

Inferring Information Frequency and Quality

John Owens and Douglas G. Steigerwald

Abstract

We develop a microstructure model that, in contrast to previous models, allows one to estimate the frequency and quality of private information. In addition, the model produces stationary asset price and trading volume series. We find evidence that information arrives frequently within a day and that this information is of high quality. The frequent arrival of information, while in contrast to previous microstructure model estimates, accords with non-model based estimates and the related literature testing the mixture-of-distributions hypothesis. To determine if the estimates are correctly reflecting the arrival of latent information, we examine intervals around earnings announcements when private information flow is likely. We find evidence that flows are more likely prior to earnings announcements, which indicates that the estimates reflect the arrival of latent information.

1. Introduction

An attraction of market microstructure models is that they allow one to assess the empirical importance of private information in security markets. The asymmetric information model developed by Glosten and Milgrom (1985) to explain the presence of bid-ask spreads over a single trading period, was extended to multiple trading periods by Easley and O'Hara (1992). With a model of multiple trading periods, succeeding papers estimated the impact of privately informed traders on price determination (see Easley, Kiefer, O'Hara and Paperman (1996), Easley, Kiefer and O'Hara (1997), Kelly and Steigerwald (2000), Easley, O'Hara and Saar (2001) and Hanousek and Podpiera (2002)). While certain features of informed traders are estimable from the model, both the frequency of (private) information arrival and the accuracy of information must be assumed. As the above mentioned empirical papers assume that information arrives at most once per day, while estimates obtained directly from buy and sell order flows (Hasbrouck

(1999)) find that information arrives many times within a day, misspecification bias of the microstructure estimates is a very real possibility. We address these issues with a model that allows us to estimate both the frequency and the quality of private information.

The microstructure model of Easley and O'Hara is inherently nonstationary, as the potential arrival of information occurs at fixed (and known) points in time. Estimation of the model requires that the researcher specify the fixed intervals that correspond to the arrival of information. Typically researchers assume that information potentially arrives at the beginning of each trading day, thereby limiting the frequency of information arrival to once per day. Such an assumption is in contrast not only to the findings of Hasbrouck but also to the assumption that underpins research about the mixture-of-distributions hypothesis (MDH). Empirical assessment of the MDH, which describes the distribution of security prices arising from the presence of informed traders, is typically based on the assumption of many information arrivals within a day (e.g. Andersen (1996)). We address the issue by constructing a model (of the type in Easley and O'Hara) in which information arrives randomly throughout each trading day (in so doing, the model is stationary as well). As a result the length of time over which information persists is random, in accord with the different types of information that enter asset markets.

In addition, we allow for information of varying accuracy. In the model of Easley and O'Hara information is perfect by assumption. Yet it may be the case that not all information is created equally. We address this issue by explicitly modeling the belief a trader has in the accuracy of information. By allowing the quality of information to vary, we also address another issue. In Easley and O'Hara the group of informed traders all receive (perfect) information simultaneously. The potential for strategic behavior by the informed is eliminated through the random arrival of traders.¹ While eliminating strategic behavior, the mechanism imparts perfect correlation at the microstructure level, as the information received by one informed trader is the same as the information received by the next informed trader. By allowing the quality of news to vary, we are able to estimate the correlation of private information and so more accurately gauge the impact of information on stochastic volatility in asset prices.

In Section 2, we present the asymmetric information microstructure model for a security market. A period of asymmetric information ends with the arrival of public news. (Not all news about the asset is privately revealed, with posi-

¹Kyle (1985) considers the strategic behavior of a single informed trader.

tive probability public news will not have been previously revealed to informed traders.) Following each public news arrival is the possible arrival of (private) information. The parameter governing the public news arrival process determines the average length of time over which informed traders exploit their information and contributes to the frequency of information arrivals. We capture the varying accuracy of information through a Markov transition matrix that governs the probability that a given information signal to a trader will be publicly revealed (and hence, accurate). The model is flexible enough to allow information accuracy to depend on whether the information reflects positively or negatively on the asset price.

We detail how to estimate the model in Section 3. As the fundamental data are the latent individual trade decisions, we first describe how to construct an observable sequence of decisions. To do so one must specify how frequently traders arrive to the market. Easley, Kiefer and O'Hara assume that the arrival frequency is fixed over time (with a trader arriving every five minutes). Unfortunately, such a specification does not accommodate the fact that trading intensity varies in predictable ways over the course of a day. To allow for these periodic effects, and so distinguish episodes of trading that follow from information arrival, we vary the arrival rate of traders over the course of a day. We also study the bias that arises from misspecification of the arrival rate and establish that the bias vanishes as the assumed arrival frequency grows. Given the sequence of trade decisions, the likelihood function is formed from the probabilities of each trade as governed by the model. We note that maximum likelihood estimation of the model of Easley and O'Hara implicitly conditions on the assumed frequency of information arrival and then detail the construction of the likelihood function for the model with the estimated frequency of information arrival. With the additional parameters, we are careful to distinguish between the information set of the market specialist and that of the econometrician, because the specialist has the additional knowledge of the timing of public news arrivals.

An empirical investigation of the impact of informed traders on a security market is contained in Section 4. We focus on 2001 data for three firms of varying liquidity that trade on the New York Stock Exchange. We estimate the frequency and accuracy of private information and find evidence that private information arrives many times within a day, in accord with the findings of Hasbrouck and empirical analysis of the MDH. We are able to detail the impact of arrival frequency misspecification on the other parameters governing the behavior of informed traders. To check if the model is capturing information arrivals, rather

than correlated arrivals of traders for other reasons, we study the period surrounding earnings announcements. We find significantly higher information activity in the period immediately preceding informative announcements, which provides a natural measure of the information “leakage” prior to public statements.

2. Microstructure Model

Trade in a market for a single stock is coordinated through a market specialist. (The market is a dealership market in that the specialist does not act as a broker, thus all orders are market orders.) There are an arbitrarily large number of potential (risk-neutral) traders, from which traders are randomly selected to meet with the specialist. Because there are an uncountable number of traders, who have only countably many opportunities to be selected to trade with the specialist, almost surely a selected trader has only one opportunity to increase utility through trade in the market. As the specialist meets with only one trader at a time, we index traders by their order of arrival, i . (We imagine that the market began at some time in the arbitrarily distant past, so i is an element of the integers \mathcal{Z} .) Concordant with the arrival of traders is the generation of a signal S_i of the intrinsic stock value.

Trade occurs over a sequence of information periods. An information period captures the interval over which private information potentially exists, so the end of an information period is characterized by agreement over all participants on the value of the stock. The end of an information period thus corresponds to public revelation of the signal, which occurs with probability $\delta \in (0, 1)$ on any arrival. If we index information periods by m , then the value of the stock at the end of period m is

$$V_m = V_{m-1} + S_{i_m},$$

where i_m denotes the last arrival in period m (thus $\Delta V_m = S_{i_m}$). Because

$$E(S_{i_m} | S_{i_{m-1}}) = 0,$$

the expected price at the next public signal conditional on the current public signal always equals the price at the current public signal.

Private information is captured through the signals generated within an information period. In detail, the signal takes one of three values: $S_i \in \{-1, 0, 1\}$. (Setting the increment amount to $k \in \mathbb{R}$, rather than 1, would simply re-scale the market.) At the beginning of an information period the price-changing signals

($S_i = 1$ and $S_i = -1$) are equally likely. Only price-changing signals correspond to news privately entering the market, so we refer to $S_i \neq 0$ as (private) news. For all arrivals during information period m , $i_{m-1} < i \leq i_m$, the arriving trader observes the signal, and so is informed, with probability α . Private signals derive their economic relevance through their implications for the next public signal.

The signal that informed traders observe may differ from the publicly revealed value and so is potentially imperfect. The evolution of the signal over trader arrivals within an information period is governed by the transition probabilities

$$M = \begin{bmatrix} \theta_1 & \theta_2 & 1 - \theta_1 - \theta_2 \\ \frac{1-\theta_3}{2} & \theta_3 & \frac{1-\theta_3}{2} \\ 1 - \theta_1 - \theta_2 & \theta_2 & \theta_1 \end{bmatrix},$$

with $P(S_i = 1|S_{i-1} = 1) = \theta_1$, $P(S_i = 0|S_{i-1} = 1) = \theta_2$ and $P(S_i = -1|S_{i-1} = 1) = 1 - \theta_1 - \theta_2$. The parameters θ_1 and θ_2 measure state persistence, and so capture the belief that informed traders attach to private information. If $\theta_1 = \theta_3 = 1$, then the signal is perfect as in Easley and O'Hara (1992). To understand how the two persistence parameters are identified, consider the effect of altering each parameter. As θ_1 increases, the likelihood of news $P(S_i \neq 0)$ increases. In addition, the quality of news increases, as it becomes more likely that the privately revealed value will be the future publicly revealed value. As θ_3 increases, the likelihood of news decreases but the quality of news is unchanged. It is this asymmetry in the behavior that identifies the two persistence parameters.

While the transition matrix M describes the evolution of the signal within an information period, we must look across information periods to determine the statistical properties of the stationary series. Each information period begins with a signal that is drawn from $\bar{\pi}$, which is the stationary (unconditional) distribution of S_i . We assume that each element of $\bar{\pi} \in (0, 1)$ so that S_i is generated by a Markov process with transition probabilities

$$\begin{bmatrix} (1 - \delta) M & \delta M \\ (1 - \delta) \Pi & \delta \Pi \end{bmatrix},$$

where the (3×3) submatrices capture the transition if: neither signal (S_i, S_{i-1}) is publicly revealed (the submatrix $(1 - \delta) M$), only S_i is publicly revealed (δM), only S_{i-1} is publicly revealed ($(1 - \delta) \Pi$, so $\Pi' = \bar{\pi} \otimes [1, 1, 1]$ reflects the fact that S_i is governed by $\bar{\pi}$ regardless of the value of S_{i-1}), both (S_i, S_{i-1}) are publicly revealed ($\delta \Pi$).

To determine the likelihood of private information, in the Appendix we derive $\bar{\pi}$. Good news ($S_i = 1$) and bad news ($S_i = -1$) are (unconditionally) equally likely with total probability

$$\nu \equiv P(S_i \neq 0) = \frac{1 - \theta_3}{\theta_2 + (1 - \theta_3)}.$$

Although the model can accommodate very general dynamics, we concentrate on the realistic case in which the signal cannot immediately switch between the good and bad states. So, in what follows, we assume that $\theta_2 = 1 - \theta_1$, and we have

$$\nu = \frac{1 - \theta_3}{(1 - \theta_1) + (1 - \theta_3)}. \quad (2.1)$$

For this case, if $\theta_1 = \theta_3$, then private information (unconditionally) arrives half the time $\nu = \frac{1}{2}$.

In order to test whether private information is perfect (that the signal remains constant until publicly revealed) we must test the hypothesis $H_0: \theta_1 = \theta_3 = 0$. Unfortunately, under this restriction $\bar{\pi}$ becomes arbitrary as the transition probabilities no longer form an ergodic Markov chain. Nevertheless, we can nest the perfect information hypothesis by recasting the model to include ν as a parameter instead of θ_3 . Notice that 2.1 is equivalent to

$$\theta_3 = 1 - \frac{\nu}{1 - \nu} (1 - \theta_1).$$

for any $\theta_3 < 1$ and any $0 \leq \nu \leq 1/(2 - \theta_1)$. Moreover, if $\theta_1 = 1$, then $\theta_3 = 1$ and ν varies freely over $[0, 1]$. When we refer to the perfect information model, we refer to the model parameterized as $\theta_1 = \theta_3 = 1$, with ν arbitrary on the interval $[0, 1]$ and $\bar{\pi} = (\frac{\nu}{2}, 1 - \nu, \frac{\nu}{2})$.

Prior to the i th arrival, the specialist sets an ask (A_i) and a bid (B_i) for one share of the stock. Let D_i represent the random variable corresponding to the decision of the i th trader, taking on d_A , d_B or d_N if the decision is to trade at the ask, the bid, or not to trade, respectively. After each arrival, $i_{m-1} < i \leq i_m$, the specialist is aware of the entire history of trades, $\{D_j\}_{j=-\infty}^i$, the entire history of the public signals, $\{s_{i_k}\}_{k=-\infty}^{m-1}$, and the structure of the market. However, because of the Markovian setting and preference assumptions, only a small subset of this information is relevant to the specialist. All relevant information, other than the structure of the market, is summarized in the public information set

$$Z_i \equiv \{V_{m-1}, D_i, D_{i-1}, \dots, D_{i_{m-1}}\}.$$

The information set of an informed trader includes the private signal and so is a finer partition of the event space at the next public signal. Because the intra-information period trade history is used by the specialist only to predict S_i , an informed trader's information set is

$$\{V_{m-1}, S_i\}.$$

The specialist does not observe the private signal and so must form beliefs about the value of the signal. Bayes rule governs the method by which the specialist (and uninformed traders) learn through observing transactions. If trader i buys the stock, then²

$$P(S_i = 1|Z_{i-1}, D_i = d_A) = P(S_i = 1|Z_{i-1}) \frac{P(D_i = d_A|S_i = 1)}{P(D_i = d_A|Z_{i-1})}.$$

With updated beliefs, the Markov transition matrix M guides prediction of the signal at the next arrival

$$P(S_{i+1} = 1|Z_i) = \theta_1 P(S_i = 1|Z_i) + \frac{1 - \theta_3}{2} P(S_i = 0|Z_i) + (1 - \theta_1 - \theta_2) P(S_i = -1|Z_i).$$

For all future values, the specialist's beliefs are contained in the vector

$$\pi_{j,i} = \begin{bmatrix} P(S_{i+j} = 1|Z_i) \\ P(S_{i+j} = 0|Z_i) \\ P(S_{i+j} = -1|Z_i) \end{bmatrix},$$

where $\pi'_{j,i} = \pi'_{0,i} M^j$.

The central problem for all agents is to determine the value of the stock at the public revelation, V_m . In predicting V_m , there are two interrelated sources of uncertainty. First, the agent must predict when the public information will arrive. To model prediction of uncertain future public news, we define a subsequence of arrivals, $\{i_m\}_{m=1}^{\infty}$, corresponding to public news. The (random) number of arrivals until a public signal is $T(i) \equiv \#\{j : i < j \leq i_m\}$ for $i_{m-1} < i \leq i_m$, where $\#$ is the number of elements in the set. Because the arrival of public news occurs randomly with probability δ , the $T(i)$ are i.i.d. geometric random variables with common distribution equivalent to T where

$$P[T = t] = (1 - \delta)^t \delta \quad t = 0, 1, 2, \dots$$

²The specialist's recursion begins with the stationary probabilities.

Note that $T(i) = 0$ corresponds to public revelation of the signal (potentially) received by the i th trader. Second, given the expected time of public information arrival, the agent must predict the value of the stock when the value is made public. Following the arrival of trader i , the specialist's valuation $E(\Delta V_m | Z_i)$ is

$$\sum_{j=0}^{\infty} P[T(i) = j] [P(S_{i+j} = 1 | Z_i) - P(S_{i+j} = -1 | Z_i)].$$

Solution of the infinite series yields (details are in the Appendix)

$$E(S_i | Z_i) \cdot \frac{\delta}{1 - (1 - \delta) [\theta_1 - (1 - \theta_1 - \theta_2)]},$$

which equals the expectation of the current signal multiplied by a factor that captures the likelihood that the signal is publicly revealed. The factor equals one only if θ_1 equals one, in which case the current signal is perfect and so is revealed with certainty. We also see that the factor is an increasing function of δ and θ_1 . Increasing δ tends to shorten the information period, while increasing θ_1 makes the current signal more informative, both of which imply that the current signal is more likely to be publicly revealed.

An informed trader receives S_i , which supersedes the public information. As S_i does not provide information about the timing of public news, an informed trader's valuation differs from the specialist's only in the prediction of the revealed signal

$$\sum_{j=0}^{\infty} P[T(i) = j] [P(S_{i+j} = 1 | S_i) - P(S_{i+j} = -1 | S_i)].$$

Solution of the infinite series yields (details are in the Appendix)

$$S_i \cdot \frac{\delta}{1 - (1 - \delta) [\theta_1 - (1 - \theta_1 - \theta_2)]},$$

which equals the current signal multiplied by the factor that captures the likelihood of public revelation of the signal. If $\theta_1 = 1$, then the valuation of an informed trader is simply

$$E(\Delta V_m | S_i) = S_i.$$

To complete the specification of the market microstructure, we define equilibrium by a sequence of bid ask pairs that result in zero expected profits for the specialist; formally, at any given arrival an equilibrium obeys

$$E[V_m - A_i | Z_{i-1}, D_i = d_A] = E[B_i - V_m | Z_{i-1}, D_i = d_B] = 0,$$

where $i_{m-1} < i \leq i_m$ for all $m \in \mathcal{Z}$. We consider this an equilibrium condition obtaining from the potential free entry of additional market specialists should the bid and ask lead to positive expected profits.

The zero-profit equilibrium imposes constraints on the quotes. First, the quotes always satisfy

$$V_{m-1} - \frac{\delta}{1 - (1 - \delta) [\theta_1 - (1 - \theta_1 - \theta_2)]} < B_i \leq A_i < V_{m-1} + \frac{\delta}{1 - (1 - \delta) [\theta_1 - (1 - \theta_1 - \theta_2)]},$$

where the lower and upper bounds are the “reservation prices” for an informed trader with $S_i = -1$ and $S_i = 1$, respectively. For example, if the ask exceeded the upper bound, then informed traders would never trade at the ask and the specialist could ensure positive profit from exclusive trade with uninformed traders. Because the quotes are bounded by the reservation prices, the decision of the informed is summarized by the following simple rule: Buy if $S_i = 1$, sell if $S_i = -1$.

We do not directly model the preferences of the uninformed, as the uninformed are assumed to trade for liquidity reasons rather than speculation. Because a trader who receives the signal $S_i = 0$ does not trade, we must allow for uninformed traders to elect not to trade, to avoid immediate revelation of a private signal. Uninformed traders elect to trade with probability ε . Of the proportion of uninformed traders who trade, half buy at the ask and half sell at the bid.

To see the specific form of the quotes, note that the ask for the stock is determined from

$$\alpha P(S_i = 1|Z_{i-1}) \cdot [E(V_m|S_i = 1) - A_i] = \frac{1}{2} (1 - \alpha) \varepsilon \cdot [A_i - E(V_m|Z_{i-1})].$$

The left side is the expected loss the specialist incurs from trade with the informed at the ask, the right side is the expected gain the specialist receives from trade with the uninformed. The corresponding equilibrium ask is

$$A_i = \frac{\alpha P(S_i = 1|Z_{i-1}) E(V_m|S_i = 1) + \frac{1}{2} (1 - \alpha) \varepsilon E(V_m|Z_{i-1})}{\alpha P(S_i = 1|Z_{i-1}) + \frac{1}{2} (1 - \alpha) \varepsilon},$$

where $E(V_m|Z_{i-1})$ equals

$$P(S_i = -1|Z_{i-1}) E(V_m|S_i = -1) + P(S_i = 0|Z_{i-1}) E(V_m|S_i = 0) \\ + P(S_i = 1|Z_{i-1}) E(V_m|S_i = 1).$$

To summarize, trading evolves as follows. After a public signal, the next signal is selected according to the dynamics of the signal process. Assuming that this signal is not public, only the informed are aware of the signal value. A trader is randomly selected to trade with the specialist and the signal is potentially revealed to the trader. The trader observes the bid and ask, and decides whether to buy or sell. After the decision of the trader, the signal is publicly revealed with probability δ . Upon observing the decision of the trader, and the possible public revelation of the signal, the specialist must set the bid and ask that will be in effect at the next arrival. The signal is then updated again, and the previously described process continues until another public signal occurs.

3. Econometric Estimation

The microstructure model yields the likelihood of each trade decision D_i as a function of the parameters $\Phi = (\alpha, \delta, \varepsilon, \theta_1, \theta_3)$. Thus the observed trade decisions are used to form the likelihood (function) without need of further distributional assumptions. As the actual sequence of trade decisions is unobserved, two transformations of the data are needed to construct the sequence.

The first transformation concerns the information content of elapsed time without trade. Within the model, the frequency of trader arrivals determines the amount of time without trade that corresponds to a no-trade decision. To form a sequence of trade decisions, we must specify the frequency of trader arrivals. In doing so we must account for the fact that, as Harris (1986), Jain and Joh (1988) and McNish and Wood (1992) verify, trade volume exhibits significant cyclic patterns both within a day and across days of the week. As the predictable patterns in trade activity are likely due to the many factors affecting trade that are not captured by the model, we must allow the arrival rate of traders to vary over time. To do so, we construct an average number of trades for each hour of the week (allowing for both day-of-week and hour-of-day effects).³ For each hour of the week the number of arrivals is assumed to be a multiple, K , of the average number of trades. (The value of K must be large enough so that the number of arrivals always exceeds the actual number of trades, and small enough so that traders do not arrive more frequently than one per second.) Thus the length of time corresponding to a no-trade interval (the time between arrivals) varies over hours.

³We use hourly effects, rather than a quadratic function of hours, to account for the additional effect of the lunch hour on trade activity for the NYSE.

While our ability to allow for varying arrival rates is an improvement over a fixed arrival rate of traders, any specification of arrival rates likely introduces misspecification bias. To determine the bias we focus on the ratio of the assumed number of arrivals to the recorded number of trades, which directly affects the estimator of Φ . For a given interval of time, the ratio depends on the (assumed) frequency of arrivals and on the timing of recorded trades. Increasing the frequency of arrivals increases the ratio of arrivals to trades, as the number of recorded trades is unaffected by the assumed frequency of arrivals. As estimation is affected by this ratio, we account for this by reporting the invariant measures that are scaled by the arrival frequency.

The timing of recorded trades also affects the number of arrivals. If trades are recorded at times other than integer multiples of the arrival frequency, then the number of arrivals is increased (for example, because two trades are recorded more closely than the assumed arrival frequency). The following theorem details the effects. Let L be the length of the time interval and let $f \leq 1$ be the assumed arrival frequency in seconds ($f = .1$ indicates a trader arrives every 10 seconds). Let T be the number of trades in the interval and A be the number of constructed arrivals, so the bias is $B = T^{-1}(A - Lf)$.

Theorem 1: The bias induced by the assumed arrival frequency f is

$$0 \leq B \leq 1.$$

Further,

$$B \rightarrow 0 \text{ as } f \rightarrow 1.$$

Proof: See Appendix.

Theorem 1 is quite intuitive. As trades are recorded to the nearest second, the assumption that a trader arrives every second eliminates bias from the misalignment between recorded trades and the arrival frequency.

While an arrival frequency of once per second eliminates bias, the resulting sequence of trade decisions is quite long. As the entire decision sequence is needed for estimation, such a long sequence is not desirable. In practice, much of the bias can be eliminated at arrival rates that are less frequent than once per second. To explore this trade-off, we construct two measures of bias relative to K . The first is simply the direct analog of the $K^{-1}B$, which upon noting that $KT = Lf$ is simply given from the formula for B as $K^{-1}\left(\frac{A}{T} - K\right)$. The second measure is obtained by replacing $\frac{T}{A}$ with the estimated unconditional probability of a trade, $\alpha\theta + (1 - \alpha)\varepsilon$ evaluated at the ML estimates, which yields

$K^{-1} \left(\left(\hat{\alpha} \hat{\theta} + (1 - \hat{\alpha}) \hat{\varepsilon} \right)^{-1} - K \right)$. To determine the optimal trade-off between bias reduction and computation efficiency, we analyze the relative bias measures for different values of K in our empirical work.

The second transformation of the data arises because all trades are assumed cleared through the specialist, so each trade is classified as buyer initiated (a trade at the ask) or seller initiated (a trade at the bid). Because transaction records do not indicate who initiates trade, a classification rule must be employed. We use a rule proposed by Lee and Ready (1991) who use the mid-point of the bid-ask spread to classify trades. A trade above the mid-point is (classified as) buyer initiated, a trade below the mid-point is seller initiated and a trade at the mid-point depends on the preceding price movement. For example, if the price of the preceding trade is above the mid-point, then the mid-point trade represents a price decline and is seller initiated. As consecutive mid-point trades are classified identically (if there is no intervening price movement), such a rule can produce artificial runs of trades on one side of the market. In fact, Lee and Radhakrishna (2000) found that while the rule correctly classified 93 percent of transactions in their test sample, only 60 percent of consecutive mid-point transactions were correctly classified. To mitigate this type of misclassification we also consider both random assignment of consecutive mid-quote trades and removal of mid-quote trades.

For the model of Section 2, in which the frequency and accuracy of private information are unknown, the sufficient statistics for Φ are the entire sequence of trade decisions and public news arrivals. (If, as in Easley, Kiefer and O'Hara (1997), one assumes that private information is perfect and can arrive only at fixed points in time, then the likelihood is considerably simplified. For this case the sufficient statistics reduce to the number of trade decisions of each type within an information period.) From a sequence of n trader arrivals (over a span of m information periods), the likelihood from the model with unknown information frequency and accuracy is

$$L \left(\Phi | D_1 = d_1, \dots, D_n = d_n, \{i_j\}_{j=1}^{m-1} \right) = \prod_{i=1}^n P(D_i = d_i | Z_{i-1}; \Phi),$$

where $P(D_1 | Z_0; \Phi)$ is the stationary probability distribution for the first decision

in the information period. In detail, for $D_i = d_A$ we have

$$\begin{aligned} P(D_i = d_A | Z_{i-1}; \Phi) &= P(S_i = 1 | Z_{i-1}, D_i = d_A; \Phi) \left[\alpha + (1 - \alpha) \frac{\varepsilon}{2} \right] \\ &\quad + P(S_i = 0 | Z_{i-1}, D_i = d_A; \Phi) [(1 - \alpha)(1 - \varepsilon)] \\ &\quad + P(S_i = -1 | Z_{i-1}, D_i = d_A; \Phi) \left[(1 - \alpha) \frac{\varepsilon}{2} \right]. \end{aligned}$$

The conditional probabilities for S_i are obtained directly from the learning rules for the specialist described in Section 2. We have

$$P(D_i = d_i | Z_{i-1}; \Phi) = \lambda'_i \pi_{1,i-1},$$

where the trade frequencies are captured by

$$\lambda_i = \begin{bmatrix} (1 - \alpha) \frac{\varepsilon}{2} \\ (1 - \alpha)(1 - \varepsilon) \\ (1 - \alpha) \frac{\varepsilon}{2} \end{bmatrix} + \alpha \begin{bmatrix} 1(D_i = d_A) \\ 1(D_i = d_N) \\ 1(D_i = d_B) \end{bmatrix}.$$

Much of the motivation for carefully extending the sequential arrival microstructure model to a stationary setting is that it is typically not possible to identify episodes of information asymmetry. Instead, the best that can be hoped for is that a process governing the ongoing evolution of information asymmetries can be identified. As a result, the econometrician has the reduced information set \tilde{Z}_{i-1} , which consists of only trade decisions without knowledge of the subsequence of arrivals corresponding to public news. Burdened by the lack of knowledge about the public news state variable, the econometrician is concerned with a state space that is twice as large as the state space confronting the specialist. The state space $\{0, 1\} \times \{1, 0, -1\}$ underpins an ergodic bivariate Markov chain. The first element, \tilde{S}_i , determines whether the signal is private (1) or public (0), the second element is S_i . (The stationary distribution is $(\delta\bar{\pi}, (1 - \delta)\bar{\pi})$, with $\bar{\pi}$ the stationary distribution for S_i .) The likelihood becomes

$$L(\Phi | D_1 = d_1, \dots, D_n = d_n) = \prod_{i=1}^n P(D_i = d_i | \tilde{Z}_{i-1}; \Phi),$$

where $P(D_1 | \tilde{Z}_0; \Phi)$ is the stationary probability distribution for the first trade in the sample. We have

$$P(D_i = d_i | \tilde{Z}_{i-1}; \Phi) = P(\tilde{S}_{i-1} = 0) \lambda'_i \bar{\pi} + P(\tilde{S}_{i-1} = 1) \lambda'_i \tilde{\pi}_{1,i-1},$$

where

$$\tilde{\pi}_{1,i-1} = \begin{bmatrix} P(S_i = 1 | \tilde{Z}_{i-1}) \\ P(S_i = 0 | \tilde{Z}_{i-1}) \\ P(S_i = -1 | \tilde{Z}_{i-1}) \end{bmatrix}.$$

To understand the compact expression, if public news arrives with trader $i - 1$, then the conditional probabilities for S_i are reset to the stationary values $\bar{\pi}$. If public news does not arrive with trader $i - 1$, then the conditional probabilities for S_i follow from the learning rules of Section 2, with the restricted information set \tilde{Z}_i . Because the public news process is i.i.d., $P(\tilde{S}_{i-1} = 1) = (1 - \delta)$ and $P(\tilde{S}_{i-1} = 0) = \delta$, so

$$P(D_i = d_i | \tilde{Z}_{i-1}; \Phi) = \delta \lambda'_i \bar{\pi} + (1 - \delta) \lambda'_i \tilde{\pi}_{1,i-1}.$$

The score function is

$$\frac{\partial}{\partial \Phi} \ln L = \sum_{t=1}^T \frac{\frac{\partial}{\partial \Phi} \{\delta \lambda'_i \bar{\pi} + (1 - \delta) \lambda'_i \tilde{\pi}_{1,i-1}\}}{L_t}$$

where

$$\begin{aligned} \frac{\partial}{\partial \Phi} \{\delta \lambda'_i \bar{\pi} + (1 - \delta) \lambda'_i \tilde{\pi}_{1,i-1}\} &= \frac{\partial \delta \lambda_i}{\partial \Phi} \bar{\pi} + \frac{\partial (1 - \delta) \lambda_i}{\partial \Phi} \tilde{\pi}_{1,i-1} + \\ &\quad \left(\frac{\partial}{\partial \Phi} \bar{\pi} \right) \delta \lambda_i + \left(\frac{\partial}{\partial \Phi} \tilde{\pi}_{1,i-1} \right) (1 - \delta) \lambda_i. \end{aligned}$$

While the terms $\frac{\partial (1 - \delta) \lambda_i}{\partial \Phi}$ and $\frac{\partial \delta \lambda_i}{\partial \Phi}$ depend only on the parameter values and trade decisions, the term $\frac{\partial}{\partial \Phi} \tilde{\pi}_{1,i-1}$ must be calculated recursively.

4. Empirical Implementation

We examine three interesting questions. First, how long is an information period? Many empirical specifications of microstructure models assume that an information period is one trading day. This is in stark contrast to much of the MDH literature that assumes significantly shorter information periods. Second, how much trade is information based? Previous researchers find that informed

traders account for less than a third of the trades in their samples of NYSE stocks. Third, how precise is private information? Signals are often assumed noiseless in the microstructure literature, raising the question of how well this assumption conforms to the data when signal quality is estimated instead of assumed.

To answer these questions, we analyze three NYSE stocks chosen according to their liquidity characteristics. We select International Business Machines (ticker symbol IBM) to represent a highly liquid stock. We select Ashland (ticker symbol ASH) to represent a moderately liquid stock.⁴ Finally, we select the Commerce Insurance Group (ticker symbol CGI), a property and casualty insurer, to represent a relatively illiquid stock.

For each stock, we extract data from the first 30 trading days of 2001 (January 2, 2001 to February 13, 2001) from the NYSE Trades and Quotes (TAQ) dataset. The TAQ dataset contains a record of every trade and quote posted on the NYSE, the American Stock Exchange and the NASDAQ National Market System for all NYSE listed securities. We filter the trade data to remove trades that were recorded out of sequence, canceled, executed with special conditions, or recorded with some other anomaly. We use quotes only from the NYSE (Blume and Goldstein (1997) find that the NYSE quote determines or matches the national best quote about 95 percent of the time). We also filter the quote data to remove recording anomalies.

Because of certain institutional details, occasionally large trades are broken up into a sequence of smaller trades, all at the same price (see Hasbrouck (1988)). In order to avoid misidentifying these sequences of same sided trades as bursts of informed trades, we aggregate all trades recorded within five seconds of each other without an intervening price change or quote revision.

The data are further filtered to remove time stamps outside of the official trading hours of the NYSE (9:30 AM to 4:00 PM). Finally, the first half-hour of each trading day is removed in order to avoid modeling the market opening of the NYSE, which is characterized by heavy activity following the morning call auction. As Harris (1986), Engle and Russell (1998), and many other authors have noted, the first half-hour of trade exhibits substantially different properties than the rest of the day.

We investigate the potential bias from sequences of mid-quote trades that are assigned the same initiator type. Fortunately, relatively few of the trades in our dataset are part of mid-quote sequences. For Ashland, the Commerce Insurance

⁴Ashland Oil Incorporated, also studied by Easley, Kiefer and O'Hara, changed its name to Ashland Incorporated in 1995.

Table 4.1: Regression of the hourly number of trades against hour-of-week regressors (SE's in parentheses).

| Coefficient | ASH | CGI | IBM |
|--------------------|----------------------|--------------------|------------------------|
| INTERCEPT | 52.406** (3.343) | 9.614** (1.358) | 755.106** (44.886) |
| TUE | -0.229 (3.233) | 0.419 (1.312) | -15.395 (43.404) |
| WED | -0.022 (3.343) | 0.161 (1.357) | 60.367 (44.886) |
| THR | 1.978 (3.343) | -0.439 (1.366) | 159.894** (44.886) |
| FRI | 2.617 (3.343) | -0.283 (1.357) | 78.672 (44.886) |
| 11AM | -14.200** (3.492) | -2.867* (1.417) | -113.900* (46.882) |
| 12PM | -21.200** (3.492) | -2.433 (1.417) | -295.767** (46.882) |
| 1PM | -17.367** (3.492) | -1.787 (1.430) | -258.467** (46.882) |
| 2PM | -13.433** (3.492) | -0.933 (1.417) | -145.233* (46.882) |
| 3PM | 5.567 (3.492) | 6.533** (1.417) | 21.533 (46.882) |
| R^2 | 0.351 | 0.264 | 0.367 |

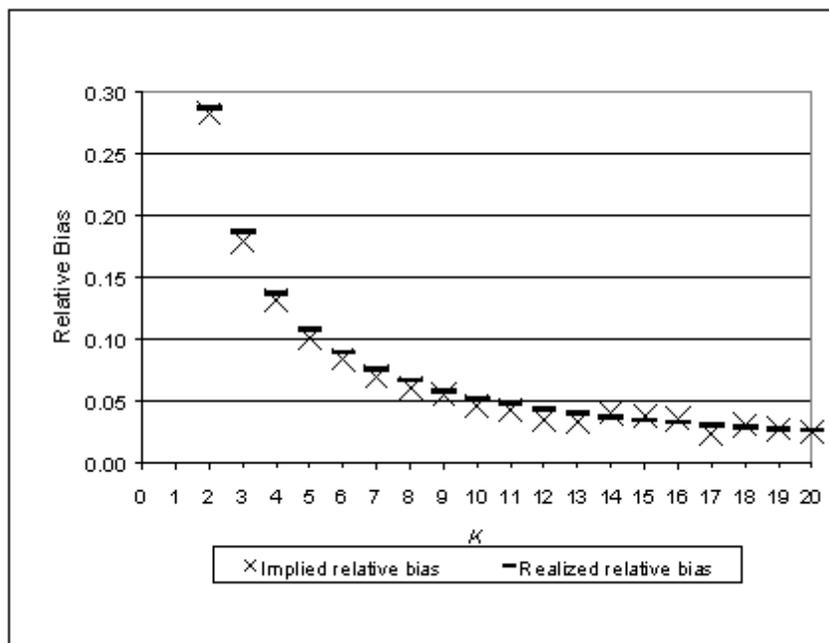
Group, and International Business Machines, 0.32, 0.26 and 0.88 percent were consecutive mid-quote trades without a price change, respectively. Neither randomly assigning these few mid-quote trades nor removing them from the sample had a significant effect on estimated parameters. Consequently, we report estimates based on the unmodified Lee and Ready approach.

After filtering the data and assigning trader initiation, we remove periodic features from the data. In detail, we first regress the number of trades for each hour on day-of-week and hour-of-day indicators. The parameter estimates are given in Table 4.1, where starred items are significant at the 5 percent level and double starred items are significant at the 1 percent level. (Hour-of-day indicators are labeled according to the beginning of the hour they correspond to, so 11AM takes on 1 over the interval 11AM to 12PM.) We determine the arrival frequency for each hour of the week by dividing the predicted number of arrivals in the hour (the predicted number of trades multiplied by K) by 3600 seconds. The arrival frequency, in turn, determines the length of time between arrivals, and whenever this length of time elapses without a trade, we record a no-trade.

Once the periodic features of the data are removed, we record the effective number of arrivals by adding the number of trades to the number of constructed no-trades. The realized relative bias is then given by $K^{-1} \left(\frac{A}{T} - K \right)$. With the ML estimates constructed from the sequence of trader decisions, the implied relative bias is $K^{-1} \left(\left(\hat{\alpha} \hat{\theta} + (1 - \hat{\alpha}) \hat{\varepsilon} \right)^{-1} - K \right)$.

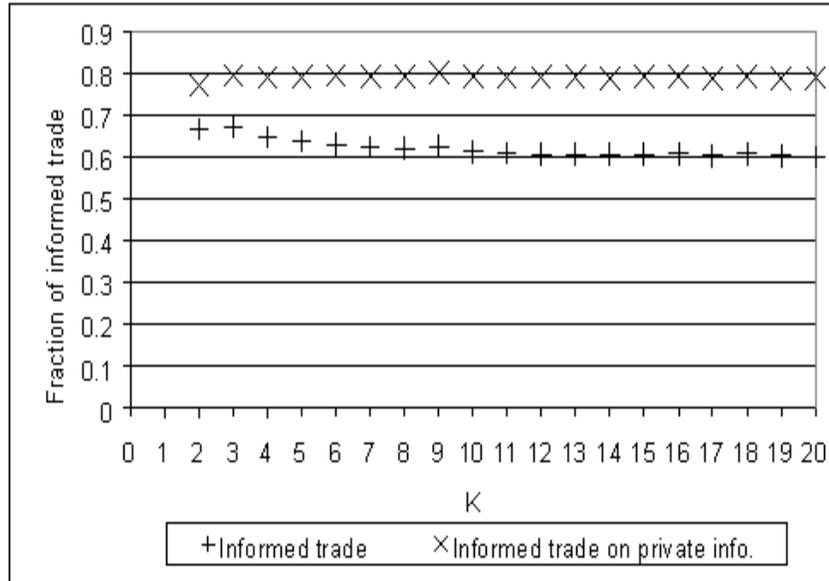
In Figure 1, we plot the realized and implied relative bias as K varies from 2 to 20 for Ashland. It is clear that the realized and implied relative biases are essentially identical, demonstrating that parameters are biased by incorrectly classifying no-trade decisions. As the figure indicates, increasing K from 2 to 10 decreases the relative bias from 25 percent to 5 percent. Further increases in K have little effect on the bias. Consequently, we choose $K = 10$ for the results reported on Ashland and the Commerce Insurance Group. For International Business Machines, we choose $K = 2$ to avoid having effective arrivals too close together (on Thursdays between 3PM and the close, trades occur, on average, less than 4 seconds apart).

Figure 4.1: Realized and implied relative biases as functions of K .



Because we are constrained in the choice of K for International Business Machines, reported results tend to over estimate the time between events. For example, estimated time between trades and the time between information arrivals are biased upward. However, many interesting estimates are largely invariant to the choice of K . Figure 2 shows that, for Ashland, the estimated fraction of informed trade as measured by $\hat{\alpha} \hat{\nu} / (\hat{\alpha} \hat{\nu} + (1 - \hat{\alpha}) \hat{\varepsilon})$ and the expected fraction of informed trade given the presence of private information, as measured by $\hat{\alpha} / (\hat{\alpha} + (1 - \hat{\alpha}) \hat{\varepsilon})$, where the hats indicate ML estimates, change little as K varies. This is an intuitive result; the biases of the estimated probability of informed trade and of uninformed trade tend to offset each other in measurements that include their ratios.

Figure 4.2: Fraction of informed trade and fraction of informed trade on private information.



Having settled the implementation issues, we next turn to estimating the model and testing information quality. Table 4.2 lists the parameter estimates for Ashland.⁵ The first item that stands out is that information quality is estimated

⁵As described in Section 2, for the unrestricted model ν is determined by $\nu = \frac{1-\theta_3}{(1-\theta_1)+(1-\theta_3)}$. The standard errors for ν reported in Table 4.2 in are derived by the delta method.

Table 4.2: ML estimates of restricted and unrestricted model.

| Coefficient | ASH | |
|----------------|----------------------|----------------------|
| | Restricted | Unrestricted |
| α | .143097 (.005008) | .143133 (.005370) |
| ν | .411310 (.021685) | .412034 (.017239) |
| ε | .042826 (.002314) | .042823 (.002453) |
| δ | .146922 (.008310) | .146738 (.018375) |
| θ_1 | 1 (N/A) | .999601 (.016371) |
| θ_3 | 1 (N/A) | .999721 (.011461) |
| Log Likelihood | -30,343.47 | -30,342.59 |
| χ^2 | | 0.88 |

to be very high. The chi-squared statistic for the likelihood ratio test of $H_0 : \theta_1 = \theta_3 = 1$ has 1 degree-of-freedom and is not significant at the 10 percent level. Moreover, the shared parameter estimates are almost identical between the two models (S.E.'s in parentheses). From the pattern that emerges, information periods correspond to short bursts of 7 ($\hat{\delta}^{-1} = 6.8$) arrivals. Roughly 40 percent of the bursts are associated with private information ($\hat{\nu} = .41$) and within an "information burst" 14 percent ($\hat{\alpha} = .14$) of the traders are informed. While the remaining 86 percent of arrivals are associated with uninformed traders, such trade is infrequent as uninformed arrivals result in trade less than 5 percent of the time ($\hat{\varepsilon} = .04$). As a result price movements during information bursts are heavily influenced by informed traders, with roughly 80 percent of trade attributed to the informed ($\frac{.41 \cdot .14}{.14 + .86 \cdot .04} = .80$). Even though information bursts occur less than half the time, the remaining periods see very little trading (as the uninformed trade infrequently). Overall, the informed account for about 60 percent of trading ($\frac{.41 \cdot .14}{.41 \cdot .14 + .86 + .04} = .60$). Finally, translating the information bursts into clock time, we find that during active times in the market (such as the first hour, 10-11 a.m., on Monday) a trader arrives roughly every 7 seconds, so an information burst lasts only 50 seconds. As a result, an information burst occurs (that is, private information arrives) roughly every 2 minutes ($\frac{49 \text{ sec}}{.41}$).

Many of the results for Ashland carry over to the case of a far less liquid stock. Estimated information quality for the Commerce Insurance Group is also

Table 4.3: ML estimates of restricted and unrestricted model for the Commerce Insurance Group.

| Coefficient | CGI | |
|----------------|----------------------|----------------------|
| | Restricted | Unrestricted |
| α | .312000 (.029530) | .312002 (.029559) |
| ν | .173258 (.027062) | .173158 (.019422) |
| ε | .055401 (.008384) | .055402 (.008387) |
| δ | .361355 (.033811) | .361358 (.034561) |
| θ_1 | 1 (N/A) | .999805 (.017933) |
| θ_3 | 1 (N/A) | .999959 (.003754) |
| Log Likelihood | -6506.96 | -6505.30 |
| χ^2 | | 1.66 |

very high (see Table 4.4). Information bursts now correspond to only 3 arrivals and are far less prevalent (occurring only 17 percent of the time). Given the less frequent arrival of information, it is perhaps not surprising that more than 30 percent of the traders are informed. As the uninformed again trade only 5 percent of the time the frequency of informed trade within an information burst is nearly 90 percent. Because information bursts are relatively rarer, the effects of the larger fraction of informed trader are partially offset and the overall fraction of informed is again nearly 60 percent. During an active market hour a trader arrives every 40 seconds, so an information burst last 2 minutes and private information arrives every 10 minutes. Thus CGI is characterized by relatively rare, but quite potent, information bursts.

In contrast, the liquid stock IBM has bursts that arrive quite frequently, 75 percent of the time, and last for 14 arrivals. Within a burst, roughly 20 percent of the traders are informed. Yet, because the uninformed trade with much greater frequency (30 percent), the overall impact of informed trade is substantially lower: 50 percent within an information burst and 35 percent overall. Finally, during an active market hour a trader arrives every 2 seconds, so information bursts last just under 30 seconds and occur (on average) every 40 seconds.

Our results are in marked contrast to previous empirical microstructure work in two important respects. We estimate information as arriving, not once a day, but

Table 4.4: ML estimates of restricted and unrestricted model .

| Coefficient | IBM | |
|----------------|----------------------|----------------------|
| | Restricted | Unrestricted |
| α | .233794 (.001777) | .259330 (.002337) |
| ν | .759209 (.006302) | .545455 (.004844) |
| ε | .294319 (.001787) | .338974 (.002311) |
| δ | .069861 (.001607) | .103751 (.002891) |
| θ_1 | 1 (N/A) | .999999 (.00001) |
| θ_3 | 1 (N/A) | .999999 (.00001) |
| Log Likelihood | -277,946.38 | -277,547.97 |
| χ^2 | | 398.41 |

many times an hour, and we estimate that well over half of trades are information based. In contrast, Easley, Kiefer and O'Hara (1997) assume that information arrives at most once per day and, consequently, they estimate that less than a third of trades are information based. If our estimates are correct, then the assumption of daily information arrival represents a substantial misspecification.

It is not hard to show that if information periods were misspecified as longer than those produced by the data generating process, then this misspecification would cause information to appear less potent than it actually is; for example, the probability of informed trade and the fraction of informed trade would both be biased downward. This is because multiple information periods in a day tend to cancel each other out.⁶ However, if information periods are misspecified as shorter than those produced by the data generating process, then the parameter estimates would not be substantially affected.

To examine the effects of misspecifying the information arrival frequency, we estimate the model with fixed information frequencies (in accord with Easley, Kiefer and O'Hara) for the Ashland data. In Table 4.5 we list the maximum

⁶If the sign of private information were correlated over the day, then the cancellation of good and bad news would be less pronounced. However, such correlations are not possible if prices follow a martingale and information shocks are symmetrically distributed. Easley, Kiefer and O'Hara include a parameter that determines the probability of bad news, thus allowing for an asymmetric distribution of the shocks. They estimate the probability of bad news as 0.502. We implicitly set the parameter to 0.500 by assuming the good and bad news are equally likely.

Table 4.5: Estimated parameters based on Easley, Kiefer and O’Hara (1997) for Ashland.

| Estimated parameter | Assumed arrival frequency | | | | |
|---------------------|---------------------------|-------------|---------------|--------------|--------------|
| | Day | Hour | 15-min | 5-min | 1-min |
| α | .047 | .034 | .050 | .074 | .112 |
| ε | .199 | .080 | .075 | .068 | .060 |
| ν | .533 | .473 | .435 | .445 | .498 |
| γ | .198 | .306 | .412 | .540 | .668 |

The column labeled “Hour” is based on the number of buys, sells, and no trades for each of the six-hour long periods of each day (the period between 9:30AM and 10:00AM is discarded). Other columns are similarly defined. The parameter $\gamma = \alpha / ((1 - \alpha)\varepsilon + \alpha)$ is the fraction of trades made by informed traders when private information is present.

likelihood estimates for five different assumed arrival frequency for the first 30 trading days in 2001. Notice that as the assumed arrival frequency increases from once a day to once a minute the estimated probability of informed trade more than doubles while the estimated probability of uninformed trade more than halves. The estimated fraction of informed trade given the presence of private information, γ , more than triples. The strong increase in the apparent potency of information as the assumed information period length declines supports the hypothesis that information arrives more frequently than once per day.

Generally, the parameter estimates based on the assumption of information arriving every minute conform to the estimates of the perfect information version of the model that allows for random information periods. However, the estimated probability of informed trade is lower for the fixed arrival model, while the estimated probability of private information is higher. To understand this discrepancy, it is important to keep in mind that, for the 1-minute version of the fixed arrival model, information is assumed to arrive every minute on the minute. If a 1-minute long information period randomly arrives on the half-minute, then the increase in informed trade associated with the period is spread over two adjacent minutes, increasing the apparent frequency of private information while decreasing its apparent effect on informed trade.

Across a variety of specifications, the picture that emerges of the market is one of constant information arrival and quick reactions to it. It is reasonable to

wonder how accurate this view of the market is. Could information really arrive as fast as every minute? Or are the bursts of one-sided trade, which are ultimately responsible for our estimation of how often information arrives, an artifact of the institutional arrangements where the securities are traded. If the model is truly responding to information instead of institutional features, then we would expect to estimate more information arriving and more informed trade occurring during times of probable information asymmetry. To help sort out the issue, we examine the period surrounding earnings announcements.

Specifically, we estimate the model for each firm using a window of the 5 trading days before the announcement date. We also calculate the announcement period return based on the closing price the day before the window begins to the closing price the day after the announcement date.

Ashland announced earnings on Wednesday January 24, 2001 and had an announcement period return of 0.017; the Commerce Insurance Group announced on Wednesday January 31, 2001 with an announcement return of .079; and International Business Machines announced on Wednesday January 17, 2001 with an announcement return of .161.

For Ashland, consistent with the relatively small earnings announcement return, we find little evidence that the earnings announcement affects trade before it is publicly released (see Table 4.6). There is a (barely) statistically significant (at the 5 percent level) decline in the probability of private information on an arrival from .058 during the entire 30 day sample to .043 during the announcement period, and there is a statistically insignificant decrease in the estimated fraction of informed trade from 61 percent to 56 percent.

For the Commerce Insurance Group, again consistent with the relatively small earnings announcement return, we find mixed evidence for informational leakage. Although there is a statistically significant (at the 5 percent level) drop in the probability of private information on an arrival from .063 during the entire 30 day sample to .018 during the announcement period, there is a large, though statistically insignificant, rise in the estimated fraction of informed trade from 59 percent to 67 percent.

For the most precisely estimated stock, International Business Machines, we find evidence for a substantial increase in private information arrival during the announcement period concordant with the large announcement period return. The probability of private information on an arrival nearly doubles during the announcement period, increasing from .053, for the estimates based on the entire 30 day sample, to .093. This increase is significant at the 1 percent level. Moreover,

Table 4.6: Estimated parameters for three NYSE stocks on the day of an earnings announcement.

| Symbol | α | ν | ε | δ |
|--------|------------------|-------------------|------------------|-------------------|
| ASH | .1734 (.0185) | .2778 (.04661) | .0454 (.0059) | .1550 (.02319) |
| CGI | .1165 (.0166) | .3683 (.0806) | .0239 (.0057) | .0497 (.0145) |
| IBM | .2323 (.0045) | .8658 (.0141) | .2150 (.0045) | .1074 (.0059) |

the fraction of informed trade increases from 44 to 55 percent (again, the increase is significant at the 1 percent level). These increases in the information arrival rate and the fraction of informed trade suggest that informative announcements affect trade before the announcements are made public.

5. Conclusion

In this paper we develop and estimate a microstructure model that allows for estimation of both the frequency and quality of private information. The frequency of private information is captured through the potential revelation of the information at each arrival. The length of time prior to revelation follows a geometric distribution and the parameter of this distribution characterizes the frequency of information. The quality of private information is captured with a Markov transition matrix; information that is likely to be publicly revealed is of high quality.

Construction of the trade decision sequence depends on two classification algorithms. We show the potential bias that results from each of the algorithms. For the no-trade classification algorithm we prove that the bias shrinks as the arrival frequency approaches the time scale on which trade is resolved. The ability to estimate the frequency and quality of information is not without cost. The sufficient statistic for the microstructure parameters is the entire sequence of trade decisions, in contrast to the case in which the frequency and quality are assumed known, for which the sufficient statistics are simply the number of trades of each type.

Our results provide evidence that asset markets are characterized by frequent short bursts of information. As the evidence is based on imbalances of order flow, and signed order flows can arise from transaction processing, these results provide an upper bound on the impact of news. In an effort to separate the sources

of order flow imbalances, we consider the period immediately before an earnings announcement, when information is a likely source. For the most newsworthy announcement (the announcement with the largest return) we find more frequent bursts of activity in the period preceding the announcement. These results pose a challenge to previous microstructure studies, in which information is assumed to arrive at most once per day. Our finding of frequent information arrival allows one to rigorously underpin tests of the mixture-of-distributions hypothesis, in which information is assumed to arrive frequently, with a microstructure model.

An interesting avenue for future research is to associate the bursts with price movements, wherein the magnitude of the price movement indicates the magnitude of the news. As the likelihood function does not deliver a sequence of trade decisions with the bursts marked, association of the bursts with prices requires an additional method of the type found in hidden Markov models.

References

- [1] Andersen, T., 1996, "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility" *Journal of Finance* 51, 169-204.
- [2] Blume, M. and M. Goldstein, 1997, "Quotes, Order Flow, and Price Discovery" *Journal of Finance* 52, 221-244.
- [3] Easley, D., N. Kiefer, and M. O'Hara, 1997, "One Day in the Life of a Very Common Stock" *Review of Financial Studies* 10, 805-835.
- [4] Easley, D., N. Kiefer, M. O'Hara, and J. Paperman, 1996, "Liquidity, Information, and Infrequently Traded Stocks" *Journal of Finance* 51, 1405-1436.
- [5] Easley, D. and M. O'Hara, 1992, "Time and the Process of Security Price Adjustment" *Journal of Finance* 47, 577-605.
- [6] Easley, D., M. O'Hara, and G. Saar, 2001, "How Stock Splits Affect Trading: A Microstructure Approach" *Journal of Financial and Quantitative Analysis* 36, 25-51.
- [7] Engle, R. and J. Russell, 1998, "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data" *Econometrica* 66, 1127-1162.
- [8] Glosten, L. and P. Milgrom, 1985, "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders" *Journal of Financial Economics* 14, 71-100.
- [9] Hanousek, J. and R. Podpiera, 2002, "Information-Driven Trading at the Prague Stock Exchange: Evidence from Intra-day Data" *Economics of Transition* 10, 747-759.
- [10] Harris, L., 1986, "A Transaction Data Study of Weekly and Intradaily Patterns in Stock Returns" *Journal of Financial Economics* 16, 99-117.
- [11] Hasbrouck, J., 1988, "Trades, Quotes, Inventory, and Information" *Journal of Financial Economics* 22, 229-252.
- [12] ——— 1999, "Trading Fast and Slow: Security Market Events in Real Time" New York University, Salomon Center Working Paper: S/99/17.

- [13] Jain, P. and G. Joh, 1988, “The Dependence between Hourly Prices and Trading Volume” *Journal of Financial and Quantitative Analysis* 23, 269-283.
- [14] Kelly, D. and D. Steigerwald, 2000, “Private Information and High-Frequency Stochastic Volatility” web manuscript, University of California, Santa Barbara.
- [15] Kyle, A., 1985, “Continuous Auctions and Insider Trading” *Econometrica* 53, 1315-1335.
- [16] Lee, C. and R. Radhakrishna, 2000, “Inferring Investor Behavior: Evidence from the TORQ Data” *Journal of Financial Markets* 3, 83-112.
- [17] Lee, C. and M. Ready, 1991, “Inferring Trade Direction from Intraday Data” *Journal of Finance* 46, 733-746.
- [18] McInish, T. and R. Wood, 1992, “An Analysis of Intraday Patterns in Bid/Ask Spreads for NYSE Stocks” *Journal of Finance* 47, 753-764.

6. Appendix

Calculation of the Stationary Distribution

The stationary probabilities for the private signal are the values of $\bar{\pi}$ such that $M'\bar{\pi} = \bar{\pi}$. If θ_1 and θ_3 do not equal 1, then the stationary probability vector is the eigenvector for M' that corresponds to an eigenvalue of 1. In detail, we first note that $|M' - I_3| = 0$, where I_3 is the 3×3 identity matrix, so that 1 is an eigenvalue of M' . The corresponding eigenvector, $\bar{\pi}$, satisfies

$$(M' - I_3) \bar{\pi} = 0.$$

From the first and third equations in the system, it follows that $\bar{\pi}_1 = \bar{\pi}_3 = c$. From the first equation

$$\bar{\pi}_2 = \frac{2\theta_2}{1 - \theta_3} c.$$

Because the three stationary probabilities sum to 1,

$$2c + \frac{2\theta_2}{1 - \theta_3} c = 1 \text{ so } c = \frac{1 - \theta_3}{2[\theta_2 + (1 - \theta_3)]}.$$

Hence $\bar{\pi}_2 = \frac{\theta_2}{\theta_2 + (1 - \theta_3)}$. Because $\bar{\pi}$ derived in the preceding displayed equation is a stationary distribution if $\theta_1 = \theta_3 = 1$ (although the stationary distribution is not unique for this case), we assume the derived $\bar{\pi}$ forms the initial distribution following public news for all values of (θ_1, θ_3) . The formula displayed in the text is arrived at by noting $\theta_2 = 1 - \theta_1$.

Stock Valuation

The specialist's valuation $E(\Delta V_m | Z_i)$ is

$$\sum_{j=0}^{\infty} P[T(i) = j] \pi'_{j,i} s$$

with $s = [1, 0, -1]'$. As noted in the text, $P[T(i) = j] = \delta(1 - \delta)^j$ and $\pi'_{j,i} = \pi'_{0,i} M^j$. Thus

$$\begin{aligned} E(\Delta V_m | Z_i) &= \sum_{j=0}^{\infty} \delta(1 - \delta)^j \pi'_{0,i} M^j s \\ &= \delta \pi'_{0,i} \sum_{j=0}^{\infty} [(1 - \delta) M]^j \cdot s \\ &= \delta \pi'_{0,i} (I - (1 - \delta) M)^{-1} s. \end{aligned}$$

Now,

$$(I - (1 - \delta) M) = \begin{pmatrix} a & b & \delta - a - b \\ c & \delta - 2c & c \\ \delta - a - b & b & a \end{pmatrix}$$

where $a = 1 - (1 - \delta) \theta_1$, $b = -(1 - \delta) \theta_2$ and $c = -\frac{1}{2}(1 - \delta)(1 - \theta_3)$. From matrix algebra $(I - (1 - \delta) M)^{-1} s = \frac{(1, 0, -1)'}{(2a - \delta + b)}$, so

$$\begin{aligned} \delta \pi'_{0,i} (I - (1 - \delta) M)^{-1} s &= \delta \cdot \frac{P(S_i = 1|Z_i) - P(S_i = -1|Z_i)}{1 - (1 - \delta) [\theta_1 - (1 - \theta_1 - \theta_2)]} \\ &= E(S_i|Z_i) \frac{\delta}{1 - (1 - \delta) [\theta_1 - (1 - \theta_1 - \theta_2)]}. \end{aligned}$$

To obtain the valuation of an informed trader, $E(\Delta V_m | S_i)$, simply replace the specialist's information set with the signal. Thus

$$\begin{aligned} E(\Delta V_m | S_i) &= \delta \cdot \frac{P(S_i = 1|S_i) - P(S_i = -1|S_i)}{1 - (1 - \delta) [\theta_1 - (1 - \theta_1 - \theta_2)]} \\ &= S_i \cdot \frac{\delta}{1 - (1 - \delta) [\theta_1 - (1 - \theta_1 - \theta_2)]}. \end{aligned}$$

Proof of Theorem 1 The arrival time of trader i is given by

$$\tau_i = \tau_{i-1} + \min(f^{-1}, g),$$

where g is the elapsed time to the first recorded trade since τ_{i-1} and τ_0 is the beginning of the interval. If all recorded trades are at integer multiples of f^{-1} , then $A = Lf$. If some recorded trades are not at integer multiples then for at most T arrivals, $\tau_i - \tau_{i-1} < f$. As a result, $A \leq Lf + T$, so

$$B = \frac{A}{T} - \frac{Lf}{T} \leq \frac{T}{T}.$$

As trades cannot be less than one second apart, if $f = 1$ then $\tau_i - \tau_{i-1} = f^{-1}$ for all arrivals and $B = 0$. ■