table 7: VCV and other information asymmetry measures

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

	VCV	PIN_{BHL}	PIN_{BH}	PIN_{EHO}	PIN_{DY}	Adj.PIN	PSOS	MIA
PIN_{BHL}	0.53							
PIN_{BH}	0.60	0.74						
PIN_{EHO}	0.53	0.62	0.68					
PIN_{DY}	0.57	0.65	0.69	0.86				
Adjusted PIN	0.52	0.58	0.71	0.64	0.72			
PSOS	0.46	0.45	0.44	0.62	0.71	0.39		
MIA	0.26	0.37	0.44	0.12	0.24	0.32	0.05	
C2	0.10	0.12	0.11	0.03	0.03	0.04	0.04	0.02

assumption that informed traders are more likely to trade in options.

For our analysis, we make use of the various PIN and MIA measures that are kindly made publicly available by the authors of previous studies. These measures include MIA estimated by Johnson and So (2018) and PIN measures estimated by Easley et al. (2010 – PIN_{EHO}); Brown, Hillegeist and Lo (2004 – PIN_{BHL}); Brown and Hillegeist (2007 – PIN_{BH}); and Duarte and Young (2006 – PIN_{DY}).[10] We compute annual firm-level observations of MIA as the annual average of the available daily observations for each firm. We derive annual stock-level observations of C2 as the estimated slope coefficient from running regressions, for each firm in each year, of daily returns on the interaction of lagged returns and lagged (detrended) turnover, while controlling for daily lagged returns (see Llorente et al., 2002, for details).

Table 7 shows the correlations between VCV and various annual firm-level information

[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.

Table 8: VCV and Adjusted PIN

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

	VCV			
	(1)	(2)	(3)	(4)
PIN_{DY}	0.448***		0.961***	0.506***
	(0.125)		(0.164)	(0.112)
Adjusted PIN	0.957***	1.186***		0.924***
	(0.158)	(0.181)		(0.162)
PSOS	0.049	0.188***	−0.129**	
	(0.055)	(0.052)	(0.059)	
Observations	37,986	37,986	37,986	37,986
Adjusted R^2	0.357	0.356	0.353	0.357
Fixed effects	Yes	Yes	Yes	Yes

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 order-level data. 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 and PSOS are positively correlated with VCV.

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, 48 Fama-French 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 there is no robust relation between VCV and PSOS, thereby supporting our claim that VCV, like adjusted PIN, is indicative of asymmetric information rather than general illiquidity.

4.4 VCV and institutional ownership

In this subsection, we study the relationship between VCV and various indicators of institutional ownership that we obtain from 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 identify groups of presumably well-informed investors: monitoring investors and dedicated investors. Following Fich et al. (2015), we define an institutional investor to be a ’monitor’ for a certain firm if that firm belongs to the top 10% of holdings in the institution’s portfolio. These monitoring investors are likely to be better informed about the firm than non-monitoring investors. Dedicated investors are those institutional investors that Bushee and Noe (2000) and Bushee (2001) classify as ’dedicated’.

[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.

Table 9: VCV and institutional ownership

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

	VCV			
	(1)	(2)	(3)	(4)
Holdings	−0.0003	−0.0004	−0.0003	−0.0004
	(0.0005)	(0.0005)	(0.0004)	(0.0005)
Breadth	−1.125***	−1.732***	−1.102***	−1.677***
	(0.070)	(0.115)	(0.072)	(0.113)
Monitors		1.011***		0.973***
		(0.155)		(0.155)
Dedicated			0.311***	0.308***
			(0.082)	(0.079)
Observations	83,339	83,339	77,225	77,225
Adjusted R^2	0.408	0.409	0.406	0.407
Fixed effects	Yes	Yes	Yes	Yes

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

[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/.

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, who each assign a relatively large fraction of their portfolio to this firm's stock (i.e. *Breadth* is low, while *Monitors* and *Dedicated* are high). Ownership of such a firm is therefore relatively concentrated in the hands of a small number of presumably well informed investors. When trading this 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 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 4 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 of VCV post-2000 suggests that asymmetric information has reduced over this period, which is consistent with recent studies that document improved market transparency, which is attributed partly to regulation, such as the enactment by the SEC of Regulation Fair Disclosure (Reg FD) in 2000 and the Sarbanes-Oxley Act in 2002 (e.g. Chen et al., 2010; Petacchi, 2015; Beaver et al., 2018; and Pawlewicz, 2018). 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

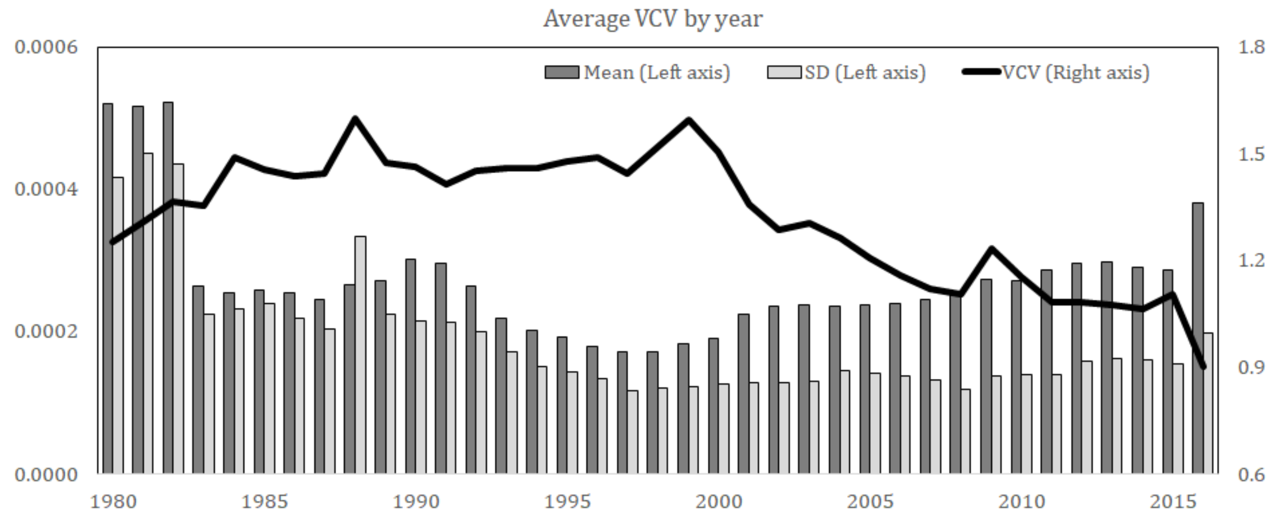

Figure 4: The black line shows the annual cross-sectional average of annual firm-level VCVs, calculated from volume market shares (Eq.(20)) over the period 1980-2016. The bars show the annual cross-sectional average of annual firm-level means and standard deviations of volume market shares.

the three volume measures.[13]

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; Li and You, 2015) consider terminations of analyst coverage due to brokerage closures as exogenous shocks to the information environment of individual 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.

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

[13]The level shift in the means and standard deviation of volume shares after 1982 occurs because of the inclusion of NASDAQ shares.

A of Kelly and Ljungqvist (2012), we identify in the IBES database a treatment sample of a total of 1,764 observations of firms that experience reductions in analyst coverage due to one of these closures.

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

The results of the difference-in-differences regression are reported in the first column of Table 10. The coefficient on the interaction term *After*×*Treated* is of primary interest. This interaction coefficient is positive and significant, meaning that the VCV of firms that face exogenous analyst reductions as a result of brokerage closures *increases* relative to the VCV of control firms that are not exposed to the brokerage closures. The coefficient on *After* is negative, which reflects that VCV is on average decreasing over time, as can be seen from Figure 4. The *Treated* coefficient indicates that there is a minor difference between the VCV of treated and control firms, prior to the event.[14]

The second and third column of Table 10 show that the interaction coefficient becomes larger when we restrict the sample to firms with low analyst coverage. The intuition behind

[14]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

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

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

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

5.2 VCV around earnings announcements

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

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

$$VCV_{XS,d} = \frac{\hat{\sigma}_{V(t=t_i+d)}}{\hat{\mu}_{V(t=t_i+d)}}, \tag{22}$$

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

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 5 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 5 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, or to more informed trading, possibly due to information leakage. 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

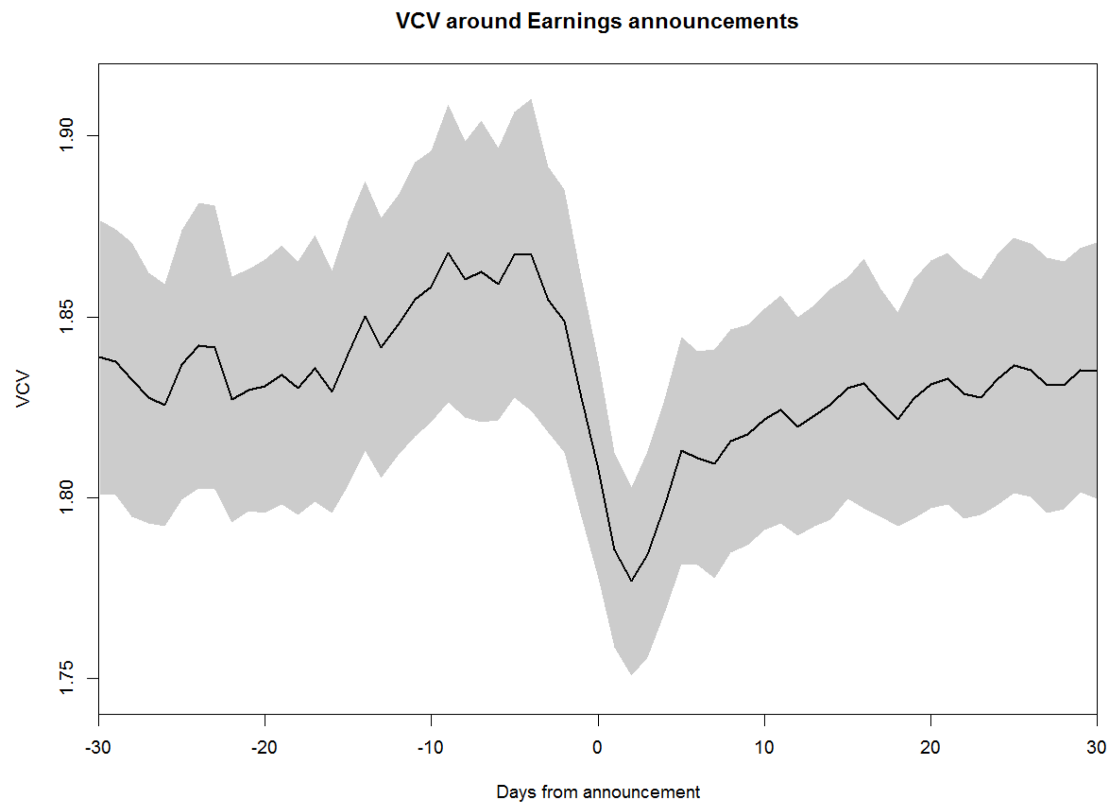

Figure 5: The black line shows the evolution of the daily cross-sectional VCV_{XS} around quarterly earnings announcements. The full sample includes all daily trading volumes over 61 days windows (day -30:30) around $N = 339,257$ quarterly announcements (sources: CRSP and COMPUSTAT). The reported VCV 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).

announcement.

Table 11 reports VCV and its components: the cross-sectional mean and standard deviation of volume shares, for each day around the announcement. The level of volume is low prior to announcements and high following announcement, which is consistent with the patterns documented by Chae (2005) and Akbas (2016). The standard deviation of volume moves in the same direction as the mean, which could be due to be the increased illiquidity and price elasticity in the days before the announcement, as documented by George et al. (1994) and Chae (2005). What we are most interested in is the pattern of VCV as a proxy for information asymmetry. Since the changes in the standard deviation are smaller in relative terms than the changes in the mean, VCV is high prior to the announcement and low afterwards. As Table 11 shows, the differences between VCV are statistically significant up to nine days before and

Table 11: Volume around earnings announcements

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

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

after the announcement.

This pattern of VCV around earnings announcements is consistent with the assertion that information asymmetries are resolved around earnings announcements, and with the behavior of alternative information asymmetry measures. Johnson and So (2018) 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. (2019) find that the volatility of order flow, driven by correlated liquidity demand, significantly increases before earnings announcements. There is mixed evidence on the behavior of PIN around announcement dates. Benos and Jochec (2007) and Duarte et al. (2019) find that PIN is in fact lower prior to earnings announcements and higher afterwards. Duarte et al. (2019) explain this puzzling result by demonstrating that the PIN measure mis-identifies asymmetric information when applied on a daily frequency, and instead simply indicates abnormal turnover. Easley et al. (2008), on the other hand, estimate a generalized PIN model in which the arrival rate of information is time-varying and find that PIN is high (low) before (after) earnings announcements, resembling the pattern of VCV in Figure 5.

In internet Appendix Figure B.1, we reproduce Figure 5 for various subsets of the data, showing a qualitatively similar pattern of VCV around earnings announcements for both NASDAQ and NYSE/AMEX stocks as well as before and after 2000. The drop in VCV around announcements has in fact become sharper post 2000. This result is consistent with Beaver et al. (2018) and Pawlewicz (2018), who find a recent increase in the information content of earnings announcements, and Weller (2018) who finds that price informativeness prior to earnings announcements has decreased, despite the presence of algorithmic trading.

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

Figure 6 reports the cross-sectional VCV around non-surprising announcement dates (ter-

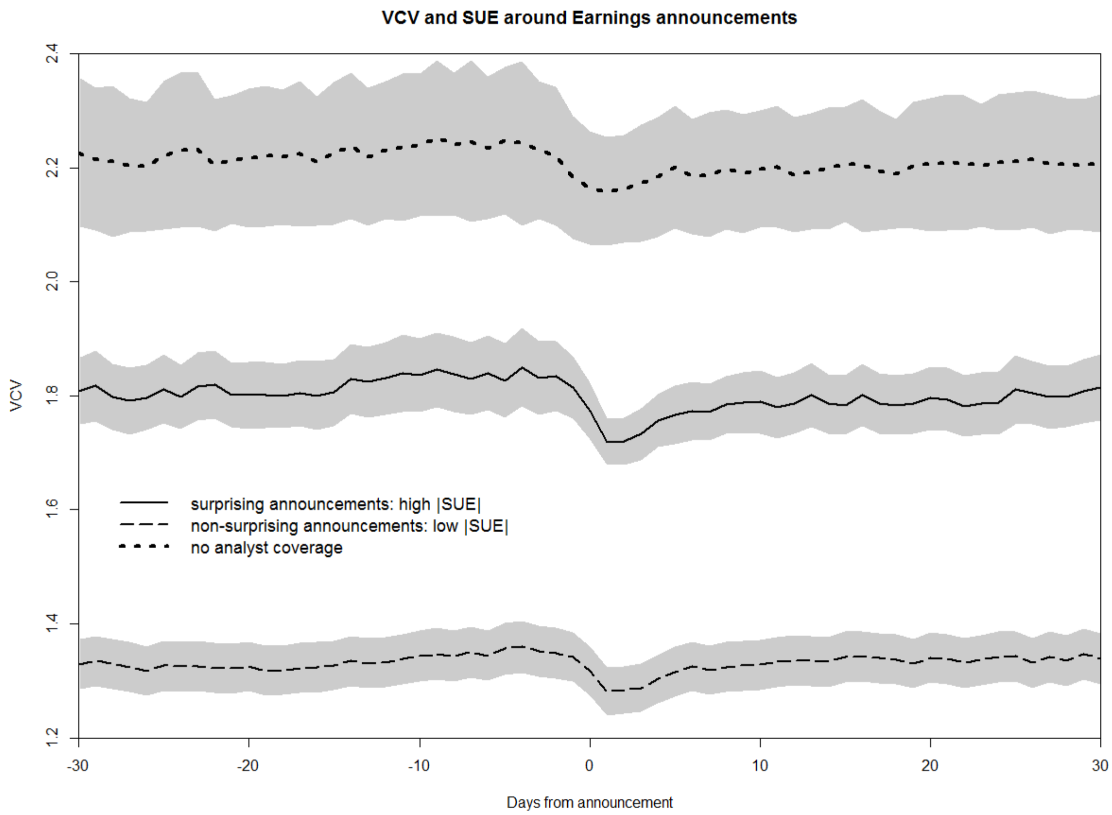

Figure 6: This figure shows the evolution of the daily cross-sectional VCV_{XS} over 61 days windows (day -30:30) around quarterly earnings announcements. The sample is divided into surprising announcements (solid line, $N = 55,340$), non-surprising announcements (dashed line, $N = 55,341$), and earnings announcement for which no analyst forecasts are reported (dotted line, $N = 173,235$). Surprising (non-surprising) announcements are defined as the announcements in the highest (lowest) tercile of announcements sorted on the absolute value of SUE: the difference between the actual and median analyst forecast of earnings, scaled by the price. See Figure 5 for details.

cile 1) and surprising announcement dates (tercile 3). The figure clearly shows for the surprising announcements both a higher level of VCV, indicating a higher degree of information asymmetry, as well as a steeper drop around the announcement date, verifying that surprising announcements are more informative than non-surprising announcements. For both of these subsamples, the level of VCV is lower than it is for the full sample in Figure 5. This is because both the surprising and non-surprising subsamples are restricted to those firms for which analyst expectations are available. For completeness, Figure 6 also displays the cross-sectional VCV around the announcements for which no analyst forecasts are reported in IBES. The level of the cross-sectional VCV for the no-forecast sample is clearly higher, consistent with the negative relation between VCV and analyst coverage as reported in Tables 4 and 10. However, the S-shape around the announcement is flatter than for the sample of surprising announcements. Overall, the breakdown in Figure 6 reveals that both the level of the cross-sectional VCV and its dynamics around the announcement dates behave as expected, depending on the informativeness of the announcement, providing further evidence that the cross-sectional VCV captures information asymmetry in event time.

6 Conclusion

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

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

Collectively, our empirical results provide broad support for the hypothesis that VCV is a measure of informed trading not only within our stylized microstructure model, but also when applied to observational data. Moreover, as we report in our internet appendix, all empirical results are qualitatively similar for subsamples of NYSE/AMEX and NASDAQ stocks, as well as for pre- and post-2000 periods, validating the robustness of VCV as a measure of information asymmetry in different market environments.

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 when there is a need to control for information asymmetry, as a sorting characteristic when studying the pricing effects of asymmetric information, or as the dependent variable of interest to compare patterns in information asymmetry across firms, countries, asset classes, or over time.

Appendix A: Variance of trading volume

Th first part of this appendix derives the variance of trading volume (Eq.5). The second part derives the variance of trading volume for the (m,n) model given in Proposition 2.

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

$$V = \tfrac{1}{2}\left(Y_I + Y_U + Y_{MM}\right). \tag{23}$$

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

$$
\begin{aligned}
Var\left(2V\right) &= Var\left(Y_I\right) + Var\left(Y_U\right) + Var\left(Y_{MM}\right) \\
&\quad +2Cov\left(Y_I, Y_U\right) + 2Cov\left(Y_I, Y_{MM}\right) + 2Cov\left(Y_U, Y_{MM}\right).
\end{aligned} \tag{24}
$$

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

$$
\begin{aligned}
Var\left(Y_I\right) &= \eta^2 M^2 \sigma^2 \left(1 - \tfrac{2}{\pi}\right) \\
Var\left(Y_U\right) &= (1-\eta) M \sigma^2 \left(1 - \tfrac{2}{\pi}\right) \\
Var\left(Y_{MM}\right) &= \left(\eta^2 M^2 + (1-\eta) M\right) \sigma^2 \left(1 - \tfrac{2}{\pi}\right).
\end{aligned} \tag{25}
$$

$Cov\left(Y_I, Y_U\right) = 0$, because the demands of informed and uninformed liquidity seekers are independent. Moreover, when M is large and $\eta > 0$, the order imbalance consists mainly of orders submitted by informed liquidity seekers. The orders of uninformed traders tend to net out against each other because of the *i.i.d* property. This implies that in the limit ($M \to \infty$), the liquidity suppliers trade exclusively to offset the imbalance from informed seekers. Therefore, $\lim_{M\to\infty} Cor\left(Y_U, Y_{MM}\right) = 0$ and $\lim_{M\to\infty} Cor\left(Y_I, Y_{MM}\right) = 1$. Given these correlations, Eq.(24) implies that when $M \to \infty$:

$$
Var\left(\tfrac{2V}{M}\right) = Var\left(\tfrac{Y_I}{M}\right) + Var\left(\tfrac{Y_U}{M}\right) + Var\left(\tfrac{Y_{MM}}{M}\right) + 2\sqrt{Var\left(\tfrac{Y_I}{M}\right) Var\left(\tfrac{Y_{MM}}{M}\right)}, \tag{26}
$$

which, given the variances in Eq.(25), results in:

$$
\begin{aligned}
Var\left(\tfrac{2V}{M}\right) &= \eta^2\sigma^2\left(1-\tfrac{2}{\pi}\right) + (1-\eta)M^{-1}\sigma^2\left(1-\tfrac{2}{\pi}\right) + \left(\eta^2 + (1-\eta)M^{-1}\right)\sigma^2\left(1-\tfrac{2}{\pi}\right) \\
&\quad +2\sqrt{\eta^2\sigma^2\left(1-\tfrac{2}{\pi}\right)}\sqrt{\left(\eta^2 + (1-\eta)M^{-1}\right)\sigma^2\left(1-\tfrac{2}{\pi}\right)} \\
&= 2\sigma^2\left(1-\tfrac{2}{\pi}\right)\left(\eta^2 + (1-\eta)M^{-1} + \eta\sqrt{\eta^2 + (1-\eta)M^{-1}}\right) \\
&= 4\sigma^2\left(1-\tfrac{2}{\pi}\right)\eta^2,
\end{aligned} \tag{27}
$$

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

$$\lim_{M\to\infty} \frac{s.d.(V)}{E\left[V\right]} = \lim_{M\to\infty} \frac{s.d.(V/M)}{E\left[V/M\right]} = \sqrt{2\pi - 4}\frac{\eta}{\eta + 1}. \tag{28}$$

Proof of Proposition 2

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

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

$$Cov(|r|, |r + \alpha s|) = 4\int_0^\infty r^2 \Phi(\frac{r}{\alpha})\phi(r)dr + \frac{2\alpha^3 - 2(\alpha^2+1)^{\frac{3}{2}}}{(\alpha^2+1)\pi} - 1, \tag{29}$$

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

Proof:

$$\begin{aligned} Cov(|r|, |r + \alpha s|) &= E[|r||r+\alpha s|] - E[|r|]E[|r+\alpha s|] \\ &= E[|r||r+\alpha s|] - \tfrac{2}{\pi}\sqrt{1+\alpha^2}. \end{aligned} \tag{30}$$

We evaluate the first term by integration:

$$
\begin{aligned}
E[|r||r+\alpha s|] &= \iint |r(r+\alpha s)| d\phi(r) d\phi(s) \\
&= \int_{-\infty}^{0} (\int_{-\infty}^{-\frac{r}{\alpha}} (r^2+\alpha rs)\phi(s)ds + \int_{-\frac{r}{\alpha}}^{\infty} (-r^2-\alpha rs)\phi(s)ds)dr \\
&\quad + \int_{0}^{\infty} (\int_{-\infty}^{-\frac{r}{\alpha}} (-r^2-\alpha rs)\phi(s)ds + \int_{-\frac{r}{\alpha}}^{\infty} (r^2+\alpha rs)\phi(s)ds)\phi(r)dr \\
&= \int_{-\infty}^{0} (r^2\Phi(-\tfrac{r}{\alpha}) - \tfrac{\alpha r}{\sqrt{2\pi}} e^{-\frac{r^2}{2\alpha^2}} - r^2(1-\Phi(-\tfrac{r}{\alpha})) - \tfrac{\alpha r}{\sqrt{2\pi}} e^{-\frac{r^2}{2\alpha^2}})\phi(r)dr \\
&\quad + \int_{0}^{\infty} (-x^2\Phi(-\tfrac{r}{\alpha}) + \tfrac{\alpha r}{\sqrt{2\pi}} e^{-\frac{r^2}{2\alpha^2}} + r^2(1-\Phi(-\tfrac{r}{\alpha})) + \tfrac{\alpha r}{\sqrt{2\pi}} e^{-\frac{r^2}{2\alpha^2}})\phi(r)dr \\
&= 2\int_{-\infty}^{0} r^2\Phi(-\tfrac{r}{\alpha})\phi(r)dr - \int_{-\infty}^{0} r^2\phi(r)dr - \tfrac{2\alpha}{2\pi}\int_{-\infty}^{0} re^{-\frac{r^2}{2\alpha^2}-\frac{r^2}{2}}dr \\
&\quad -2\int_{0}^{\infty} r^2\Phi(-\tfrac{r}{\alpha})\phi(r)dr + \int_{0}^{\infty} r^2\phi(r)dr + \tfrac{2\alpha}{2\pi}\int_{0}^{\infty} re^{-\frac{r^2}{2\alpha^2}-\frac{r^2}{2}}dr \\
&= 2\int_{0}^{\infty} r^2\Phi(\tfrac{r}{\alpha})\phi(r)dr - \tfrac{1}{2} + \tfrac{\alpha^3}{(\alpha^2+1)\pi} \\
&\quad -2\int_{0}^{\infty} r^2(1-\Phi(\tfrac{r}{\alpha}))\phi(r)dr + \tfrac{1}{2} + \tfrac{\alpha^3}{(\alpha^2+1)\pi} \\
&= 4\int_{0}^{\infty} r^2\Phi(\tfrac{r}{\alpha})\phi(r)dr + \tfrac{2\alpha^3}{(\alpha^2+1)\pi} - 2\int_{0}^{\infty} r^2\phi(r)dr \\
&= 4\int_{0}^{\infty} r^2\Phi(\tfrac{r}{\alpha})\phi(r)dr + \tfrac{2\alpha^3}{(\alpha^2+1)\pi} - 1.
\end{aligned}
\tag{31}
$$

Substitute (31) into (30) to obtain the lemma (29).

To evaluate the variance of the trading volume we use:

$$
\begin{aligned}
Var(2V) &= \sum_n var(|u_i|) + var(|\sum_m x_j|) + var(|z|) \\
&\quad +2Cov(|\sum_m x_j|, |z|) + 2\sum_n Cov(|u_i|, |z|),
\end{aligned}
\tag{32}
$$

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

$$
\begin{aligned}
\sum_n Var(|u_i|) + Var(|x|) + Var(|z|) &= n\sigma_u^2(1-\tfrac{2}{\pi}) + mn\sigma_u^2(1-\tfrac{2}{\pi}) \\
&\quad +(1+m)n\sigma_u^2(1-\tfrac{2}{\pi}) \\
&= 2(m+1)n\sigma_u^2(1-\tfrac{2}{\pi}).
\end{aligned}
\tag{33}
$$

The first covariance term in (32) can be evaluated as follows:

$$Cov(|\sum_m x_j|, |z|) = Cov(|x|, |x + u|), \tag{34}$$

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

$$Cov(|x|, |x + u|) = mn\sigma_u^2(4\int_0^\infty x^2\Phi(\sqrt{m}x)\phi(x)dx + \frac{2(1-(m+1)^{\frac{3}{2}})}{\pi\sqrt{m}(m+1)} - 1). \tag{35}$$

The final covariance terms in (32) are all identical, and can be evaluated as:

$$Cov(|u_i|, |z|) = Cov(|u_i|, |u_i + z_{-i}|), \tag{36}$$

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

$$\begin{aligned} Cov(|u_i|, |z|) &= \sigma_u^2(4\int_0^\infty x^2\Phi(\tfrac{x}{\sqrt{mn+n-1}})\phi(x)dx \\ &\quad + \tfrac{2(mn+n-1)^{\frac{3}{2}} - 2(mn+n)^{\frac{3}{2}}}{\pi(mn+n)} - 1). \end{aligned} \tag{37}$$

Combining the variance terms (33), and the covariance terms (35) and (37) gives, after rearranging, the variance of trading volume ($Var(V)$) as given in Proposition 2.

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.

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

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. (2018). Identifying information asymmetry in securities markets. *Review of Financial Studies*, 31, 2277-2325.

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.

Bayraktar, E., and Munk, A. (2017). Mini-flash crashes, model risk, and optimal execution. Working paper.

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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., Hu, E. and Young, L. (2018). A comparison of some structural models of private information arrival? *Journal of Financial Economics*, forthcoming.

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

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

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

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

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

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

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

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

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

Easley, D., 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.

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

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

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.

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

May 29, 2019

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, to demonstrate the robustness of VCV. Tables A.1 and A.2 provide detailed simulation results for the benchmark model (Table 1 in the paper). The remaining tables present simulation results for various modifications of our benchmark model, including Non-Gaussian demand, stochastic liquidity demand and provision, heterogeneous information, and endogenous trading intensity. For all model specifications considered in this appendix, we continue to find that VCV is strictly increasing in the proportion of informed trade η.

Non-Gaussian demand and random proportion of informed trade

Table A.3 provides detailed results of the simulations with non-Gaussian demand distributions (Table 2 Panel A in the paper), including a Uniform demand distribution. Table A.4 provides detailed simulation results for the model with a Bernoulli distributed proportion of informed trade (Table 2 Panel B in the paper), as well as results from the model with a Uniformly distributed proportion of informed trade.

Stochastic liquidity

Next, we demonstrate the robustness of VCV to random variation in liquidity demand and liquidity provision. Table A.5 and A.6 provide simulations for a variation of our model in which liquidity demand (the number of market participants M and their trading intensities σ) are varying randomly (Uniform distribution) across observations. In Table A.7, we analyze the distribution of trading volume in the presence of stochastic liquidity provision: different from the benchmark model in Section 2, we assume that in each trading session liquidity providers absorb the order imbalance $\sum_M y_i$ with probability p and do not absorb the imbalance with probability (1-p). Total trading volume can thus be written as:

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

That is, with probability p, liquidity provision is high and the counting of volume is identical to the benchmark model in our paper (Eq.1 in the paper): double-counted volume includes the trades among liquidity seekers, as well as the trades between the liquidity providers and unmatched liquidity seekers. With probability (1-p), liquidity is low and the volume only includes liquidity seeking demands that can be matched among each other.

Heterogeneous information

Table A.8 provides simulation results for a version of our model in which there is heterogeneous information (or differences of opinion) among the informed investors. Instead of all ηM informed investors making the same order based on the same information, we divide the informed investors into two groups that trade on independent private signals and therefore make independent orders. That is, the first $\frac{1}{2}\eta M$ informed traders each submit identical orders and the second $\frac{1}{2}\eta M$ informed traders each submit identical orders that are uncorrelated to the orders by the first group. The demands of each of the $(1-\eta)\,M$ uninformed liquidity seekers remain uncorrelated as in Section 2.

Endogenous informed trading intensity

Table A.9 provides results of simulations in which the informed trading intensity is endogenized, by making it explicitly dependent on the volume generated by the uninformed traders. The demands of the $(1-\eta)M$ uninformed investors are normally distributed with mean zero and standard deviation of one, as in our benchmark model, while the standard deviation of the ηM informed investors' demand is conditional on the demand by the uninformed investors. Specifically, the standard deviation of informed demand is increasing in the absolute order imbalance generated by the uninformed investors, such that informed investors attempt to 'hide' their orders in the uninformed order flow, as in Kyle (1985). The ηM informed investors observe in each trading session t the uninformed order flow $\sum_{\eta M+1\ldots M} y_{i,t}$ and adjust their trading intensity accordingly:

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

Hence, the unconditional expected value of the informed investor's trading intensity σ_t is unity, as for the uninformed investors, but the informed investors trade more (less) aggressively when the uninformed net order flow is relatively high (low). We obtain qualitatively similar results when the informed trading intensity is dependent on the cumulative size of the uninformed orders ($\sum_{\eta M+1\ldots M} |y_{i,t}|$), instead of on the uninformed order imbalance.

Table A.1: Simulation results: Benchmark model

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

Panel A: VCV											
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$M = 10$											
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

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

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

					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	0.30	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

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

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

Panel A: VCV											
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Uniform distribution											
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

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

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

Panel A: VCV											
Uniform distribution: Informed investors $\sim U[0, X]$											
Uninformed	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
X	0	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
Bernoulli 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
Bernoulli 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

Table A.5: Simulation results: Random trading intensity σ

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

					Panel A: VCV						
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
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

Table A.6: Simulation results: Random number of market participants M

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

					Panel A: VCV						
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
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

Table A.7: Simulation results: Stochastic Liquidity Provision

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

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

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

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

Table A.8: Simulation results: Heterogeneous information

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

					Panel A: VCV						
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
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

Table A.9: Simulation results: Endogenous informed trading

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

					Panel A: VCV						
η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
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	.

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 regressions around brokerage closures (Table 10 in the paper), using VCV computed over a period of 6 months, and using a selective 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

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).

	NASDAQ			NYSE/AMEX		
	VCV_{USD}	$VCV_{\%}$	VCV_{TO}	VCV_{USD}	$VCV_{\%}$	VCV_{TO}
Observations	$77,629$	$77,629$	$77,629$	$59,893$	$59,893$	$59,893$
N	$11,556$	$11,556$	$11,556$	$5,877$	$5,877$	$5,877$
T	34	34	34	37	37	37
Mean	1.53	1.51	1.46	1.14	1.13	1.12
s.d.	0.84	0.85	0.78	0.68	0.70	0.66
s.d. (CS)	0.81	0.81	0.76	0.64	0.66	0.62
s.d.(TS)	0.57	0.57	0.53	0.45	0.46	0.44
$q_{0.1}$	0.74	0.70	0.71	0.51	0.47	0.51
$q_{0.25}$	1.02	1.00	0.97	0.68	0.65	0.67
Median	1.37	1.35	1.30	1.00	0.98	0.98
$q_{0.75}$	1.81	1.80	1.73	1.41	1.40	1.37
$q_{0.9}$	2.42	2.40	2.33	1.90	1.90	1.84
ρ	0.12	0.12	0.13	0.22	0.23	0.23
Correlations						
$VCV_{\%}$	0.98			0.98		
VCV_{TO}	0.96	0.95		0.98	0.96	
	Pre-2000			**Post-2000**		
	VCV_{USD}	$VCV_{\%}$	VCV_{TO}	VCV_{USD}	$VCV_{\%}$	VCV_{TO}
Observations	$75,607$	$75,607$	$75,607$	$61,915$	$61,915$	$61,915$
N	$12,213$	$12,213$	$12,213$	$8,331$	$8,331$	$8,331$
T	20	20	20	17	17	17
Mean	1.47	1.46	1.42	1.23	1.20	1.18
s.d.	0.71	0.73	0.67	0.88	0.88	0.81
s.d. (CS)	0.68	0.69	0.65	0.86	0.87	0.80
s.d. (TS)	0.49	0.49	0.46	0.52	0.52	0.49
$q_{0.1}$	0.79	0.77	0.78	0.51	0.47	0.50
$q_{0.25}$	1.03	1.02	1.00	0.66	0.62	0.64
Median	1.34	1.33	1.28	0.99	0.96	0.95
$q_{0.75}$	1.72	1.72	1.66	1.52	1.48	1.44
$q_{0.9}$	2.24	2.25	2.17	2.20	2.16	2.10
ρ	0.07	0.08	0.09	0.09	0.09	0.10
Correlations						
$VCV_{\%}$	0.98			0.99		
VCV_{TO}	0.97	0.96		0.97	0.96	

Table B.2: Summary statistics of firm-characteristics

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

	Size	BM	Age	Vol.	TO	Illiq	Bid-Ask	Roll	Cov.
Observations	137,522	134,770	137,522	137,522	137,522	137,522	112,644	137,522	69,228
N	15,918	15,778	15,918	15,918	15,918	15,918	13,643	15,918	9,934
T	37	37	37	37	37	37	34	37	26
Mean	12.36	0.67	15.47	3.51	0.49	−16.44	2.35	0.97	2.04
s.d.	2.03	0.55	15.33	2.32	0.70	3.27	3.13	3.04	0.81
s.d.(CS)	1.90	0.49	15.17	2.00	0.53	2.90	2.39	2.71	0.81
s.d.(TS)	0.67	0.35	3.67	1.29	0.28	1.30	1.46	2.29	0.33
$q_{0.1}$	9.83	0.18	2	1.47	0.08	−20.91	0.08	−2.09	0.69
$q_{0.25}$	10.87	0.32	4	2.00	0.14	−18.95	0.28	−1.03	1.39
Median	12.22	0.54	11	2.92	0.28	−16.33	1.38	0.84	2.08
$q_{0.75}$	13.71	0.85	21	4.35	0.58	−13.87	3.22	2.29	2.64
$q_{0.9}$	15.06	1.27	36	6.23	1.09	−12.15	5.81	4.38	3.14

Table B.3: VCV and other firm characteristics

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).

	NASDAQ VCV	NYSE/AMEX VCV	Pre-2000 VCV	Post-2000 VCV
Size	−0.54	−0.67	−0.54	−0.76
BM ratio	0.20	0.20	0.17	0.18
Age	−0.12	−0.31	−0.26	−0.33
Volatility	0.28	0.35	0.29	0.48
Turnover	−0.31	−0.24	−0.29	−0.34
Illiquidity	0.62	0.68	0.62	0.76
Bid-Ask Spread	0.60	0.64	0.51	0.73
Roll's measure	0.29	0.07	0.27	0.23
Coverage	−0.49	−0.57	−0.59	−0.57

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

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).

	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

Table B.5: VCV and weekly reversals

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).

	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

Table B.6: Summary statistics of information asymmetry measures

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

	PIN_{BHL}	PIN_{BH}	PIN_{EHO}	PIN_{DY}	Adj.PIN	PSOS	MIA	C2
Observations	76,465	76,919	32,576	37,986	37,986	37,986	21,420	128,470
N	11,322	11,340	4,369	4,639	4,639	4,639	3,504	15,390
T	18	18	19	22	22	22	20	33
Mean	0.19	0.21	0.19	0.21	0.17	0.28	0.41	0.02
s.d.	0.10	0.11	0.07	0.09	0.08	0.15	0.10	0.10
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	0.09	0.11	0.11	0.09	0.13	0.27	−0.11
$q_{0.25}$	0.11	0.13	0.15	0.14	0.12	0.17	0.35	−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	0.23	0.24	0.21	0.33	0.47	0.08
$q_{0.9}$	0.31	0.36	0.28	0.33	0.28	0.50	0.52	0.13

Table B.7: VCV and other information asymmetry measures

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).

	NASDAQ VCV	NYSE/AMEX VCV	Pre-2000 VCV	Post-2000 VCV
PIN_{BHL}	0.43	0.52	0.39	0.62
PIN_{BH}	0.52	0.65	0.46	0.69
PIN_{EHO}	0.39	0.53	0.51	0.72
PIN_{DY}	0.40	0.57	0.52	0.73
Adjusted PIN	0.40	0.52	0.47	0.71
PSOS	0.29	0.46	0.43	0.57
MIA	0.23	0.32	0.11	0.29
C2	0.05	0.06	0.14	0.06

Table B.8: VCV and Adjusted PIN

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).

	NASDAQ VCV	NYSE/AMEX VCV	Pre-2000 VCV	Post-2000 VCV
PIN_{DY}	0.340	0.466^{***}	0.320^{***}	1.315^{***}
	(0.271)	(0.142)	(0.120)	(0.203)
Adjusted PIN	0.762^{***}	0.962^{***}	0.913^{***}	0.861^{***}
	(0.282)	(0.163)	(0.174)	(0.219)
PSOS	-0.259^{*}	0.063	0.054	0.092
	(0.137)	(0.057)	(0.064)	(0.058)
Observations	1,726	36,260	29,734	8,252
Adjusted R^2	0.341	0.357	0.320	0.499
Fixed effects	Yes	Yes	Yes	Yes

Table B.9: Summary statistics of institutional ownership characteristics

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

	Holdings	Breadth	Monitors	Dedicated
Observations	83,339	83,339	83,339	77,225
N	11,862	11,862	11,862	11,253
T	26	26	26	24
Mean	46.67	5.06	3.56	6.71
s.d.	29.80	7.04	6.05	9.96
s.d.(CS)	27.61	7.05	6.02	5.45
s.d.(TS)	11.60	1.15	1.96	5.91
$q_{0.1}$	7.02	0.41	0	0
$q_{0.25}$	21.00	1.01	0	0
Median	46.28	2.81	1.37	1.82
$q_{0.75}$	70.48	5.97	4.69	11.11
$q_{0.9}$	85.81	12.08	9.70	20

Table B.10: VCV and institutional ownership

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).

	NASDAQ	NYSE/AMEX	Pre-2000	Post-2000
	VCV	VCV	VCV	VCV
Holdings	-0.0004	-0.0002	0.002^{***}	-0.001^{*}
	(0.001)	(0.0003)	(0.001)	(0.0004)
Breadth	-1.987^{***}	-1.651^{***}	-1.273^{***}	-1.888^{***}
	(0.323)	(0.147)	(0.085)	(0.129)
Monitors	1.201^{***}	0.939^{***}	0.350^{***}	1.321^{***}
	(0.207)	(0.172)	(0.118)	(0.145)
Dedicated	0.267^{***}	0.726^{***}	0.175^{***}	0.946^{***}
	(0.080)	(0.134)	(0.028)	(0.190)
Observations	45,077	32,148	32,999	44,226
Adjusted R^2	0.330	0.463	0.342	0.457
Fixed effects	Yes	Yes	Yes	Yes

Table B.11: Brokerage closures: 6 month VCV

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 occurring in month 0.

	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

Table B.12: Brokerage closures: Matched control group

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 firms matched by firm size and analyst coverage in the calendar year prior to the event. For all 7,056 stocks, we compute VCV over the months [-14,-3], and over the months [3,14], with the brokerage closure occurring in month 0.

	Full sample	Coverage < 10	Coverage < 5
	VCV	VCV	VCV
After	-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

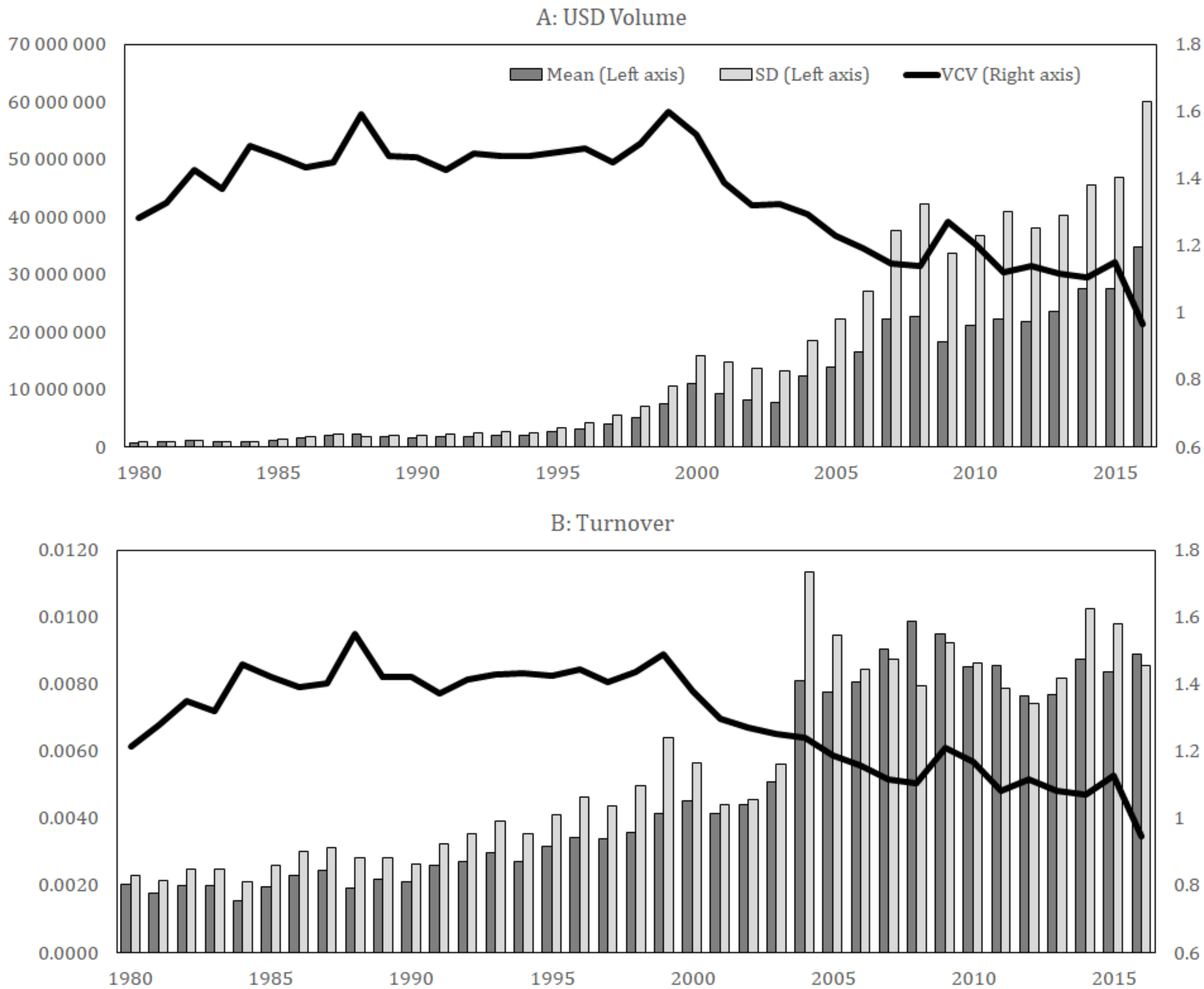

Figure B.1: This figure reproduces Figure 4 in the paper, with the exception that the annual cross-sectional average of firm-level estimates of the mean, standard deviation, and coefficient of variation (VCV) of volume are not based on volume market shares, but on US dollar volume (panel A) and turnover (panel B). Although the means and standard deviations (gray bars) are very different for the three volume measures, VCV (black line) is highly similar.

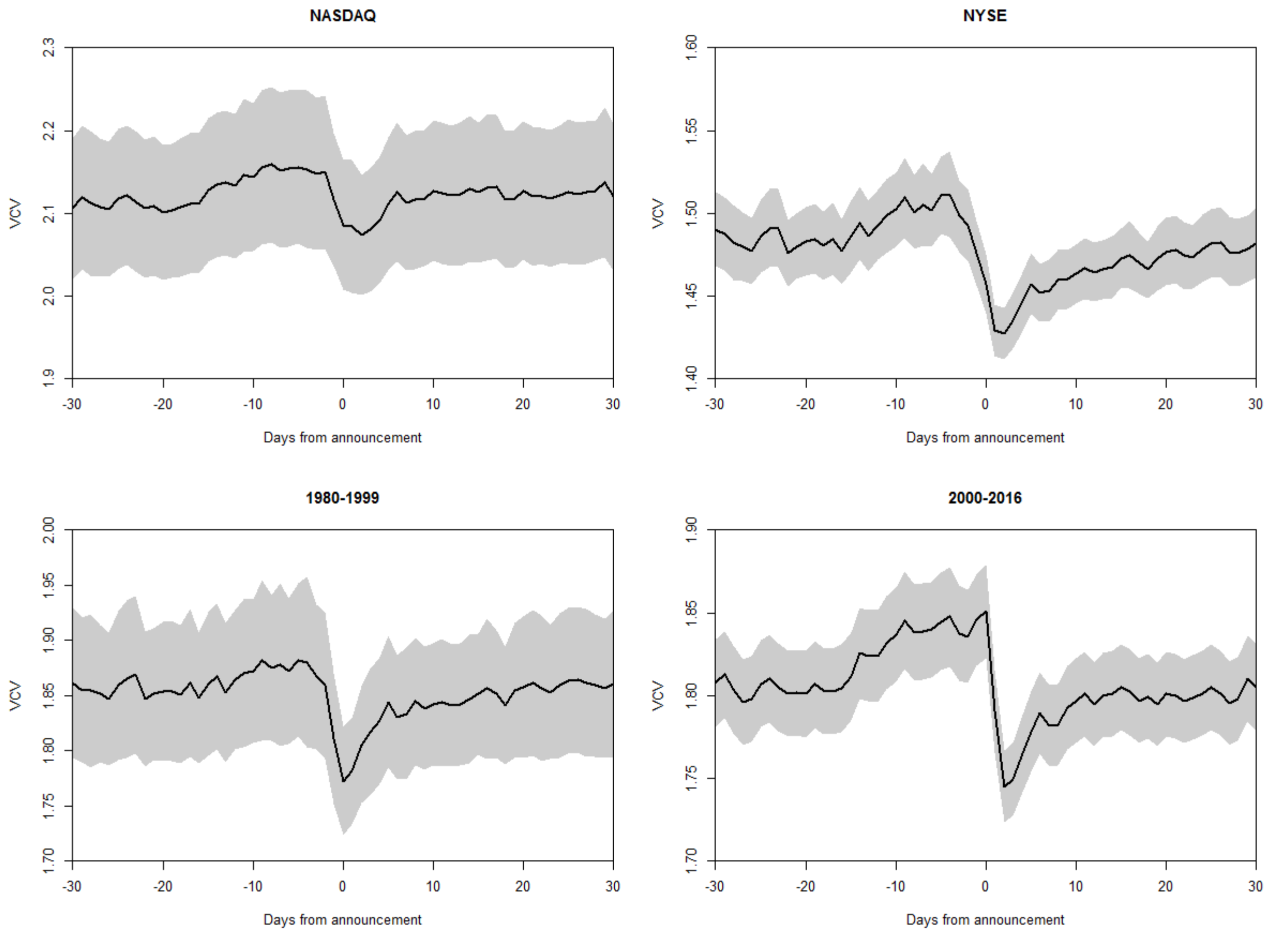

Figure B.2: This figure reproduces Figure 5 in the paper, for subsamples of stocks listed on NASDAQ and stocks listed on NYSE/AMEX, and for subsamples of observations prior to 2000 (1980-1999) and post 2000 (2000-2016).