Estimating the Proportion of Informed Trade in Call Auctions

Juan Raposo, Fabrice Riva, Francisco Sogorb, Jos van Bommel

Estimating the Proportion of Informed Trade in Call Auctions

Preliminary version: February 2011

Juan Raposo[#], Fabrice Riva[#], Francisco Sogorb[*], and Jos van Bommel[+]

[#] *Université Paris-Dauphine, Paris, France*

[*] *Universidad CEU Cardenal Herrera, Valencia, Spain*

[+] *Luxembourg School of Finance, University of Luxembourg, Luxembourg*

We show that the distribution of trading volume in call auctions is indicative of the proportion of informed trade. We use the Kyle (1985) model to predict the shape of the volume distribution as a function of the proportion of informed trade. If most liquidity demanders are uninformed, there is likely to be significant wash trade among them, resulting in Normal-like volume distributions. If most liquidity demanders are informed, their trades are correlated so that the volume distribution is skewed and has a higher coefficient of variation.

JEL Classification: C13, C22, G12

Keywords: market microstructure, liquidity, call auctions

This research is partly funded by the Spanish Ministry of Education and Culture, project ECO2008-03058.

1. Introduction

Financial economists and investors are often interested in the cost of adverse selection that comes with trading securities. In this paper we show that the distribution of the trading volume in call auctions gives us an indication of the proportion of informed traders in the market.

Easley, Kiefer, O´Hara and Paperman (1996) develop a measure for the probability of informed trading in continuous trading, the well-known PIN measure. Their model is based on the Glosten and Milgrom (1985) market microstructure model where liquidity suppliers quote bid and ask quotes and liquidity seekers can either be informed or uninformed. Easley *et al.* (1996) assume that market order submitters can either be informed or uninformed, and estimate the proportion of informed traders from the dynamics of the signed order process.

In this paper we investigate how to estimate the proportion of informed trading in the other seminal market micro structure model: Kyle (1985). The double sided auctions that are used in security markets closely resemble Kyle auctions: informed and uninformed liquidity seekers submit orders and competing liquidity providers (market makers) absorb the imbalance and set the price at the expected value given the observed net order flow. We show that we can estimate the proportion of informed vis-à-vis uninformed trades from the distribution of transaction volume over a time series of Kyle-auctions.

If, in the absence of information, we assume that the number of liquidity seekers is sufficiently large, the total demand for the asset follows the same distribution as the total supply: Due to the Central Limit theorem it demand and supply will follow approximate

Normal distributions on R^+. The observed trading volume however follows a different distribution. It consists of wash trades among liquidity seekers and net order flow trades with market makers who take up the slack. The total trading volume in the absence of information is then the maximum of two Normally distributed random variables. Clark (1961) showed that this maximum follows a slightly skewed, bell curved, distribution on R^+. We show that the distance of the bell curve from the origin, measured in standard deviations, increases in the number of individual uninformed liquidity seekers.

The informed trading volume on the other hand is likely to follow a severely skewed distributions. Kyle (1985) models the informed orderflow to follow a Normal around zero, so that the observed trading volume due to informed traders follows a truncated Normal. Even if Kyle´s Normality assumption does not hold, we expect the informed volume to follow a very skewed distribution, with a mode close to the origin.

If we have both informed and uninformed traders, the observed trading volume in case of good (bad) news is the maximum of the sum of the uninformed and informed demand (supply) and the uninformed supply (demand). We analyze the distribution of this maximum under very general assumptions and show that the trading volume's skewness increases in the proportion of informed traders, and that the distance of the expected volume from the origin, measured in standard deviations, decreases in the proportion of informed traders increase. Hence we propose the observed volume distribution´s skewness and its coefficient of variation (*CV*) as measures of adverse selection in call auctions.[1]

[1] The coefficient of variation is the standard deviation – mean ratio.

Since there is no closed form solution for the relation between the proposed measures and the proportion of informed trade, we conduct Monte Carlo analyses of Kyle (1985) auctions and analyze the generated volume distribution as a function of the proportion of informed trade. We find that the skewness coefficient increases concavely in the proportion of informed traders, and reaches its maximum at relatively low proportions of informed traders.

Also the *CV* is an increasing concave function of the proportion of informed investors, but the concavity is much less severe. Moreover, we find the *CV* to be virtually independent of the number of individual traders. These desirable properties, and the fact that the *CV* can be estimated more precisely than the skewness coefficient, make the latter measure the preferred adverse selection measure.

Before taking our measure to the data, and comparing the *CV*'s of the time series volume distributions of different stocks, we observe that in reality there is also a systematic factor that affects the trading volume. For this reason we propose looking at the time series distributions of the stocks' volume market shares. We expect high asymmetric information stocks to have more skewed market share distributions.

We test this hypothesis on a sample of French Euronext stocks that are traded during an opening call auction, a continuous trading session and a closing call auction. We find that the time series distributions of small capitalization stocks see higher coefficients of variation than the distributions for large capitalization stocks, for both the opening and closing auctions. We also find that opening auction volumes see higher CVs than closing auctions, suggesting that asymmetric information reduces during the trading day. Finally, the cross sectional correlation of opening auction *CV*s and closing auction *CV*s is positive and statistically significant, suggesting that stocks that see a high *CV* at the open also have a high *CV* at the close. We interpret these findings as support for our theory, as it is reasonable to believe that small cap stocks see higher proportions of informed traders and that information asymmetries diminish throughout the trading day.

The rest of this paper is organized as follows. Section 2 presents the theory. Section 3 contains a calibration with a Monte Carlo simulation. Section 4 presents the empirical results, and section 5 summarizes and concludes.

The next step in our research is to compare the volume skewness coefficients, and more importantly the *CV*s with the PIN coefficients computed from intraday trading. Finally, we intend to investigate whether the *CV* is a "priced" risk factor.

2. Theory

We assume that there are *M* liquidity seekers of which a proportion θ is informed. We denote $\tilde{y}_i$ the individual demands of the η*M* informed traders and $\tilde{y}_u$ the demands of the (1-η)*M* uninformed. The demand of every liquidity seeker, whether informed or uninformed, is distributed with mean zero and variance σ^2.[2] The uninformed traders´ demands are *i.i.d.* because uninformed trades are, by definition, uncorrelated. The informed traders´ demands on the other hand are perfectly correlated, and hence equal.[3] We shall denote the total *net* liquidity demand that is absorbed by the market makers

[2] The assumption that individual uninformed and informed demands follow the same distribution is without loss of generality. Letting informed traders´ demands be distributed $N\left(0, k\sigma^2\right)$, and letting $\frac{\eta}{k-\eta}M$ traders be informed leads to the same conclusions.

[3] Notice that we define "informed" not with respect to some "true" liquidation value *v* that is to be disclosed at the end of the trading, but with respect to the ex-post equilibrium price.

$\tilde{y} = \sum_{\eta M} \tilde{y}_i + \sum_{(1-\eta)M} \tilde{y}_u$. It follows a Normal distribution around zero, as postulated by, among others, Kyle (1985). In particular, if we have η = 0, the net liquidity demand follows $N\left(0, M\sigma^2\right)$. If all liquidity seekers are informed we have $\tilde{y} \sim N\left(0, M^2\sigma^2\right)$.

The problem is that we cannot observe net liquidity demand because we cannot distinguish liquidity seekers from liquidity suppliers. Instead we can only observe the total order flow, which is given by:

$$\tilde{V} = \max(\tilde{y}_{buy}, \tilde{y}_{sell}) = \min(\tilde{y}_{buy}, \tilde{y}_{sell}) + |\tilde{y}| \quad , \qquad (1)$$

where $\tilde{y}_{buy}$ and $\tilde{y}_{sell}$ are the accumulated buy and sell orders demanded by liquidity seekers, given by:

$$\tilde{y}_{buy} = \sum_{\eta M} \max\left(0, \tilde{y}_i\right) + \sum_{(1-\eta)M} \max\left(0, \tilde{y}_u\right) \qquad (2)$$

and $$\tilde{y}_{sell} = \sum_{\eta M} \min\left(0, \tilde{y}_i\right) + \sum_{(1-\eta)M} \min\left(0, \tilde{y}_u\right) \qquad (3)$$

Since clearly σ^2 only affects the *scale* of $\tilde{V}$ we will henceforth investigate how parameters θ and *M* affect the *shape* of the distribution of $\tilde{V}$.

The distribution of $\tilde{V}$ is complex and in general, no analytical solution exists, except for the special cases with θ = 0 and θ = 1. For the latter case, we know that $\tilde{y}_{buy}$ and $\tilde{y}_{sell}$ follow truncated Normal distributions augmented with a probability mass of ½ at zero. Clearly, if $\tilde{y}_{buy}$ > (=) 0, $\tilde{y}_{sell}$ = (<) 0, and vice versa, so that if η = 1, $\tilde{V}$ follows a truncated normal without the probability mass at zero. Letting Where φ(*y*) denote the density function of the standard Normal, we find that the first three moments of the trading volume if η = 1 are given by:

$$E\left[\tilde{V}\middle|\eta=1\right]=2M\sigma\int_0^\infty y\varphi(y)dy=M\sigma\sqrt{\frac{2}{\pi}} \tag{4}$$

$$E\left[\left(\tilde{V}-E\left[\tilde{V}\right]\right)^2\middle|\eta=1\right]=2M^2\sigma^2\int_0^\infty\left(y-\sigma\sqrt{\frac{2}{\pi}}\right)^2\varphi(y)dy=M^2\sigma^2\frac{\pi-2}{\pi} \tag{5}$$

$$E\left[\left(\tilde{V}-E\left[\tilde{V}\right]\right)^3\middle|\eta=1\right]=2M^3\sigma^3\int_0^\infty\left(y-\sigma\sqrt{\frac{2}{\pi}}\right)^3\varphi(y)dy=M^3\sigma^3\frac{(4-\pi)\sqrt{2}}{\pi\sqrt{\pi}} \tag{6}$$

If there is no information in the market (η = 0), the $\tilde{V}$ distribution is more complex. This is due to the fact that $\tilde{y}_{buy}$ and $\tilde{y}_{sell}$ are sums of severely skewed random variables: each individual $\max(0,\tilde{y}_i)$ is distributed as a truncated Normal augmented with a probability mass at zero. Still, for large *M*, the distribution of $\tilde{y}_{buy}$ and $\tilde{y}_{sell}$ will approximate the Normal due to the Central Limit Theorem. Moreover, $\tilde{y}_{buy}$ and $\tilde{y}_{sell}$ are negatively correlated, also if η = 0. Clearly, if in a given population, there happen to be more uninformed buyers, there will to be fewer uninformed sellers. It can be shown that, for large *M*, the joint distribution of $\tilde{y}_{buy}$ and $\tilde{y}_{sell}$ is approaches a bivariate Normal:

<u>Lemma</u>: *If individual demands are i.i.d. distributed with* $\tilde{y}_i$ ~ N(0, σ^2), w*e have that the joint distribution of* $\tilde{y}_{buy}\equiv\sum_M\max(0,\tilde{y}_i)$ *and* $\tilde{y}_{sell}\equiv-\sum_M\min(0,\tilde{y}_i)$, *follows a bivariate Normal distribution with*:

$$\lim_{M \to \infty} \begin{pmatrix} \frac{\tilde{y}_{buy}}{M} \\ \frac{\tilde{y}_{sell}}{M} \end{pmatrix} \sim N\left(\begin{pmatrix} \frac{\sigma}{\sqrt{2\pi}} \\ \frac{\sigma}{\sqrt{2\pi}} \end{pmatrix}, \begin{pmatrix} \frac{\sigma^2(\pi-1)}{2\pi} & \frac{-\sigma^2}{2\pi} \\ \frac{-\sigma^2}{2\pi} & \frac{\sigma^2(\pi-1)}{2\pi} \end{pmatrix} \right)$$

Normality follows from the Central Limit Theorem. The means and variances follow from integration. To see that $\tilde{y}_{buy}$ and $\tilde{y}_{sell}$ are negatively correlated, notice that in an auction where there happen to be more uninformed buyers, there are likely to be fewer uninformed sellers.

Hence, for large *M* we can approximate the distribution of $\tilde{V}$ with the formulas of Clark (1961), who gives expressions for the first four moments of the maximum of two Normally distributed random variables. In particular, we find:

$$\lim_{M \to \infty} \frac{E\left[\tilde{V} \middle| \eta = 0\right]}{M} = \frac{\sigma}{\sqrt{2\pi}} \tag{10}$$

$$\lim_{M \to \infty} \frac{E\left[\left(\tilde{V} - E\left[\tilde{V}\right]\right)^2 \middle| \eta = 0\right]}{M} = \frac{\sigma^2(\pi-2)}{2\pi} \tag{11}$$

$$\lim_{M \to \infty} \frac{E\left[\left(\tilde{V} - E\left[\tilde{V}\right]\right)^3 \middle| \eta = 0\right]}{M\sqrt{M}} = \frac{\sigma^3(4-\pi)}{4\pi\sqrt{2\pi}} \tag{12}$$

We see that if $\eta = 0$, the volume follows a slightly skewed bell-curve distribution on the R^+, which for large *M* can lie infinitely many standard deviations away from the origin. This distribution is very different from the distribution which obtains when $\eta = 1$, which is the much more skewed truncated Normal.

It can be shown that for intermediate percentages of informed investors, $0 < \eta < 100\%$, the trading volume distributions takes on intermediate shapes. To characterize the $\tilde{V}$ distributions, we consider the coefficient of variation, $CV = \frac{\hat{\sigma}}{\hat{\mu}}$, and the skewness coefficient, $\hat{\gamma} = \frac{\frac{1}{N}\sum_N (V_i - \hat{\mu})^3}{\hat{\sigma}^3}$. From the above analysis we find:

$$\lim_{M \to \infty} E[CV|\eta = 0] = \lim_{M \to \infty} \frac{\sqrt{M}\sqrt{\pi - 2}}{(M - \sqrt{M})} = 0 \tag{13}$$

$$\lim_{M \to \infty} E[CV|\eta = 1] = \sqrt{\frac{\pi - 2}{2}} \approx 0.75551 \tag{14}$$

$$\lim_{M \to \infty} E[\hat{\gamma}|\eta = 0] = \frac{(4 - \pi)}{2(\pi - 2)\sqrt{\pi - 2}} \approx 0.35188 \tag{15}$$

$$\lim_{M \to \infty} E[\hat{\gamma}|\eta = 1] = \frac{(4 - \pi)\sqrt{2}}{(\pi - 2)\sqrt{\pi - 2}} \approx 0.99527 \tag{16}$$

It can be shown that intermediate levels of η result in intermediate skewness coefficients and intermediate *CV*s. In particular,

Proposition 1: *Consider M liquidity seeking traders who submit market orders in a call auction where the net order flow is absorbed by liquidity suppliers. The probability distribution of the resulting trading volume depends on the proportion of informed traders. In particular,*

i) the skewness coefficient increases in the proportion of informed traders.

ii) the coefficient of variation (CV) increases in the proportion of informed traders.

Summarizing, our analysis shows that the uninformed liquidity seeking demand generates a bell curve shaped volume distribution on the positive axis. This is because the uncorrelated orders generate significant wash trades, while relatively low orderflow is taken up by liquidity providers. For large *M* the observed trade volume will not stray far from its expectation, resulting in a low skewness coefficient and a low *CV*.

The informed liquidity seekers on the other hand submit perfectly correlated orders, and generate a much more skewed distribution with a significant likelihood of small volumes because information in an efficient market is symmetric and has a zero mean. If we assume information to follow a Normal around zero, the observed trading volume will follow a truncated Normal hugging the y-axis, resulting in a higher skewness coefficient and a high *CV*. As the proportion of informed traders increase, the volume distribution will become more skewed and have a higher *CV*.

3. Monte Carlo Simulation

In this section we analyze the volume distribution 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 ηM, and represents the aggregate informed demand.[4] The remaining variables represent the individual uninformed demands. We take the positive and negative parts of these variables, and sum them to obtain buy and sell demands $\tilde{y}_{buy}$ and $\tilde{y}_{sell}$. The maximum of these values is the observed volume $\tilde{V}$. For each (M, η) combination we generate 10,000,000 $\tilde{V}$ observations.

[4] We choose (η, M), so that ηM are integers.

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

The upper graph of Figure 1 displays six histograms of simulated volumes for $M = 1000$, and different values of η. The simulation confirms our analysis: 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. As per proposition 1, for intermediate values, the distributions take intermediate shapes.

To compare our simulated results with real data, we also give, in the lower graph, the empirically observed histograms of observed €-transaction volume, scaled by the average, for two Euronext stocks: one largecap, AirFrance, and one smallcap, Biomerieux. We clearly see that the smallcap has a higher skewness coefficient and *CV*, which according to the analysis in the previous section is indicative of a higher proportion of informed investors trading this stock. The differences in distribution is typical for large and smallcap firms.

Table I and Figure 2 give the results of additional analyses, for different values of *M*. From table I we see that, for a given *M*, the skewness coefficient of the simulated volume distribution indeed increases in η. However, we also see that the skewness coefficient as a measure of adverse selection is problematic, as the relationship between η and the skewness is concave, and the maximum (of 0.99527) is reached for relatively low levels of η. This implies that it will be difficult to estimate η from $\hat{\gamma}$.

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

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

The coefficient of variation however behaves much nicer. Not only does it increase continuously in θ, but more importantly, for most parameter values it is very insensitive to *M*. Only for very small values of θ does the number of liquidity seekers affect the *CV*. The insensitivity to *M* is because the volume generated by the informed, follows a distribution of which the shape is independent of *M*, carries most of the weight in the total volume, because it does not contain wash trades, as does the uninformed trade.

The finding that the *CV* is insensitive to *M* is encouraging as it implies that there is little concern for confounding a high θ with a low *M*. We allude to the fact that often the number of order submitters in a call auction is unknown.[5] Our analysis shows that incorrectly attributing adverse selection to a low trading activity only happens for very small θ and *M*.

4. Empirical Evidence

In this section, we analyze, for a sample of 63 stocks traded on Euronext Paris between January 2006 and December 2007, the volume distributions during the opening and closing auctions, and compare the γ and *CV* measures with two incumbent measures of adverse selection, estimated from high frequency data from continuous trading sessions: the MRR-θ, suggested by Madhavan et al. (1997), which estimates the price adjustment in response to the unexpected component of the order flow, and the PIN developed by Easley *et al.* (1997), which estimates the proportion of informed traders from the market orders time series.

[5] Although some data-providers identify the individual transactions assigned during a call auction, this number is different from the number of order-submitters because orders may be broken up.

4.1. Data.

The stocks are divided into three size groups: Large capitalization stocks (or largecaps), which are drawn from the CAC40, medium capitalization stocks (midcaps), that are drawn from the NEXT20, and the MID100 index, and small capitalization stocks drawn from the SMALL90 index.

4.2. The new adverse selection measures.

Before computing the *CV*s and γ's of different stocks, we first have to decide on how to measure trading volume. The theory of session 2 does not say anything about the unit of measurement, which can be the number of shares or monetary units (euros). In our analysis we do both, although we favour looking at numbers of shares instead of euros, so as to control for market capitalization.

A natural concern is that the trading volume varies exogenously over time, e.g. Fridays may see larger trades or more participants than Wednesdays, or summers may see less trading than winters. Naturally, also other events affect the market wide trading volume.

To control for the *systematic* trading activity variation, we also weight the volume (both in shares and in euros) by the total trading volume in the market. This standardization assumes that the long term market shares of trading volumes are constant over time and that time series variation of market share is indicative of differences in adverse selection.

Table II gives the average skewness coefficients and *CV*s for the opening and closing auctions, for raw volumes and market shares. We find that average *CV*s vary across stocks in the intuitive direction: Largecaps have lower *CV*s than mid cap stocks, and smallcaps see the highest *CV*s. This is consistent with small capitalization stocks being subject to more adverse selection, an often made premise in the empirical corporate finance literature. Moreover, opening auctions have higher *CV*s than closing auctions. Also this is reassuring, since adverse selection is likely to decrease over the trading day. (See, e.g. …. Hoffmann and van Bommel (2011).

We also see that differences in average *CV*s are larger than the differences in average skewness coefficients, which are not statistically significant.[6] The low power of the skewness estimator is consistent with the Monte Carlo analysis of the previous section. Similarly, the standard errors confirm that that taking turnover market shares leads to more powerful tests than looking at raw turnovers.

We also check how the *CV* and skewness coefficients during the two daily auctions are correlated: do stocks that have high *CVs* in the opening auctions have high *CVs* in the closing auction. We compute the cross sectional Pearson correlation coefficients for the 63 stocks and (for the time being) eight adverse selection measures. Table III gives the result of this analysis.

Consistent with earlier findings we find that the *CV* seems to behave best. That is, stocks that have a high *CV* for opening auction volumes tend to have a high *CV* for closing auction volumes.

4.3. Comparison with incumbent adverse selection measures

In this section, we investigate how the shape of the trading volume distribution as a measure of informed trading compares to other measures. Before us Easley, O'Hara,

[6] Notice that the skewness coefficients are not in the right order: the average skewness of largecap volumes is higher than that of the midcaps.

Kiefer and Paperman (1996) and Madhavan, Richardson and Roomans (1997) developed ways to estimate of the proportion of informed trading by analysing high frequency data.

Easley *et al.* (1996) propose a structural model of the trading process to back out the proportion of informed trades. Their model posits that during each trading day an information event is likely to occur with probability α. On an event day, the price of a stock can take on a low or a high value with probabilities δ and $(1-\delta)$. Informed traders observe a signal that allows them to infer the true value of the asset and trade accordingly. Conditional on the occurrence of an event, their arrival rate is modelled as a Poisson process with intensity μ. Both on event and non-event days, uninformed traders come to the market according to an independent Poisson process with intensity ε and they buy or sell with equal probabilities. Prices are set by a Glosten and Milgrom (1985) market maker who posts quotes based on the trading history. For a trading day of length T, the probability of observing *B* buys and *S* sells is given by:

$$\begin{aligned} L(B,S|\alpha,\delta,\varepsilon,\mu) = &\ (1-\alpha)e^{-2\varepsilon}\frac{\varepsilon^{S+B}}{B!S!} \\ &+ \alpha\delta e^{-(\mu+2\varepsilon)}\frac{(\mu+\varepsilon)^{S}\varepsilon^{B}}{B!S!} \\ &+ \alpha(1-\delta)e^{-(\mu+2\varepsilon)}\frac{(\mu+\varepsilon)^{B}\varepsilon^{S}}{B!S!} \end{aligned} \tag{17}$$

Assuming independence across days, the parameters $\{\alpha, \delta, \varepsilon, \mu\}$ can be estimated from a sample of trading days with a maximum likelihood procedure.

From the estimated parameters, the probability of trading with an informed trader, or *PIN*, can be computed as:

$$PIN = \frac{\alpha\mu}{2\varepsilon+(1-\alpha)\mu} \tag{18}$$

The second measure of informed trading is based on the decomposition of the effects of trades on transaction prices, as suggested by Madhavan *et al.* (1997). Their return generating model is given by:

$$p_t - p_{t-1} = (\phi + \theta)x_t - (\phi + \rho\theta)x_{t-1} + \varepsilon_t \qquad (19)$$

Where p_t is the trading price time *t*, x_t is the sign of the order flow at time *t* (+1 for buys, -1 for sells), ϕ is the liquidity suppliers' cost of providing liquidity, ρ is the autocorrelation in the order flow, and θ is the price adjustment in response to the unexpected component of the order flow, i.e. the liquidity suppliers' perception about the degree of asymmetric information. The θ parameter, as estimated from high frequency trading data with a GMM procedure, will serve as our second benchmark estimate of the proportion of informed traders.

Both the measures we use to compute the proportion of informed traders require the identification of the direction of trades. Naturally, we classify a trade as a buy, if the price equals the best ask price, and as a sell if the price equals the best bid. Due to the nature of trading on Euronext and the precise time stamping, virtually all transactions can be unambiguously classified.[7]

Since the theory in section 2 is concerned with the impact of informed liquidity seekers on the distribution of trading volume during a call auction, our estimation strategy must

[7] Due to the precise time stamping we do not have trades that are recorded ahead of quotes, nor grouped trades. Cross-trades are very few in number and marked with a special flag in the data. For a trade classification in the presence of non-synchroneity, transactions grouping and unmarked cross-trades, see Lee and Ready for an algorithm.

be such that it allows us to recover the θ and the PIN values that prevail at the market open. The PIN expression in equation **Erreur ! Source du renvoi introuvable.** is computed using the unconditional parameter set $\{\alpha, \delta, \mu, \varepsilon\}$. It thus implicitly measures the probability of information-based trading at the open. I don't understand this. Do you only use data during the first hour or so?

The situation is more complex regarding the θ parameter. As pointed out by Madhavan et al., the GMM technique requires that the number of observations be large enough to get reliable parameter estimates. Yet, using data over the entire trading session to perform the estimation will certainly lead to θ estimates that do not necessarily reflect the proportion of informed traders during the opening call auction. However, since informational asymmetries can possibly persist over the entire trading day, θ estimated using data that cover the whole trading session can be indicative of the proportion of informed traders at the beginning of the day. Our approach to addressing these issues is the following: for each stock year we first estimate θ using all the trades recorded during the trading session (daily θ). We then estimate intradaily θ parameters by splitting the trading session into five sub periods: 9:00-10:00am, 10:00:12:00am, 12:00-2:00pm, 2:00:4:00pm, and 4:00:5:25pm. Can't we do this for the PIN too? (Intradaily PINs).

Table IV reports various summary statistics regarding the features of our data and their evolution over time. A salient feature of the data is the dramatic increase in trading activity: daily trading volumes increase by 34% from 2005 to 2006 and by 36% from 2006 to 2007, which yields a significant 83% increase over the sample period. Volumes traded at the opening call auction also increase in absolute numbers, but the percentage of volume traded at the open decreases over time. This is consistent with the findings of Hoffmann and van Bommel (2011), a.o.

This lower contribution is accompanied by changes in the shape of their distribution: all of the four distributional measures we employ exhibit decreasing values over our sample period with a negative significant drop (from -31.78% for the adjusted skewness to -13.17% for the raw *CV*) from 2005 to 2006. In light of our theory, this result suggests that the proportion of informed investors who trade during the call auction may have been decreasing over time. Consistent with this story, the evolution of *PIN* parallels the one we observe on both the skewness and the SDMR of call auction volumes (either adjusted or not). However, we do not observe such positive correlation for the θ parameter. Yet, recall that daily θ estimates may not necessarily reflect the situation that prevails regarding informed trading during the opening call auction. We thus computed the evolution of intradaily θ. The results are reported in Panel B of Table IV. As in Madhavan et al., we find that the θ values are not constant throughout the trading session and that information asymmetry is the highest at the beginning of the trading day. In addition, the θ values that are estimated over the time period that is closest to the market open (9:00 – 10:00am) exhibit a decreasing (though statistically insignificant) pattern over the years 2005 to 2007 and thus parallel the evolution that we document on both PIN and our distributional statistics on call auction volumes.

We further explore the relationship between our various measures by examining their correlation. Table V reports the Pearson correlation matrix computed from the sample of pooled stock year measures. All correlation coefficients are significantly positive at the 5% level but for the correlation between daily theta estimates and the two measures of skewness. In particular we observe that, consistent with our theory, both the skewness and the SMDR of the distribution of the call auction volumes significantly increase in presence of higher informed trading as measured by PIN and the opening θ. The correlations are the highest for the shape measures based on the adjusted volume, which gives support to our intuition that, though simple, the proposed adjustment can effectively remove variations in trading volumes that are unrelated to information. In addition the two measures based on the SDMR are those that achieve the highest correlation with PIN and θ, which is not surprising since the SDMR is less sensitive than the skewness to extreme values in trading volume. PIN is more correlated than θ with the skewness whereas θ is the measure that achieves the highest correlation with the SDMR. As far as informed trading is reflected in the shape of volumes distribution, these differences suggest that our two measures probably capture different features of how adverse selection risk impacts the trading process. Consistent with this view, the PIN and θ estimates are not perfectly correlated. Yet, the correlation is as high as 0.57 and highly significant. Recalling that the two measures are developed in completely different frameworks, this result supports the idea that both can be used as reliable estimates of information-based trading.

Though the previous results provide support to our theory that information-based trading induces an increase in both the skewness and the SDMR of the distribution of call auction volumes, they do not account for the heterogeneity in our sample stocks. For example, Easley et al. (2002) show that the risk of informed trading as measured by the PIN is lower for actively traded stocks and large cap firms. We control for these factors as well as others in a multivariate setting. The results of various specifications are reported in table VI.

Consistent with the evidence in table IV, the value of the four distributional measures are significantly decreasing over time as indicated by the negative coefficient on the year dummies.

Questions:

- It seems that there is a ‘secular’ trend toward less dispersion in volumes traded during the call: why? Intuitively, the increased sophitiscation in the way investors trade makes the opening call less important, especially if they are engaged in VWAP-benchmarked strategies
- Our four distributional measures are also significantly decreasing with the amount of average traded volume: any idea?

After controlling for year / cap / vol, we get an almost significant coefficient for opening theta and SDMR, which is not so bad because the number of pooled stock year measures is just 123.

Table IV – Summary statistics

This table reports the cross-sectional mean values of various statistics over the years 2005-2007 for the 47 stocks in our sample. We compute for each stock day the euro volume traded during the opening call auction, the total daily volume and the proportion of the total daily volume traded during the opening call auction. For each stock, on each year, we compute the raw skewness (raw SDMR) as the skewness (SDMR) of the time series of daily volumes traded during the opening call. We also compute the adjusted skewness (adjusted SDMR) of the time series of daily adjusted volumes traded during the call auction where the adjusted volume is computed as the euro volume divided by the total euro volume traded on Euronext over the corresponding day. PIN measures the probability that the first trade of the continuous trading session is information-based. Theta measures the contribution of asymmetric information to price changes at a transaction level. We test the null hypothesis of equality of means across years. *, **, *** denote rejection of the null at the 1%, 5% and 10% levels, respectively. Asterisks in column 2006 denote rejection that the 2005 and the 2006 values are equal. The first group of asterisks in column 2007 denotes rejection that the 2007 and the 2005 values are equal whereas the second group denotes rejection that the 2007 and the 2006 values are equal.

		2005	2006		2007	
Panel A: Daily estimates						
PIN		0.204	0.195	***	0.184	***/***
Theta (x 10^5)		8.766	9.314	***	9.014	/
Raw skewness		4.654	3.417	***	3.156	*** /
Adjusted skewness		5.043	3.441	***	3.115	***/*
Raw SDMR		1.187	1.031	***	0.988	***/
Adjusted SDMR		1.233	1.029	***	0.968	***/*
Market value (in billion €)		22.817	29.385	***	36.677	***/**
Daily trading volume (in million €)		97.320	130.382	***	177.827	***/***
Call auction volume / Daily volume		0.015	0.012	***	0.012	***/
Panel B: Theta intradaily estimates						
Theta (x 10^5)						
	9:00-10:00 am	14.721	14.557		14.022	
	10:00-12:00 am	8.330	9.120		8.830	
	12:00 am-2:00 pm	7.939	8.437		8.168	
	2:00pm-4:00pm	7.718	8.497	**	8.327	
	4:00pm-5:25pm	7.871	8.209		8.041	

Table V – Correlation matrix of informed trading measures

This table reports the Pearson correlation matrix of our various measures for the intensity of informed trading for our sample of 47 stocks over the period 2005-2007. We compute for each stock day the euro volume traded during the opening call auction, the total daily volume and the proportion of the total daily volume traded during the opening call auction. For each stock, on each year, we compute the raw skewness (raw SDMR) as the skewness (SDMR) of the time series of daily volumes traded during the opening call. We also compute the adjusted skewness (adjusted SDMR) of the time series of daily adjusted volumes traded during the call auction where the adjusted volume is computed as the euro volume divided by the total euro volume traded on Euronext over the corresponding day. PIN measures the probability that the first trade of the continuous trading session is information-based. Theta measures the contribution of asymmetric information to price changes at a transaction level. Daily theta is estimated using the time series of trades over the entire trading session. Opening theta is estimated using only the trades made during the first hour of trading following the opening call auction. We test the null hypothesis that the correlations are zero. We report the p-values in parenthesis. Boldface figures indicate rejection of the null at the 5% level.

	Raw skewness	Adjusted skewness	Raw SDMR	Adjusted SDMR	PIN	Daily theta	Opening theta
Raw skewness	1.000	**0.888**	**0.839**	**0.785**	**0.267**	0.121	**0.213**
		(0.00)	(0.00)	(0.00)	(0.00)	(0.19)	(0.02)
Adjusted skewness		1.000	**0.761**	**0.829**	**0.295**	0.116	**0.243**
			(0.00)	(0.00)	(0.00)	(0.21)	(0.01)
Raw SDMR			1.000	**0.941**	**0.344**	**0.320**	**0.400**
				(0.00)	(0.00)	(0.00)	(0.00)
Adjusted SDMR				1.000	**0.401**	**0.369**	**0.476**
					(0.00)	(0.00)	(0.00)
PIN					1.000	**0.456**	**0.567**
						(0.00)	(0.00)
Daily theta						1.000	**0.852**
							(0.00)
Opening theta							1.000

Table VI – Multivariate analysis of call auction volumes distribution

This table reports the regression analysis of volume distributional measures. All variables are computed on a year stock basis and the regression is performed on the pooled sample. The reference year is 2005. Year 2006 and Year 2007 are dummy variables that take on the value on the corresponding year and zero otherwise. log(Avg Mkt Value) is the average market value, log(Avg Trading Vol) is the daily average euro trading volume, PIN estimates the probability of informed trading at the market open and Opening theta estimates the contribution of asymmetric information to price changes over the first hour of trading following the opening call auction. Heteroscedastic-consistent t-statistics are reported in parenthesis.

	Raw skewness		Adjusted skewness		Raw SDMR		Adjusted SDMR	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Intercept	18.143	14.513	15.521	2.304	5.053	3.921	4.980	2.447
	(2.52)	(1.65)	(2.41)	(0.31)	(3.64)	(-2.60)	(4.09)	(-1.90)
Year 2006	-1.063	-0.988	-1.503	-1.526	-0.094	-0.103	-0.138	-0.163
	(-2.48)	(-2.29)	(-3.95)	(-4.13)	(-1.37)	(-1.54)	(-2.04)	(-2.60)
Year 2007	-1.153	-0.980	-1.734	-1.720	-0.075	-0.089	-0.132	-0.174
	(-2.40)	(-2.11)	(-4.16)	(-4.54)	(-0.93)	(-1.21)	(-1.91)	(-2.89)
log (Avg Mkt Value)	0.491	0.442	0.696	0.620	-0.010	-0.014	0.055	0.048
	(1.32)	(1.17)	(1.93)	(1.74)	(-0.18)	(-0.25)	(0.99)	(-0.90)
log(Avg Trading vol)	-1.248	-1.112	-1.304	-0.715	-0.193	-0.140	-0.267	-0.148
	(-2.72)	(-1.94)	(-2.98)	(-1.46)	(-2.72)	(-1.67)	(-3.69)	(-2.08)
PIN	-12.164		-16.007		-0.666		-1.025	
	(-0.87)		(-1.29)		(-0.28)		(-0.51)	
Opening theta		-0.010		0.074		0.009		0.022
		(-0.13)		(1.06)		(0.71)		(1.95)
Adjusted R^2	0.145	0.142	0.249	0.251	0.216	0.220	0.292	0.316

Conclusion

References

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

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

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

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.

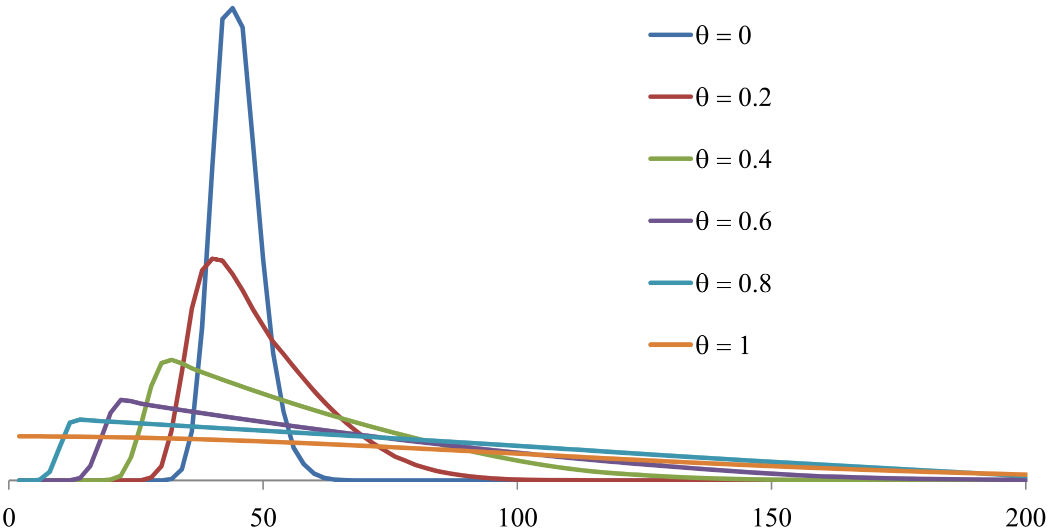
Kyle, Albert S. (1985) Continuous auctions and insider trading, *Econometrica*, 53, 6, 1315-1335.

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

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.

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

We simulate trade by letting θ*M* informed liquidity seekers trade with (1-θ)*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. In the top graph we depict 6 histograms of 10,000,000 simulated trading volume observations for different θ´s. In the bottom graph we depict the histograms of average-scaled volumes for two French stocks, AirFrance, a large capitalization stock and Biomerieux, a smallcap.



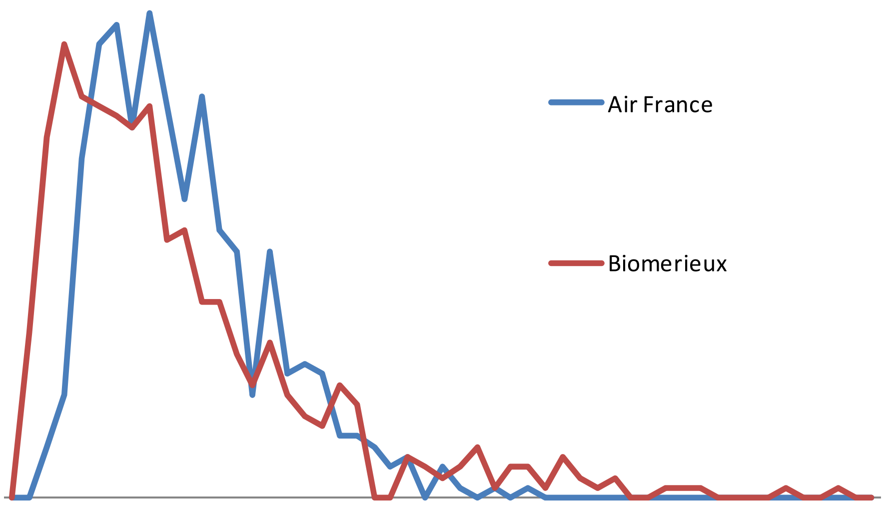


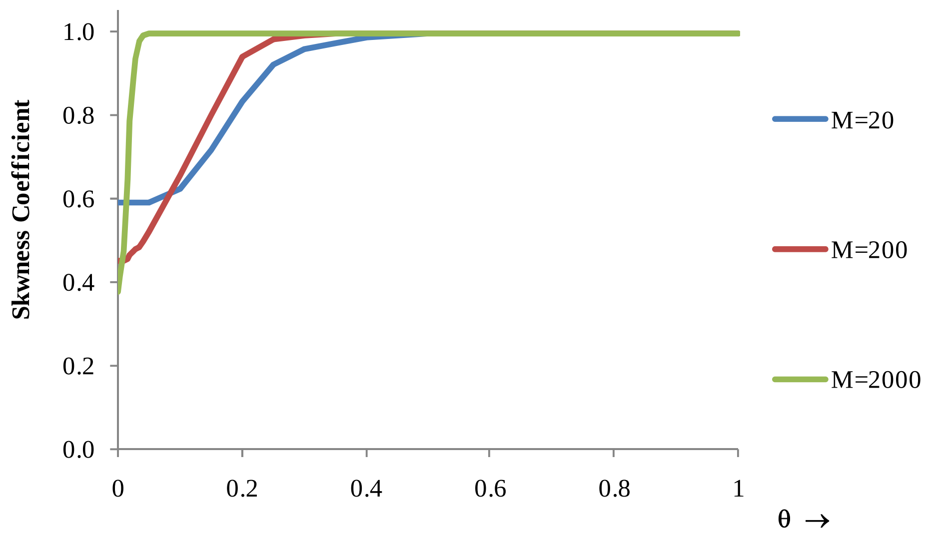
Table I: distribution statistics of simulated trading volumes for different *M* and θ.

We simulate trade by letting θ*M* informed liquidity seekers trade with (1-θ)*M* uninformed liquidity seekers and market makers, who take up the net demand. Each liquidity seeker´s demand is drawn from N(0,1). We report characteristics of the volume distributions of 10,000,000 simulations. For θ = 0, we also report the mean and standard deviation as predicted by Clark (1961).

		M							
θ		10	20	50	100	200	500	1000	2000
0	mean	5.26	9.77	22.75	43.90	85.41	208.45	411.62	815.64
	mean-Clark	5.25	9.76	22.77	43.88	85.43	208.39	411.56	815.73
	st.dev.	1.52	2.08	3.20	4.46	6.22	9.73	13.62	19.21
	st.dev.-Clark	1.35	1.91	3.01	4.26	6.03	9.53	13.48	19.06
	CV	0.289	0.213	0.140	0.102	0.073	0.047	0.033	0.024
	skew	0.667	0.590	0.508	0.479	0.452	0.392	0.387	0.374
0.05	mean				44.27	86.61	212.72	422.24	841.70
	st.dev.				5.02	8.00	16.77	31.40	61.56
	CV				0.113	0.092	0.079	0.074	0.073
	skew				0.577	0.657	0.881	0.968	1.005
0.1	mean		9.85	23.30	45.43	89.45	221.04	440.55	879.17
	st.dev.		2.22	4.01	6.92	12.85	30.69	60.38	121.41
	CV		0.225	0.172	0.152	0.144	0.139	0.137	0.138
	skew		0.621	0.707	0.848	0.958	1.022	1.000	1.043
0.2	mean	5.37	10.25	24.68	48.49	96.47	240.37	479.98	958.91
	st.dev.	1.71	2.85	6.39	12.22	24.25	60.65	121.44	241.83
	CV	0.319	0.278	0.259	0.252	0.251	0.252	0.253	0.252
	skew	0.750	0.832	0.963	1.010	1.007	1.010	1.026	1.001
0.4	mean	5.85	11.48	28.27	56.17	111.72	279.93	558.20	1120.76
	st.dev.	2.56	4.95	12.18	24.18	48.17	120.42	241.41	483.62
	CV	0.437	0.432	0.431	0.431	0.431	0.430	0.432	0.432
	skew	0.949	1.023	1.002	1.003	1.016	0.975	1.010	0.984
0.6	mean	6.52	12.91	32.10	63.86	128.03	318.60	638.39	1274.65
	st.dev.	3.68	7.29	18.08	36.03	72.49	181.03	361.32	721.22
	CV	0.564	0.565	0.563	0.564	0.566	0.568	0.566	0.566
	skew	1.018	0.995	0.994	1.014	0.997	1.004	1.004	1.004
0.8	mean	7.25	14.34	36.06	71.73	144.07	359.84	719.26	1441.89
	st.dev.	4.87	9.60	24.06	48.10	96.89	240.95	482.63	971.19
	CV	0.671	0.669	0.667	0.671	0.673	0.670	0.671	0.674
	skew	0.994	1.022	0.983	1.000	1.004	0.977	0.984	1.000
1	mean	7.97	15.97	39.91	79.82	159.34	398.24	795.88	1594.36
	st.dev.	6.04	12.05	30.21	60.40	120.62	302.05	602.81	1203.53
	CV	0.758	0.755	0.757	0.757	0.757	0.758	0.757	0.755
	skew	1.000	1.008	0.998	0.996	1.007	1.008	0.992	1.000

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

We simulate trade by letting θM informed liquidity seekers trade with (1-θ)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. In the upper graph we give the skewness coefficient of the simulated histogram as a function of θ for different values of M. In the lower graph we depict the distribution´s standard deviation mean ratio (CV) as a function of θ.



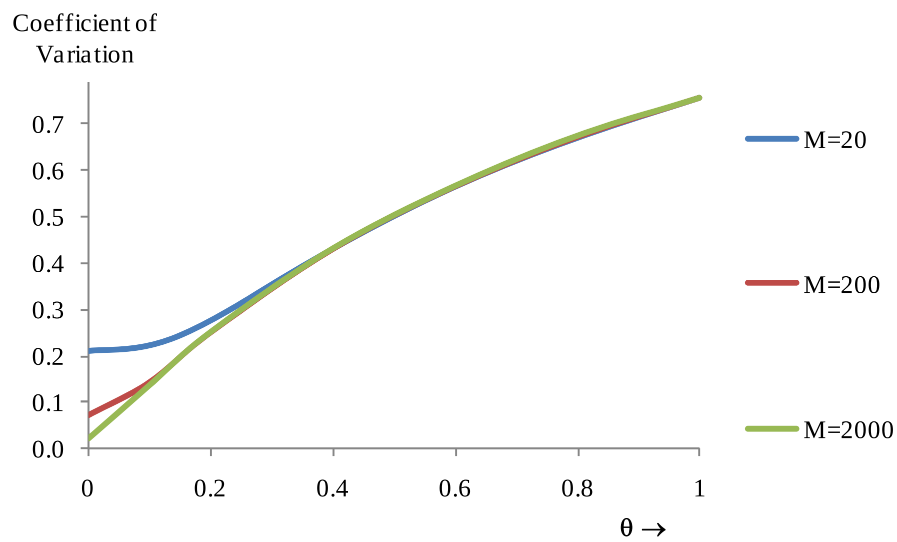


Table II: volume skewness coefficients and SDMRs for large, mid, and small capitalization stocks.

We compute, for each stock day, the trading volume in euros, the volume in shares, the market share in euros and the trading volume in shares relative to the market. From each time series of approximately 500 daily observations, we compute the skewness coefficients and the standard deviation – mean ratios (*CV*s). We report the averages for the three subsamples of 21 large capitalization stocks, 21 midcaps and 21 smallcaps. Standard errors are in italic.

Opening Auction

	Average Skewness Coefficient				Average Coefficient of Variation			
	€	shares	€ mktshr	shr mktshr	€	shares	€ mktshr	shr mktshr
Large Caps	8.02	7.72	6.00	5.80	1.52	1.52	1.28	1.31
s.e.	*1.14*	*1.01*	*0.64*	*0.59*	*0.17*	*0.18*	*0.13*	*0.14*
Mid Caps	6.70	6.45	6.59	6.61	1.66	1.63	1.58	1.58
s.e.	*0.58*	*0.61*	*0.52*	*0.54*	*0.10*	*0.08*	*0.09*	*0.08*
Small Caps	6.74	7.07	6.91	7.25	1.85	1.90	1.82	1.90
s.e.	*0.82*	*0.84*	*0.82*	*0.86*	*0.13*	*0.13*	*0.13*	*0.14*

Closing Auction

	Average Skewness Coefficient				Average Coefficient of Variation			
	€	shares	€ mktshr	shr mktshr	€	shares	€ mktshr	shr mktshr
Large Caps	5.39	5.21	4.38	4.17	0.83	0.90	0.59	0.64
s.e.	*1.26*	*1.33*	*0.84*	*0.88*	*0.12*	*0.15*	*0.07*	*0.09*
Mid Caps	3.75	3.55	3.89	3.69	1.03	1.05	1.08	1.05
s.e.	*0.38*	*0.26*	*0.65*	*0.53*	*0.06*	*0.05*	*0.11*	*0.09*
Small Caps	6.44	6.66	6.76	6.94	2.41	2.46	2.35	2.40
s.e.	*0.84*	*0.92*	*0.89*	*0.98*	*0.52*	*0.51*	*0.43*	*0.42*

Table III: Correlation of skewness coefficients and coefficients of variations

We compute, for each stock day, the trading volume in euros, the volume in shares, the market share in euros and the trading volume in shares relative to the market. For each time series of approximately 500 daily observations, we compute the skewness coefficients and the coefficients of variation (CVs). This gives us 63 data points for each measure. We report the cross sectional correlation of the measures between the opening and closing auction. E.g. the largest correlation, between the CV of the market shares, indicates stocks that have large market share CV for the opening tend to have large market share CV for the closing auction. * and ** indicate statistical significance at the 5% and 1% levels.

	Correlation Open/Close
€ volume skewness	0.176
share turnover skewness	0.134
€ Market share skewness	0.034
relative share turnover skewness	-0.002
€ volume CV	0.378 *
share turnover CV	0.376 *
€ Market share CV	0.740 **
relative share turnover CV	0.731 **