Determinants of Intra-Day Stock Price Change and Asymmetric Information*

CHRISTOPHER TING†

Lee Kong Chian School of Business
Singapore Management University
469 Bukit Timah Road, Singapore 259756

*This project is funded by the Office of Research, Wharton-SMU Research Center.
†email: christopherting@smu.edu.sg; tel: +65 68220364

Abstract

Trade sign and trade size are two variables used in the literature to capture informed trading. This paper presents a synthesized model of asymmetric information. Our empirical analysis of more than 1,400 NYSE common stocks shows that trade direction is more important than volume in reflecting the asymmetry. Despite its relative importance, existing models that use the trade sign only to capture information asymmetry has an upward bias in the estimation of adverse selection costs. This bias is economically significant. There is also evidence to suggest that signed duration reflects informed trading activity. This paper emphasizes that insiders are not the only source of information asymmetry. If insiders were the only source, then it would suggest that illegal insider trading is quite rampant in the U.S markets. Following Harris (2003), we adopt the view that a trader is asymmetrically better informed if she can arrive at reliable conclusions about the fair values. Even public news can create information asymmetry because traders have different interpretations. The motivation for this broader interpretation of informed trading is to better understand the conflicting results in the literature.

A measure of asymmetric information arises naturally from the synthesized model. This model relies on four variables (lagged trade sign, signed volume, signed duration and duration) that are generally considered to reflect information asymmetry in the literature. They jointly explain the innovations in the efficient price. After controlling for firm size and book-to-market ratio, the analysis demonstrates that our asymmetric information measure is negatively related to the number of analysts following a firm and whether there is an exchange-traded equity option written on the firm's stock. An implication of our findings is that firms can reduce information asymmetry by attracting more analysts and option writers.

The information structure of a firm is determined largely by the information channels through which the state of the firm is revealed to the world. For listed firms, the primary channel is the stock market. Through the stock market, traders convey and exchange information. Better informed traders convey more information, and stand a better chance of making a profit in the market.

Before the release of material information, those who know it and transact ahead of public disclosure are privately informed traders. Investors who do not know the private information is referred to as uninformed traders. In particular, market makers are also uninformed and they trade in a way that minimizes losses to insiders. For example, Copeland and Galai (1983), Kyle (1985), Glosten and Milgrom (1985) and many others draw a crisp line between those who have private information and those who do not.

We find this definition somewhat narrow and inadequate for two reasons. Firstly, trading on material information before public announcement is illegal in many countries and certainly in the U.S. where insider trading laws are enforced. In the absence of evidence to the contrary, it is reasonable to believe that most insiders are law-abiding with respect to these laws. Nonetheless, Easley, Hvidkjaer, and O'Hara (2002) find that the probability of informed trading (PIN) is 19.1% on average for the NYSE firms. Namely, in 19 out of 100 trading days, informed traders are trading on the NYSE. If informed traders were all insiders, it would appear that despite the laws, insider trading is quite rampant and unchecked in the U.S. market. It would also imply that law enforcement agencies such as the SEC were not doing enough to curtail illegal insider trading. Secondly, in markets populated by professional dealers only, adverse selection costs are still significant. An example is the inter-dealer Treasury market analyzed by Green (2004). Intriguingly, Green finds that the level of information asymmetry as measured by a modified model of Madhavan, Richardson, and Roomans (1997) increases *after* the scheduled release of macroeconomic data. How could information asymmetry prevail among these professional traders with regard to *public* information?

Of course, one can turn the empirical evidence around to suggest that either the structural PIN model and the measure used by Green are misspecified. As a matter of fact, all empirical analyses can only *infer* the existence of asymmetric information from the data, as these data do not provide any indicator as to which orders and trades are motivated by private information. It could well be that the models are prone to infer that there is information asymmetry even when privately informed traders are not trading. Several papers document evidence that casts doubts on existing econometric models of information asymmetry. For example, Neal and Wheatley (1998) conclude that the adverse selection components of closed-end funds are larger than expected. They suspect that some of the popular models of asymmetric information might be misspecified. Van Ness, Van Ness, and Warr (2001) also ascertain that existing measures of adverse selection costs relate inconsistently to the corporate finance proxies for the information structure of a firm.

Finally, the econometric models themselves appear to be in conflict with each other. For example, in the basic model proposed by Glosten and Harris (1988), the carrier of asymmetric information is the current signed volume, whereas Huang and Stoll (1997) and Madhavan, Richardson, and Roomans (1997) argue that it is the lagged trade sign that is asymmetrically informative. When the econometric models disagree on the variables that reflect adverse selection, conflicting interpretations of empirical results are inevitable.

Instead of rescinding the existing models, we take a constructive approach by first acknowledging that the customary definition of informed trading is inadequate. This paper suggests that one can better understand the empirical results documented in Easley, Hvidkjaer, and O'Hara (2002) and Green (2004) when a broader framework of informed trading is employed. Within this broader framework, this paper offers a synthesized model to resolve the conflicts. Instead of treating the variables as mutually exclusive, a natural route is to take both signed volume and past trade sign as explanatory variables. It is noteworthy that using one without the other will lead to the problem of variable omission, which will result in biased parameter estimates. Therefore, it may well be that the adverse selection costs estimated in the literature are biased because not all relevant variables are included. This could potentially explain why Neal and Wheatley (1998) and Van Ness, Van Ness, and Warr (2001) find the estimates inconsistent with other proxies for asymmetric information.

The main thrust of this paper is to provide evidence that the information structure of a firm relates consistently to a novel measure that emerges from the joint estimation. Specifically, we find that analysts and equity option markets play a significant role in reducing the level of information asymmetry in the stock market. Since information asymmetry is related to the cost of capital, our findings indicate that it makes sense for firms to attract more analysts and option writers.

In addition to trade sign and signed volume, we include trade-to-trade duration and a signed version of this duration in our specification. Our findings indicate that signed duration also captures asymmetric information. Another contribution is the recognition that a broader framework is needed to clarify the notion of informed trading. We categorize market players into four groups according to their ability in acquiring and digesting information speedily. Whenever there is some sort of a divide that differentiates traders, there will be information asymmetry. Private signal is not the only source. As in Green (2004), different interpretation of the same *public* information could lead to information asymmetry among dealers of U.S. Treasury securities. This broader framework is useful for reconciling some of the conflicting findings in the literature.

The paper is organized as follows. The next section discusses the information structure of a firm and clarifies the notion of informed trading. Three popular models of adverse selection costs are reviewed. A synthesis of these models results in a natural measure of asymmetric information. Data used in our empirical study are described in Section II. Section III reports the empirical findings based on daily parameter estimates. Section IV documents the relations of our measure of informed trading with variables that are used as proxies for the information channels of a firm. In Section V, we summarize the findings and conclude the paper.

I. Informed Traders and Structural Spread Models

A variety of material information is generated by firms' managers, financial service providers and government agencies. In the event of delisting or takeover, the terminal value of a firm is material information. Quarterly financial results and forward guidance, major contracts, changes in dividend payout, capital reductions, private placements, stock splits and so on are important news, especially when the element of surprise is substantial. In addition, index reconstitutions as well as some macroeconomic numbers are also monitored by traders.

With regard to these news, which do not occur everyday for a given stock, this section discusses the information structure and three popular models of adverse selection costs. Motivated by the apparent inconsistency among these existing measures of asymmetric information, we attempt a classification of traders that helps to fine-tune the terms "informed" and "uninformed" used in the literature. The motivation for this discussion is to emphasize that private information is not the only source of information asymmetry. A broader interpretation of informed trading is needed to better understand the implications of, for example, Neal and Wheatley (1998), Easley, Hvidkjaer, and O'Hara (2002) and Green (2004), as well as our empirical findings. Finally, we propose a synthesized model to resolve the conflict and to measure the extent of informed trading.

A. Who are Informed and Who are Uninformed?

In Harris (2003), a trader is said to be informed if she can arrive at reliable conclusions about whether financial instruments are fundamentally overvalued or undervalued. Informed traders understand intrinsic values better than other traders because they have better access to fundamental data and can better analyze the implications from their data. Based on this definition, we consider four types of traders.

Obviously, company insiders are the most informed. Since their transactions are regulated under the insider trading laws, they are deterred from trading on material information. Of course, unscrupulous insiders can tip others and trade through proxies. In addition, espionage on material information by company employees, associates or outsiders as well as inadvertent leakage prior to public announcements can happen. For convenience, this category of informed traders are referred to as insiders, which includes connected persons and people who obtain pre-release material information. Officers of government agencies who prepare and report key macroeconomic statistics that have bearing on the financial markets are also insiders. Similarly, executives responsible for index reconstitutions are in this category as well. The signal they receive is most precise. A common feature of this group of traders is that they possess private information before it is made public.

The second type of informed traders is institutional traders. These traders know that their trading activity can move prices. They also know that their analysts' forecasts, ratings and recommendations are influential. Therefore, even in the absence of firm-specific information, their discretionary portfolio rebalancings affect market prices. Typically, they have access to information systems and news feeds that allow them to gain a better understanding of not only the firms but also the macroeconomic conditions and real-time trades and quotes. Included in this category are designated market makers with inventories. They too attempt to obtain information from news and reports to form reliable valuation of the companies they specialize in. If institutional traders and market makers were not diligent in information gathering to gain deeper insights on the firms and the overall market, they would incur losses and go out of business. These traders are constantly on the look out for insiders' private information in the order flows.

The third type of informed traders are small traders. They are mindful that their trades will not move prices. Nevertheless, they too strive to obtain public announcements of firm-specific information as soon as they are released. More importantly, they also monitor the trades and quotes closely to make real-time trading decisions. Skillful day traders who trade for a living fall under this category. In addition, managers of small funds whose trades are not large enough to create an impact are also considered as small traders. They become informed by reading charts, company and analysts' reports, investment magazines as well as newspapers.

Traders who are not in these three categories are noise traders. They do not understand fundamental values better than other traders because even if they have access to fundamental data, they cannot decipher the implications reliably. Almost surely, every trader will examine past and prevailing prices before they submit an order. Therefore, every market participant is informed to a certain extent. But, the lack of sufficient resources and real-time analytical expertise makes a difference. Their trades are noise because their opinions do not constitute a reliable valuation of the securities and market conditions. Retail investors who cannot devote full time to monitor market pulses are noise traders most of the time. On the contrary, if they spend time and effort to form reliable opinions about the value and the price trend, they will be in the league of small investors discussed earlier. Conversely, institutional and small investors may at times make investment mistakes so that their trades are effectively no different from noise. In other words, every trader except the insider can become a noise trader. Retail followed by small investors are more likely to be in this fourth category.

We stress that institutional traders including the market makers have vested interests to know the fundamental values and the market conditions. They trade strategically for liquidity and profit. Small investors are as motivated to stay informed but they do not have the economy-of-scale advantage of institutional traders.

More importantly, even when insiders abstain from trading, information asymmetry still exists in the other three categories of traders. The main reason is that these outsiders have different capabilities and speed to acquire and process public information. Obviously, traders who trade for a living will expend greater effort learning whether a signal (announcement) has occurred. If it does occur, they will analyze it to determine whether it is a good or bad signal before the trading session. Noise traders, in contrast, are not able to form a correct interpretation of the signal even if they know that the signal exists.

Therefore, it is plausible that the ability to create order imbalance brings about a dichotomy between the informed and the uninformed. The results in Easley, Hvidkjaer, and O'Hara (2002) may be interpreted consistently under this broader framework of informed trading. The average value of about 19% of informed trading found by the PIN model may be reflecting the information asymmetry primarily among the institutional traders, small investors and noise traders, and not so much whether some of them have insider information. The higher-than-expected adverse selection costs in Neal and Wheatley (1998) may also be attributable to traders' different abilities in evaluating closed-end funds' premiums or discounts. Changes in the dividend payout and stock splits may also cause traders to interpret the signals emitted by closed-end funds differently.

Transactions per se are also informative in reflecting the forces of supply and demand. In our framework, market makers are informed on the net demand for buy and sell, which helps them to set the quotes accordingly. Indeed, Green (2004) suggests that primary dealers of Treasury bonds have different interpretations of newly released macroeconomic statistics. They also have different order flows from their clients. Together, these differences give rise to information asymmetry among the dealers.

In the context of this broader interpretation of informed versus uninformed, we turn to the discussion of several market microstructure theories. The motivation is to identify variables that reflect informed trading. Most theories pitch the informed traders against the (institutional) market makers. For example, in the sequential framework of Glosten and Milgrom (1985), informed traders maximize the profit by trading as often as possible. An implication is that signed duration reveals information asymmetry. Their trading direction will affect the price when it is still not at the full-information level. Glosten and Milgrom (1985) further argue that the order size can be normalized to one. Thus, trade sign rather than volume is asymmetrically informative.

This proposition is contrary to the model in Kyle (1985). Under the batch auction trading, the monopolist market maker strives to infer whether some of the orders are submitted by an insider prior to setting a price. The linear equilibrium solution of the Kyle model suggests that signed volume is informative in determining the transaction price. He and Wang (1995) smoothen the sharp dichotomy between insiders and outsiders with the notion of differential information. In their framework, information includes not only new private signals but also public announcements and market prices. Even the private information is differential. Each informed trader has some information that others do not know. An implication of their theory is that signed volume reflects information asymmetry.

Easley and O'Hara (1987) develop a structural model with the market makers and the uninformed traders not knowing whether an information event has occurred. In their theory, the signal is assumed to occur before the trading day begins. Only the informed traders observe the signal. The equilibrium structure of the Easley-O'Hara model is more complicated, but the implication is quite the same as the previous two theoretical models. Both trade size and trade direction reveal asymmetric information.

As a summary, these theories suggest that trade sign, signed duration and volume are variables that reflect asymmetric information.

B. Empirical Models of Asymmetric Information

Most empirical models that attempt to measure asymmetric information are motivated by these microstructure theories. In this subsection, we discuss three popular empirical models. These are the Huang-Stoll (1997), Madhaven-Richardson-Roomans (1997) and Glosten-Harris (1988) models. We show that the Huang-Stoll (HS) and the Madhaven-Richardson-Roomans (MRR) models are essentially the same from the model specification viewpoint. The Glosten-Harris (GH) model is different from these two econometric specifications.

Despite the differences, these three models originate from Roll (1984)'s formulation of the transaction price:

$$P_i = M_i + C\,Q_i\,. \tag{1}$$

In words, the transaction price P_i of the ith trade is postulated to equal the sum of the unobservable efficient price M_i and the transitory transaction cost C. The trade sign Q_i is one or minus one for buyer- and seller-initiated transactions, respectively.

By definition, the efficient price is a random walk process given by

$$M_i = M_{i-1} + u_i \, . \tag{2}$$

Since u_i affects the efficient price directly, it contains information. Depending on the assumptions made for the random component u_i, we show that these three models are obtainable from equation (1) as variants of the following canonical form:

$$\Delta P_i = C \, \Delta Q_i + u_i \, , \tag{3}$$

where Δ is the first-order difference operator. To make estimation possible, all the three models rely on this equation with different postulates for u_i.

After reviewing these three models, we propose a framework that synthesizes the key insights. The synthesis leads to a normalized quantity that can be interpreted as a measure of information asymmetry. The proposed model starts from equation (3) without postulating a mechanical scheme that dictates how designated market makers and informed traders behave. A by-product of this approach is that it extends the applicability of these three existing econometric models to markets where market makers are not appointed.

B.1. Huang-Stoll Model

In the basic model of Huang and Stoll (1997), volume per se has no information. It is the trade sign Q_{i-1} that is informative. Motivated by the theoretical analysis of Glosten and Milgrom (1985), the HS model assumes the existence of a designated market maker who learns from the order flows. From the preceding trade, the market maker infers the likelihood of trading with an informed trader. She then updates her opinion about the current efficient price M_i based on the trade sign Q_{i-1} of the previous transaction. Volume traded V_{i-1} does not matter because both small and large trades are assumed to contain the same information that is condensed in the trade sign Q_{i-1} alone. Hence, the innovation u_i in the efficient price process is postulated to be $A_{\text{HS}} \, C Q_{i-1} + \epsilon_i$, where A_{HS} is the adverse selection cost as a percentage of the implicit transaction cost C, and ϵ_i is pure noise. Hence, the efficient price M_i is

$$M_i = M_{i-1} + A_{\text{HS}} \, C \, Q_{i-1} + \epsilon_i \, . \tag{4}$$

The coefficient A_{HS} provides a measure of adverse selection costs in percent. Obviously, a large A_{HS} means that the market is highly suspicious of informed trading. On the other hand, $A_{\text{HS}} \approx 0$ indicates that the information each trader has is more or less symmetric. From equation (3), the price change ΔP_i in the HS model is given by

$$\Delta P_i = C \Delta Q_i + A_{\text{HS}}\, C\, Q_{i-1} + \epsilon_i\,. \tag{5}$$

B.2. Madhaven-Richardson-Roomans Model

Starting from a premise different from Huang and Stoll (1997), Madhavan, Richardson, and Roomans (1997) arrive at

$$\Delta P_i = (\phi + \theta) Q_i - (\phi + \rho\,\theta) Q_{i-1} + \epsilon_i\,. \tag{6}$$

Here, ρ is the first-order correlation in the signed volume, ϕ is the frictional component and θ the asymmetric component. In percent, the adverse selection component of the MRR model is

$$A_{\text{MRR}} = \frac{\theta}{\theta + \phi} \tag{7}$$

After a re-arrangement of terms, the MRR model is formally the same as the HS model:

$$\Delta P_i = (\phi + \theta) \Delta Q_i + (1 - \rho) \theta\, Q_{i-1} + \epsilon_i\,. \tag{8}$$

In a way, the MRR model is an improvement over the HS model in that the autocorrelation of trade sign is taken into account. From the standpoint of Roll's formulation, the implied assumption for the efficient price in the MRR model is

$$M_i = M_{i-1} + \theta \Big(Q_{i-1} - E\big[Q_i | Q_{i-1} \big] \Big) + \epsilon_i\,. \tag{9}$$

Similar to Hasbrouck (1991a) and Hasbrouck (1991b), the postulate is that it is the surprise defined as $Q_{i-1} - E\big[Q_i|Q_{i-1}\big]$ in the trade sign that contains asymmetric information. In other words, the expected component in the trade sign, $E\big[Q_i|Q_{i-1}\big]$, does not contain asymmetric information as market makers anticipate some amount of serial correlation in the trade sign. Therefore, this anticipated component is discounted from the nominal adverse selection component $\theta\, Q_{i-1}$.

Since the conditional expectation $E\big[Q_i|Q_{i-1}\big]$ is $\rho\, Q_{i-1}$, equation (9) becomes

$$M_i = M_{i-1} + (1-\rho)\theta\, Q_{i-1} + \epsilon_i\,. \tag{10}$$

To obtain an alternative derivation of equation (8), we let $C = \phi + \theta$ in equation (1). Applying the first-order difference operator on both sides of equation (1) and substituting in equation (10), the MRR model ensues.

Moreover, it can be shown that the adverse selection component of the MRR model expressed in equation (7) as A_{MRR} is proportional to that of the HS model $\big(A_{\text{HS}}\big)$ as follows:

$$A_{\text{MRR}} = \frac{A_{\text{HS}}}{1-\rho}\,. \tag{11}$$

This result is obtained when we compare equation (5) with equation (8). Therefore, the econometrics of these two models are quite the same. The insight gained from this analysis is that two important models of asymmetric information converge to the same measure, up to a factor of $1-\rho$. Above all, both Huang and Stoll (1997) and Madhavan, Richardson, and Roomans (1997) assert that Q_{i-1} is the carrier of asymmetric information.

B.3. Glosten-Harris Model

In the GH model, the random component u_i is postulated to contain information with A_{GH} being the adverse selection parameter:

$$u_i = A_{\text{GH}}\, C\, X_i + \epsilon_i\,, \tag{12}$$

In contrast to Huang and Stoll (1997) and Madhavan, Richardson, and Roomans (1997), Glosten and Harris (1988) assert that the signed volume X_i conveys asymmetric information. In other words, the efficient price is influenced by the signed volume:

$$M_i = M_{i-1} + A_{\text{GH}}\, C\, X_i + \epsilon_i\,. \tag{13}$$

This specification for the efficient price is rooted in the notion that informed traders want to trade a large amount so as to maximize the value of their private information as in Kyle (1985) and Easley and O'Hara (1987). Thus, their transaction orders X_i impound the information they possess into the efficient price. If informed traders know the good news, they will buy as many shares as they could. As a result, the efficient price increases when $X_i \geq 0$, as evident from equation (13). Conversely, the efficient price will decline if they sell. Under this assumption, the main model estimated in Glosten and Harris (1988) is

$$\Delta P_i = C\,\Delta Q_i + A_{\mathrm{GH}}\,C\,X_i + \epsilon_i\,. \tag{14}$$

The parameter of interest is the adverse selection cost A_{GH} as a percentage of C.

The GH model is quite different from the HS and the MRR models because the carrier of asymmetric information is the signed volume X_i rather then the trade sign Q_{i-1}. Since signed volume is different from trade sign, it is apparent that there is a conflict in these models.

C. Asymmetric Information Measure of Informed Trading

If one regards equation (3) as the main econometric specification, then it boils down to identifying the relevant explanatory variables for u_i. As a summary, we present the above three models' postulates as follows:

- The HS model: $u_i = A_{\mathrm{HS}} C Q_{i-1} + \epsilon_i$
- The MRR model: $u_i = (1-\rho) A_{\mathrm{MRR}} C Q_{i-1} + \epsilon_i$
- The GH model: $u_i = A_{\mathrm{GH}} C X_i + \epsilon_i$

The residual ϵ_i is of course different from one model to the other. If, for example, the signed volume X_i is also a carrier of information, then the residual in the HS model should contain this variable.

To eschew misspecification, it is natural to admit both Q_{i-1} and X_i as explanatory variables rather than favoring one over the other. Therefore, we postulate that the innovation u_i of efficient price in equation (2) is

$$u_i = b_1 Q_{i-1} + b_2 X_i + b_3 Q_i \Delta T_i + b_4 \Delta T_i + \epsilon_i\,. \tag{15}$$

The first variable is motivated by the HS and the MRR models, and the second by the GH model. We also examine whether inter-trade duration[3] ΔT_i and signed duration $Q_i \Delta T_i$ reflect informed

[3]If the efficient price M_i follows geometric Brownian motion, then the duration ΔT_i is a relevant variable. In this context, b_4 is the drift rate of the efficient price.

trading. These two variables are motivated by Glosten and Milgrom (1985), Diamond and Verrecchia (1987) and Easley and O'Hara (1992). Their theoretical results suggest that durations are indicative of informed trading. Moreover, Dufour and Engle (2000) demonstrate that durations do affect price updates.

In view of this specification, the three models discussed earlier have the problem of omitted variables. In the HS and the MRR models, the signed volumes X_i as well as the duration ΔT_i and signed duration $Q_i \Delta T_i$ are omitted. For the GH model, the same problem occurs. From the standpoint of econometrics, when all the relevant variables are not specified, the estimates of asymmetric information from these models are likely to be biased.

On the other hand, one may suspect that some variables in equation (16) are irrelevant. However, in this case, there is no bias in the parameter estimates even if some variables are irrelevant. The drawback, however, is that the estimators for the parameters are inefficient. Nonetheless, the estimator $\widehat{\sigma}^2$ for the residual variance determined by the residual ϵ_i is unbiased. This property is crucial in ensuring that the R^2 of equation (16) is also unbiased, since it is a linear function of $\widehat{\sigma}^2$ (see Maddala (1977) and Maddala (2001)).

With the innovation u_i specified as in equation (15), we obtain from equation (3) the following specification to explain intra-day stock price change:

$$\Delta P_i = C \Delta Q_i + b_1 Q_{i-1} + b_2 X_i + b_3 Q_i \Delta T_i + b_4 \Delta T_i + \epsilon_i \,. \tag{16}$$

This econometric model follows the path pioneered by the HS, MRR and GH models. The inclusion of duration and signed duration extends these structural models to capture information asymmetry that is revealed in trade frequency.

Equation (16) also allows us to investigate which of the variables have more explanatory power from their Newey and West (1987) t-statistics. This test is useful in ascertaining whether trade size is more important than trade sign in capturing asymmetric information. Put differently, one may take equation (16) as the GH model with the other variables $Q_{i-1}, Q_i \Delta T_i$ and ΔT_i as controls. If b_2 is still significantly positive, then one may conclude that signed volume X_i is a carrier of information asymmetry after controlling for lagged trade sign Q_{i-1}, duration ΔT_i and sign duration $Q_i \Delta T_i$. Similarly, equation (16) provides a robustness check for the HS and the MRR models with X_i, $Q_i \Delta T_i$ and ΔT_i as controls instead.

If the stock price change is purely frictional and void of information, the innovation u_i of equation (3), which is nested in equation (16), will be pure noise. Namely, u_i and ϵ_i in equation (16) do not yield statistically different residual sum of squares in this circumstance. Their respective R^2 values, denoted as R_c^2 for equation (3) and R_{all}^2 for equation (16), will not be much different as a result. Conversely, when u_i is a function of these four variables, then R_{all}^2 will be larger than R_c^2.

To quantify the joint explanatory power of X_i, Q_{i-1}, $Q_i \Delta T_i$ and ΔT_i in equation (16) we consider the following relation (see Maddala (2001)):

$$1 - R_{\text{all}}^2 = (1 - R_{\text{asy}}^2)(1 - R_c^2)\,, \tag{17}$$

where R_{asy}^2 is the additional goodness-of-fit contributed jointly by the four variables predicated to reflect informed trading. This equation is basically an identity of linear regression rather than an empirical relation. As equation (3) is nested in equation (16), the unexplained portion $1 - R_c^2$ in equation (3) will be reduced by a factor given by $1 - R_{\text{asy}}^2$ when four variables are added. The product of these two quantities equals $1 - R_{\text{all}}^2$.

Instead of using the estimates for b_1, b_2, b_3 or b_4 as (absolute) adverse selection costs, one could use R_{asy}^2 to quantify information asymmetry. This measure summarizes the explanatory power jointly contributed by the four variables. The motivation for this asymmetric information measure is twofold. First, if we use only, for example, the coefficient b_1 of the lagged trade sign to estimate adverse selection costs, then the asymmetric information components captured by b_2, b_3 and b_4 are not accounted for. On the other hand, one cannot simply add these components together because each reflects a different dimension of information asymmetry. Moreover, b_1 is in dollars, b_2 in dollars per share, b_3 and b_4 are in dollars per unit time. R_{asy}^2 has the added advantage that its value is between zero and unity, which makes it possible to interpret the proportion of price change that is jointly explained by the four variables of informed trading. By using R_{asy}^2 , this paper resolves the conflict between models that depend on Q_{i-1} only and models that use X_i only to quantify adverse selection costs.

Second, R_{asy}^2 is complementary to PIN and a relative measure proposed in Hasbrouck (1991b). PIN is estimated with the numbers of buys and sells. Price-sensitive information encoded in volume and duration is not incorporated in the PIN model. Also, it is not possible to obtain PIN estimates on the daily basis and one has to assume that the parameters needed to construct PIN remain constant over the sample period. Hasbrouck (1991b) measures asymmetric information in trade innovation with vector autoregression methodology. This is different from our framework based on Roll's definition, equation (1). Although VAR's approach has the advantage that the estimation is robust to the specification of transitory component, there is no solid reason to dispute the sensibility of equation (1). From the practical standpoint, the data requirement is less demanding with our approach than the VAR approach. This is especially true for less liquid stocks, which do not have sufficient observations to make VAR feasible. These potential limitations provide a motivation to use R^2_{asy} as a complementary measure to quantify information asymmetry.

II. Data

The sample period of our study is from January 2, 2003 through December 31, 2003, a total of 252 trading days. The reason for choosing year 2003 is that data such as the number of shareholders needed in our analysis are not as widely obtainable for earlier years. Another reason is the need to control for microstructure noise generated by price discreteness. After February 2001, the effect of price discreteness should be considerably smaller compared to the pre-decimal tick-size regimes. Moreover, by 2003, traders should have become accustomed to trading with the minimum tick size of one cent.

From CRSP database, a sample of NYSE common stocks is taken. Firms with negative stock prices or market capitalizations are excluded. For firms with more than one class of security, we use TAQ's master file to cross-examine and choose the one that is representative of the firm according to the CUSIP as well as the security's name. With all these filters, the final sample size is 1,461 stocks. Panel A of Table I reports the descriptive statistics for these NYSE common stocks.

Intra-day trades and quotes are obtained from TAQ. We sign the trades based on the algorithm of Lee and Ready (1991) but without the five-second rule. In the same panel, we tabulate the descriptive statistics for the trading activity. Seller-initiated transactions are indicated with negative statistics. Annual data for the number of shareholders and book values are extracted from Compustat. The number of analysts following a firm is derived from I/B/E/S. Panel A of Table I also provides the descriptive statistics for these data. In Panel B, we report the correlation coefficients. As expected, market capitalization correlates positively with both shareholders and analysts but negatively with book-to-market ratio. We also find that the book-to-market ratio relates negatively with the number of analysts but insignificantly with shareholders. As anticipated, the latter two numbers correlate positively.

Finally, information regarding whether a firm has exchange-traded equity options is obtained from the Options Clearing Corporation.

III. Estimations and Analysis

For each trading day and each stock, we perform a GMM regression based on equation (16) and use (17) to obtain the asymmetric information measure R^2_{asy}. In total, there are 329,187 stock-days in our sample. We plot the time series of daily cross-sectional average R^2_{asy} values in Figure 1. It is evident that average R^2_{asy} fluctuates from one trading day to the other.

The distribution of 329,187 R^2_{asy} values is plotted in Panel A of Figure 2. On average, this measure of informed trading is 19.7% with a standard deviation of 9.2%. In other words, after controlling for transitory component C, the last four explanatory variables in equation (16) jointly explain about 20% of the variation in the high-frequency price change. It follows that 20% of the trades in our sample of NYSE firms are associated with informed trading in the broader sense discussed in Section I.A. Panel B displays the histogram of the annual average R^2_{asy} values for our 1,461 sample stocks.

A. Relative Importance of Explanatory Variables

To examine the relative importance of informed trading variables, we count the numbers of statistically significant coefficients in equation (16). The results are reported in Table II. We find that b_1, the coefficient of preceding trade sign Q_{i-1}, is the main contributor. At the two-tail 5% and 1% levels, b_1 is, respectively, 92.65% and 87.33% of the times significant. Therefore, the HS or the MRR model accounts for a larger portion of the variation in the non-transitory price change.

Results also show that signed volume X_i and signed duration $Q_i\Delta T_i$ are the next two most important explanatory variables. While the role of X_i has been well studied, it appears that the importance of signed duration has been overlooked in the literature on adverse selection costs. The coefficient of the duration ΔT_i, however, is significant only 5.77% of the times at the 1% level. This is three times lower than signed duration's 19.25%. There is therefore some evidence to suggest that it is not the duration per se but the signed duration that captures informed trading. As a robustness check, intra-day observations are pooled over a month. Same regressions are performed, but the results do not change the conclusions.

As shown in Panel C of Table II, among the parameter estimates significant at the 1% level, more than 93% of b_1, b_2 and b_3 are positive. The positivity of these estimates for the lagged trade sign, the signed volume and the signed duration, respectively, lends support for their roles in revealing asymmetric information. In particular, positive loading on the signed duration is consistent with the notion that after a long duration, a trade has more impact in moving the stock price in the direction of its trade sign. An interpretation for this result is as follows. When the market is quiet, it is likely that traders are waiting for fresh news. During a long duration with no trade, there is no news. So when a trade occurs after a long period, that trade is deemed to set the direction for the market to follow and thus contains more asymmetric information than other transactions. Recall that in this paper's broader interpretation of informed trading, insiders are not the only source of information asymmetry. Transactions themselves generate (public) information. When there has not been a trade since the last transaction, the new trade creates a blip that is likely to be asymmetrically informative in moving the stock price.

B. Bias in the Huang-Stoll Model

With Q_{i-1} found to be the most important carrier of informed trading, we examine the amount of bias in A_{HS} when the other relevant variables, X_i, $Q_i\Delta T_i$ and ΔT_i are omitted. Specifically, we write b_1 as AC and compare A with A_{HS} of equation (5) to check if these two estimates (the adverse selection costs as percentages of the transaction costs) are statistically different. The specification in equation (16) may be viewed as the HS model with the signed volume and duration-related variables as controls.

We estimate daily A and A_{HS} values using intra-day data for each stock. To ascertain whether there is a bias in A_{HS} due to variable omission, we examine a subsample of 1,409 stocks that have more than 10 days of daily estimates. A two-population mean test is performed for these stocks. Table III reports the results by market capitalization quintile. Within each quintile, the first to 99-th percentile statistics for A_{HS}, A and the test results are tabulated. We find that estimated A and A_{HS} values are larger for smaller capitalization stocks, as anticipated. Also, within each quintile, A_{HS} is larger than A for the tabulated percentiles except the two 99-th percentiles in the first and second quintiles.

Even at the first percentile, the t-statistic of the two-population mean test in each quintile is above 2. This result suggests that the means of A_{HS} and A are not the same, and $A_{\text{HS}} - A$ is statistically positive. We also conduct the non-parametric Wilcoxon signed rank tests. The Z-statistics for all quintiles indicate that the p values are zero. Therefore, we conclude that the medians of A_{HS} and A are not the same. All these results provide evidence to suggest that A_{HS} has an upward bias[4].

One may argue that contributions from the other parameters in equation (16) should be added to A. But, this is not possible because, as mentioned earlier, b_2 depends on the unit used for trade size, and b_3 and b_4 on the unit employed to measure time. In Glosten and Harris (1988), a unit of a thousand shares is used. However, there is no a prior reason why a thousand shares is a suitable aggregate unit. If the number of shares is used instead, b_2 will be a thousand times smaller because the regression is linear. As a result, the asymmetric information cost contributed by b_2 will be negligibly small compared to the frictional component. The same argument applies with respect to b_3 and b_4. This observation provides further motivation for using a unit-independent quantity such as the R^2_{asy} as a measure of asymmetric information.

IV. Asymmetric Information Measure

This section examines the determinant of R^2_{asy}. We study the relations between R^2_{asy} with firm size, book-to-market and other variables that serve as proxies for the information channels. To make the exposition clearer, we define

$$\text{AIM}_j \equiv \text{Average } R^2_{\text{asy}} \text{ of firm } j \text{ over the sample period} \tag{18}$$

[4]Analogous analysis is performed to estimate the bias in A_{MRR} of the MRR model. Since we have demonstrated that it is proportional to A_{HS}, an upward bias is also found as anticipated.

The distribution of 1,461 AIM$_j$ is plotted in Panel B of Figure 2. Average AIM is 20.5%, which suggests that this proportion of the intra-day price changes is attributable to asymmetric information. This value seems to be quite large if the asymmetry in information is caused only by insiders who trade ahead of public disclosures. But, if one uses the broader framework discussed in Section I.A, where some traders are better informed by virtue of their superior ability in pricing the fair value of a stock, then 20.5% appears to be a reasonable estimate.

As anticipated, larger firms do not have as much asymmetric information among traders as smaller firms. We also find that higher book-to-market or value firms have a larger proportion of informed trading. However, the more robust relation is with the number of analysts and whether there are exchange-traded equity options on firms' shares.

A. Informed Trading, Firm Size and Book-to-Market Ratio

Everything else being equal, it is reasonable to expect larger market capitalization firms to have smaller amount of asymmetric information. This is because news on large firms are more frequently reported in the mass media and more investors are trading these stocks. It follows that AIM$_j$ should be negatively related to firm size.

Firms with large book-to-market ratios are considered to have prices driven down by a string of bad news, and may be near financial distress (Fama and French (1995)). Better informed traders will know whether these value firms are either on the verge of bankruptcy, or are on the way to come back. Therefore, AIM$_j$ should be higher for larger book-to-market ratios.

We find that AIM$_j$ is correlated with firm size in logarithmic levels. The coefficient of correlation is -0.585 with a t-statistic of -23.9. For the book-to-market ratio in logarithmic levels, the correlation coefficient is 0.209 with a t-statistic of 7.87. Although not highly correlated, these two statistically significant results are consistent with the intuitive propositions.

B. Relations with three Channels of Information

In the beginning of this paper, we enunciate the information structure of a firm as channels through which information about the firm is revealed to the market, which is reflected in the stock price. The number of traders in this channel determines the channel capacity in revealing the information. Viewed from the perspective of information structure, it is natural to reckon that analysts constitute another channel for the information flow. The bandwidth of this channel is determined by the number of analysts.

At any given time, the stock price represents a consensus valuation by traders. It follows that the equity option market provides yet another channel, as the option price and the stock price are related. Therefore, we consider the following variables that characterize the information structure of firm j:

- Shareholders_j: Number of firm j's shareholders in logarithmic levels[5];
- Analysts_j: Number of analysts following firm j in logarithmic levels;
- Option_j: An indicator variable that equals one if the stock of firm j is an underlying instrument of some options traded in some option exchanges, and zero otherwise.

The number of shareholders is used as a proxy for the number of traders in the stock market. In general, a larger number of shareholders is likely to correlate with a greater degree of heterogeneity in their stock valuations. Regression A in Table IV provides a test for this heterogeneity hypothesis. The tabulated t-statistics in the parentheses are adjusted by Newey-West (1987)'s algorithm to account for potential heteroskedasticity and inter-stock correlations. As anticipated, we find that AIM_j relates negatively to Shareholders_j with a coefficient of -1.06.

Next, we consider the information channel characterized by Analysts_j. It is reasonable to assume that analysts' reports are informative in general. Their reports and forecasts at least raise the public awareness of the firms they are following. Consequently, investors may become more receptive to trade the stocks of these firms. More analysts tend to narrow the information gap between the informed and the uninformed. The level of information asymmetry should be lower when there are more analysts providing their assessments and recommendations. Regression B in Table IV tests the relation between AIM_j and Analysts_j. We obtain a Newey-West t-statistic of -27.6, which implies that their relation is negative and statistically significant.

We find that Analysts_j is more powerful in explaining the cross-sectional variation in AIM_j. The goodness-of-fit for Regression B is more than 45%. In contrast, Shareholders_j in Regres-

[5]The logarithmic scale is used because the number of shareholders varies over a few orders of magnitude in our sample.

sion A explains only 11.5%. This could be attributable to the possibility that Shareholders_j is a noisy proxy for the number of traders.

The existence of an option market with the stock as the underlying instrument provides an alternative channel for information flow. Traders who are confident about their valuation may find options attractive owing to the leverage they provide. These informed traders may choose to make a profit not from the stock market but from the option market instead. Consequently, fewer traders are asymmetrically informed on the stock market, meaning that Option_j is negatively related to AIM_j, as it draws some informed traders away from the stock market.

In Regressions C and D, the indicator variable Option_j is included in the specification. Consistent with the hypothesis that informed traders may choose to take advantage on the equity option market instead, we find that AIM_j is negatively related to Option_j. The double-digit t-statistics for Option_j in these two regressions strongly suggest that a stock that attracts option writers and option traders have a smaller amount of information asymmetry on the stock market.

It is noteworthy that when Option_j is included, the loadings on Shareholders_j and on Analysts_j are reduced. The coefficient for Shareholders_j is -0.55 in Regression C as compared to -1.06 in Regression A. Similarly, the coefficient for Analysts_j changes from -4.32 to -2.80 when Option_j is included. We perform Regression E to further examine the relative effectiveness of the three variables, Shareholders_j, Analysts_j and Option_j in explaining cross-sectional AIM_j. All the coefficients for the three information channel variables are found to be negatively significant.

Finally, we regress AIM_j on these three information channel variables along with Cap_j (the market capitalization) and B/M_j (the book-to-market ratio) in logarithmic levels. The result is as follows:

$$\begin{aligned} \text{AIM}_j \;=\; & \underset{(23.6)}{30.1} - \underset{(-2.46)}{0.166}\,\text{Shareholders}_j - \underset{(-11.1)}{2.383}\,\text{Analysts}_j - \underset{(-13.1)}{5.026}\,\text{Option}_j - \underset{(-2.81)}{0.357}\,\text{Cap}_j - \underset{(-2.89)}{0.501}\,\text{B/M}_j\,. \end{aligned} \tag{19}$$

The Newey-West t-statistics are shown in the parentheses. The adjusted R^2 of this regression is 58.0%, which is marginally better than 57.1% of Regression E. More importantly, Shareholders_j is significant at the 5% level, while Analysts_j and Option_j are consistently significant at the 1%

level for this specification, and for all the regressions reported in Table IV. Therefore, we have evidence to suggest that more traders and analysts following the firm help to reduce information asymmetry, and the option market diverts informed traders from the primary information channel, the stock market.

V. Summary and Concluding Remarks

An understanding of informed trading has implications for financial market regulators, institutional traders and investors at large. In addition, the design of information structure is important in making corporate decisions on how material information should be managed. Since a firm with higher asymmetric information costs tends to have higher capital costs, a market microstructure analysis of informed trading is potentially useful for corporate finance.

We show that existing models of asymmetric information suffer from an omitted variable problem. They are also inconsistent with each other with regard to either the trade size or the trade sign as the variable that captures asymmetric information. This paper jointly examines four variables predicated to reflect informed trading. These variables are lagged trade sign, current signed volume, as well as duration and signed duration. Overall, the results indicate that lagged trade sign reflects information asymmetry better than the other three variables. In view of this finding, we conclude that the econometrics in Huang and Stoll (1997) and Madhavan, Richardson, and Roomans (1997) provide a better measure to quantify adverse selection costs. However, the estimates of adverse selection costs produced by these two models tend to have an upward bias.

We quantify the explanatory power of these four determinants of intra-day stock price change. On average, they jointly explain 19.7% of the variation that is attributable to information asymmetry. This paper provides a more generic framework to interpret informed trading. In our framework, not only are the insiders informed, small investors who can arrive at reliable conclusions concerning the fundamental values of firms are also deemed to be better informed than noise traders. There is information asymmetry even when the trades are not motivated by private information. Transactions per se are informative as well. Our empirical results show that trades after a long duration tend to move stock price in the direction of their trade signs.

The empirical findings with a subsample of 1,207 NYSE common stocks in year 2003 support the hypotheses that analysts and equity option markets are helpful in reducing the information asymmetry among traders in the stock market. Specifically, after controlling for firm size and book-to-market value, we find that the asymmetric information measure (AIM) relates negatively to the number of analysts following a firm. AIM will also be lower if there are exchange-traded options with the firm's common stock as the underlying instrument. These results have practical implications for firms' managers. To improve the information structure, our findings suggest that it is worthwhile to consider ways to attract more analysts and option writers.

References

Copeland, Thomas E., and Dan Galai, 1983, Information effects on the bid-ask spread, *Journal of Finance* 38, 1457–1469.

Diamond, Douglas W., and Robert E. Verrecchia, 1987, Constraints on short-selling and asset price adjustment to private information, *Journal of Financial Economics* 18, 277–311.

Dufour, Alfonso, and Robert F. Engle, 2000, Time and the price impact of a trade, *Journal of Finance* 55, 2467–2498.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2002, Is information risk a determinant of asset returns?, *Journal of Finance* 57, 2185–2221.

Easley, David, and Maureen O'Hara, 1987, Price, trade size, and information in securities markets, *Journal of Financial Economics* 19, 69–90.

Easley, David, and Maureen O'Hara, 1992, Time and the process of security price adjustment, *Journal of Finance* 47, 577–605.

Fama, Eugene F., and Kenneth R. French, 1995, Size and book-to-market factors in earnings and returns, *Journal of Finance* 50, 131–155.

Glosten, Lawrence R., and Lawrence E. Harris, 1988, Estimating the components of the bid/ask spread, *Journal of Financial Economics* 21, 123–142.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a market-maker market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.

Green, Clifton T., 2004, Economic news and the impact of trading on bond prices, *Journal of Finance* 59, 1201–1233.

Harris, Larry, 2003, *Trading and exchanges: Market microstructure for practitioners*. (Oxford University Press, New York, New York).

Hasbrouck, Joel, 1991a, Measuring the information content of stock trades, *Journal of Finance* 46, 179–207.

Hasbrouck, Joel, 1991b, The summary Informativeness of stock trades: An econometric analysis, *Review of Financial Studies* 4, 571–595.

He, Hua, and Jiang Wang, 1995, Differential information and dynamic behavior of stock trading volume, *Review of Financial Studies* 8, 919–972.

Huang, Roger D., and Hans R. Stoll, 1997, The components of the bid-ask spread: A general approach, *Review of Financial Studies* 10, 995–1034.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1335.

Lee, Charles M. C., and Mark J. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733–746.

Maddala, G. S., 1977, *Econometrics*. (McGraw-Hill, London, England).

Maddala, G. S., 2001, *Introduction to econometrics, third edition*. (John Wiley, West Sussex, England).

Madhavan, Ananth, Matthew Richardson, and Mark Roomans, 1997, Why do security prices change? A transaction-level analysis of NYSE stocks, *Review of Financial Studies* 10, 1035–1064.

Neal, Robert, and Simon M. Wheatley, 1998, Adverse selection and bid-ask spreads: Evidence from closed-end funds, *Journal of Financial Markets* 1, 121–149.

Newey, Whitney, and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–798.

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

Van Ness, B. F., R. A. Van Ness, and R. S. Warr, 2001, How well do adverse selection components measure adverse selection?, *Financial Management* 30, 77–98.

Table I
Annual Statistics for 1,461 NYSE Firms

Panel A tabulates the number of common shares outstanding, market price and capitalization as at end of December 2002. These statistics are obtained from CRISP. From Compustat, annual book values are extracted to arrive at the book-to-market ratios. The number of shareholders is also from Compustat whereas the number of analysts is derived from I/B/E/S. Not all our sample firms had transactions on each of the 252 trading days in year 2003. Firms in the first percentile had only 55.1 days. Annual statistics for the trading activity are based on TAQ. We report seller-initiated transactions in negative numbers. Panel B tabulates the correlation coefficients. All the correlation coefficients are significant at the 5% level except the correlation between the book-to-market ratio and the number of shareholders.

Panel A: Descriptive Statistics

	Unit	Mean	Standard Deviation	Percentile 1st	25th	50th	75th	99th
Common Shares	Billions	0.19	0.53	0.005	0.03	0.05	0.14	2.37
Price	$	23.86	25.53	1.269	10.81	20.62	31.89	77.45
Capitalization	Billion $	5.29	17.66	0.026	0.35	1.03	3.16	83.34
Book-to-Market $\times 10^3$		0.74	0.71	0.055	0.37	0.59	0.86	3.57
Number of Shareholders	Thousands	36.30	162.21	0.045	1.24	4.67	17.50	667.47
Number of Analysts		11.17	7.94	1	5	10	16	34
Days of Trading		244.1	33.0	55.1	251	252	252	252
Number of	Thousands	105.5	109.4	0.488	25.89	69.30	154.22	505.8
Trades		-87.8	-94.3	-0.415	-20.71	-56.11	-124.94	-440.3
Volume	Millions	1.06	2.09	0.004	0.11	0.36	1.13	10.01
Traded		-0.77	-1.57	-0.003	-0.09	-0.26	-0.80	-7.46
Dollar Volume	Million $	30.4	63.5	0.015	1.73	7.74	29.98	314.1
Traded		-22.2	-47.7	-0.014	-1.36	-5.60	-21.68	-236.1

Panel B: Correlation Coefficients

	Capitalization	Book-to-Market	Shareholders	Analysts
Capitalization	1	-0.152	0.370	0.388
Book-to-Market		1	-0.032	-0.252
Shareholders			1	0.277
Analysts				1

Table II
Relative Importance of Variables that Capture Asymmetric Information

This table reports the summary statistics for the coefficients of the following regression:

$$\Delta P_i = b_0 + C\Delta Q_i + b_1 Q_{i-1} + b_2 X_i + b_3 Q_i \Delta T_i + b_4 \Delta T_i + \epsilon_i \,.$$

The dependent variable is the trade-to-trade price change. The explanatory variables of interest are lagged trade sign Q_{i-1}, signed volume X_i, signed duration $Q_i\Delta T_i$ and duration ΔT_i. In total, 329,187 GMM regressions have been performed. In Panels A and B, the percentage is computed with this number as the denominator. Panel C uses the numbers in Panel B to arrive at the percentages of positive estimates that are significant at the 1% level.

Panel A: 5% Significance Level

	b_1	b_2	b_3	b_4
Number	304,989	143,759	107,362	49,156
Percent	92.65	43.67	32.61	14.93

Panel B: 1% Significance Level

	b_1	b_2	b_3	b_4
Number	287,475	89,958	63,376	18,996
Percent	87.33	27.33	19.25	5.77

Panel C: Positive Estimates (1% Level)

	b_1	b_2	b_3	b_4
Number	287,424	87,714	59,174	7,188
Percent	99.98	97.51	93.37	37.84

Table III
Summary Statistics for Adverse Selection Costs

This table reports the summary statistics for the parameter estimates in the following two specifications:

$$\begin{aligned} \Delta P_i &= a_0 + C_{\mathrm{HS}}\,\Delta Q_i + C_{\mathrm{HS}}\,A_{\mathrm{HS}}\,Q_{i-1} + \varepsilon_i\,; \\ \Delta P_i &= b_0 + C\,\Delta Q_i + C\,A\,Q_{i-1} + b_2 X_i + b_3 Q_i \Delta T_i + b_4 \Delta T_i + \epsilon_i\,. \end{aligned}$$

In these two specifications, ΔP_i is the trade-to-trade price change, Q_i the trade sign, X_i the signed volume and ΔT_i the trade duration. Residuals are denoted by ε_i and ϵ_i, respectively. Quantities of interest are the adverse selection cost A_{HS} proposed by Huang and Stoll (1997) in the first specification, and the corresponding construct A in the second specification. These costs are presented as percentages of the implicit transaction costs denoted by C_{HS} and C, respectively. The second specification provides a platform to examine the bias in A_{HS} when the three variables $X_i, Q_i\Delta T_i$ and ΔT_i are omitted. For each sample stock, a two-population mean test for the null hypothesis of $A_{\mathrm{HS}} - A = 0$ is performed. The stocks are grouped into quintiles by market capitalization. The first quintile contains 293 smallest market capitalization stocks. For the other four quintiles, each has 292 stocks. Within each quintile, the percentile (per) statistics for A_{HS}, A and the t-statistics are reported. Overall, the results provide evidence that irrespective of firm size, A_{HS} has an upward bias.

First Quintile: Smallest Market Capitalization Stocks

	1st per	25th per	50th per	75th per	99th per
A_{HS} (in %)	17.20	46.28	58.71	71.66	107.02
A (in %)	5.85	38.05	51.95	66.74	109.79
t-statistic	2.01	2.97	3.74	4.78	7.73

Second Quintile

	1st per	25th per	50th per	75th per	99th per
A_{HS} (in %)	21.15	48.04	59.51	71.09	103.31
A (in %)	11.20	40.72	53.29	66.20	104.95
t-statistic	2.26	3.79	4.65	5.48	9.12

Third Quintile

	1st per	25th per	50th per	75th per	99th per
A_{HS} (in %)	18.68	44.75	55.80	66.56	94.31
A (in %)	11.93	38.12	49.77	61.36	93.12
t-statistic	2.37	4.58	5.37	6.42	12.84

Fourth Quintile

	1st per	25th per	50th per	75th per	99th per
A_{HS} (in %)	20.46	42.12	51.15	59.88	83.04
A (in %)	14.03	35.60	45.15	54.40	79.44
t-statistic	2.99	5.66	7.05	8.19	12.87

Fifth Quintile: Largest Market Capitalization Stocks

	1st per	25th per	50th per	75th per	99th per
A_{HS} (in %)	18.68	39.39	47.43	55.22	75.25
A (in %)	10.98	32.15	40.40	48.61	71.10
t-statistic	4.59	7.76	9.72	11.87	24.22

Table IV
AIM and Information Structure

This table reports the summary statistics for five cross-sectional regressions labeled as A, B, C, D and E. The dependent variable is the asymmetric information measure (AIM). The number of sample stocks is not the same for these five regressions. This is because not all the firms in our sample have a record for the number of shareholders (Shareholders) in Compustat, and for the number of analysts following the firm (Analysts) in I/B/E/S. Information concerning the indicator variable Option is obtained from the Options Clearing Corporation. The t-statistics in the parentheses are adjusted according to Newey and West (1987). The adjusted R^2 values in this table indicate the goodness-of-fit for these five regressions.

Regression	Intercept	Shareholders	Analysts	Option	Number of stocks	Adjusted R^2
A	21.9	-1.06			1,275	11.5%
	(-96.1)	(-13.7)				
B	28.9		-4.32		1,279	45.2%
	(74.2)		(-27.6)			
C	27.3	-0.55		-8.68	1,275	48.7%
	(79.0)	(-8.65)		(-22.0)		
D	29.5		-2.80	-5.24	1,279	56.2%
	(84.1)		(-16.1)	(-14.5)		
E	29.4	-0.28	-2.59	-5.08	1,207	57.1%
	(81.2)	(-4.80)	(-14.4)	(-14.0)		

Figure 1. Time Series of Cross-Sectional Daily Average Values.

This figure plots the cross-sectional average value of R^2_{asy} on a daily basis. This asymmetric information measure is the joint contribution of four variables in equation (16) in explaining intra-day price changes. These four variables are lagged trade sign, signed volume, signed duration and duration. On each trading day in our sample period of year 2003, we estimate R^2_{asy} for each sample stock. The daily average is taken cross-sectionally and plotted as time series.

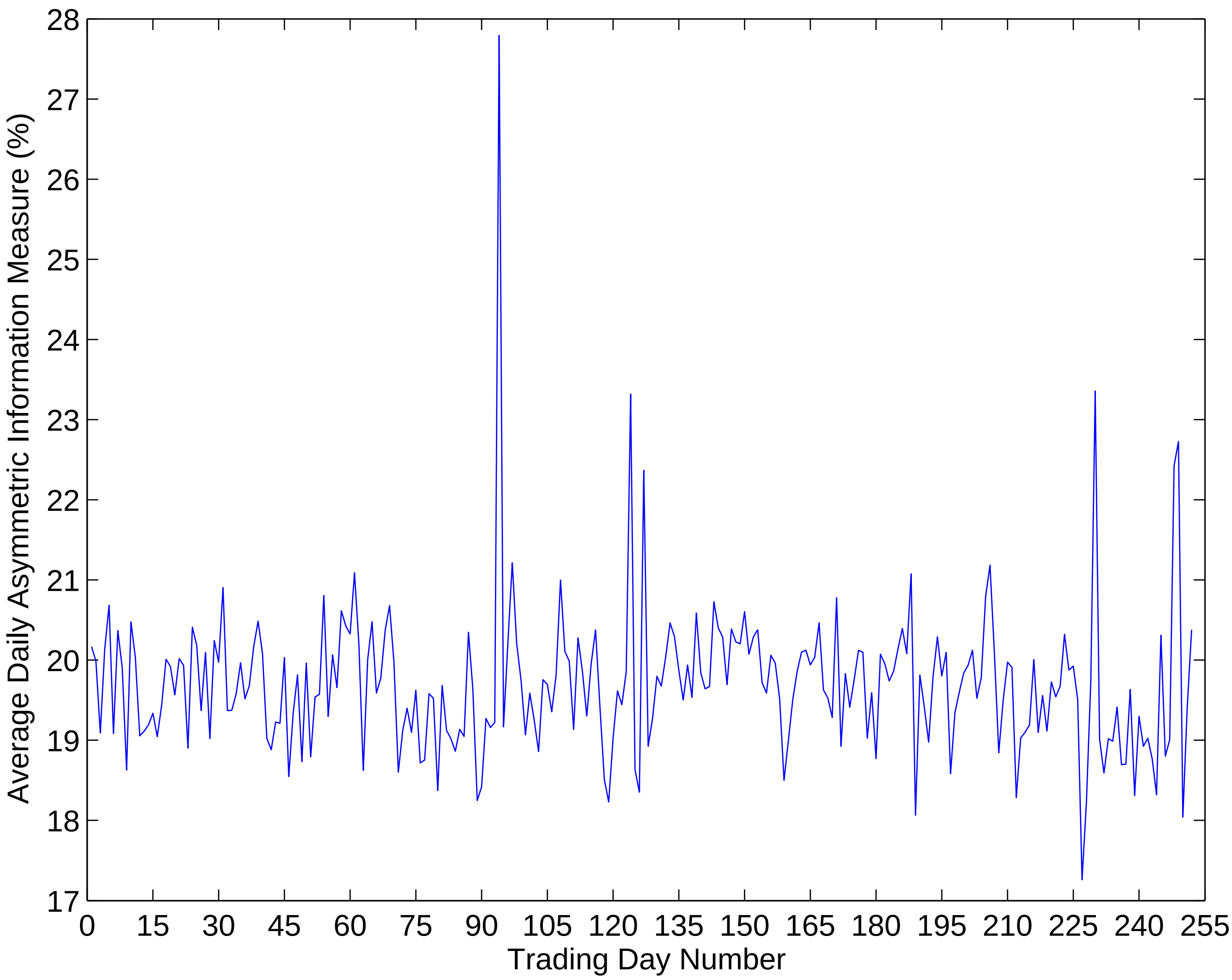
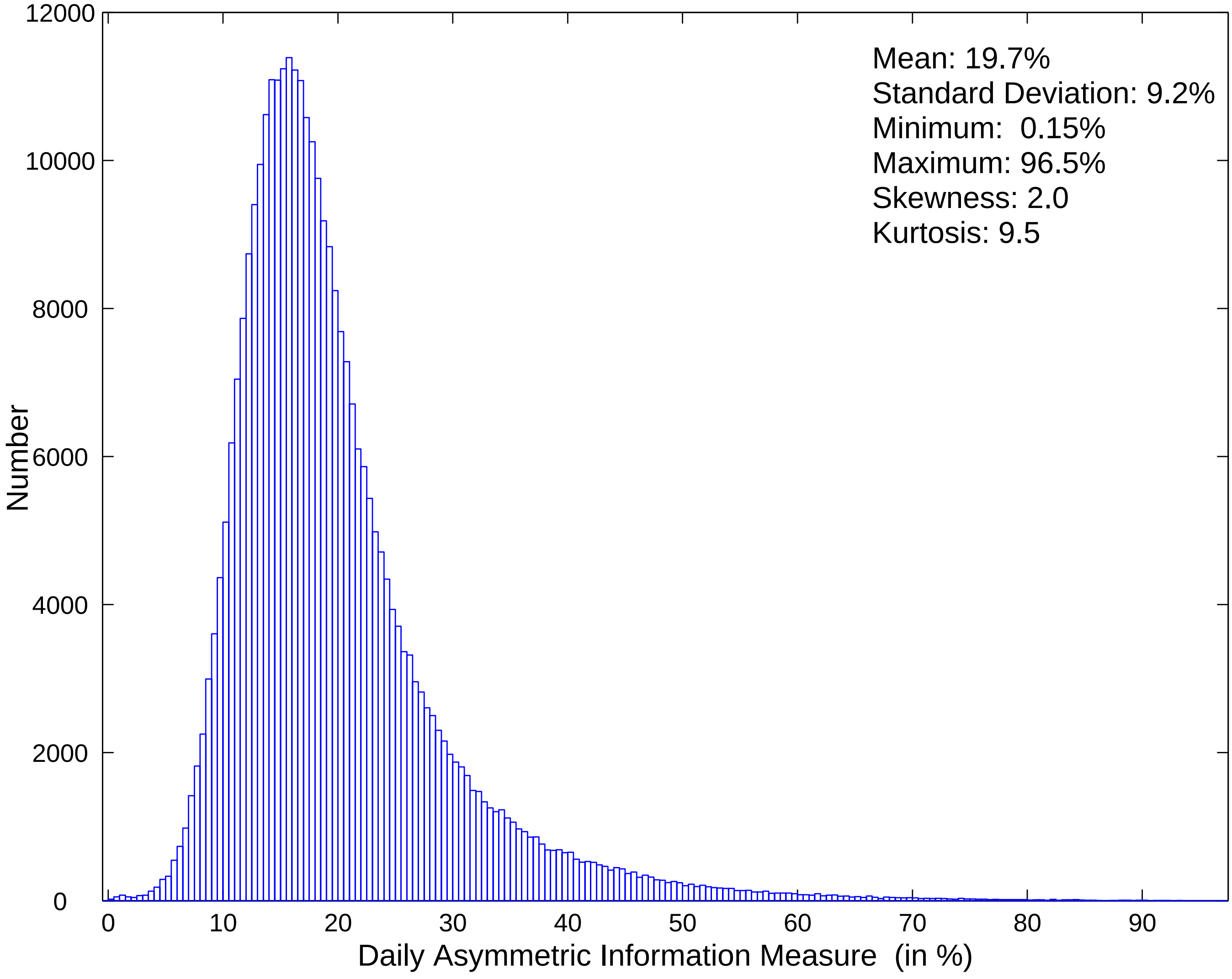


Figure 2. Distributions of Asymmetric Information Estimates.

This figure plots the histograms for the estimates of asymmetric information from high-frequency data. The asymmetric information measure is the joint contribution of four variables in equation (16) in explaining intra-day price changes. These four variables are lagged trade sign, signed volume, signed duration and duration. Our sample is 1,461 NYSE common stocks and our sample period is year 2003 (252 trading days). In total, we have 329,187 stock-days. Panel A shows the histogram for this number of daily estimates. Panel B displays the histogram of 1,461 values averaged over the sample period.

Panel A: Daily Estimates



Panel B: Daily Estimates Averaged Over the Sample Period