One Day in the Life of a Very Common Stock

David Easley
Cornell University

Nicholas M. Kiefer
Cornell University, CLS, and University of Aarhus

Maureen O'Hara
Cornell University

Using the model structure of Easley and O'Hara (Journal of Finance, 47, 577–604), we demonstrate how the parameters of the market-maker's beliefs can be estimated from trade data. We show how to extract information from both trade and no-trade intervals, and how intraday and interday data provide information. We derive and evaluate tests of model specification and estimate the information content of differential trade sizes. Our work provides a framework for testing extant microstructure models, shows how to extract the information contained in the trading process, and demonstrates the empirical importance of asymmetric information models for asset prices.

The theoretical market microstructure literature abounds with structural models of the market-maker's price-setting decision problem in securities markets. These models [Glosten and Milgrom (1985); Kyle

We are grateful to seminar participants at Carnegie Mellon, Cornell, the London Business School, Maryland, McGill, Minnesota, Northwestern, Princeton, Rice, Rochester, Southern Methodist, Virginia Tech, Florida, and Aarhus University, and the European Finance Association meetings. We also thank the editor, Robert Korajczyk, an anonymous referee, Joel Hasbrouck, Rick Green, Bruce Lehmann, and Tal Schwartz for helpful comments. P. S. Srinivas and Joe Paperman provided valuable programming assistance. This research was supported by National Science Foundation grant SBR93-20889.
Address correspondence to Maureen O'Hara, Johnson Graduate School of Management, Malott Hall, Cornell University, Ithaca, NY 14853.

The Review of Financial Studies Fall 1997 Vol. 10, No. 3, pp. 805–835

(1985); Easley and O'Hara (1987) to name but a few] predict the price process by analyzing the learning problem confronting market makers. Central to this learning problem is the trade process. The market maker watches the timing and sequences of trades, inferring both the motivation of traders and their private information about asset values. Viewed from this perspective, the trading process contains the information that subsequently appears in prices.

Yet the vast majority of empirical work in finance analyzes only price data, and even theoretical models devote scant attention to what information the trade process should have, or even could have. Does the market maker learn only from the net imbalance of buys and sells as in a Kyle model? Is it the number of trades that provides information as in the Glosten–Milgrom model, or does their arrival rate matter as well? Can total volume, or transaction size, or the timing of trades provide meaningful information to the market? Does the existence of trades at all provide information relevant for discerning the underlying true asset value?

The importance of resolving these issues is illustrated by recent intriguing empirical research by Jones, Kaul, and Lipson (1994). Those authors find that the relation between volume and volatility so frequently analyzed in finance actually reflects the positive relation between volatility and the number of transactions. Using daily data, these authors show that it is the number of trades that appears to provide virtually all the explanation for the volatility phenomena, with volume (and trade size) playing little role. Indeed, the authors go on to conclude that "our evidence strongly suggests that the occurrence of transactions per se contains all of the information pertinent to the pricing of securities." Yet, is this really the case? Might not other features of the trade process also be informative? And if they are, why? Jones, Kaul, and Lipson argue that more theoretical work is needed to resolve this issue, but we argue in this article that this is not the case: what is needed is an empirical methodology for using the structure of existing microstructure models in empirical research.[1]

In this article we develop such a framework for analyzing the information in the trading process, and by extension for analyzing the behavior of security market prices. Using the structure of the Easley and O'Hara (1992) theoretical microstructure model, we empirically estimate the model's parameters from a time series of trade data. These

[1] Jones, Kaul, and Lipson note that the Easley and O'Hara (1992) model demonstrates that the number of trades will be positively correlated with absolute price changes, but that since the model does not explicitly contain trade sizes it cannot explain their results on the noneffect of trade size and volume. We use results from Easley and O'Hara (1987) to extend the (1992) model to include trade size, and thus we can address these issues explicitly.

parameters are the "primitives" underlying the market-maker's learning and pricing problem, and as such are the probabilities the market maker attaches to the underlying information structure. The primitives form the probabilities along each branch of an extensive-form game tree, and it is this entire tree that we estimate. Having extracted the information from the trade process, we then have precise estimates of the parameters of the market-maker's decision problem. These parameter values tell us, for example, how likely it is the market maker believes there is informed trading in the security. These parameter values also tell us the information conveyed by various features of the trading process. For example, we can explicitly determine the information content conveyed by the size of trades.

As we demonstrate in this article, it is these underlying parameters that can explain the behavior of the price process; the number of trades influences these parameters, and that is why Jones, Kaul, and Lipson found trades such an important predictor of price volatility. But there is much more that affects these underlying parameters, and the estimation technique developed in this article provides a way to empirically determine the information content of various pieces of market information. As we show, this approach allows us to predict the behavior of the price process in a new, and we believe important, way using trade data.

To set the stage, and to illustrate the importance of trades in understanding price behavior, we consider the relation between daily closing prices and trades. The regressions we report are for 30 days of trading in Ashland Oil (more on the data below). Our dependent variable is the CRSP closing price (p_d for day d). The regressors are the number of buys (Buy_d), number of sells ($Sell_d$), and the lagged price (p_{d-1}). The theoretical model we use emphasizes the number of buy and sell trades, and thus, our choice of regressors (these trades are not shares but actual trades). The lagged price is included since the model explains the effect of an information event over the course of a day——in some sense the lagged price provides a base value from which the current price can move. Our first regression is

$$\underset{}{p_d} = \underset{(1.41)}{3.27} + \underset{(0.006)}{0.025}\, Buy_d \underset{(0.009)}{-0.031}\, Sell_d \underset{(0.047)}{+0.89}\, p_{d-1} \qquad d = 1, \ldots, 30$$

with R^2 (adjusted) = 0.93. Buys and sells are clearly important on the basis of their standard errors (reported in parentheses below the parameter estimates). Their contribution to fit can be seen by comparison to the regression on the lagged price alone; there the adjusted R^2 drops to 0.80. Adding the buy volume and sell volume (number of shares of each) marginally improves the fit (R^2 goes to 0.94; the

F (2,23) is 3.75, p-value 0.039), but clearly trades provide most of the action. The R^2 with buy and sell volume alone is 0.87, the F for buy and sell trades with volumes included is 21. Thus, there is some action in trades that requires explanation. Note that we are not proposing that this is a full analysis of the dynamics of prices, we are reporting this regression as a compilation of summary statistics indicating that the topic is interesting. However, we can report that the result on the importance of trades is robust to various minor respecifications and to use an alternative price series (ISSM close versus CRSP).

In the analysis that follows, we provide one explanation for why it is that trades have the explanatory power exhibited above. We also demonstrate how trade size affects price determination, and explain when this variable will (and will not) be informative. What underlies our analysis is a microstructure model, and it is our goal in this work to show how such a model can be used in a well-defined statistical framework to guide empirical work. To demonstrate our methodology, we focus on the trade process of one stock, but it should be clear that this is merely illustrative; our methodology applies to any security with sufficient high-frequency observations.

One issue we stress at the outset is the role played by the structural model in our approach. It would, of course, be naive to assume that any individual consistently acted in the simple mechanical fashion depicted in the model. But our concern is not with the elegance or complexity of the model per se. Instead, the question at issue is whether the model provides a useful interpretation and description of actual market behavior.[2] Such a focus can also be found in the recent work of Bernhardt and Hughson (1993) and Foster and Viswanathan (1995) who examine the empirical implications of the Kyle model. Our approach here differs dramatically from those articles, but our purposes are allied: we seek to evaluate the efficacy of theoretical models in the hope of improving our ability to understand empirical market behavior. To do so, we estimate the order arrival process, and using the structure of an asymmetric information market microstructure model, we interpret the parameter values. To the extent that the model does not capture important features of the trading process it will fail to empirically explain market behavior. The implementation technique we develop in this article, however, allows us to test how well the model does, and this, in turn, allows us to investigate the efficacy of alternative model specifications.

[2] This important distinction between the model and the world is important in several areas in which empirical and theoretical modeling have been complements, including search models of the labor market [see the review by Devine and Kiefer (1991)], applied dynamic programming, and real business cycle modeling in macroeconomics.

The article is organized as follows. In the next section we provide the basic structure of the Easley and O'Hara (1992) theoretical microstructure model. Section 2 details the econometric estimation and discusses some underlying specification issues. Section 3 describes the data and presents our maximum likelihood estimation results. We develop a number of statistical tests to evaluate the fit of the model, and the reasonableness of the model's assumptions. We also investigate the interpretation and implications of our parameter estimates. In Section 4, we consider extensions and generalizations to our model, and in particular investigate the role played by trade size. We then estimate this more general model, and we determine the effects of trade size on the market-maker's beliefs. In Section 5, we return to the regressions reported above, but this time we use our model's parameter estimates to explain return behavior. This provides a simple de facto check on the validity of our trade-based approach. Section 6 provides a summary of our results, and concludes with a discussion of the applicability of our approach and some suggestions for future research.

1. The Theoretical Model

In this section, we provide a brief description of the sequential trade microstructure model that we estimate in this article.[3] We first set out a simple framework in which trade size does not enter. We then introduce trade size as an explicit variable in the model and show how the approach easily generalizes to include this and other extensions. In a sequential trade model, potential buyers and sellers trade a single asset with a market maker. The market maker is risk neutral and competitive, and quotes prices at which she will buy or sell the asset. Traders arrive individually at the market according to a probabilistic structure, and trade or chose to not trade at the quoted prices. Following each arrival, the market maker revises her quotes based on information revealed by the trading process.

In this model, the asset being traded has a value at the end of the trading day represented by the random variable, V.[4] An information

[3] This model is based on the Easley and O'Hara (1992) model which is similar in spirit to Glosten and Milgrom (1985).

[4] We do not index values by day in order to keep the notation simple. All that matters for our model of the trade process and our empirical implementation of it is that those informed of good news buy and those informed of bad news sell. Later in the article, when we look at prices over many days, we do index values by day. Then we interpret good news as a signal that the value of the asset has shifted up from its previous close by $\overline{\Delta V}$; similarly, bad news is interpreted as a signal that the value has shifted down by $\underline{\Delta V}$. The obvious martingale property of prices then requires that V^* be the previous closing price, or that $\delta\underline{\Delta V} = (1-\delta)\overline{\Delta V}$.

event is the arrival of a signal, Ψ, about V. The signal can take on one of two values, L and H, with probabilities δ and $1-\delta$. The value of the asset conditional on bad news, L, is $\underline{V}$; similarly, conditional on good news, H, it is $\bar{V}$. Information events need not occur, reflecting the fact that new information does not always arise. If no new signal has occurred, we denote this as $\Psi = 0$, and the value of the asset simply remains at its unconditional level $V^* = \delta\underline{V} + (1-\delta)\bar{V}$. We assume that the probability that an information event has occurred before the start of a trading day is α, with $1-\alpha$ the corresponding probability that there has been no new information. This assumption that information events occur only prior to the start of a trading day is clearly an abstraction; what underlies our analysis is the notion that new information arises at discrete intervals. From a theoretical perspective, this is most easily captured by adopting the fiction of a trading day.

Trade in this market arises from uninformed and informed traders. An informed trader is assumed to be risk neutral and to take prices as given. This assumption rules out any strategic behavior by the informed and results in a simple trading strategy: If an informed trader has seen a high signal, he will buy the stock if the current quote is below $\bar{V}$; if he has seen a low signal, he will sell if the quote is above $\underline{V}$. The uninformed trader's behavior is more complex. As is well known, the presence of traders with better information dictates that an uninformed trader trading for speculative reasons would always do better not trading at all. To avoid this no-trade equilibrium, at least some uninformed traders must transact for nonspeculative reasons such as liquidity needs or portfolio considerations. For the uninformed as a whole, we make the realistic assumption that one-half are potential buyers and one- half are potential sellers. It would be reasonable to expect that uninformed traders demands would depend on history and quotes, but as uninformed traders know that prices are conditional expected values they know that they trade at the right price. We assume that when an uninformed trader checks the quote, the probability that he will trade is $\varepsilon > 0$.[5]

The assumptions of market-maker risk neutrality and competitive behavior dictate that her price quotes yield zero expected profit conditional on a trade at that quote. Since the informed traders profit at the market-maker's expense, the probability that a trade is actually informed is important for determining these prices. We assume that if

[5] Allowing ε to depend on history and quotes is feasible in the theoretical model, but empirical implementation would require a specific functional form for this dependence. The simple functional form we use here is clearly an abstraction, but it seems a reasonable representation of noise trader behavior.

an information event occurs then the market maker expects the fraction of trades made by the informed to be μ. Note that this need not correspond identically to the fraction of the trader population who receives any signal, as the trading intensity of informed traders may differ from that of uninformed traders.

Trades occur throughout the trading day. We divide the trading day into discrete intervals of time, denoted $t = 1, 2, \ldots$. Each interval is long enough to accommodate one trade. This timing specification is designed to capture the possibility that during some intervals no-trades may occur. In the estimation, the exact length of the time interval will be an important variable, and we will discuss its implications further. At this point it is useful to note, however, that the length of the interval is intended to capture the general pattern of trades in the stock.

Trades take place sequentially, meaning that at each time interval some trader is randomly selected according to the probabilities given above and given the opportunity to trade. At each time t, the market maker announces the bid and ask prices at which she is willing to buy or sell one unit of the asset. Similarly, at each time t, the trader selected to trade has the option of buying one unit at the market-maker's ask price, selling one unit at the market-maker's bid price, or not trading at all. Following the trade outcome, the market maker has the opportunity to set new prices for the next trading interval, and a new trader is selected to trade.[6]

This trading structure for a trading day is depicted in the tree diagram given in Figure 1. In the tree, the first node corresponds to nature selecting whether an information event occurs. If there is an information event (which occurs with probability α), then the type of signal is determined at the second node. There is a δ probability that the signal is low and a $1-\delta$ probability that the signal is high. These two nodes are reached only at the beginning of the trading interval, reflecting the model's assumption that information events occur only between trading days.

From this point, traders are selected at each time t to trade based on the probabilities described previously. If an information event has occurred, then we are on the upper portion of the tree and an informed trader is chosen to trade with probability μ. Whether the trader buys or sells depends upon the signal he has seen. With probability $(1-\mu)$ an uninformed trader is chosen, and the trader is equally likely to be

[6] It is possible to reformulate the statistical model in terms of Poison arrivals in continuous time. The discrete-time formulation is used here to accord most closely with the theoretical model. The market-maker's updating equations for beliefs and the resulting pricing equations are considerably more complex in a continuous-time formulation. Of course, which "works better," a discrete-time or continuous-time version, is an empirical question.

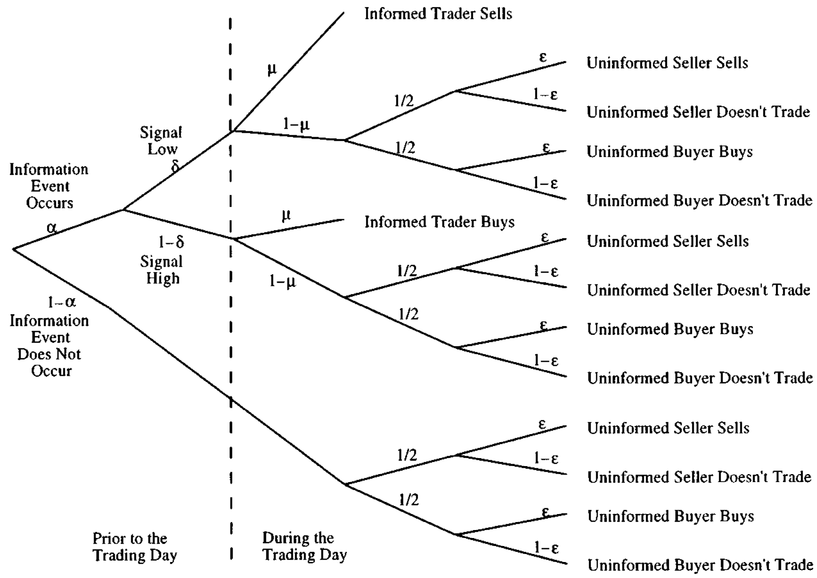

Figure 1
Tree diagram of the trading process.
α is the probability of an information event, δ is the probability of a low signal, μ is the probability that the trade comes from an informed trader, 1/2 is the probability that an uninformed trader is a seller, and ε is the probability that the uninformed trader will actually trade. Nodes to the left of the dotted line occur only at the beginning of the trading day; nodes to the right are possible at each trading interval.

a potential buyer or seller. An uninformed trader will buy with probability ε, and will not trade with probability $1 - \varepsilon$. If no information event has occurred, then we are on the lower part of the tree and all traders are uninformed. A trader selected to trade may thus buy, sell, or not trade with the indicated probabilities. For trades in the next time period, only the trader selection process is repeated, so the game proceeds from the right of the dotted line on the tree diagram. This continues throughout the trading day.

There are several aspects of this diagram that are important for our analysis. First, the probabilistic structure of the tree is completely described by the parameters α, δ, μ, and ε. Given those values, we could calculate the probability of any trade outcome. We will demonstrate shortly that the market-maker's price-setting decision problem will also depend on these variables, so the estimation process in the remainder of the article will essentially focus on determining these parameter values. Second, the outcome in any trading interval can only be a buy, a sell, or a no-trade observation. The probabilities of observing each of these events differs depending upon where we are in the tree. Since a no-trade outcome is more likely to occur if there has been no information event, observing a no-trade may lead the market maker to think it more likely we are on the bottom part of the tree. The market maker will thus use the trade outcome to infer where on the tree diagram she is, and thus how likely it is that an information event has actually occurred. Third, the iterative structure of the game means that we have more opportunities to extract information on the μ and ε terms than we do on the α and δ terms. In particular, since events to the left of the dotted line happen only at the beginning of the day, there is really only one "draw" from that distribution per day. Events to the right of the dotted line happen every time interval, so observing trade outcomes throughout the day provides the potential at least to observe multiple draws from these trader-related distributions. These differences in observability will play a major role in our estimation procedure.

The tree diagram depicts what we refer to as the *trade process*. The market maker is assumed to know this trade process, and hence she knows the parameter values α, δ, μ, and ε. What she does not know is whether an information event has occurred, whether it is good or bad news given that it has occurred, and whether any particular trader is informed. However, the market maker is assumed to be a rational agent who observes all trades and acts as a Bayesian in updating her beliefs. Over time, these observations allow the market maker to learn about information events and to revise her beliefs accordingly. It is this revision that causes quotes, and thus prices, to adjust.

The trade process can thus be viewed as an input to the *quote process*, or how it is that the market maker sets her prices to buy and sell every period. To determine this quote process, we begin by considering the market-maker's quotes for the first trade of the day. Recall that the assumptions of risk neutrality and competitive behavior dictate that the market maker set prices equal to the expected value of the asset conditional on the type of trade that will occur. This requires determining the conditional probability of each of the three possible values for the asset; here we provide calculations for the conditional probability of the low value $\underline{V}$. If no signal has occurred, then this probability remains unchanged at δ. If a high signal occurred, then the true probability is zero, while if a low signal occurred the true probability is one. The market-maker's updating formula given a trade observation Q is then

$$\begin{aligned}\delta(Q) &= \Pr\{V=\underline{V} \mid Q\}\\ &= 1\cdot\Pr\{\psi=L \mid Q\}+0\\ &\quad \cdot\Pr\{\Psi=H \mid Q\}+\delta\Pr\{\Psi=0 \mid Q\}. \qquad (1)\end{aligned}$$

As the market maker is a Bayesian, these conditional probabilities are given by Bayes' rule:

$$\Pr\{\Psi = X \mid Q\} = \frac{\Pr\{\Psi = X\}\Pr\{Q \mid \Psi = X\}}{\Pr\{\Psi = L\}\Pr\{Q \mid \Psi = L\} + \Pr\{\Psi = H\}\Pr\{Q \mid \Psi = H\} + \Pr\{\Psi = 0\}\Pr\{Q \mid \Psi = 0\}}. \tag{2}$$

The explicit probabilities can be derived from the tree diagram in Figure 1. For example, the probability that there was no information event (i.e., $\Psi = 0$) given that a sale occurred ($Q = S_1$) is

$$\Pr\{\Psi = 0 \mid S_1\} = \frac{(1-\alpha)1/2\varepsilon}{(\delta\alpha\mu + (1-\alpha\mu)1/2\varepsilon)}. \tag{3}$$

The market-maker's conditional probability of $\underline{V}$, given a sale, is then given by

$$\delta_1(S_1) = \delta\left[\frac{\alpha\mu + \varepsilon 1/2(1-\alpha\mu)}{\delta\alpha\mu + \varepsilon 1/2(1-\alpha\mu)}\right] > \delta. \tag{4}$$

Hence, the market maker increases the probability she attaches to $\underline{V}$ given that someone wants to sell to her. The amount of this adjustment depends on the probability of information-based trading ($\alpha\mu$) and on the trading sensitivities of the uninformed traders (the ε).

Given these conditional expectations, the market-maker's bid and ask prices can be calculated:

$$E[V \mid S_1] = b_1 = \frac{\delta\underline{V}(\alpha\mu + \varepsilon 1/2(1-\alpha\mu)) + (1-\delta)\bar{V}\varepsilon 1/2(1-\alpha\mu)}{\delta\alpha\mu + \varepsilon 1/2(1-\alpha\mu)} \tag{5}$$

$$E[V \mid B_1] = a_1 = \frac{\delta\underline{V}(\varepsilon(1/2)(1-\alpha\mu)) + (1-\delta)\bar{V}(\alpha\mu + \varepsilon(1/2)(1-\alpha\mu))}{(1-\delta)\alpha\mu + \varepsilon(1/2)(1-\alpha\mu)}. \tag{6}$$

These equations give the market-maker's initial quotes for the first trading interval of the day. Following the trading outcome, the market maker will revise her beliefs given the information she learns and set new quotes for the next trading interval. To describe the quote process, therefore, we must determine how the market-maker's beliefs evolve over the trading day.

At each time t, there are three possible trading outcomes: a buy (B), a sale (S), or a no-trade (N). Let $Q_t \in$ [B, S, N] denote this trade outcome at time t. Then as the day progresses the market maker observes the trade outcomes and by the beginning of period t she has seen the history $Q^{t-1} = (Q_1, Q_2, Q_3, \ldots, Q_{t-1})$. Her beliefs at the beginning of period t are given by Bayes' rule and are represented by $\rho_{L,t} = \Pr\{\Psi = L \mid Q^{t-1}\}$, $\rho_{H,t} = \Pr\{\Psi = H \mid Q^{t-1}\}$, and $\rho_{0,t} = \Pr\{\Psi = 0 \mid Q^{t-1}\}$.

As these beliefs will be crucial in our estimation, it may be useful to illustrate their derivation with a simple example. Suppose that in the past t intervals the market maker has observed N no-trades, B buys, and S sales. Then her posterior probability that no information event has occurred is

$$\begin{aligned}\Pr\{\psi=0|Q^t\} = {} & (1-\alpha)(1/2\varepsilon)^{S}(1/2\varepsilon)^{B}[(1-\alpha)(1/2\varepsilon)^{S}(1/2\varepsilon)^{B} \\ & +(1-\mu)^{N}[\alpha\delta(\mu+(1-\mu)1/2\varepsilon)^{S}((1-\mu)(1/2)\varepsilon)^{B} \\ & +\alpha(1-\delta)((1-\mu)1/2\varepsilon)^{S}(\mu+(1-\mu)(1/2)\varepsilon)^{B}]]^{-1}. \quad (7)\end{aligned}$$

Since beliefs depend only on (N, B, S) it follows that quotes will also depend on these variables. Easley and O'Hara demonstrate that the trade-tuple {buys, sells, and no-trades} is a sufficient statistic for the quote process. Hence, to describe the stochastic process of quotes, we need only know the total numbers of *buys*, *sells*, and *no-trades*; the quote process does not depend on any other variables. This property will be important for the estimation in the next section. The market maker's quotes at time $t+1$ are then the expected value of the asset conditional on the cumulative trading and no-trade outcomes to time t, (N, B, S), and the trade at time $t+1$, or

$$\begin{aligned}b_{t+1} = {} & \Pr\{\psi = L|\mathrm{N}, \mathrm{S}+1, \mathrm{B}\}\underline{V} + \Pr\{\psi = H|\mathrm{N}, \mathrm{S}+1, \mathrm{B}\}\bar{V} \\ & + \Pr\{\psi = 0|\mathrm{N}, \mathrm{S}+1, \mathrm{B}\}V^{*} \quad (8)\end{aligned}$$

and

$$\begin{aligned}a_{t+1} = {} & \Pr\{\psi = L|\mathrm{N}, \mathrm{S}, \mathrm{B}+1\}\underline{V} + \Pr\{\psi = H|\mathrm{N}, \mathrm{S}, \mathrm{B}+1\}\bar{V} \\ & + \Pr\{\psi = 0|\mathrm{N}, \mathrm{S}, \mathrm{B}+1\}V^{*}. \quad (9)\end{aligned}$$

The theoretical model has thus far described the evolution of the trade process and the quote process. There remains, however, the derivation of the price process. The price process is the stochastic process of transaction prices, and as should now be apparent, it is determined by the quote process and the trade process. At each time t, the trade process determines whether there will be an actual trade, and if there is, the transaction price is either the bid quote or the ask quote. Note, however, that a trade need not occur in every trade interval. But if a no-trade outcome occurs it, too, will change the market-maker's beliefs and prices, a movement that will not be reflected in a transaction price. Thus the price process is a *censored sample* of the quote process.

The model is thus complete. Given the trade process and some prior beliefs, the market maker forms her expectations of the asset's expected value conditional on the trades that can occur, and these expectations are the market-maker's initial quotes. The market maker learns from trade outcomes, so given the trade history, the market maker revises her beliefs and sets price quotes for the next trading interval. Over the course of the trading day, beliefs, quotes, and transaction prices evolve. The day ends, and the market maker begins the entire process over again the next day.

2. The Estimation of the Trade Process

Suppose now we turn the problem around and consider it from the point of view of an econometrician. The model developed in the previous section is a structural model of the security price formation process. It gives the specific rules used to set quotes at every trade interval. Since we know these rules, if, like the market maker, we also knew the trading process parameters α, δ, μ, and ε and the history of trades, then presumably we too could calculate the quote for the next trade interval.

We know the trading history, but we must estimate the structural parameters. Information on these parameters is contained in both the trade and quote histories. Our model tells us how to use trades to estimate these parameters; it does not tell us how to use quotes. This is not a problem for two reasons. First, the quotes are the outcome of the market-maker's decision problem, so for her they do not carry information. The market maker must have learned the parameters from trades and perhaps other data that we do not have. Second, we show in this section that we can estimate reasonably precisely the structural parameters from the available data on trades.

The estimation of the trade process requires recovering the parameter structure depicted in the tree diagram in Figure 1. As noted earlier, the extensive form game depicted there is actually a composite of two trees, one relating to the existence and type of information events, and the other detailing the trader selection process. A significant difference between these games is their frequency of occurrence. The information event moves (which involve α and δ) occur only once a day, while the trader selection (which depends on μ and ε) occurs many times throughout the day.

This difference has an important implication for our ability to estimate these underlying parameters. In the course of one day, there could easily be 100 or more trade outcome observations. These observations are independent draws from a distribution parameterized by μ and ε, but they share a single draw of α and δ. Hence, while it may be possible to estimate μ and ε from a single day's trade outcomes, it is not possible to estimate α and δ. Instead, multiple days of data are needed to identify these two parameters.

To construct the likelihood function we first consider the likelihood of trades on a day of known type. Conditional on a known day, trades are independent. Thus we have a standard estimation problem in which the total number of buys B, sells S, and no-trades N forms a sufficient statistic for the daily trade data. Consider a good event day. The probability of a buy, sell, or no-trade at any time during this day can be read off the good event branch of the tree in Figure 1. The probability of B buys, S sells, and N no-trades on a good event day is thus proportional to[7]

$$\Pr\{B, S, N|\psi = H\} = [\mu + (1-\mu)1/2\varepsilon]^{B}[(1-\mu)1/2\varepsilon]^{S}[(1-\mu)(1-\varepsilon)]^{N}. \tag{10}$$

Similarly, on a bad event day the probability of (B,S,N) is proportional to

$$\Pr\{B, S, N|\psi = L\} = [(1-\mu)1/2\varepsilon]^{B}[\mu + (1-\mu)1/2\varepsilon]^{S}[(1-\mu)(1-\varepsilon)]^{N}. \tag{11}$$

Finally, on a day in which no event has occurred the probability of (B, S, N) is proportional to

$$\Pr\{B, S, N|\Psi = 0\} = [1/2\varepsilon]^{B+S}(1-\varepsilon)^{N}. \tag{12}$$

While Equations (10), (11), and (12) give the distributions of buys, sells, and no trades on known days, for our analysis we need the unconditional probability that B buys, S sells, and N no-trades occurs. This probability is just a mixture of Equations (10), (11), and (12) using the probabilities of the days: $\alpha(1-\delta)$, $\alpha\delta$, and $(1-\alpha)$. Because the likelihood function is this probability regarded as a function of the parameter values α, δ, μ, and ε, it is useful to write this dependence explicitly, so the likelihood function for a single day is proportional to

$$\begin{aligned}\Pr\{B, S, N|\alpha, \delta, \mu, \varepsilon\} = {} & \alpha(1-\delta)[[\mu + (1-\mu)1/2(\varepsilon)]^{B} \\ & \cdot [(1-\mu)1/2(\varepsilon)]^{S} \cdot [(1-\mu)(1-\varepsilon)]^{N}] \\ & +\alpha\delta[[(1-\mu)1/2(\varepsilon)]^{B} \cdot [\mu + (1-\mu)1/2(\varepsilon)]^{S} \\ & \cdot [(1-\mu)(1-\varepsilon)]^{N}] \\ & +(1-\alpha)[[1/2(\varepsilon)]^{B+S}(1-\varepsilon)^{N}]. \end{aligned} \tag{13}$$

To calculate this likelihood function over multiple days, note that the assumption of independence of information events between days

[7] The probability is the expression given in Equation (10) times the combinatorial factor expressing the number of ways of choosing B buys, S sells, and N no-trades out of a sample of size B+S+N. This factor involves only data, not parameters, and has no effect on estimated parameter values.

means that this probability is the product

$$\Pr\{(B_d, S_d, N_d)_{d=1}^{D} | \alpha, \delta, \mu, \varepsilon\} = \prod_{d=1}^{D} \Pr\{(B_d, S_d, N_d) | \alpha, \delta, \mu, \varepsilon\} \quad (14)$$

where (B_d, S_d, N_d) is the outcome on day d, $d = 1, \ldots, D$.

The likelihood function, as usual, is an efficient description of all of the data information about the parameters. The form of this likelihood function has a number of implications for our estimation. First, using only one day's data, it is clear that α and δ are not identified. The likelihood in Equation (13) is bilinear in α and δ, so the maximum likelihood estimators will be zeros or ones. This situation is analogous to estimating a Bernoulli probability from one trial. For any given day, our sufficient statistic is at best three-dimensional, so it is possible to estimate at most three parameters. But if $\mathrm{B} + \mathrm{S} + \mathrm{N}$ is approximately a constant (as may be the case for many stocks) or is ancillary (as is the case if we are considering fixed time intervals such as seconds or minutes), then our statistic is actually two-dimensional, limiting our estimation ability accordingly. In any case, our statistic is sufficient to allow estimation of the two parameters μ and ε from daily data.

Second, in our model, information on μ and ε accumulates at a rate approximately equal to the square root of the number of trade outcomes, while information on α and δ accumulates at a rate approximately equal to the square root of the number of days. Hence, using many days of data greatly enhances our ability to estimate more parameters. For our estimation problem, this means that using a multiday sample can provide sufficient information to estimate α and δ.

What is also apparent, however, is that while it may be sensible to use large sample methods to estimate μ and ε, it is less so for α and δ. The presumed stationarity of information is unlikely to be true over a long sample period, dictating a natural limit to the number of days we can sensibly employ. The difference in information accumulation rates also dictates that the precision of our μ and ε estimates will exceed the precision of our α and δ estimates. Of course, this is reflected in the standard errors of the estimators.

Having defined the likelihood function for our model, we can now calculate the parameter values of α, δ, μ, and ε that maximize this function for a given stock. Roughly, what our procedure does is to classify days into buy-led high-volume days, sell-led high-volume days, and low-volume days (as in Figure 1). In our structural interpretation, these are good-event days, bad-event days, and no-event days, respectively. The likelihood function given in Equation (13) is a mixture of trinomials, with specific terms reflecting the numbers of buys, sells, and no-trades and with restrictions across the components of the mixture. For tractability in the estimation, we use a log transformation of our likelihood function, which after simplifying and dropping a constant term, is given by

$$\sum_{d=1}^{D} \log\left[\alpha(1-\delta)\left(1+\frac{\mu}{x}\right)^{\mathrm{B}} + \alpha\delta\left(1+\frac{\mu}{x}\right)^{\mathrm{S}} + (1-\alpha)\left(\frac{1}{1-\mu}\right)^{\mathrm{S+B+N}}\right] + \sum_{d=1}^{D} \log[((1-\mu)(1-\varepsilon))^{\mathrm{N}} x^{\mathrm{S+B}}], \tag{15}$$

where $x = (1-\mu)\frac{1}{2}(\varepsilon)$. Note that the weights on the trinomial components reflect the information event parameters α and δ.[8] Hence, a simple test of the information model is to compare the goodness-of-fit of the model's likelihood function with that of the simpler trinomial in which α and δ do not appear. In this specification, each day is the same as far as information is concerned. As will be apparent, maximizing the likelihood function of Equation (15) provides parameters yielding a direct measure of the effect of information on trades, quotes, and prices.

3. The Data and Maximum Likelihood Estimation

The estimation described in the last section requires trade outcome data for a specific stock over some sample period. As our focus here is on the estimation approach rather than on any specific performance measure, which particular stock is analyzed is not important, nor is the sample period of any particular concern. Of course, since we intend our methods to be quite generally applicable, we do not wish to select a bizarre stock or unusual period for illustration. For our analysis, we selected Ashland Oil to be our sample stock. This selection was dictated both by convenience, and by the relatively active trading found in the stock. This latter characteristic is important given that it is the information contained in trade data that is the focus of our work.

Trade data for Ashland Oil were taken from the ISSM transactions database for the period October 1, 1990, to November 9, 1990.[9] A 30 trading-day window was chosen to allow sufficient trade observations for our estimation procedure. The ISSM data provide a complete listing

[8] This mixture of trinomials is reminiscent of the work of Clark (1973), who postulated that the distribution of security prices could be represented by such mixtures. Here our focus is on the trade process, but this in turn affects the price process.

[9] Over this period there was an earnings announcement in day 10 of our sample, but the data suggest little impact of this on the stock.

of quotes, depths, trades, and volume at each point in time for each traded security. For our analysis, we require the number of buys, sells, and no-trades for each day in our sample. Since these are not immediately obtainable from the data, a number of transformations were needed to derive our data.

The first of these involves the calculation of no-trade intervals. In the theoretical model, no-trade observations are more likely if there has been no information event, and hence the time between trades is an important input. How this no-trade interval should be measured, however, is not obvious. In any trading system, there are frictions such as delays in order submission and execution, time lags in the posting of quotes (or trades), or even physical constraints on order placement that result in time arising between trades. Moreover, while the most active stocks may trade every minute, the vast majority of stocks trade much less frequently, suggesting that at least on average some time period will elapse between trades.

Based on the average general trading pattern in the stock, we chose five minutes as a reasonable measurement of a no-trade interval. Over our sample period, the number of Ashland Oil's daily transactions ranged from a low of 20 to a high of 73, so that a five- minute interval seemed long enough to exclude market frictions, while being short enough to be informative. Using the time and trade information in the ISSM data, we define a no-trade outcome if at least five minutes has elapsed since the last transaction. The total number of no-trade outcomes in a trading day is thus the total number of 5-minute intervals in which no transaction occurred.

This choice of a five-minute interval is, of course, arbitrary. As with any discretization, the only way out of the arbitrariness is to move to continuous time. To check whether our particular discretization matters, we have also processed the data based on alternative no-trade intervals ranging from 30 seconds to 10 minutes. For all sufficiently small no-trade intervals, we obtain similar results. We focus on the five-minute interval, but we also present our results for other intervals. As we will show later in this section, our results are remarkably robust to this specification issue.

A second transformation to the data involves the classification of buy and sell trades. Our model requires these to be identified, but the ISSM data record only transactions, not who initiated the trade. This classification problem has been dealt with in a number of ways in the literature, with most methods using some variant on the uptick or downtick property of buys and sells. In this article, we use a technique developed by Lee and Ready (1991). Those authors propose defining trades above the midpoint of the bid-ask spread to be buys and trades below the midpoint of the spread to be sells. Trades at the midpoint are classified depending upon the price movement of the previous trade. Thus, a midpoint trade will be a sell if the midpoint moved down from the previous trade (a downtick) and will be a buy if the midpoint moved up. If there was no price movement then we move back to the prior price movement and use that as our benchmark. We applied this algorithm to each transaction in our sample to determine the daily numbers of buys and sells.[10]

The resulting trade outcome data are given in Table 1. As is apparent, the number of transactions and no-trades varies across days, but over the entire sample buys and sells are approximately equal. The variability in the number of trades and no-trades across days reflects both differences in daily trade volume and frequency. In particular, while no-trades occur only once every five minutes, buys and sells are recorded whenever they occur, so the overall trade outcome totals need not be constant across days.

3.1 The maximum likelihood estimation

Using the data in Table 1, we can now estimate our log-likelihood function given in Equation (15). The likelihood function is well-behaved, and a gradient method (GRADX from the GQOPT package) went directly to the maximum from a variety of starting values. The resulting maximum likelihood estimates of our parameters are as follows:

Parameter	Standard error
$\mu = 0.172$	0.014
$\varepsilon = 0.332$	0.012
$\alpha = 0.750$	0.103
$\delta = 0.502$	0.113

The log-likelihood value is -3028.

The estimation results provide an intriguing picture of the underlying information structure in the trade data. If our structural model is correct then the market maker believes it fairly likely that information events do occur in the stock, attaching a 75% probability to there being new information in the stock at the beginning of trading each day. If an information event has occurred, the market maker believes that approximately 17% of the observations (of trades and no-trades) are information based, with the remaining 83% coming from uninformed traders. Since δ is approximately one-half (i.e., 0.502), the

[10] This technique allows us to classify trades throughout the day, but it is not useful for the opening trade. Since the opening trade results from a different trading mechanism than is used during the rest of the day (and differs from that derived in our model), we exclude the opening trade in our sample.

Table 1
Trade process data

Trading day	Buys	Sells	No-trade intervals
1	39	12	61
2	39	31	55
3	11	27	65
4	29	15	58
5	33	20	53
6	8	11	71
7	11	24	63
8	9	24	64
9	10	41	60
10	55	7	59
11	27	27	61
12	18	38	62
13	27	16	63
14	12	31	63
15	38	34	54
16	24	22	59
17	15	29	61
18	8	26	66
19	11	20	66
20	6	33	65
21	19	33	59
22	35	14	63
23	22	21	61
24	55	12	54
25	20	12	64
26	13	15	65
27	16	17	67
28	5	33	61
29	21	20	61
30	37	5	63
Mean	20.8	20.9	61

Trade data for Ashland Oil from October 1, 1990–November 9, 1990. The number of buys and sells is determined from transactions data using the Lee–Ready algorithm, excluding the first trade of the day. A five-minute interval is used to determine the number of no-trade intervals.

market maker believes that good and bad news are generally equally likely over our sample period. This seems reasonable given the relatively balanced order flow over the period. Finally, the probability that an uninformed trader actually trades given an opportunity is 33%.

The standard errors reveal the expected property that our μ and ε estimates are much more precise than the α and δ estimates. Of more importance is that all of our parameter estimates are reasonably precisely estimated. Consequently, using maximum likelihood estimation we have been able to identify and determine the underlying parameter values of our theoretical model.

Table 2
No-trade intervals and estimated parameters

Estimated parameter	No-trade interval (minutes)					
	1/2	1	2	5	8	10
μ	.023	.044	.083	.174	.238	.277
ε	.039	.076	.148	.333	.476	.556
α	.769	.768	.758	.753	.722	.704
δ	.494	.495	.497	.502	.515	.510
γ	.376	.377	.379	.387	.394	.408
β	.328	.326	.330	.333	.332	.334

This table gives our estimated parameter values for 30 days of trade data defined over different no-trade filters. The parameters μ, ε, α, and δ are defined in the model and are, respectively, the probability of an informed trade, the probability an uninformed trader trades, the probability of an information event, and the probability information events are bad news. The parameter $\gamma = \mu/((1-\mu)\varepsilon+\mu)$ is the fraction of trades made by informed traders when an information event occurs. The parameter $\beta = 1-(1-\varepsilon_T)^{5/T}$ is the probability of at least one trade during a five-minute interval on a nonevent day.

3.2 Specification issues

But how good is this specification? Determining this is complex since specification issues arise with respect to a number of areas. Among the most important of these are stability of the parameter estimates, alternative model specification and testing, and the independence assumptions underlying the model's estimation. We now consider these issues in more detail.

3.2.1 Parameter stability. We first investigate the sensitivity of our results to the choice of the time filter used to develop the no-trade data. Table 2 reports estimated parameter values for no-trade filters between 30 seconds and 10 minutes. The first striking result is that the estimates of α (the probability of an information event) and δ (the probability the news is bad) do not depend heavily on the choice of the filter. This makes sense: it is day-to-day differences in the distribution of trades that identify these parameters. Although the distributions change with the filter, the variation in distributions across days apparently does not.

The estimates of μ and ε do depend on the choice of the filter. This also makes sense: the effect of changing the filter is primarily to change the number of no-trades, and thus the number of observations in a day. The parameter μ, for example, reflects the trading intensity or presence of informed traders and essentially is the fraction of observations (including no trades) made by informed traders when an information event occurs. Since informed traders always trade, the estimate of μ must fall when the number of no- trades rises (i.e., the filter is smaller) since the number of trades is not affected.

What is of more economic interest is the fraction of trades (buys plus sells) made by informed traders on information event days. This is given by a new composite parameter γ, where $\gamma = \mu/((1-\mu)\varepsilon+\mu)$. Remarkably, this parameter estimate (reported in row 5 of Table 2) is nearly constant as the size of the filter varies. It is sensitive only to changes in very long filters, which can be expected to produce distortions. Thus, our inference on the overall fraction of trades made by informed traders (obtained by multiplication by α) of approximately 29% is robust to changes in the filter.

Similarly, ε is the fraction of observations (trades plus no-trades) that are buys or sells on nonevent days. Clearly, as the filter (and hence the number of no-trades) changes, ε also changes. Row 6 of Table 2 reports values of a new parameter, β, which is the estimated probability of at least one trade in a five-minute interval on nonevent days.[11] This parameter estimate is nearly constant over filters. Thus, our inference on the propensity of the uninformed to trade is also robust to filter changes.

3.2.2 Model specification. While the above discussion highlights the stability of our estimated parameter values, there remains the more fundamental question of whether the model per se is correctly specified. In particular, does asymmetric information really affect the trade process or are trades merely artifacts of some more general random process? One way to address this issue is to compare these results with an estimation based on a simpler trinomial model in which the probabilities of buy, sell, or no-trade are constant over the entire sample period. Such a model corresponds to assuming a fixed information structure every day. If our theoretical model is "better" in the sense of explaining the data more completely, then we would expect the log likelihood to be greater for our model than for the simple model. Consequently, our approach provides a natural mechanism for testing and comparing the efficacy of alternative models.

The results of this estimation confirm the value of our model. The log-likelihood value for the simple trinomial model falls to −3102, dictating that the simple specification does not model the data as well as our theoretical model. Moreover, the likelihood ratio statistic of 148, a very surprising value on one degree of freedom, provides strong evidence that the two model specifications statistically differ. These results suggest, therefore, that the theoretical model provides some

[11] If we let ε_T be the estimate of ε using a no-trade filter of T minutes, then $\beta = 1 - (1 - \varepsilon_T)^{5/T}$.

economic insight into the nature of the trade process. Perhaps more important, these results suggest that models incorporating asymmetric information may capture important economic phenomena affecting the trade process. This, in turn, raises the issue of the specification of the underlying information event structure.

3.2.3 Independence and information event specification. The above specification test maintains the hypothesis that information events are independent from day to day in both the behavioral model and in the simple alternative. This independence assumption is certainly restrictive, but whether it is an unreasonable approximation seems a natural question to ask. Note that the assumption is difficult to test, since the occurrence of information events is unobserved. The model does suggest a method, however, of checking this independence assumption. First, note from Figure 1 that the total number of trades (buys plus sells) has the same distribution whenever an information event occurs and a different distribution (with fewer trades) when an event does not occur. Therefore, it is possible to classify days as event days and nonevent days according to the number of trades. Given this classification, we use a runs test to look for dependence in the sequence of events and nonevents. A runs test is appropriate here as this type of test is nonparametric and has power against a very wide class of alternatives, that is, against many types of dependence. Partly as a consequence of this power against many alternatives, runs tests have relatively low power against many specific, restricted classes of alternatives. Below, we consider other tests against specific alternatives.

From Table 1, we calculate the total trades by adding buys and sells—to fix ideas, the trades for the first 15 days are

$$51, 70, 28, 44, 53, 19, 35, 33, 51, 62, 54, 56, 43, 43, 72, \ldots .$$

Now, our estimate of α is 0.75, so we classify the lower one-fourth of these numbers as nonevent days and the upper three-fourths as event days. Days with 34 or fewer trades are nonevent days. Denoting event occurrence by one and nonoccurrence by zero, we have a series whose first 15 observations are

$$\begin{array}{ccccccccccccccccc} 1, & 1, & 0, & 1, & 1, & 0, & 1, & 0, & 1, & 1, & 1, & 1, & 1, & 1, & 1, & & \ldots . \\ & 1, & 2, & & 3, & 4, & 5, & 6 & & & & & & & & 7, & \ldots . \end{array}$$

The numbers below the observations give the cumulative number of runs. Let e be the total number of days in which events occur and n be the total in which no events occur. Here, $e = 22$ and $n = 8$. It can be shown [see, e.g., Moore (1978)] that the total number of runs under the null hypothesis that the series is independent is approximately normally distributed with mean $m = 2en/(e + n) + 1$ and variance $\sigma^2 = 2en(2en - e - n)/((e + n)^2(e + n - 1))$. Our mean is 12.7 and variance 4.34, so our observed number of runs, 11, is not at all surprising and the hypothesis of independence is not brought into question by this test.

Given this result, we can ask whether, when events occur, good and bad events are independent. We simply take the 22 observations for which events occurred and classify them as good event days or bad event days on the basis of whether there were more buys or more sells (recall that δ was approximately 1/2). One observation was lost with a tie. We do a runs test on this sequence. We find 11 runs and the approximate normal distribution has mean 10.9 and variance 4.4, so our observation is not at all unusual under the null. Once again, the hypothesis of independence is consistent with our observables.

As a final, somewhat crude check on our independence assumption, we look at the linear time-series structure of the buy, sell, and no-trade series across days separately. These should be independent across days. We first simply examine the autocorrelations and partial autocorrelations (looking for MA or AR structures), then turn to a chi-square test that the first six autocorrelations are zero. Buys exhibit marginally significant autocorrelation at lag 4. We see no economic rationale for taking this seriously.[12] The chi-square value is 8.6; not a surprising value on six degrees of freedom ($P \approx .2$). Sells exhibit no evidence of autocorrelation and the $\chi^2 = 4.4$ ($P \approx .62$). No-trades also appear not to have any linear dependence, with $\chi^2 = 3.98$ ($P \approx .68$) and all autocorrelations and partial autocorrelations inside a two-standard error band. Finally, we consider the time series of daily net trades (the number of buys minus the number of sells). These also fail to exhibit any autocorrelation ($\chi^2 = 0.88$). Regression of net trades on six lagged values of net trades is

$$\underset{}{Net_t} = \underset{(0.265)}{-0.043}\ Net_{t-1} \underset{(0.250)}{- 0.051}\ Net_{t-2} \underset{(0.233)}{- 0.129}\ Net_{t-3} \underset{(0.233)}{- 0.502}\ Net_{t-4} \underset{(0.266)}{+ 0.033}\ Net_{t-5} \underset{(0.245)}{+ 0.068}\ Net_{t-6} \underset{(4.11)}{- 2.85} \tag{16}$$

where Net_t is the number of net buys for day t (standard errors are reported in parentheses). The single significant coefficient occurs at a lag of four. Again, we see no plausible explanation for this result and the group is jointly insignificant. Thus, on balance we see no evidence that the independence assumption is violated in this dataset.

Although there is no significant evidence against independence in our dataset, it is reasonable to expect that other datasets may ex-

[12] In particular, since autocorrelations are binomial, the probability under the null of having one autocorrelation significantly positive is .74.

hibit dependence (this is suggested by external evidence on volatility clustering and on the time-series properties of volume). It is useful to consider how dependence might bias our estimates. We conjecture that the within-day parameters, ε and μ, will not be significantly affected by dependence. As long as days are classified correctly, the within-day parameters are tightly determined by within-day trade. The primary effect of dependence will be on α and δ. With dependence, α and δ must be reinterpreted as marginal means, for example, $\alpha\delta$ is the unconditional probability that a day is a bad-event day. These parameters are likely to be fairly well estimated, as $\alpha\delta$ for example corresponds to the fraction of sell-led high-volume days. However, the asymptotic standard errors, computed under independence, will be incorrect and are likely to be underestimated.

Our model thus survives a number of within-sample specification checks. Although the model is clearly highly stylized, it performs well in terms of internal consistency with the assumptions we have made. What is now useful to consider is how the model can be extended to include more complex features of the trading process. In particular, allowing trade size to enter seems a natural step in that it allows us to determine the role played (or more precisely, the information conveyed) by traders' order strategies. We now turn to this issue.

4. Extensions and Generalizations: The Role of Trade Size

Incorporating trade size (or, for that matter, other features of the trading process) into our analysis requires a model in which a richer set of variables enter into the market-maker's learning problem. This extension is not trivial; if trade size variables are informative to the market maker, then the sufficient statistics underlying the quote process also change, and with it, the price process. The methodology we have developed, however, provides the general framework in which to investigate these broader issues, and as we shall demonstrate, the simple model we have derived thus far can be viewed as a restricted version of this larger, more general model. Thus, a natural testing procedure arises in which we can explicitly test for trade size effects (or those of other relevant variables) by examining the performance of restricted and unrestricted versions of the general model.

In earlier work, Easley and O'Hara (1987), we investigated theoretically the differential information content of large versus small trades. In that work we allowed both informed and uninformed traders to transact in a large trade size and a small trade size. We continue this same characterization in this model, so that the possible trade outcomes are now denoted SB for a small buy, LB for a large buy, SS for a small sale, LS for a large sale, and N for a no-trade outcome.

For an equilibrium to exist in which trade size is not instantly revealing, it must be the case that at least some uninformed traders transact the large trade quantity.[13] To capture this, we denote by φ the probability that an uninformed buyer or seller who chooses to trade, trades the large amount, with $1 - \varphi$ being the corresponding probability that he trades the small amount. Note that for simplicity we have not distinguished this trade probability between uninformed buyers and sellers; this can be done with the addition of two more parameters to the model. Informed traders may also trade large and small quantities. We denote by ω the probability that an informed trader trades the large trade size, and by $1 - \omega$ the probability that he trades the small trade size.[14]

The addition of these two probabilities (φ and ω) to the model dictates a more complex structure to the game being analyzed than previously. Figure 2 depicts this new structure. As is apparent, the model now depends upon the hextuple $\{\alpha, \mu, \delta, \varepsilon, \varphi, \omega\}$. It is this parameter set that we will estimate from our trade data. Second, the outcome in any trading interval is now the quintuple {SB, LB, SS, LS, N}, and it is this quintuple that we use in our estimation. The addition of alternative trade sizes does not affect the role played by the no-trade outcome, but because the informed may split differentially from the uninformed by trade sizes, large trades and small trades may have different information content. We will be able to test for this in our empirical estimation. Third, as was also the case with our μ and ε trade probabilities, the iterative structure of the game means that we can use intraday observations to estimate the additional φ and ω probabilities. One difficulty we note at the outset is that specifying a reasonable cutoff to define the large trades may result in too few trades for meaningful estimation. A solution to this problem is to increase the time period used in the model's estimation. In the estimation that follows, we used data from the 60 trading-day period October 1, 1990, to December 22, 1990.

[13] This is because if only informed traders trade the large quantity then the fact that a trader wants to buy a large amount tells the market maker that he is informed of good news, and conversely if he wants to sell. Since prices in this model are "regret free," this means that the market-maker's quotes for the large quantities are simply the polar values $\underline{V}$ for sells and $\bar{V}$ for buys. Thus, informed traders could not profit from large trades and there would be no large trades. Provided there are some uninformed who will trade the large quantity, this perverse outcome will not occur.

[14] As demonstrated in Easley and O'Hara (1987), the informed traders could choose to trade the large trade size (a separating equilibrium) or they could split across both trade sizes (the pooling equilibrium). The structure we introduce here allows for either outcome to occur, with the exact estimates of the parameters providing evidence of the type of equilibrium prevailing. Our empirical model restricts the parameters ψ and ω to be constant throughout our sample. The equilibrium values of these parameters are likely to vary over time. Thus, our empirical model is best viewed as an approximation to the equilibrium trade process.

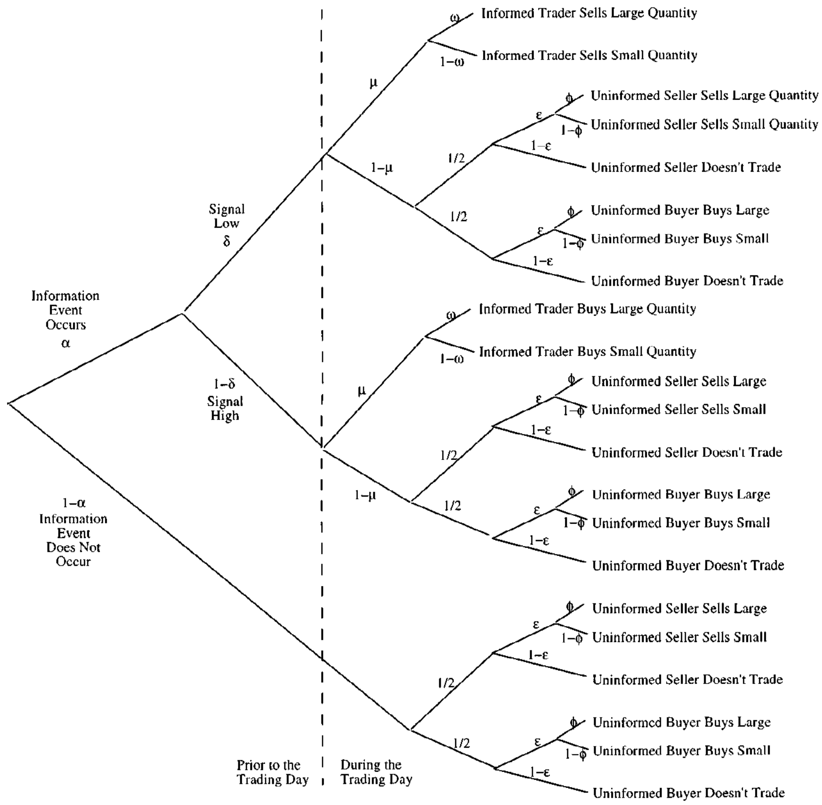

Figure 2
Tree diagram of the trading process.
α is the probability of an information event, δ is the probability of a low signal, μ is the probability that the trade comes from an informed trader, $\frac{1}{2}$ is the probability that an uninformed trader is a seller, ε is the probability that the uninformed trader will actually trade, ω is the probability that an informed trader trades the large amount, and ϕ is the probability that the uninformed trader trades the large amount. Nodes to the left of the dotted line occur only at the beginning of the trading day; nodes to the right are possible at each trading interval.

The model depicted in Figure 2 explicitly allows trade size to enter. Notice that the model derived earlier in this article is essentially just a restricted version of this more general model in which $\varphi = \omega$. Thus, if trade size does not convey differential information, then the restricted version of the model will perform no differently than the unrestricted model. This provides a direct test of the trade size effect. In addition, the estimation of the trade size probabilities φ and ω provides a means to estimate exactly how much trade size matters. If $\omega > \varphi$, then the information content of the large trade size is greater than that of the small trade size. If this relation does not hold, trade size does not provide additional information to the market maker beyond that conveyed by the transaction itself.

Estimating the general model proceeds essentially as before, with the difference being that the sufficient statistics for the quote and price process are now {SB, LB, SS, LS, N}. It is straightforward to derive the updating formulas and to show how the market-maker's quotes evolve from this expanded trade data. For our purposes, what is of more importance is the derivation of the likelihood function, and it is to this task we now turn. As before, this likelihood can be derived by looking at the likelihoods on days of known type, and then using the probabilities of the underlying information structure to weight the components across all types of days. This likelihood function for a single day is proportional to

$$\begin{aligned}
&\Pr\{SB, LB, SS, LS, N|\alpha, \delta, \mu, \varepsilon, \varphi, \omega\}\\
&= \alpha(1-\delta)\left[\mu(1-\omega) + (1-\mu)\tfrac{1}{2}\varepsilon(1-\varphi)\right]^{SB}\\
&[\mu\omega + (1-\mu)\tfrac{1}{2}\varepsilon\varphi]^{LB}\left[(1-\mu)\tfrac{1}{2}\varepsilon(1-\varphi)\right]^{SS}\left[(1-\mu)\tfrac{1}{2}\varepsilon\varphi\right]^{LS}\\
&\times\left[(1-\mu)(1-\varepsilon)\right]^{N}\\
&+\alpha\delta\left[(1-\mu)\tfrac{1}{2}\varepsilon(1-\varphi)\right]^{SB}\left[(1-\mu)\tfrac{1}{2}\varepsilon\varphi\right]^{LB}\\
&\times\left[\mu(1-\omega)+(1-\mu)\tfrac{1}{2}\varepsilon(1-\varphi)\right]^{SS}[\mu\omega+(1-\mu)\tfrac{1}{2}\varepsilon\varphi]^{LS}\left[(1-\mu)(1-\varepsilon)\right]^{N}\\
&+(1-\alpha)\left[\tfrac{1}{2}\varepsilon(1-\varphi)\right]^{SB+SS}[\tfrac{1}{2}\varepsilon\varphi]^{LB+LS}[1-\varepsilon]^{N}.
\end{aligned} \tag{17}$$

The likelihood function for multiple days is then given by

$$\begin{aligned}
&\Pr\{(SB_d, LB_d, SS_d, LS_d, N_d)_{d=1}^{D}|\alpha, \delta, \mu, \varepsilon, \varphi, \omega\}\\
&\quad= \prod_{d=1}^{D}\Pr\{(SB_d, LB_d, SS_d, LS_d, N_d)|\alpha, \delta, \mu, \varepsilon, \varphi, \omega\},
\end{aligned} \tag{18}$$

where $(SB_d, LB_d, SS_d, LS_d, N_d)$ is the outcome on day d.

Table 3 reports our estimated values for this function using data for Ashland Oil for 60 trading days. In this estimation, a crucial issue is the definition of the "large" trade size. We report estimates from a 1000 share cutoff, but similar results obtain using a 200 or 500 share cutoff (cutoff levels above 1000 were not feasible due to the relatively rare occurrence of such large trades). As is apparent, all estimated coefficients are statistically significant.

The estimation reveals an intriguing result: trade size provides no information content beyond that conveyed by the underlying transactions. The likelihood statistics show that the restricted model (in which trade size is not included) is identical to that of the unrestricted model (in which trade size explicitly enters). This equivalence means that the restriction is not binding, thus resulting in no reduction in fit

Table 3
The information content of trade size

Parameters	Unrestricted model	Restricted model
μ	.15 (.03)	.15 (.03)
ε	.28 (.02)	.28 (.02)
α	.59 (.18)	.59 (.19)
δ	.58 (.19)	.57 (.19)
ω	.34 (.04)	.32 (.02)
ϕ	.28 (.09)	
Log-likelihood statistic	−6294	−6294

This table provides estimates from the restricted and unrestricted versions of our model. The restricted version requires the probability of informed and uninformed trading in the large quantity to be the same. The unrestricted version allows these to vary. The parameters are defined as follows: μ is the probability the trade comes from an informed trader, ε is the probability that an uninformed trader actually trades, α is the probability of an information event, δ is the probability of a low signal, ω is the probability the informed trader trades the large amount, and ϕ is the probability the uninformed trader trades the large amount.

when the trade size variable is omitted. The estimated coefficients also bear out this conclusion. Although the coefficient on the probability of informed large trading (.34) exceeds the corresponding probability for uninformed large trading (.28), it is clear from the standard errors that we cannot reject the hypothesis that these two variables are the same. For at least this particular stock, therefore, trade size is not informative to market participants.

One way to interpret this result is that informed traders are trading both large and small quantities, and so trade size is not informative to the market maker. Such an outcome arises in a "pooling equilibrium" [see Easley and O'Hara (1987)] in which some informed traders submit orders for the small quantity and some informed traders submit orders for the large amount.[15] With the informed trading in every trade size, it is the transaction, more than its size, that is informative to market participants. We stress, however, that this outcome need not always occur. In a "separating equilibrium," the preponderance of informed trading in the large quantity imparts information content to order size, and our estimated trade probabilities would be expected to reflect this. Indeed, in other research [see Easley, Kiefer, and O'Hara (1997)] we have found exactly such trade size effects for other stocks.

Our conclusion on information and trade size is model specific. This is unavoidable, as information is not directly observed and its effects must be inferred. The raw empirical result driving our conclusion is that the probability of a large trade, conditional on a trade occur-

[15] In a pooling equilibrium, the informed would still be more likely to trade the large quantity, but this differential need not be large. Our estimates show exactly this, ω (the probability of informed trading large) exceeds φ (the probability of uninformed trading large), but the difference is not statistically significant.

ring, is the same on high- and low-volume days. It is our structural interpretation that high-volume days are information days that leads to our conclusion that trade size is not informative. Thus, while our estimates for Ashland Oil are consistent with the findings of Jones, Kaul and Lipson (1994) regarding the lack of information in trade size, we believe that additional research is needed to determine if this holds true more generally.

5. Trades and Prices

We opened with a regression illustrating the importance of trades in determining prices. We have proposed a model explaining theoretically a mechanism by which trades could determine price movements. It makes sense to assess whether our proposed mechanism is plausible empirically as the means by which trades work on prices. We have seen that the model makes sense as an explanation of trades, the data underlying the estimation. Prices have not been used in computing our estimates—in line with our observation that trades provide a separate, complementary source of information about what goes on in financial markets.[16]

In order to use price data, we need to introduce some additional notation. We let p_{t-1} denote the closing price for day $t-1$. As is standard, prices are assumed to follow a martingale. That is, the expectation of p_t, conditional on information available at the end of day $t-1$ is p_{t-1}. Define $\overline{\Delta V}$ to be the increase in the value of the asset on a good news day; that is, $\overline{\Delta V} = \bar{V}_t - p_{t-1}$.[17] Similarly, the decrease in value on a bad news day is $\underline{\Delta V} = p_{t-1} - \underline{V}_t$. Finally, on a no-event day the closing price should not be expected to change from its previous value. The martingale property then requires that the unconditional expectation of the day t closing price is

$$E[p_t] = \Pr_t(N)p_{t-1} + \Pr_t(G)\bar{V}_t + \Pr_t(B)\underline{V}_t. \qquad (19)$$

Using our definitions of price changes we then have

$$E[p_t] = p_{t-1} + \Pr_t(G)\overline{\Delta V} - \Pr_t(B)\underline{\Delta V}, \qquad (20)$$

suggesting a regression of p_t on p_{t-1}, $\Pr_t(G)$, and $\Pr_t(B)$. The probabilities can be calculated for each day using Equation (7) and anal-

[16] That is, our estimates of the market-maker's beliefs are derived from the Bayesian updating that occurs from watching trades. The Lee–Ready algorithm uses the relation of execution prices relative to quotes to assign whether a trade is a buy or a sell, but the actual prices of trades are irrelevant for our analysis.

[17] Now asset values are indexed by day as we view new information relative to the previous closing price.

ogous equations for good and bad event probabilities together with our estimated parameters.

Doing the regression gives

$$p_1 = \underset{(1.51)}{4.49} + \underset{(0.051)}{0.84 p_{t-1}} + \underset{(0.209)}{0.53 \Pr_t(G)} - \underset{(0.211)}{0.61 \Pr_t(B)} \tag{20}$$

with an adjusted R^2 of 0.92. Thus $\Pr(G)$ and $\Pr(B)$, which are nonlinear functions of buys and sells, have significant explanatory power. Adding buy and sell volumes does not significantly increase the fit [$F(2, 23) = 1.7$, $P = .2$]. Testing the joint hypothesis that the coefficient on lagged price is one, that the constant is zero, the coefficients on the two probabilities are equal in magnitude and opposite in sign (recall our estimated δ, the probability of a bad event, is nearly .5), and that the volume coefficients are zero gives $F(5, 23) = 2.68$ with a P-value of .0476. Though this joint null can be marginally rejected at .05, we nevertheless find this supportive of the model, as prices were not even used in our estimation. The informational role of trades is significant and economically important. When trades are included in the form implied by the model, volumes are insignificant.[18]

6. Conclusions

We have shown in this article how an asymmetric information dynamic model of market-maker behavior can be empirically estimated, and we have demonstrated how the information in trade data can be extracted and analyzed. One contribution of this research is to suggest a new approach for empirical analyses of security price behavior.[19] We model the price-setting problem facing one agent and determine using trade data the underlying parameters of that agent's optimal policies. On the substantive side, we apply our methods to 30 days of trading in a typical common stock, Ashland Oil. Our model has two prior parameters: the market-maker's probability that an information event occurs overnight and the prior probability that an information event is good news, given that it occurs. Interpreting the data through the lens of our model, we find that the market maker thinks information

[18] Our regressors, $\Pr(G)$ and $\Pr(B)$, are functions of estimated coefficients and thus an error in variables problem may arise. We have fit the equation by instrumental variables with exogenous variables p_{t-1}, buy and sell volumes, numbers of buys and sells, and the number of no-trades. The joint test of the significance of volumes gives $F = 1.36$ ($P = .28$). The full joint null has $F = 2.39$, $P = .07$, not indicating rejection.

[19] Related applications of such an approach can be found in labor economics, resource economics, and replacement investment. Our setup differs from the usual labor applications in which information for many individuals and few time periods is coordinated through the dynamic programming and homogeneity assumptions.

events are likely, occurring overnight with .75 probability. She thinks good news and bad news are equiprobable. Our model has two environmental parameters, the fraction of trades that are information based, given that an information event occurs, and the probability that an uninformed trader trades upon entering the market. We find that on days in which information events occur, about 38% of trades are information based. The probability that some uninformed trader will trade in any five-minute interval is found to be about one-third. Using an extension of our model, we show that trade size provides no information content beyond that contained in the underlying transaction.

Our model holds up well against a variety of specification checks. Of course, no one could think that such a simple model is an accurate description of the world, but it does seem to be a satisfactory start as a model of the role of information reflected in trades. In particular, the reduced-form relationship between closing prices and trades, which on its own is strong but difficult to interpret, is closely mimicked by the implications of our information-based model, which gives a natural economic interpretation of the role of trades in determining prices. Clearly, much more work needs to be done on price determination and the role of information, but we feel that our model has made an important start in linking theoretical, information-based models of financial markets with the extensive empirical literature.

References

Bernhardt, D., and E. Hughson, 1993, "Intraday Trade in Dealership Markets," working paper, Queens University.

Clark, P., 1973, "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, 41, 135–155.

Devine, T. J., and N. M. Kiefer, 1991, *Empirical Labor Economics: The Search Approach*, Oxford University Press, New York.

Easley, D., N. Kiefer, and M. O'Hara, 1997, "The Information Content of the Trading Process;" forthcoming in *Journal of Empirical Finance*.

Easley, D., and M. O'Hara, 1987, "Price, Trade Size, and Information in Securities Markets," *Journal of Financial Economics*, 19, 69–90.

Easley, D., and M. O'Hara, 1992, "Time and the Process of Security Price Adjustment," *Journal of Finance*, 47, 577–604.

Foster, D. F., and S. Vishwanathan, 1995, "Can Speculative Trading Explain the Volume-Volatility Relation?," *Journal of Business and Economic Statistics*, 13, 379–396.

Glosten, L., and P. Milgrom, 1985, "Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 13, 71–100.

Jones, C., G. Kaul, and M. Lipson, 1994, "Transactions, Volume, and Volatility," *Review of Financial Studies*, 7, 631–651.

Kyle, A. P., 1985, "Continuous Auctions and Insider Trading," *Econometrica*, 53, 1315–1336.

Lee, C., and M. Ready, 1991, "Inferring Trade Direction from Intraday Data," *Journal of Finance*, 733–746.

Moore, P. G., 1978, "Nonparametric Statistics: Runs," in W. J. Kruskal and J. Tanur (eds.), *International Encyclopedia of Statistics*, Free Press, New York, pp. 655-651.