Asymmetric Information and the Distribution of Trading Volume[1]

Matthijs Lof
Aalto University School of Business
Helsinki, Finland.
matthijs.lof@aalto.fi

Jos van Bommel
Luxembourg School of Finance
University of Luxembourg, Luxembourg
jos.vanbommel@uni.lu

December 2016

ABSTRACT

We introduce a new measure of information asymmetry in security markets: the trading volume coefficient of variation (VCV). Using the Kyle (1985) model, we show that this easily computable ratio is a function of the proportion of informed trade. We compute the VCV for a cross-section of stocks from time-series of daily volume market shares and find that the VCV correlates strongly with extant measures of informed trade. We also compute time series of VCV from the cross-sections of relative trading volumes to estimate information asymmetries in event- time, and find a clear spike of the volume coefficient of variation around earnings announcements.

JEL Classification: D82, G12, G14

Keywords: trading volume, informed trading, adverse selection, earnings announcements.

[1] We thank David Easley, Eliezer Fich, Jillian Popadak, Pete Kyle, Carlos Forner Rodríguez, Enrique Sentana, Conall O'Sullivan, Christian Westheide and participants at the Spanish Finance Association meetings, Multinational Finance Society, German Finance Association, Aalto University, the University of Amsterdam, and the University of Maastricht for constructive comments.

1. Introduction

In this paper we show that the proportion of informed trade affects the distribution of trading volume within the context of the seminal market micro structure model of Kyle (1985). We model a trading interval as Kyle-auction, where liquidity seekers who are either informed or uninformed, submit orders to competitive liquidity providers (market makers) who match buys and sells, and absorb the order imbalance conditional on which they set the clearing price. Liquidity seeking orders can either be buys or sells. Because many uninformed orders cancel out against each other, the total observed trading volume is likely to be significantly larger than the net order flow. The double-counted trading volume is the sum of all buy and sell orders submitted by the liquidity seekers, plus the net orderflow absorbed by the market makers. This observation, and the assumptions that uninformed orders are uncorrelated while informed trade is correlated, leads to simple expressions for the first two moments of the trading volume as functions of the proportion of informed trade and of the number of liquidity seekers. We show that the coefficient of variation of the total trading volume continuously increases in the proportion of informed trade orders. We introduce the volume coefficient of variation ('VCV') as a new and very easily computable measure of the proportion of information trade.

The intuition of our measure is that the distribution of the trading volume depends on the correlation of individual orders. If all liquidity seeking market participants are uninformed and uncorrelated, most orders are netted out against each other, and net order flow is relatively low compared to the observed trading volume, which follows a Normal-like, slightly skewed distribution on $\boldsymbol{R}^{+}$. In the presence of informed (and correlated) liquidity seeking demand, the net orderflow will take up a higher and increasing share of the total trading volume, which will become more skewed. We analyze the volume distribution as a function of the proportion of informed trade, and find that the shape of the trading volume distribution is a tell-tale of the proportion of informed trade.

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 (1987, 1992), Admati and Pfleider (1989), Foster and Viswanathan (1994), and many others have increased our understanding of the strategic behavior of asymmetrically informed traders and their effect on security markets. There has been no shortage of subsequent papers that aim to measure information asymmetries in security markets. Stoll (1978), Huang and Stoll (1997), and Madhavan *et al.* (1997) provided methods to extract the adverse selection component from bid ask spreads, and other microstructure data.

Easley, Kiefer, O´Hara and Paperman (1996) develope a measure for the probability of informed trading, the well-known PIN measure.[2] The PIN has been widely used to study information asymmetries in the finance literature.[3] The measure has also received criticism,

---

[2] Easley et al. (1996) 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 estimator is computed from individual orders that are classified into buys and sells with a computationally costly maximum likelihood procedure.

[3] Easley et al. (1997a and 1997b) analyze the information content around trade lags and trade size. The PIN is used to used to study the impact of analyst coverage on informational content (Easley *et al.*, 1998)), stock splits (Easley *et al.*, 2001), dealer vs. auction markets (Heidle and Huang, 2002), trader anonymity (Gramming *et al.*, 2001), information disclosure (Vega, 2006), 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), a.o.

among others from Duarte and Young (2009), who argue that the (unadjusted) PIN is more a measure of illiquidity and the lack of symmetric order orderflow rather than of the proportion of informed trade.[4] Other critiques focus on the trade classification (Boehmer *et al.* 2007) and the estimation robustness, particularly in high-turnover stocks (Lin and Ke (2011) and Yan and Zhang (2012). In response to these latter critiques, and the advent of high frequency trading, Easley et al. (2012a) developed the volume synchronized PIN, or VPIN. This estimator captures not only information asymmetry but also order flow toxicity, the risk of unbalanced orderflows. The VPIN is computed over volume-time rather than clock time and used bulk-classification of trades to decrease the computational burden. More recently, Bongaerts *et al.* (2016) suggested the XPIN, a measure that looks at the relative product of the price impact and order imbalance.[5]

One difficulty of the PIN and its variations is that they are often difficult to estimate and interpret, because they need signed orderflows derived from microstructure data. Our VCV measure only requires total observable trading volume, and can be computed from time-series data, if we are interested in the proportion of informed trade of a cross section of securities or portfolios, or from a cross section of trading volume data, if we want to study information asymmetry in event- or calendar time. In this paper we illustrate that both cross-sectional VCVs and time-series VCVs have significant statistical power.

[4] Other papers in the debate on the validity of the PIN include Easley et al. (2010) and Akay et al. (2012).

[5] Bongaerts et al. (2016) develop a model in which informed traders take into account their price impact and when buying (selling) rebalance their portfolios with offsetting trades of which the aggregate price impact. They test their private information measure by looking at return reversals, a.o.

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

To further analyze the relation between the proportion of informed trading and the volume distribution, and to gain insight in the small sample properties of the VCV, we conduct a Monte Carlo analysis. We see that the generated volume distributions change markedly as a function of the proportion of informed trade, in that with more informed (correlated) orders, the volume distribution becomes more skewed. We confirm that the simulated VCV increases in the proportion of informed trade in a continuous fashion, and that it is virtually independent of the number of liquidity seekers.

For our empirical analysis, we compute the VCVs from daily volumes of all NYSE and AMEX stocks from 1982 until 2014, obtained from CRSP. Our cross-sectional analysis shows that the VCV, computed from the daily trading volume of individual stocks, is significantly correlated with various PIN measures. We find in particular a strong correlation with the adjusted PIN measure by Duarte and Young (2009), which measures information asymmetry net of other sources of illiquidity. In addition, we find the VCVs to be lower for stocks that have higher Amihud liquidity ratios, expected trading volume, analyst coverage, and institutional ownership. The latter result is consistent with Boone and White (2015), who find that institutional ownership is associated with improved disclosure of information, and therefore lower information asymmetry.[6]

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

[6] Chordia et al. (2002) also investigate the coefficient of variation of trading volume, without relating the measure to informed trading. Contrary to their priors, they find that it is a negatively priced factor.

when the resolution of information asymmetry between insiders and outsiders is most significant.

Finally, we show VCV predicts the post-earnings-announcement drift: Consistent with Sadka (2006), this drift is found to be stronger when there is a relatively high proportion of uninformed noise traders (i.e. when VCV is low).

The remainder of this paper is organized as follows. Section 2 presents the theory. Section 3 contains a Monte Carlo simulation exercise. Section 4 compares the VCV empirically to alternative measures of asymmetric information in the cross section of US stocks. In section 5, we study the time series dynamics of the VCV around earnings announcements. Section 6 summarizes and concludes.

2. Theory

To develop our measure of informed trade we postulate three kinds or traders in the market: (i) informed liquidity seekers, (ii) uninformed liquidity seekers, and (iii) competitive market makers. We assume that there are M individual liquidity seekers, of which a proportion η is informed. Both M and ηM are integers. The M individual traders submit orders to the market where they are matched and where the left over orders (henceforth referred to as 'order imbalance' or 'net order flow') are taken up by the market makers who set the price. The model thus closely resembles that of Kyle (1985), with the only difference that we consider the individual orders of the liquidity seekers, and analyze the total trading volume, not only the net orderflow.

To be more precise, we denote $\tilde{y}_i$ the individual demands of the liquidity seekers. We assume these individual demands follow Normal distributions. Without loss of generality, we let the demand of every liquidity seeker, whether informed or uninformed, to have zero mean and variance σ^2.[7]

The order imbalance, which follows a Normal distribution around zero as in Kyle (1985), is often not observable because it is difficult to empirically distinguish liquidity seekers from liquidity suppliers. This is particularly the case when only the aggregate volume data is available rather than transaction-level trading data. Total volume is much easier and more precisely to measure, and often readily available. In our model the total trading turnover can be written as:

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

The term in brackets is the "double counted transaction volume", where both buys and sells are counted. In addition to the trades amongst liquidity seekers, where every order is counted, to arrive at the double counted volume, we have the trades of the market makers who take up the order imbalance, $\sum_M \tilde{y}_i$.

---

[7] The assumption that individual uninformed and informed demands are of the same order of magnitude is innocuous. Letting informed traders´ demands be distributed $N\left(0, k\sigma^2\right)$, and letting $\frac{\eta}{k} M$ individual traders be informed leads to the same conclusions. For this reason we refer to η being the proportion of informed *trading*, rather than *traders*, in the market.

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

We now derive the first two moments of the total trading volume as a function of η. We first observe that the order imbalance follows a Normal distribution, with mean zero, and variance $\eta^2 M^2 + (1-\eta)M$. Using the properties of the Half Normal distribution we then find:

$$E\left[\sum_M |\tilde{y}_i|\right] = M\sigma\sqrt{\frac{2}{\pi}} \tag{2}$$

$$E\left[\left|\sum_M \tilde{y}_i\right|\right] = \sigma\sqrt{\eta^2 M^2 + (1-\eta)M}\sqrt{\frac{2}{\pi}} \tag{3}$$

The variance of the two components are given by:

$$\mathrm{var}\left(\sum_M |\tilde{y}_i|\right) = M\sigma^2\left(1-\frac{\pi}{2}\right) \tag{4}$$

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

We see that, for large *M*, the expected value of the trading volume increases linearly in $(1+\eta)\sigma M$, while the standard deviation increases linearly in $\eta\sigma M$. It can thus be shown that, for large *M*, the volume coefficient of variation only increases in η.

<u>Proposition</u>

Consider a market where M liquidity seeking traders submit market orders with standard deviation σ and where the net order flow is absorbed by liquidity suppliers. If a proportion η of the traders are informed,

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

ii) For large M, the relationship converges to:

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

From the above analysis we can also conclude that we can find a consistent estimator for η by rewriting (6). In particular, we have:

<u>Proposition 2</u>

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

$$VCV \equiv \frac{\hat{\sigma}_V}{\hat{\mu}_V} \qquad (7)$$

is an estimator of which the expectation increases monotonically in η.

$$\hat{\eta} \equiv \frac{\hat{\sigma}_V}{\hat{\mu}_V \sqrt{2\pi - 4} - \hat{\sigma}_V} \qquad (8)$$

is, for large M, a consistent estimator of η.

To better understand the small scale properties of the VCV and $\hat{\eta}$ estimators, we conduct a Monte Carlo simulation. The analysis in the next section reveals that although $\hat{\eta}$ is a consistent estimator for distributed orders that follow the Normal distribution and large *M*, the VCV is the more reliable indicator for the proportion of informed trade for small and heterogenous samples. The main problem for the $\hat{\eta}$ estimator is that for small samples or excessively noisy observations, the denominator can become close to zero or negative.

3. Monte Carlo Simulation

In this section we analyze the trading volume distribution for our trading model, for different values of *M* and η. To do this, we draw $1+(1-\eta)M$ random variables from $N(0,1)$ to simulate the individual demands. The first variable is multiplied by η*M*, and represents the aggregate informed demand. The remaining variables represent the individual uninformed demands. We compute the observed trading value volume $\tilde{V}$ that follows from (1). For each (*M*, η) pair we generate 10,000,000 $\tilde{V}$ observations.

---- Figure 1 around here ----

Figure 1 displays six histograms of simulated volumes for *M* = 1,000 and different values of η. The simulation confirms the analysis in the previous section: if there are no informed traders, the volume distribution follows a slightly skewed bell-curve, while in the case of only informed traders the volume follows a truncated normal.[8] For intermediate values, the distributions take intermediate shapes.

At first sight, one may be inclined to compare the skewness coefficients of the volume distribution. We find however that the skewness coefficient is a very poor indicator of η. See Table 1 and Figure 2, where we report relevant measures of the simulated volume distribution for different values of *M*. From table 1 we see that, for a given *M*, the skewness coefficient reaches its maximum value for relatively low levels of η.[9]

---- Table 1 around here ----

---- Figure 2 around here ----

The sample VCV and $\hat{\eta}$ however behave much better. Not only do both estimators increase monotonically in η, but more importantly, for most parameter values they are very insensitive to *M*. Only for very small values of η does the number of liquidity seekers *M* affect the VCV and $\hat{\eta}$. The insensitivity to *M*, which is due to the relative weight of the informed volume, of which the shape is independent of *M*, is encouraging as it implies that

[8] The slightly skewed bell-curved volume distribution for $\eta=0$ converges to the distribution of the maximum of two Normally distributed random variable, which was first described by Clark (1961): For large *M*, the total buy demand and the total sell demand approach Normal distributions due to the Central Limit Theorem. Clark (1961) showed that the Maximum of two Normals is not Normal and gives the first four moments.

[9] It can be shown that the skewness coefficient converges to $\frac{(4-\pi)\sqrt{2}}{(\pi-2)\sqrt{\pi-2}} \approx 0.995$.

there is little concern for confounding a high η with a low *M*. The insensitivity to M is desirable because often the number of order submitters in market sessions is unknown.[10]

We recognize that in practice η is unlikely to be constant across observations, and that we are typically interested in estimating the average proportion of informed trade, over either a time series, or over a cross section of observations. To gauge the precision of our estimators we conduct a second Monte Carlo analysis where we let η be random across observations. In addition, we are interested in the small sample properties of our estimators. In the second Monte Carlo analysis we investigate the reliability of our estimators for small samples, and non-constant η.

We simulate 1,000,000 samples of either 100 or 10 volume observations, and compute the VCVs and $\hat{\eta}$s. The first panel of table 2 gives our benchmark case where the sample size is 100 observations, each one generated from 100 market participants where η is fixed across (generated) observations. From the first two panels of Table 2, we see that both VCV and $\hat{\eta}$ are reliable estimators. Both measures increase in η, as expected, and the standard deviations (which can be interpreted as standard errors) are low compared to the differences in η-scenarios. Naturally the standard deviations across generated samples is higher for smaller sample. In addition we detect small biases of the $\hat{\eta}$ estimator for samples of 10 observations.

[10] Although some data-providers identify individual transactions, this number is different from the number of order-submitters because orders may be broken up.

When we analyze more realistic scenarios where η varies across observations, a very different picture emerges. Panel B of table 2 gives the results for the case where we pick the number of informed liquidity seekers from a Uniform distribution and in panel D, we let the informed order flow be zero or five times the benchmark case with probability ⅕.

---- Table 2 around here ----

As can be seen from panels C and D of table 2, for non-constant η, the $\hat{\eta}$ diverges dramatically from the true average η. In addition, the $\hat{\eta}$'s precision is significantly lower than that the of the VCV, and may easily take on negative values. It is not difficult to see that this is due to the fact that the $\hat{\eta}$'s denominator can easily take on small positive or negative numbers, which makes the estimator very imprecise.

In the remainder of this paper, we therefore focus on the VCV as measure of proportion of informed trade.

4. The Cross-Section of VCV

After having established from theoretical and numerical analysis a positive monotonic relation between VCV and the proportion of informed traders, we now turn to empirics. In this section, we describe cross-sectional variation in VCV for US stocks, while in the next section we study the time-series behavior of VCV. We compute annual Volume Coefficients of Variation (VCV) for US stocks and compare these figures to alternative measures of informed trading and illiquidity. We obtain daily trading volumes (number of shares traded multiplied by the closing price) from the CRSP daily stock file for all common stock listed on NYSE and AMEX over the period 1982-2014 (we exclude

NASDAQ firms from our sample in order to avoid biased caused by the differences in market structure). We disregard the most infrequently traded stocks by only considering stocks that had positive trading volume on at least 200 days in the previous calendar year.

The VCV measure is computed by dividing the annual standard deviation of daily trading volume by the annual mean of daily trading volume. We recognize that our theory assumes that observations come from the same underlying distribution, where the expected trading volume and the proportion of informed trade is constant. Our Monte Carlo analysis showed that a violation of this assumption decreases the estimator's precision. Therefore, if we want to gauge the proportion of informed trading in a security, we need to look at the volume distribution of a time series of observations where expected volume and proportion of informed trading is relatively constant. For this reason, we also compute the annual VCV of daily volume *shares* (defined as daily volume of a stock divided by total market volume on the same day) and daily *turnover* (defined as daily volume divided by market capitalization). For these three variables (daily volume, volume shares, turnover) we also compute the annual skewness coefficient for all three volume definitions.

Tables 3 show summary statistics and correlations for these six measures of the volume distribution. The average of the three VCV measures are very close to each other, and they are also highly correlated. In the remainder of this paper, our measure of informed trading VCV is defined as the annual coefficient of variation of daily volume shares. Similar results are obtained for the other VCV measures. Table 3 also shows substantial variation of our VCV measure, both in the cross section, as in the time series. We also find that in the time series VCV measures are highly autocorrelated, indicating that they are meaningful characteristics of the stocks in our sample.

---- Table 3 around here ----

---- Table 4 around here ----

Table 4 shows the correlations between VCV and various annual PIN measure for US stocks. We do not calculate PIN measures ourselves, but make use of the various annual PIN measures kindly made publicly available by the authors of previous studies. PIN (EHO) refers to the PIN measures estimated by Easley et al. (2010).[11] PIN (BHL) refers to the PIN measure estimated by Brown, Hillegeist and Lo (2004). PIN (BH) refers to the PIN measure estimated by Brown and Hillegeist (2007).[12] PIN (DY) refers to the PIN measure estimated by Duarte and Young (2006), who also derive two related measures: Adjusted PIN, which is according to Duarte and Young (2006) a cleaner measure of asymmetric information; and PSOS (probability of symmetric order-flow shock), which is a measure of illiquidity unrelated to asymmetric information.[13] Table 4 shows that our VCV measure is positively correlated to all these PIN related measures, suggesting that VCV is, like PIN, indicative of informed trading. Compared to these PIN measures, however, our VCV measure is far easier to compute and does not require intraday data on the order process.

---

[11] Annual stock-specific observations of PIN (EHO) are made available by Søren Hvidkjær (https://sites.google.com/site/hvidkjaer/data)

[12] Annual stock-specific observations of PIN (BH) and PIN (BHL) are made available by Stephen Brown (http://scholar.rhsmith.umd.edu/sbrown/pin-data)

[13] Annual stock-specific observations of PIN (DY) and Adjsuted PIN and PSOS are made available by Jefferson Duarte (http://www.owlnet.rice.edu/~jd10/)

---- Table 5 around here ----

Table 5 shows the correlation between VCV and various other stock characteristics. Overall, these correlations lend support to the notion that VCV is indicative of the presence of informed trading: VCV is negatively correlated to Size (market capitalization at the last trading day of June) and Turnover and positively correlated to the Illiquidity measure by Amihud (2002). We compute illiquidity, size and turnover using CRSP data. These results are not surprising since information asymmetry is likely to be predominant in smaller stocks and asymmetric information reduces liquidity. Table 6 also shows the correlation between analyst coverage, defined as the logarithm of one plus the number of distinct analysts covering a stock in a given year (Source I/B/E/S). Analyst coverage is likely to reduce information asymmetries, which is confirmed by the negative correlation with VCV. Finally, Table 5 includes various indicators of institutional ownership (Source: 13 F filings – we obtain similar results when we obtain holding data from the CRSP mutual fund database). These measures include institutional holdings (defined as the percentage of shares of a firm held by institutional investors at the end of the year) and breadth of ownership (defined as the number of institutional investors holding shares in the firm, as a percentage of the total number of institutional investors at the end of each year – Chen et al., 2002). Boone and White (2015) claim that institutional ownership leads to higher transparency and therefore lower information asymmetry. We find indeed that these measures on mutual fund ownership are negatively correlated to VCV. In addition, we consider the amount of monitoring institutions: Following Fich et al (2015), institutional investor x is a monitor for firm y if firm y belongs to the top 10% of holdings in the institution's portfolio. The variable monitors in table 5 is the number of monitoring institutions in the firm as a percentage of the number of total institutional investors. These monitoring investors are presumably better informed relative to non-monitoring investors. Although VCV is negatively correlated with the number of monitoring investors, it becomes clear below that the partial correlation is strongly positive when controlling for the other institutional ownership characteristics.

---- Table 6 around here ----

In Table 6 we reconsider these correlations in a regression context, while controlling for multiple firm characteristics and year fixed effects. The first column in Table 7 shows the result of regressing VCV on size, illiquidity and turnover and industry. In the second column, the three PIN type measures by Duarte and Young (2006) are added to the regression. Duarte and Young show that the standard PIN indicator measures both asymmetric information and illiquidity unrelated to asymmetric information. They decompose the PIN measure into PSOS, which is a liquidity measure unrelated to asymmetric information, and *adjusted* PIN, which is a supposedly cleaner measure for asymmetric information net of other sources of illiquidity. The regression results indicate that VCV is strongest associated with adjusted PIN and is not significantly related to PSOS, thereby supporting our claim that VCV, like adjusted PIN, measures asymmetric information rather than general illiquidity. The third column shows that VCV is lower for stocks with high breadth of ownership, consistent with Boone and White (2015), while the VCV is higher when there is a high fraction of monitoring institutions among the shareholders. The final column shows that VCV is lower for stocks with high analyst coverage.

---- Table 7 around here ----

5. VCV around earnings announcements

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

We obtain quarterly earnings announcement dates from COMPUSTAT for all NYSE and AMEX listed US firms. We compute the VCV over 10 day windows before and after the announcement, while skipping the 5 days closely around the announcement. Hence, the before window includes days -12:-3 and the after window includes days 3:12, where day 0 is the announcement date. [14] Table 7 shows some summary statistics of the VCV in each window. The table reports that the median VCV in the before window is higher than in the after window. Also the average difference is negative and statistically significant. The table further shows that the number of firms for which the VCV decreases following the announcement is around 52%. Although this number is only slightly higher than 50% (which it would be if earnings announcements have no effect on information asymmetry),

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

the difference is statistically significant: A reduction in VCV (information asymmetry) around earnings announcements is significantly more common than an increase in VCV (information asymmetry). For each earnings announcement date, we also obtain a *placebo date*, which is a date within 100 trading days before or after the actual announcement, randomly drawn from a uniform distribution. Around these placebo dates, the fraction of firms that see a reduction in VCV is approximately 50%.

---- Table 8 around here ----

The second and third column of Table 8 show the same statistics for surprising and non-suprising announcement separately. We choose a market-based definition of surprise: An announcement is classified as surprising when the absolute value of the cumulative abnormal return over days -1, 0 and 1 around the announcement exceeds its median: $|CAR_{-1:+1}| > median(|CAR_{-1:+1}|)$. Abnormal returns are defined as the stock's return minus the market return on the same day. As expected, patterns in VCV around announcement as described above around announcement dates are strongest when the announcement is surprising. That is, when the announcement is surprising, the announcement is more informative and is more effective in reducing information asymmetry, which is reflected in stringer variation in the VCV.

Computing the VCV over 10 day windows is not optimal. First of all, the small sample causes the estimates to be noisy, which explains the small (although significant) changes in Table 8. More worrying is that table 8 is showing that the distribution of trading volume changes around the announcement, hence there is no a priori reason to assume that the distribution does not change *within* the 10 day windows before and after the announcement.

For this reason, we compute the so-called *cross-sectional VCV* for each day around the announcement day. For all announcements we calculate the coefficient of variation using this cross section of daily trading volumes (all volumes are as before volume shares, i.e. volumes as a percentage as total trading volume on that calendar date). This cross sectional VCV is then computed for all days over the interval -30 days before the announcement to +30 days after the announcement. The black line in Figure 3 shows the pattern of VCV over this interval, while the shaded areas indicate 90% confidence bounds. Clearly, the VCV increases when the VCV increases, as uninformed investors are delaying their trading activity. When information asymmetries are resolved at the announcement date, the VCV sharply declines and stays relatively low for around 10-15 trading days (i.e. 2-3 weeks). The VCV 30 trading days after the announcement is roughly equal to the announcement 30 trading days prior to the announcement. The lower panel shows the same graph for the placebo earnings dates also used in Table 8. This panel shows a flat line as expected.

6. VCV, momentum and the post-earnings announcement drift.

In Table 9 we show that VCV predicts the Post Earnings Announcement Drift (PEAD). Our measure of PEAD is the cumulative abnormal returns over days +2 to +21 after the earnings announcement: $CAR_{+2:+21}$. The first column of Table 9 shows that these post announcement returns are positively predicted by the cumulative returns on the 3 days around the announcement: $CAR_{-1:+1}$. This is the basic idea of PEAD: after positive (negative) news, measured by positive (negative) abnormal returns on the days around the announcement, returns drift upward (downward) for the next month. The second column shows that this drift is significantly mitigated when VCV is high, i.e. when there are more informed traders (and therefore less uninformed noise traders).

---- Table 9 around here ----

This result is consistent with prior literature relating PEAD to information asymmetry. Hung et al (2014), show that the PEAD reduces globally after countries adopted the International Financial Reporting Standards (IFRS) in 2005 leading to improved financial reporting quality. Sadka (2006) who also finds that returns on a PEAD strategy are lower when the number of informed traders is high relative to the number of noise traders (measured by the exposure of returns the so-called permanent-variable component of liquidity risk).

Sadka (2006) finds the same pattern in momentum profits: past winners outperform past losers, in particular for stocks with a relatively low ratio of informed traders. Our results are consistent with this finding. In Table 10, we regress one month excess returns, on the excess returns over the prior year, while skipping the last month (i.e. the prior year includes month -12 to -2). The positive coefficient indicates the existence of stock return momentum. After we include a dummy variable indicating a high VCV (equal to one if VCV over the past 12 months exceeds the median VCV over the past 12 months and zero otherwise), and an interaction term to the regression, we find that the momentum profits (i.e. persistence in stock returns) is *lower* for firms with a *high* VCV. Just like PEAD, momentum is highest when the stock is traded mostly by uninformed noise traders.

---- Table 9 around here ----

6. Conclusions

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

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

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

References

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

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

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

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

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

Black, F. (1986). Noise. The journal of finance, 41(3), 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.

Boone, A. and Joshua T. White (2015), The effect of institutional ownership on firm transparency and information production, *Journal of Financial Economics*, 117 (3), 508-533.

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

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

Brown, S., Hillegeist, S. A., & Lo, K. (2004). Conference calls and information asymmetry. *Journal of Accounting and Economics*, *37*(3), 343-366.

Chae, J. (2005). Trading volume, information asymmetry, and timing information. *The journal of finance*, 60(1), 413-442.

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

Chen, J., Hong, H., & Stein, J. C. (2002). Breadth of ownership and stock returns. *Journal of financial Economics*, 66(2), 171-205.

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

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

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

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

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

Easley, D., Hvidkjaer, S., and O'Hara, M. (2010). Factoring information into returns. *Journal of Financial & Quantitative Analysis*, *45*(2).

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., 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., Lopez de Prado, M. and O'Hara, M. (2012). Flow toxicity and liquidity in a high frequency world. , 25, 1457-1493.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Milgrom, P., & Stokey, N. (1982). Information, trade and common knowledge. *Journal of Economic Theory*, *26*(1), 17-27.

Sadka, R. (2006), Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk. *Journal of Financial Economics*, *80*(2), 309-349.

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

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

Wang, J. (1994), A model of competitive stock trading volume. *Journal of political Economy*, 102 (1) 127-168.

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

TABLE 1: This table summarizes the results of a Monte Carlo simulation of Kyle auctions where ηM informed liquidity seekers trade with (1-η)M uninformed liquidity seekers and market makers who absorb the order imbalance. We draw a random number from the standard Normal and multiply it with ηM to model the informed demand, and (1-η)M random variable to model the uninformed orders, and record the trading volume. We do this 1,000,000 times for every (η,M) pair and record the first three moments as well as the the coefficient of variation, and the $\hat{\eta} = \frac{stdev(V)}{\sqrt{2\pi-4}\times mean[V]-stdev(V)}$.

		η										
		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
M = 10	Mean	5.25	5.25	5.37	5.59	5.86	6.17	6.51	6.87	7.22	7.59	7.98
	Stdev	1.52	1.52	1.71	2.08	2.56	3.09	3.67	4.25	4.83	5.42	6.03
	Skew	0.681	0.681	0.744	0.871	0.963	0.996	1.005	1.002	1.002	0.996	1.002
	VCV	0.290	0.290	0.319	0.373	0.437	0.501	0.563	0.618	0.669	0.714	0.756
	$\hat{\eta}$	0.238	0.238	0.267	0.327	0.407	0.496	0.594	0.692	0.794	0.895	1.000
M = 100	Mean	43.88	45.39	48.64	52.31	56.14	60.02	64.01	67.90	71.84	75.82	79.79
	Stdev	4.44	6.90	12.32	18.19	24.17	30.15	36.19	42.23	48.27	54.29	60.30
	Skew	0.476	0.836	1.006	1.004	0.998	0.995	1.000	1.002	1.003	0.997	0.993
	VCV	0.101	0.152	0.253	0.348	0.430	0.502	0.565	0.622	0.672	0.716	0.756
	$\hat{\eta}$	0.072	0.112	0.201	0.299	0.398	0.498	0.598	0.700	0.801	0.901	1.001
M = 1000	Mean	411.6	440.7	479.4	519.0	558.9	598.6	638.2	678.9	718.1	758.8	796.8
	Stdev	13.7	60.7	120.6	180.8	241.3	301.3	361.3	422.5	482.2	542.6	602.4
	Skew	0.391	1.004	0.997	1.014	1.002	0.995	1.001	1.002	0.993	0.999	0.994
	VCV	0.033	0.138	0.251	0.348	0.432	0.503	0.566	0.622	0.671	0.715	0.756
	$\hat{\eta}$	0.022	0.100	0.200	0.300	0.400	0.499	0.599	0.700	0.800	0.898	1.001

TABLE 2: This table summarizes the results of a Monte Carlo simulation of trading sessions where informed and uninformed liquidity seekers submit orders to market makers. We draw a random number from the standard Normal and multiply it with the number of informed traders to model the informed demand, and additional random numbers for each informed trader. We generate 1,000,000 samples of 100 res. 10 volume observations and record the mean, median, standard deviation, and 5% and 95% percentiles of the VCV and $\hat{\eta}$. In Panel A the number of informed investors is constant across observations.

Panel A: η constant across observations

		uninformed traders	90	80	70	60	50	40	30	20	10
		informed traders	10	20	30	40	50	60	70	80	90
		η	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Sample size 100	VCV	Mean	0.151	0.252	0.349	0.432	0.501	0.568	0.620	0.670	0.714
		Median	0.151	0.252	0.349	0.431	0.500	0.567	0.619	0.668	0.713
		stdev	0.012	0.019	0.025	0.029	0.034	0.039	0.043	0.047	0.051
	$\hat{\eta}$	Mean	0.111	0.201	0.301	0.401	0.498	0.605	0.701	0.803	0.903
		Median	0.111	0.200	0.300	0.400	0.495	0.600	0.694	0.793	0.892
		stdev	0.010	0.018	0.028	0.038	0.051	0.067	0.083	0.102	0.123
Sample size 10	VCV	Mean	0.146	0.242	0.336	0.415	0.483	0.550	0.603	0.653	0.699
		Median	0.143	0.239	0.331	0.410	0.477	0.542	0.594	0.644	0.688
		stdev	0.038	0.059	0.078	0.094	0.108	0.122	0.136	0.148	0.161
	$\hat{\eta}$	Mean	0.108	0.193	0.291	0.390	0.488	0.600	0.708	0.828	0.962
		Median	0.105	0.188	0.281	0.373	0.462	0.559	0.648	0.742	0.835
		stdev	0.031	0.057	0.089	0.126	0.170	0.230	0.389	0.428	1.087

TABLE 2 (continued): We draw one random number from the standard Normal and multiply it with the number of informed traders to model the informed demand, and additional random numbers for each informed trader. We generate 1,000,000 samples of 100 volume observations and record the mean, median, standard deviation, and 5% and 95% percentiles of the VCV and $\hat{\eta}$. In panel C the number of informed traders is uniformly distributed. In panel D the number of informed investors is either zero or five times the benchmark.

Panel B: number of informed traders follows a Uniform distribution, sample size 100

	uninformed	90	80	70	60	50	40	30	20	10
	informed	U[0,20]	U[0,40]	U[0,60]	U[0,80]	U[0,100]	U[0,120]	U[0,140]	U[0,160]	U[0,180]
	E[η]	0.097	0.19	0.28	0.36	0.45	0.54	0.63	0.73	0.84
VCV	Mean	0.189	0.336	0.471	0.588	0.684	0.778	0.851	0.922	0.984
	Median	0.188	0.335	0.470	0.586	0.681	0.775	0.848	0.919	0.980
	stdev	0.020	0.033	0.044	0.050	0.058	0.064	0.070	0.075	0.081
$\hat{\eta}$	Mean	0.143	0.287	0.456	0.643	0.836	1.078	1.316	1.611	1.944
	Median	0.142	0.285	0.451	0.635	0.821	1.054	1.278	1.552	1.844
	stdev	0.017	0.036	0.062	0.091	0.131	0.186	0.254	0.349	0.481

Panel C: number of informed traders either zero or (with prob. 1/5) five times the benchmark, sample size 100

	uninformed	90	80	70	60	50	40	30	20	10
	informed	B(1/5,50)	B(1/5,100)	B(1/5,150)	B(1/5,200)	B(1/5,250)	B(1/5,300)	B(1/5,350)	B(1/5,400)	B(1/5,450)
	E[η]	0.07	0.11	0.14	0.15	0.17	0.18	0.18	0.19	0.20
VCV	Mean	0.423	0.788	1.113	1.399	1.650	1.879	2.078	2.268	2.452
	Median	0.423	0.792	1.115	1.394	1.640	1.866	2.062	2.243	2.419
	stdev	0.070	0.097	0.117	0.128	0.143	0.169	0.199	0.238	0.289
$\hat{\eta}$	Mean	0.395	1.130	3.210	-13.145	-60.061	-5.608	-4.081	-3.240	-2.783
	Median	0.389	1.102	2.813	7.643	-8.869	-5.235	-3.740	-3.063	-2.664
	stdev	0.090	0.288	1.414	37.728	34.909	3.175	1.218	0.752	0.569
	5%-ile	0.250	0.699	1.541	-55.570	-61.910	-13.659	-6.646	-4.747	-3.898
	95%-ile	0.546	1.646	6.194	68.543	52.930	-3.213	-2.639	-2.272	-2.027

TABLE 3: Panel A reports summary statistics of the Annual Volume Coefficient of Variation (*VCV*) of daily dollar trading *Volume*, daily volume shares (*Volume %* - daily dollar volume as a percentage of total market dollar volume), and *Turnover* (dollar volume as a fraction of market capitalization). The table reports N; the number of distinct stocks in the sample, T; the number of time-series observations (years), mean, standard deviation, *sd (CS)*, the time-series average of annual cross-sectional standard deviations, *sd (TS)*, the cross-sectional average of stock-specific time-series standard deviations, min, max, median and 1st order autocorrelation (ρ).

Panel B reports the different VCV measures. The upper diagonal entries show the correlations for levels. The lower diagonal entries show within-year rank correlations. Sample: 1982-2014. Source: CRSP.

A: Summary statistics

	N	*T*	*mean*	*sd*	*sd (CS)*	*sd (TS)*	*min*	*max*	*median*	*ρ*
Volume	5939	33	1.17	0.68	0.66	0.46	0.24	12.98	1.03	0.50
Volume %	5939	33	1.15	0.70	0.67	0.47	0.16	14.00	1.02	0.51
Turnover	5939	33	1.14	0.66	0.63	0.45	0.23	10.18	1.01	0.50

B: Correlation

	Volume	*Volume %*	*Turnover*
Volume		0.98	0.97
Volume %	0.98		0.95
Turnover	0.97	0.96	

TABLE 4: This table shows the correlation between Annual stock-specific Volume Coefficients of Variation (*VCV*) of daily dollar volume and various annual PIN measures (probability of informed trading). *PIN (DY)*, *Adjusted PIN*, and *PSOS* are the PIN, liquidity-adjusted PIN and liquidity measure PSOS estimated by Duarte and Young (2009). *PIN (EHO)* is the PIN measure estimated by Easley, Hvidkjaer, and O'Hara (2010). *PIN (BHL)* is the PIN measure estimated by Brown, Hillegeist and Lo (2004). *PIN (BH)* is the PIN measure estimated by Brown and Hillegeist (2007). The upper diagonal entries show the correlations for levels. The lower diagonal entries show within-year rank correlations. Sources: CRSP and cited authors' websites.

	VCV	*PIN (DY)*	*adjusted PIN*	*PSOS*	*PIN (EHO)*	*PIN (BHL)*	*PIN (BH)*
VCV		0.41	0.40	0.32	0.39	0.40	0.52
PIN (DY)	0.55		0.67	0.73	0.76	0.51	0.67
adjusted PIN	0.51	0.70		0.34	0.60	0.44	0.69
PSOS	0.44	0.70	0.38		0.55	0.32	0.42
PIN (EO)	0.51	0.85	0.63	0.62		0.50	0.67
PIN (BHL)	0.51	0.62	0.55	0.43	0.62		0.68
PIN (BH)	0.64	0.70	0.70	0.45	0.67	0.72	

TABLE 5: This table shows the correlation between Annual stock-specific Volume Coefficients of Variation (*VCV*) of daily dollar volume and various annual firm characteristics: *Size* is the log of market capitalization at the last trading day of June. *ILLIQ* is the Amihud Illiquidity measure. *Turnover* is the average of daily trading volume as a percentage of market capitalization. *Coverage* refers to analyst coverage and is equal to log(1+number of analysts) (Source: I/B/E/S). *Institutional holdings %* is the fraction of shares held by mutual funds, *Breadth* is the percentage of mutual funds holding the stock, and Monitors is the fraction of institutional investors in each firm for which the firm is in the top 10% of the institution's holdings (Source: 13F). The upper diagonal entries show the correlations for levels. . The lower diagonal entries show within-year rank correlations.

	VCV	*Size*	*Illiq*	*Turnover*	*Coverage*	*Institutional holdings %*	*Breadth*	*Monitors*
VCV		-0.55	0.60	-0.20	-0.45	-0.43	-0.41	-0.34
Size	-0.66		-0.95	0.26	0.75	0.58	0.73	0.67
Illiq	0.69	-0.96		-0.41	-0.74	-0.67	-0.65	-0.59
Turnover	-0.22	0.29	-0.46		0.25	0.42	0.05	0.07
Coverage	-0.57	0.80	-0.83	0.38		0.41	0.66	0.54
Holdings %	-0.38	0.51	-0.56	0.45	0.42		0.32	0.31
Breadth	-0.67	0.93	-0.94	0.39	0.82	0.64		0.81
Monitors	-0.53	0.78	-0.77	0.26	0.63	0.53	0.78	

TABLE 6: This table shows the results from regressing Volume Coefficients of Variation (*VCV*) of daily dollar volume on various annual firm characteristics: (i) *Size (*log of market capitalization*), Amihud Illiquidity* and *turnover* (ii) *PIN*, *PSOS* and *adjusted PIN* (Duarte and Young, 2009), (iii) I*nstitutional holdings*, *Breadth* and *Monitors* of ownership, (iii) Analyst *Coverage*. All regressions year fixed-effects. T-statistics (based on standard errors clustered at the firm level) in italics.

	VCV	*VCV*	*VCV*	*VCV*
Size	0.13 ***	0.12 ***	0.15 ***	0.14 ***
	12.19	*10.21*	*11.90*	*14.10*
Illiq	0.22 ***	0.20 ***	0.23 ***	0.23 ***
	25.56	*22.11*	*27.81*	*25.28*
Turnover	0.02 ***	0.03 ***	0.02 ***	0.02 ***
	7.95	*8.93*	*7.82*	*7.88*
PIN (DY)		0.17		
		1.35		
PSOS		-0.16 **		
		-2.42		
Adjusted PIN		0.90 ***		
		10.96		
Holdings			0.00	
			-0.59	
Breadth			-1.01 ***	
			-9.09	
Monitors			0.90 ***	
			5.90	
Coverage				-0.02 ***
				-2.76
Observations	58998	39971	47220	39293
Adjusted R-squared	0.38	0.32	0.42	0.4
Year fixed effects	yes	yes	yes	yes

TABLE 8. This table shows nonparametric statistics on the pattern of *VCV* computed over 10 day windows before and after quarterly earnings announcements, excluding the 5 days window closely around the announcement (i.e. the before window includes days -12:-3 and the after window includes days 3:12, where day 0 is the announcement date.) The first two rows show the median *VCV* during windows the before and after the announcement. The third and fourth row show the average change in *VCV* following the announcement, and a t-statistic for the hypothesis that the change is zero. The final rows show the number of firms for which the *VCV* decreases following the announcement and a t-statistic for the hypothesis that this percentage is 50%. The final rows show these percentages for randomly chosen placebo dates within 100 trading days before or after the actual announcement. Column 2 and 3 differentiate surprising and non-surprising announcements. An earnings announcement is considered surprising when the absolute cumulative announcement window return exceeds the median absolute return around earnings announcements ($|CAR_{-1:+1}| > median(|CAR_{-1:+1}|)$).

	All announcements	*Surprising announcements*	*Non-surprising announcements*
Median VCV before announcement	0.53	0.54	0.52
Median VCV after announcement	0.52	0.52	0.51
Mean change in VCV	-0.010	-0.014	-0.005
T-statistic (H0: no change)	-11.55	-12.22	-4.31
% VCV before > VCV after	0.52	0.52	0.51
T-statistic (H0: 50%)	15.11	12.70	8.70
Placebo dates:			
% VCV before > VCV after	0.500	0.500	0.501
T-statistic (H0: 50%)	0.31	-0.15	0.58
Observations	198500	96962	101538

TABLE 9: Post Earnings Announcement Drift. This table shows the result from regressing cumulative abnormal returns following earnings announcement ($CAR_{+2:+21}$) on $CAR_{-1:+1}$, VCV and an interaction term. T-statistics (based on standard errors clustered at the quarter level) in italics.

	CAR(+2:+21)		*CAR(+2:+21)*	
CAR(-1,+1)	0.03	**	0.05	***
	2.13		*3.62*	
VCV			0.00	
			-0.81	
*VCV*CAR(-1,+1)*			-0.03	***
			-2.86	
Quarter fixed effects	yes		yes	
Observations	184731		184731	
Adjusted R-squared	0.01		0.01	

TABLE 10: Momentum. This table shows the result from regressing monthly stock returns (in devation form market returns) on the stock's excess returns over the rpior year (skipping the last month), a dummy indicating *VCV* over the past 12 months higher than the median, and an interaction term. T-statistics (based on standard errors clustered at the month level) in italics.

	R_t	R_t
$R_{t-2:t-12}$	0.004 *	0.007 *
	1.66	*1.89*
High VCV		0.000
		0.11
High VCV$*R_{t-1}$		-0.004 *
		-1.71
Month fixed effects	yes	yes
Observations	590903.00	590903.00
Adjusted R-squared	0.03	0.03

Figure 1: simulated trading volumes for different proportions of informed trading

We simulate trade by letting 100η informed liquidity seekers trade with 100(1-η) uninformed liquidity seekers and market makers. Each liquidity seeker´s demand is drawn from N(0,1). The demands of the informed are perfectly correlated, the demands of the uninformed are independent. In the top graph we depict 6 histograms of 10,000,000 simulated trading volume observations for different η´s.

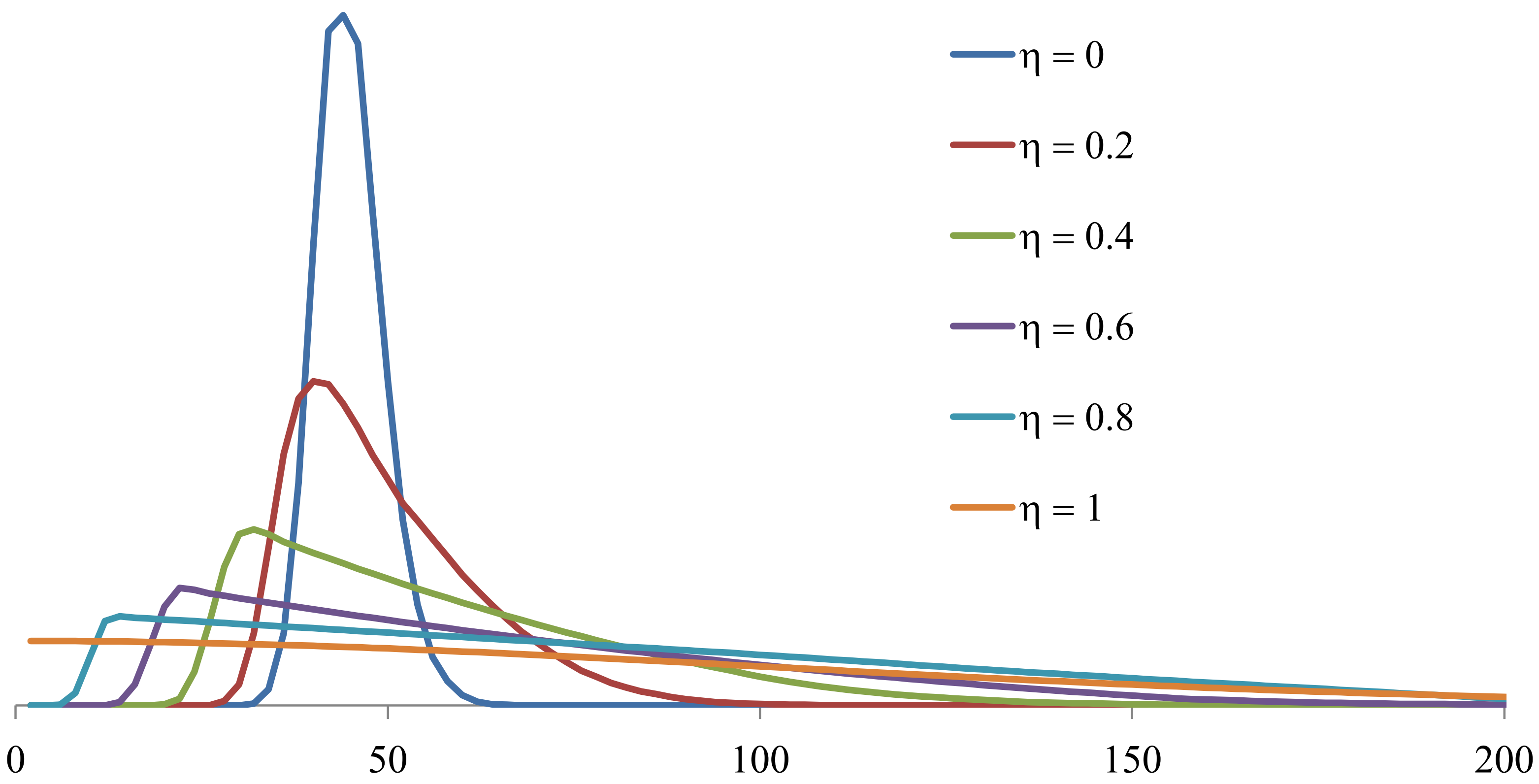

Figure 2: simulated trading volumes for different proportions of informed trading η

We simulate trade by letting η*M* informed liquidity seekers trade with (1-η)*M* uninformed liquidity seekers and market makers. Each liquidity seeker´s demand is drawn from N(0,1). The demands of the informed are perfectly correlated, the demands of the uninformed are independent. The upper black dots and lines give the skewness coefficients, the gray lines depicts the VIPIN and the bottom black lines give the VCV as a function of η for different values of *M*.

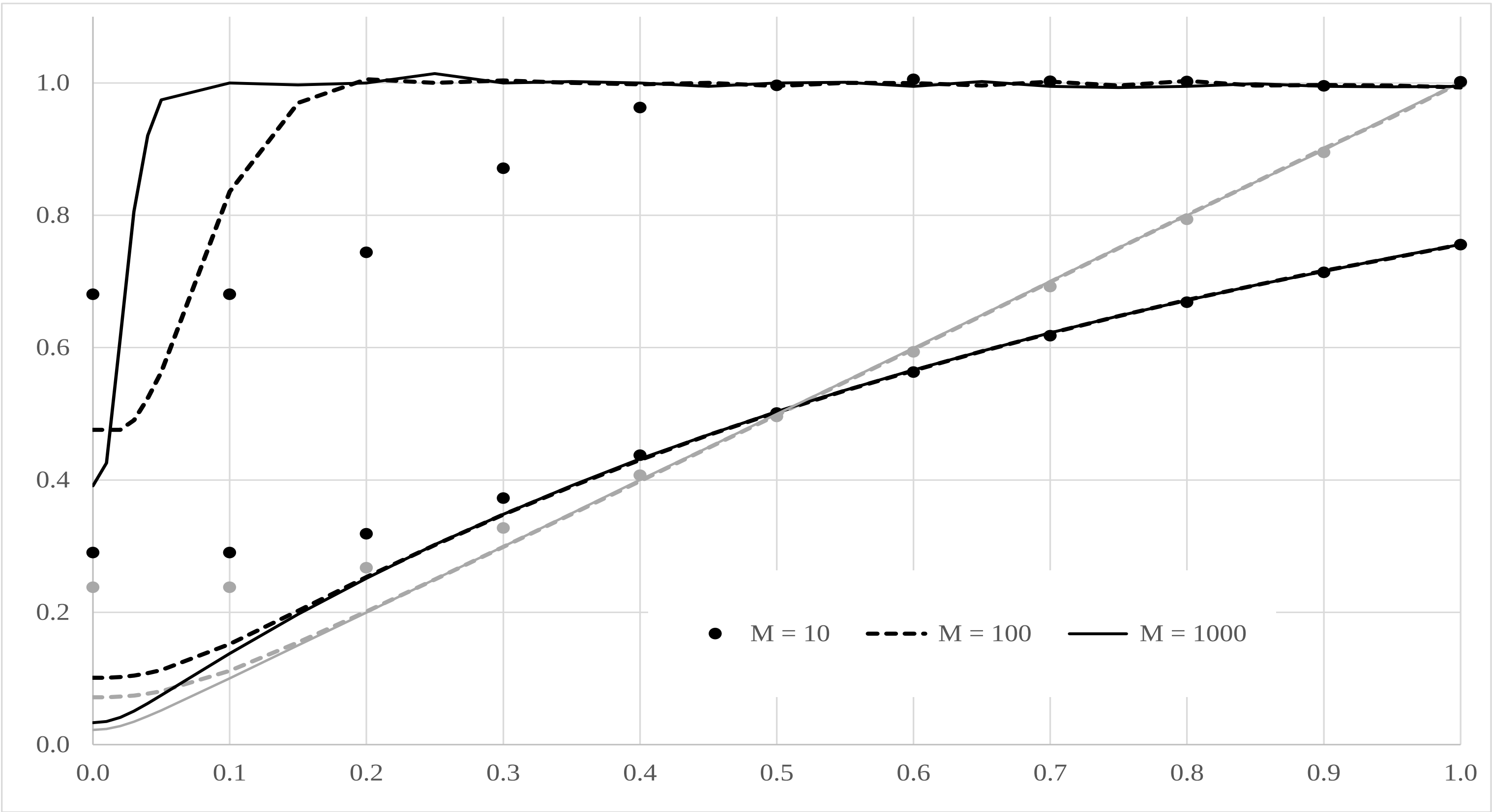

Figure 3: The first panel shows the evolution of the daily cross-sectional VCV on days -30 to 30 around quarterly earnings announements. That is, at day *d*, the sample includes the stock's trading volume on day *d* days after the annoucement, for all quarterly announcements in COMPUSTAT. The black line shows the coefficient of variation for this sample. Grey areas show 90% bootstrapped confidence intervals.

The second panel shows the same figure for randomly chosen placebo dates within 100 trading days before or after the actual announcement.