The Summary Informativeness of Stock Trades: An Econometric Analysis

Joel Hasbrouck
New York University

In a security market with asymmetrically informed participants, trades are signals of private information. In this article, new measures of trade informativeness are proposed based on a decomposition of the variance of changes in the efficient price into trade-correlated and -uncorrelated components. The trade-correlated component has a natural interpretation as an absolute measure of trade informativeness. The ratio of this component to the total variance is a relative measure (i.e., a proportion normalized with respect to the total public information). For a sample of NYSE-listed companies, trades are found to be more informative for small firms in both absolute and relative senses. From an analysis of intraday patterns, it appears that trades are in absolute terms more informative at the beginning of trading, but slightly less informative in relative terms.

This article describes an empirical assessment of trade informativeness in a security market where agents

I am grateful to the editor and Duane Seppi (the referee) for numerous suggestions. I have also benefited from the comments of Kerry Back, Philip Dybvig, Thomas George, Lawrence Glosten, Lawrence Harris, Ravi Jaganathan, Steven Manaster, Maureen O'Hara, Timothy Opler, Kenneth Singleton, and seminar participants at Rutgers University, Ohio State University, University of Iowa, University of Oregon, University of Washington, Washington University, and the Western Finance Association. I am especially indebted to John Campbell for initially bringing the random-walk decomposition techniques to my attention. All errors are my responsibility. Address reprint requests to Joel Hasbrouck, Department of Finance, Leonard N. Stern School of Business, New York University, 90 Trinity Place, New York, NY 10006.

The Review of Financial Studies 1991 Volume 4, number 3, pp. 571–595

trade by executing buy or sell orders against preexisting ask and bid quotes. Numerous studies have modeled transaction and price dynamics in such markets when traders possess varying levels of information.[1] In these models, the underlying (and generally unobservable) characteristics of the informational asymmetries are prevalence, precision, and timeliness of the private signals, and the extent of competition among the informed traders. Agents' beliefs concerning these characteristics determine two observable features of the market: the bid–ask spread and the impact of a trade on the security price. The related empirical literature explores both attributes. McInish and Wood (1989), Chiang and Venkatesh (1988), and Foster and Viswanathan (1990b, 1990c) use the stated bid–ask spread as a summary measure of informational asymmetry. Studies of the price impact of a trade include Glosten and Harris (1988), Stoll (1989), Foster and Viswanathan (1990b, 1990c), and Hasbrouck (1988, 1991).

Both the stated spread and the estimated impact of a trade on price have merit as summary measures of the market's assessment of the informational asymmetry. However, both share certain drawbacks. First, both are conditioned on trades of a certain size. The stated spread is often valid only for small trades. The estimated impact of a trade on price is typically conditioned on the particular size of the trade. Neither measure summarizes the effect of trades of all sizes. A second more subtle limitation is that both the spread and the price impact are absolute measures of the informativeness of an incoming trade. As most useful private information is essentially advance knowledge of public information, it is also of interest to quantify the information contained in trades relative to the total information bearing on the security's value.

In this article, two new summary measures of trade informativeness are proposed and implemented. The approach may be described intuitively as follows. Many microstructure models fit into a paradigm that decomposes actual prices or quotes into a "true" or "efficient" price (generally, the conditional expectation of the end-of-trading security value) and a second "disturbance" component, which impounds various microstructure imperfections. Changes in the efficient price may be further decomposed into a component that is attributable to trades and a component that is not. Analogously, the variance of the efficient price changes may be decomposed into trade-correlated and -uncorrelated components. This variance decomposition leads directly to an absolute measure of trade informativeness

[1] See Bagehot (1971), Copeland and Galai (1983), Glosten and Milgrom (1985), Kyle (1985), Easley and O'Hara (1987), Glosten and Harris (1988), Glosten (1987, 1989), Admati and Pfleiderer (1988), and Foster and Viswanathan (1990a).

(the efficient price variance attributable to trades) and a relative measure (the proportion of efficient price variance attributable to trades).

To estimate these measures in situations where the efficient price is unobservable, the vector autoregression (VAR) model of transactions data discussed in Hasbrouck (1991) is combined with results from the random-walk decomposition literature. The latter work deals with techniques whereby a nonstationary time series may be decomposed into random-walk and stationary components.[2] Here, the decomposition is applied to security price changes, and the random-walk component is equated with the efficient price. This leads to a decomposition of the efficient price variance into a component due to trades and a component that is orthogonal to trades.

This article is organized as follows. In Section 1, the econometric techniques are fully discussed. Since these techniques are primarily descriptive, it is useful to examine them for a specific microstructure model. In Section 2, an analysis of a model that exhibits both inventory control and asymmetric information effects is presented. In Section 3, estimation and, in particular, the bias in cross-sectional analyses are discussed. In the remainder of the article, an empirical analysis of a sample of NYSE issues is presented. In Section 4, the transactions data used in the study are described. For illustration purposes, a detailed analysis of a particular stock is presented in Section 5. In Section 6, a cross-sectional analysis aimed at discerning differences in trade informativeness across firms with different equity capitalizations is presented. In Section 7, intraday patterns in the trade informativeness are examined. A brief summary concludes the article in Section 8.

1. The Variance of the Efficient Price Changes and Its Decomposition

In this analysis, it is assumed that trades are motivated by private information and/or exogenous liquidity needs. The impact of trades on the security price reflects various effects, some of which (such as inventory control) may be transient. The permanent price impact of a trade is due to agents' beliefs about the private information content of the trade. The present approach is a technique for measuring this permanent impact.

Market events (quote revisions and transactions) are defined on a sequence of discrete points indexed by t. It is natural to define this

[2] Random-walk decompositions have been studied primarily in macroeconomic series. See Nelson and Beveridge (1981), Watson (1986), and Campbell and Mankiw (1987). Stock and Watson (1988) provide a good introduction to the subject. Hasbrouck (1990) illustrates other applications of these techniques to short-term security returns.

sequence either as a set of wall-clock times, or alternatively as a set of event times (i.e., a counter of trades and quote revisions). While the distinction is not generally relevant for theoretical models [an exception is Easley and O'Hara (1990)], the stationarity considerations discussed later in this section favor event sequencing.

The price variable is q_t, the midpoint of the prevailing bid and ask quotes. Transactions may result from market orders hitting these quotes, or from intraquote crosses (although most of the theoretical models do not explicitly admit this possibility). The trade variable is the signed trade volume (positive if the market order is a purchase and negative if a sale), denoted x_t. In Section 4, problems that arise when attribution of trade direction is uncertain are discussed. The progression of trades and quote revisions that characterizes both the theoretical models and most actual markets is preserved in the following ordering convention. The prevailing quotes at the beginning of time t are those set in the preceding period (q_{t-1}). A trade is executed (x_t); nontrade public information arrives; and finally the market-maker posts new quotes (q_t). In the absence of any trade, x_t may be taken as zero.

The quote midpoint is the sum of two unobservable components:

$$q_t = m_t + s_t, \tag{1}$$

where m_t is interpreted as the efficient price in the sense of the expected end-of-trading security value conditional on all time-t public information. The second component may be provisionally taken as a residual discrepancy term. Two assumptions impose structure on (1). The first is that the efficient price evolves as a random walk:

$$m_t = m_{t-1} + w_t, \tag{2}$$

where $Ew_t = 0$, $Ew_t^2 = \sigma_w^2$, and $Ew_t w_\tau = 0$ for $\tau \neq t$. The innovations (w_t) reflect updates to the public information set (including, if relevant, the most recent trade). The second assumption is that s_t is a zero-mean nondeterministic stochastic process that is jointly covariance stationary with w_t. Intuitively, s_t is assumed to embody all the transient microstructure imperfections that cause the quote midpoint to deviate from the efficient price (inventory control, price discreteness, etc.). In keeping with this generality, s_t may be serially correlated and (optionally) partially correlated with the w_t history.

The assumption of covariance stationarity implies that $E_t s_{t+k} \rightarrow Es_{t+k} = 0$ as $k \rightarrow \infty$. This reflects the transience of microstructure imperfections. It is reasonable because microstructure imperfections are not ordinarily assumed to cumulate over time, leading to arbitrarily large divergences between market prices and underlying economic values. The stationarity assumption does rule out, on the other hand, price level effects. Among these neglected effects are the apparent inverse relation between price level and volatility [Christie (1982)] and discreteness (since the eighth-dollar price grid is not equally spaced in logarithms).

The assumed joint stationarity of $\{s_t, w_t\}$ also implies homoskedasticity. In view of the well-documented intraday and intraweek patterns in return variances, this assumption requires reinterpretation or modification. The first point is a formal one. Specifically, the assumed stationarity can be viewed as applicable to data sequenced by an index t that preserves the ordering of events, but from which the natural time and date stamps have been suppressed. When the data are viewed thusly, the model is interpreted as an unconditional one that reflects the average behavior of the data over all times and days. This view is formally correct, but it may be unattractive because it ignores information available to both market participants and the econometrician.

A second strategy involves defining the sequencing index in a fashion that is likely to mitigate the nonstationarity. Intraday return variances (per unit time) are elevated at the beginning and end of trading. This pattern also characterizes the frequency (per unit time) of market events (transactions and quote revisions). This suggests that variances per event are likely to be more constant than variances per unit time. In consequence, some relief may result from defining t as an event counter, rather than an index of natural time. The empirical analysis in this article adopts this approach. The third and best approach is to model explicitly the conditional nonstationarity. The simplest technique is to estimate the specification over various intraday intervals, a method that the present article pursues. The more elegant approach of constructing a single comprehensive specification that incorporates natural time effects remains as a task for future research.

A related specification issue is whether the per-share price variables (q_t, m_t, s_t) should be defined in levels or logarithms. Because disturbances are likely to have both multiplicative and additive components (e.g., from price discreteness), neither position is completely satisfactory. As a practical matter, estimates using data samples of the duration considered here (three months) are not highly sensitive to the choice. In comparing statistics across firms with differing share prices, however, proportional measures are somewhat easier to report and discuss. In consequence, the remainder of the analysis assumes that the price variables are specified as logarithms.

Trade informativeness is measured in the following fashion. The market's signal of private information is the current trade innovation, defined as $x_t - E[x_t | \Phi_{t-1}]$, where Φ_{t-1} is the public information set prior to the trade. In most of the theoretical models, this innovation is simply the trade itself. The generalization is useful because it permits serial correlations in the trades of the sort that might be induced, for example, by inventory control, price discreteness, or price-dependent trading strategies. The impact of the trade innovation on the efficient price innovation is $E[w_t|x_t - E[x_t|\Phi_{t-1}]]$. Therefore, two natural summary measures of trade informativeness, both in the absolute and relative to the total public information, are

$$\mathrm{Var}(E[w_t|x_t - E[x_t|\Phi_{t-1}]])$$

and (3)

$$\mathrm{Var}(E[w_t|x_t - E[x_t|\Phi_{t-1}]])/\mathrm{Var}(w_t).$$

The random-walk decomposition (1) on which these measures are based is unobservable. The connection to the observable data follows from the vector autoregressive (VAR) model described in Hasbrouck (1991). Defining the quote revision as $r_t = q_t - q_{t-1}$, this model is

$$\begin{aligned} r_t &= a_1 r_{t-1} + a_2 r_{t-2} + \cdots + b_0 x_t + b_1 x_{t-1} + \cdots + v_{1,t}, \\ x_t &= c_1 r_{t-1} + c_2 r_{t-2} + \cdots + d_1 x_{t-1} + d_2 x_{t-2} + \cdots + v_{2,t}. \end{aligned} \tag{4}$$

In the present discussion, x_t is the signed trade variable described above. More generally, x_t is a column vector of trade attributes, and b_i and d_i are conformable coefficient matrices. The innovations are zero-mean, serially uncorrelated disturbances with $\mathrm{Var}(v_{1,t}) = \sigma_1^2$, $\mathrm{Var}(v_{2,t}) = \Omega$, and $Ev_{1,t}v_{2,t} = 0$. The inclusion of the contemporaneous trade (x_t) in the quote-revision (r_t) specification imposes a recursive structure that reflects the ordering at time t of the trade and quote revision discussed above. It is worth noting that since transaction prices and quantities are determined jointly, the specification would be inappropriate if transaction prices were used in lieu of quote midpoints. Making an assumption of invertibility, the trades and quote revisions may be expressed as a linear function of current and past innovations.[3] The vector moving average (VMA) representation corresponding to (4) is

$$\begin{aligned} r_t &= v_{1,t} + a_1^* v_{1,t-1} + a_2^* v_{1,t-2} + \cdots + b_0^* v_{2,t} + b_1^* v_{2,t-1} + \cdots, \\ x_t &= c_1^* v_{1,t-1} + c_2^* v_{1,t-2} + \cdots + v_{2,t} + d_1^* v_{2,t-1} + d_2^* v_{2,t-2} + \cdots. \end{aligned} \tag{5}$$

Judge et al. (1985, p. 657) describe the transformation.

[3] For an autoregressive model to be invertible, it must be feasible to recover the disturbances as convergent linear combinations of past realizations. In the present application, invertibility may be violated by certain extreme forms of inventory control. Specifically, if the counterparty to all trades is a monopolistic dealer, and trades are observed without error, then over sufficiently long intervals all trades will eventually be reversed. [For inventory to remain (on average) constant, a dealer purchase must be eventually offset by a dealer sale.] This complete and total offset is

From the VMA, it is possible to identify the trade informativeness measures in (3) under the further assumptions that the public information set is the trade and quote history (i.e., $\Phi_t = \{x_t, r_t, x_{t-1}, r_{t-1}, \ldots\}$) and that the expectation operators are linear projections. With this additional structure, the absolute and relative measures in (3) become

$$\mathrm{Var}(E[w_t \mid x_t - E[x_t \mid \Phi_{t-1}]]) = \mathrm{Var}(E^*[w_t \mid v_{2,t}]]) \equiv \sigma^2_{w,x}$$

and

$$\mathrm{Var}(E[w_t \mid x_t - E[x_t \mid \Phi_{t-1}]])/\mathrm{Var}(w_t) = \sigma^2_{w,x}/\sigma^2_w \equiv R^2_w, \qquad (6)$$

where the E^* notation denotes linear projection. The R^2 notation in the relative informativeness measure stems from its interpretation as a coefficient of determination in a regression of w_t on the trade innovation. The following proposition (proved in Appendix A) gives the computational details.

Proposition 1. *For the trade/quote-revision VMA given in (5), the random-walk variance and the contribution of trades to this variance defined in (6) are given by*

$$\sigma^2_w = \left(\sum_{i=0}^{\infty} b_i^*\right)\Omega\left(\sum_{i=0}^{\infty} b_i^{*\prime}\right) + \left(1 + \sum_{i=1}^{\infty} a_i^*\right)^2 \sigma_1^2$$

and (7)

$$\sigma^2_{w,x} = \left(\sum_{i=0}^{\infty} b_i^*\right)\Omega\left(\sum_{i=0}^{\infty} b_i^{*\prime}\right).$$

An intuitive interpretation of Proposition 1 may be gleaned from the VMA representation for quote revisions in (5). Assume that at time $t = 0$ the system is initially in a stable state with all lagged innovations equal to zero. Suppose that we impart to the system an initial trade shock given by $v_{2,0}$. Consider the ultimate effect of this trade on the security price. The initial quote revision will be $r_0 = b_0^* v_{2,0}$; the quote revision one period later will be $r_1 = b_1^* v_{2,0}$; and so on. By cumulating all the future quote revisions, the ultimate effect of the initial trade on the price can be computed as $(\Sigma_{i=0}^{\infty} b_i^*) v_{2,0}$. This is the impulse response function used in Hasbrouck (1991). Computation of its variance directly yields the expression for $\sigma^2_{w,x}$ given in Proposition 1.

This analysis suggests a structure for analyzing trade informativeness in comparative analyses. At the primary level, Σb_i^* is the multi-

incompatible with the recursion assumption implicit in an autoregression in trades. If the data were this perfect, however, one could still conduct the analysis using inventory levels in lieu of trades.

plier used to compute the ultimate impact of a given trade. In Section 7, for example, this quantity is used to assess whether a 1000-share trade conveys more information in the morning than at other times. The size of Ω indicates trade intensity, the magnitude of the trade innovations. Both terms appear in the expression for $\sigma^2_{w,x}$ given in (7), which reflects the comprehensive contribution to efficient price variance by trades of all sizes. Finally, R^2_w is a comprehensive relative measure of trade informativeness.

Aside from data considerations (discussed in Section 4), the limitations of this analysis may be grouped for purposes of discussion into those problems that plague VARs in general and difficulties that arise from the specific interpretation placed on the VAR in the present framework. Chief among the generic difficulties is that VAR estimations have low power in resolving weak effects that persist over intervals of the same order as the sample size or longer. The empirical analyses presented here are based on three months of trading data, and so would not be expected to capture slowly varying monthly and annual components that typically require decades of data to detect. This problem also leads to difficulties in determining lag length.

The problems with the economic interpretation placed on the VAR in the present application stem from the assumed linearity of the specification and the assumption that the relevant public information set is the trade and quote revision history. The linearity may be generalized to allow specifications that are linear in arbitrary predetermined functions of the trades and quote revisions, provided that the transformed variables satisfy the stationary assumptions. (The estimations presented in later sections adopt this practice.) This must be viewed, however, as an empirical approximation to an unknown functional form.

Misspecification of the information set is more troublesome. In the narrowest econometric sense, the measures presented here quantify the explanatory power of trades in a regression in which the dependent variable is the innovation in the random-walk price implicit in the reported quote. If the trades were correlated with an omitted public information variable, the measures would overstate the (incremental) informativeness of trades. Conversely, if the time stamps for transactions were subsequent to the actual time that the transaction was widely reported, the present measures would tend to understate the true informativeness of trades.

The specification of the information set also bears on the stationarity assumption in general, and the stationary-increment random-walk assumption for m_t in particular. While trades and quote revisions may be jointly covariance stationary in an unconditional sense, they are almost certainly nonstationary when conditioned on the true public information set. Considerations such as heteroskedasticity associated with time of day are potentially amenable to modeling. Other nonstationarities may not be so apparent. The effects of these errors on the trade informativeness measures are unknown.

2. Illustration: A Simple Model

To illustrate the analysis, consider the following model suggested by Lawrence Glosten (in personal communication) and discussed in Hasbrouck (1991). Assume the efficient stock price, m_t, follows

$$m_t = m_{t-1} + zv_{2,t} + v_{1,t}, \tag{8}$$

where $v_{1,t}$ and $v_{2,t}$ are mutually and serially uncorrelated disturbance terms with $\mathrm{Var}(v_{1,t}) = \sigma_1^2$ and $\mathrm{Var}(v_{2,t}) = \Omega$. From an economic viewpoint, $v_{1,t}$ captures public nontrade information; $v_{2,t}$ is the innovation in trades; and the z coefficient reflects the private information conveyed by the trade innovation. The quote midpoint, q_t, follows

$$q_t = m_t + a(q_{t-1} - m_{t-1}) + bx_t, \tag{9}$$

where x_t is the signed trade at time t, and a and b are adjustment coefficients with $0 < a < 1$ and $b > 0$. The evolution of trades is

$$x_t = -c(q_{t-1} - m_{t-1}) + v_{2,t}, \tag{10}$$

where $c > 0$ defines a downward sloping demand schedule and $v_{2,t}$ is the innovation (unexpected component) of the trade.

The focus of interest in this article is the variance in $w_t = m_t - m_{t-1}$ that is explained by trades. Since the innovations are uncorrelated,

$$\sigma_w^2 = z^2\Omega + \sigma_1^2, \tag{11}$$

implying $\sigma_{w,x}^2 = z^2\Omega$ and $R_w^2 = \sigma_{w,x}^2/\sigma_w^2$. The computational details of the following analysis are given in Appendix B. The VAR representation is

$$\begin{aligned} r_t &= (z + b)x_t + [zbc - (1 - a)b]x_{t-1} \\ &\quad + a[zbc - (1 - a)b]x_{t-2} + \cdots + v_{1,t}, \\ x_t &= -bcx_{t-1} - abcx_{t-2} - a^2bcx_{t-3} + \cdots + v_{t,2}. \end{aligned} \tag{12}$$

the corresponding VMA is

$$\begin{aligned} r_t &= v_{1,t} + (z + b)v_{2,t} + (-b^2c + (a - 1)b)v_{t-1} + \cdots, \\ x_t &= v_{2,t} - bcv_{2,t-1} + (b^2c^2 - abc)v_{2,t-2} + \cdots. \end{aligned} \tag{13}$$

Appendix B also shows that the sum of the moving average coefficients of $v_{2,t}$ in the r_t equation is equal to z. This implies that a one-unit trade innovation ($v_{2,t} = 1$) leads to a persistent change in the quote midpoint of magnitude z. From Proposition 1, it also implies $\sigma_w^2 = z^2\Omega + \sigma_1^2$ and $\sigma_{w,x}^2 = z^2\Omega$. This verifies that the VMA for the observable data is sufficient to compute the variance decomposition in (11).

3. Statistical Issues

As noted in Hasbrouck (1991), the VAR model in (4) may be estimated consistently under general stationarity assumptions. If r_t and x_t are jointly covariance stationary, then they possess a VMA representation (5). If this VMA representation is also invertible, then the process has a VAR representation as well, which may be estimated consistently using ordinary least squares. The estimated VAR representation may be inverted to get the estimated VMA coefficients, and from these one may compute a variance decomposition and estimates for $\sigma_{w,x}^2$ and R_w^2 by using the VMA estimates in lieu of the true values in Proposition 1. Since all parameter transformations are continuous, these estimates are consistent.

The estimates are not, unfortunately, unbiased in finite samples. The biases depend both on the parameters of the true model and on the sample size.[4] The sample-size dependencies of the biases are of particular concern in the present application because one goal of the exercise is a summary measure that can then be used in comparative and cross-sectional analyses. These applications will almost certainly involve comparisons of estimates between firms with differing numbers of observations (since firms differ greatly in their transaction frequencies).

To correct for bias in the estimates, a bootstrap procedure is employed in this article. The bootstrap technique uses repeated simulation of the estimated specification to approximate the small-sample distribution of the parameter estimates. Efron and Gong (1983) provide a review. In the present application, bootstrapping is only approximately correct. Specification and estimation of the original VAR system in (4) assume only that the disturbances are serially uncorrelated. The bootstrap is formally correct only when the disturbances are serially independent—a stronger assumption. In this application, it is almost impossible to claim independence. Typically, independence will be violated because of discreteness of the variables, such as a quote revision that always lies on multiples of \$.125, or a discrete buy/sell trade indicator variable. Such considerations will generally cause serial dependencies in moments of order greater

[4] Illustrative simulation studies for the model discussed in the last section are available from the author upon request.

than 2. The effects of these higher-order dependencies on the bias estimates are unknown. Further work is under way to address this concern, but at this point the procedure implemented here must be viewed as plausible and probably preferable to ignoring a possibly significant problem.

4. Data

The results reported in this article are based on a sample of transactions data from the Institute for the Study of Security Markets (ISSM). These data comprise a sequenced trade and quote record from the New York and American Stock Exchanges, and the consolidated regional exchanges, over the 62 trading days in the first quarter of 1989, with a one-second time resolution. The sample of firms for the present study was selected in the following manner. For all issues present on the ISSM and CRSP tapes, I computed market capitalizations as of the end of 1988, formed quartile market value subsamples, and sorted each quartile by sequence on the ISSM tape (roughly alphabetical). I applied all analyses to the first 50 firms in each quartile that had at least 500 transactions, average price greater than $5, and less than a 10 percent change in the number of shares outstanding over the sample period (to rule out stock splits and large stock dividends). Only 27 firms satisfied these criteria in the lowest market capitalization quartile.

As a result of stationarity considerations discussed in Section 1, the data are viewed as an ordered sequence of observations in which the time subscript t is incremented each time a transaction or a quote is posted. There are several exceptions to this rule. First, I ignore quotes posted by the regional exchanges, which tend to follow those of the principal exchange closely. Second, if a quote revision occurs within five seconds following a trade, I assign it the same t subscript. Third, trades that occur within five seconds of each other with no intervening quote are cumulated as a single observation. This alleviates the problem of reporting fragmentation. (The current NYSE procedure is to report sales: a 300-share transaction with one seller and three buyers of 100 shares is reported as a single transaction; however, if there were three sellers and one buyer, the report would consist of three transactions.)

Differences in the reporting of quote revisions and transactions lead to further complications. While most quote revisions are immediately keyed into post workstations by the specialists' clerks, most transactions are recorded manually on cards that are then fed into electronic readers. The relative slowness of the latter process often leads to an apparent reversal of reported trades and quotes. Accordingly, I assume that a quote posted within five seconds prior to a trade is a reporting anomaly, and resequence the events to place the quote revision after the trade. While I believe that this procedure generally leads to a more accurate data record, it is admittedly approximate. Using this rule, if a quote revision and a transaction are mistakenly resequenced, part of the quote revision may be erroneously attributed to a trade that actually followed it. Some results are sensitive to this resequencing, and will be so noted in the discussion. [Lee and Ready (1991) provide further discussion of this problem.]

Another feature of the reporting process is the possibility that the quotes may not be reported promptly. This is most likely to occur during periods of rapid trading. If the quotes are systematically reported with a delay (relative to the trades), this can be viewed as a form of misspecification in which the amount of trade information that apparently precedes a quote is richer than the trade information actually available. This will tend to inflate the explanatory power of the trade history.

Determination of the direction of a trade is made solely from the prevailing quote. A transaction price above the prevailing quote midpoint is assumed to be a purchase (positive sign), and vice versa for prices below the midpoint. If the transaction price is exactly at the midpoint, the direction of the trade is indeterminate, and x_t is set to zero. In the overall sample, roughly 78 percent of the trades and 81 percent of the volume could be signed (i.e., were not at the midpoint). To put these figures in perspective, it should be emphasized that these are for transactions aggregated over five-second intervals, and opening trades are excluded from these figures. [The opening procedure (a call-auction variant) differs substantially from the mechanism used at other times. Particularly in high-volume stocks, the opening volume may constitute a substantial fraction of the daily total.] When the trade/quote resequencing was not used, classifications for the trades and volume fell to 73 percent and 68 percent, respectively.

The analyses presented here exclude the overnight return and the opening trade. In computing the autoregressions, lagged trades and quote revisions prior to the first observation of the day are assumed to be zero. The justification for this is the assumption that the process starts up anew at the beginning of each day. This convention is followed to simplify the intraday analyses. Lagging the data across overnight boundaries does not materially alter the conclusions. However, it should be noted that overnight boundaries are likely to be relatively more important for firms with few trades, and inclusion in the sample is subject to cutoffs based on the number of observations (500 for the cross-sectional analysis and 500 per period in the intraday analysis).

It was noted in Section 1 that one source of misspecification was the presumed linearity in the relation between trades and efficient price innovations. While a priori considerations provide no precise characterizations of the exact functional form, the requirement of linearity is probably too restrictive. A useful yet still tractable generalization is a specification that is linear in nonlinear functions of the trade variables. By analogy with the practice of approximating arbitrary functions by polynomials, it is logical to include powers of the transaction volume. More precisely, if x_t denotes the signed volume, the signed trade volume of power k is defined as $x_t^k = \text{sign}(x_t)\,|x_t|^k$. Thus, x_t^0 is an indicator variable that takes on values $\{-1, 0, +1\}$, $x_t^1 = x_t$, and x_t^2 is a quadratic trade variable. With the exception of the illustrative estimation discussed in the next section, all estimations reported in this article are based on a four-variable VAR model of $\{r_t, x_t^0, x_t^1, x_t^2\}$. As a matter of notation, this is tantamount to replacing x_t in (4) by the column vector $[x_t^0, x_t^1, x_t^2]'$.

The VAR model described in (4) involves infinite lags. It is necessary in practice to make a judicious truncation. The present application truncates specifications at five lags (again, excepting the illustration reported in the next section). Based on spot-checks of individual estimates using portmanteau goodness-of-fit statistics [Newbold (1983)] and estimations of more generous specifications, five lags appears adequate. It is worth emphasizing, however, that in a three-month sample of transactions data, these tests have little power in detecting weak monthly, weekly, or daily dependencies in the data. For example, with 62 daily return observations (the number of trading days in the sample), the magnitude of an autocorrelation estimate must be .25 before it exceeds the conventional twice-standard-error bound about zero. Finally, a finite VAR generally possesses a VMA representation of infinite order. Here, the VMA was truncated after 30 lags, beyond which point the implied moving average coefficients were of negligible magnitude.

5. An Illustrative Analysis for One Stock

To illustrate the estimation technique, this section presents full results for Ames Department Stores (ADD). For brevity in reporting the results, the VAR uses only quote revisions and the signed trade indicator variable $\{r_t, x_t^0\}$, and is truncated at lag 3. Based on 3866 observations, the VAR estimates are

$$r_t = \underset{(-8.08)}{-.1333 r_{t-1}} - \underset{(-1.28)}{.0216 r_{t-2}} - \underset{(-.72)}{.0120 r_{t-3}} + \underset{(15.91)}{.000912 x_t^0}$$

$$
\begin{aligned}
&+ \underset{(7.28)}{.000437x^0_{t-1}} + \underset{(.73)}{.000045x^0_{t-2}} + \underset{(1.54)}{.000094x^0_{t-3}} + v_{1,t}, \\
&R^2 = .110, \qquad \hat{\sigma}^2_1 = .00000556; \qquad\qquad (14) \\
&x^0_t = \underset{(-14.12)}{-63.73r_{t-1}} - \underset{(-6.77)}{31.79r_{t-2}} - \underset{(-1.74)}{8.107r_{t-3}} \\
&+ \underset{(10.36)}{.172x^0_{t-1}} + \underset{(8.00)}{.135x^0_{t-2}} + \underset{(5.33)}{.091x^0_{t-3}} + v_{2,t}, \\
&R^2 = .090, \qquad \hat{\sigma}^2_2 = .438.
\end{aligned}
$$

Asymptotic t-statistics are given in parentheses. This estimation displays several features typical of trade/quote revision VARs. In the quote-revision equation, the effects of current and lagged trades are positive. Positive lagged trade effects are inconsistent with liquidity-related "overshooting" and inventory-control mechanisms. As a matter of methodology, the lagged effects may be due in part to my practice of considering the trade and quote revision concurrent only if occurring within five seconds. Therefore, a quote revision that occurs later will appear to lead the trade. Another consideration is price discreteness. Since the quotes cannot continuously adjust, the quote-setter may wait until a string of trades of the same sign have occurred before making an adjustment. It is important to note that by construction the implicit efficient price does not depend on the lagged trades. Hasbrouck (1988) and Madhavan and Smidt (1990) provide further evidence concerning the absence of inventory control effects on prices. Other characteristics of the estimations are negative first-order autocorrelation in the r_t, positive autocorrelation in trades, and causality running from lagged r_t to trades (cf. the significant r_t coefficients in the x^0_t equation). These features are discussed more fully in Hasbrouck (1991).

The corresponding VMA representation (also truncated at lag 3) is

$$
\begin{aligned}
r_t &= v_{1,t} - .191v_{1,t-1} - .052v_{1,t-2} - .029v_{1,t-3} \\
&\quad + .00091v_{2,t} + .00042v_{2,t-1} + .00011v_{2,t-2} + .00020v_{2,t-3}, \\
x^0_t &= -63.7v_{1,t-1} - 30.5v_{1,t-2} - 12.6v_{1,t-3} \qquad (15) \\
&\quad + v_{2,t} + .114v_{2,t-1} + .099v_{2,t-2} + .096v_{2,t-3}.
\end{aligned}
$$

From the sum of the $v_{2,t}$ coefficients in the VMA equation for r_t, a buy order causes an ultimate log price charge of .0016. The trade innovation variance is $\Omega = .438$ (per event). This may be approximately restated on a per-hour basis as .438 × (3866 observations)/(62 days × 6.5 hours/day) = 4.20 per trading hour. That is, if one were to cumulate buy and sell order innovations (counting +1 for a buy and

-1 for a sale) in a typical trading hour, the variance of this sum would be 4.20.

From Proposition 1, the random-walk variance decomposition is $\sigma^2_{w,x} = (.0016)^2(.438) = 1.17 \times 10^{-6}$, and $\sigma^2_w = 4.12 \times 10^{-6}$. From these values, $R^2_w = .284$. In other words, roughly 28 percent of the variance in the random-walk component of the stock price is attributable to trades. Note that this is over twice as large as the simple R^2 reported for the original r_t regression. Placing the variance measures on a per-trading-hour basis (as described above) gives $\sigma^2_{w,x}$/hour $= 1.12 \times 10^{-5}$ and σ^2_w/hour $= 3.95 \times 10^{-5}$.

Based on a bootstrap of 200 replications, the estimated bias in the $\sigma^2_{w,x}$ estimate was found to be 3.5×10^{-8} and that of the R^2_w estimate was found to be .005. In other words, both estimates are biased upward by about 2%. The bootstrap standard errors of these estimates were found to be 1.5×10^{-7} and .030, respectively.

As a final point, different VAR specifications lead to different estimates. The VAR model used for all estimations in the rest of the article ($\{r_t, x^0_t, x^1_t, x^2_t\}$, through five lags) yield $\sigma^2_{w,x} = 1.69 \times 10^{-6}$, $\sigma^2_w = 4.07 \times 10^{-6}$, $R^2_w = .415$. Note that while the total variance of the random walk component (σ^2_w) changes only slightly, the more comprehensive model attributes substantially more of the efficient price variance to trades.

6. Cross-sectional Analysis by Market Capitalization

For each of the 177 stocks in the sample, the four-variable VAR model described in Section 4 was estimated. Table 1 shows summary statistics for the total sample and for the market-value subsamples. As described in the previous section, all variances are restated as hourly standard deviations. Pairwise Spearman (rank-order) correlation coefficients are given in Table 2. For convenience in presentation, only aggregate summary statistics are reported. The statistical tests, which assume that estimates obtained for different firms are independent, may overstate the statistical significance in the presence of common influences.

Across all firms, the average proportional spread is 1.521 percent of the stock price. (For each firm, this variable is computed as a time-weighted average of all intraday spreads.) In agreement with the finding of numerous earlier studies, it declines markedly with increasing market value. (From Table 2, the rank-order correlation between spread and equity capitalization is $-.861$.) The spread decline is sometimes viewed as prima facie evidence that informational asym-

Table 1
Cross-sectional summary statistics for 177 NYSE issues, first quarter of 1989

	Total sample	Market value subsamples: 1 (lowest)	2	3	4 (highest)
No. of firms	177.	27.	50.	50.	50.
Equ. cap. ($MM)	1136.	27.	93.	413.	3502.
	[2266.]	[7.]	[30.]	[173.]	[3221.]
No. of obs.	4434.	1338.	1889.	3533.	9552.
	[4868.]	[611.]	[1157.]	[2785.]	[6031.]
Impulse (×100)	.255	.631	.324	.162	.078
	[.218]	[.200]	[.139]	[.065]	[.044]
σ_x/hour	144.247	26.804	53.926	136.554	305.679
	[161.737]	[16.978]	[42.640]	[124.926]	[183.859]
σ_w/hour (×100)	.721	1.189	.796	.621	.493
	[.397]	[.602]	[.298]	[.274]	[.146]
$\sigma_{w,x}$/hour (×100)	.412	.727	.470	.345	.275
	[.226]	[.244]	[.207]	[.144]	[.084]
R^2_w	.343	.426	.348	.320	.318
	[.107]	[.131]	[.099]	[.101]	[.084]
Spread (×100)	1.521	2.630	1.982	1.281	.701
	[.863]	[.771]	[.588]	[.538]	[.295]

Values are averages for the total sample of 177 firms and subsamples based on equity-capitilization quartiles. Standard deviations are given in brackets. Equity capitalizations are as of the end of 1988. All other figures refer to the 62 trading days in the first quarter of 1989. The spread is the time-weighted average proportional spread. Impulse, σ_x/hour, σ_w/hour, $\sigma_{w,x}$/hour, and R^2_w are based on the quadratic VAR model described in Section 6. Impulse is the implied impact on the log quote-midpoint of a 1000-share buy order; σ_x/hour is the standard deviation of the innovation in signed trade volume x_t (measured in 100-share lots), a measure of trading intensity; σ_w/hour is the standard deviation of the change in the efficient price; $\sigma_{w,x}$/hour is the square root of the trade-correlated component of the efficient price-change variance; and R^2_w is the proportion of variance in efficient price changes that is explained by trades.

metries are worse—and hence trading is more informative—for low-value firms.

As noted in the introductory section, however, such an interpretation suffers from numerous shortcomings. Even ignoring inventory costs, clearing costs, and dealer monopoly power, the adverse selection component of the spread depends positively on the public information intensity, the total return variance. From Table 1, it is clear the random-walk standard deviation (σ_w/hour) also declines monotonically: small firms have larger return variances. The spread pattern does not by itself, therefore, uniformly imply more extreme informational asymmetries for smaller firms.

The present analysis affords a more detailed dissection of the trade informativeness. Consider first the persistent impact of a trade on the security price. When there is a single trade variable (as in the preceding section), the sum of its moving average coefficients is a convenient scalar measure of this persistent impact. The present specification involves three trade variables, however, and so this sum is a vector. One convenient way of obtaining a single number is to com-

Table 2
Bivariate rank-order (Spearman) correlations

	Impulse ($\times$ 100)	σ_x/hour	σ_w/hour ($\times$ 100)	$\sigma_{w,x}$/hour ($\times$ 100)	R^2_w	Spread ($\times$ 100)
Equ. cap. ($MM)	−.878*	.779*	−.630*	−.620*	−.299*	−.861*
Impulse ($\times$ 100)		−.726*	.718*	.756*	.459*	.796*
σ_x/hour			−.267*	−.271*	−.252*	−.550*
σ_w/hour ($\times$ 100)				.951*	.263*	.734*
$\sigma_{w,x}$/hour ($\times$ 100)					.520*	.671*
R^2_w						.136

Equity capitalizations are as of the end of 1988. All other figures refer to the 62 trading days in the first quarter of 1989. The spread is the time-weighted average proportional spread. Impulse, σ_x/hour, σ_w/hour, $\sigma_{w,x}$/hour, and R^2_w are based on the quadratic VAR model described in Section 6. Impulse is the implied impact on the log quote-midpoint of the 1000-share buy order; σ_x/hour is the standard deviation of the innovation in signed trade volume, a measure of trading intensity; σ_w/hour is the standard deviation of the change in the efficient price; $\sigma_{w,x}$/hour is the square root of the trade-correlated component of the efficient price-change variance; and R^2_w is the proportion of variance in efficient price changes that is explained by trades.

* Statistically significant at the .01 level.

pute the persistent impact of a given trade innovation. The particular trade selected as being well within customary order sizes for all firms in the sample was a 1000 share buy order (10 100-share round lots; i.e., $x^0_t = 1$, $x^1_t = 10$, and $x^2_t = 100$ at $t = 0$). The persistent impact of this trade on the (log) quote midpoint is given in Table 1 as Impulse. The sample average value implies that a 1000 share buy order moves the quote midpoint by about .255 percent. This declines monotonically across market-value subsamples, indicating that same-sized trades convey more information for smaller firms, in conformance with Hasbrouck (1991).

The problem with simply taking the impulse figures as direct summary measures of trade informativeness is that larger firms tend to have more trading activity. Table 1 reports as σ_x/hour the standard deviation of the innovation in the x^1_t equation. (The units of x^1_t are 100-share round lots.) The subsample averages are indeed much higher for the high-capitalization subsamples.

The $\sigma_{w,x}$/hour estimates, summarizing the absolute contribution of all trading activity to the efficient price movements, provide additional insights. In Table 1, the average values of $\sigma_{w,x}$/hour are larger for small firms. For small firms, therefore, the larger impact of a same-sized trade on price is only partially offset by the lower level of trading activity in these firms. More relevant evidence is given by the R^2_w figures, which are also higher for smaller firms. These results imply that in both absolute and relative senses, trades are more informative for small firms.

The rank-order correlations in Table 2 provide evidence for the statistical significance of the effects cited above. Of additional interest

Table 3
Bias-corrected summary statistics

	Total sample	Market value subsamples 1 (lowest)	2	3	4 (highest)
Impulse (×100)	.254	.613	.324	.164	.079
	[.212]	[.189]	[.141]	[.068]	[.047]
σ_x/hour	144.872	27.530	53.792	137.560	306.629
	[163.681]	[17.903]	[41.082]	[129.630]	[186.449]
σ_w/hour (×100)	.668	1.003	.783	.547	.491
	[.533]	[.918]	[.301]	[.581]	[.145]
$\sigma_{w,x}$/hour (×100)	.356	.494	.448	.273	.272
	[.390]	[.616]	[.228]	[.501]	[.086]
R^2_w	.330	.395	.334	.311	.311
	[.116]	[.139]	[.106]	[.114]	[.085]

Values are averages for the total sample of 177 firms and subsamples based on equity-capitalization quartiles. Standard deviations are given in brackets. Values correspond to those in Table 1, except that values in this table are bias-corrected using the bootstrap procedure discussed in Section 3. All values refer to the 62 trading days in the first quarter of 1989. The spread is the time-weighted average proportional spread. Impulse, σ_x/hour, σ_w/hour, $\sigma_{w,x}$/hour, and R^2_w are based on the quadratic VAR model described in Section 6. Impulse is the implied impact on the log quote-midpoint of a 1000-share buy order; σ_x/hour is the standard deviation of the innovation in signed trade volume, a measure of trading intensity; σ_w/hour is the standard deviation of the change in the efficient price; $\sigma_{w,x}$/hour is the square root of the trade-correlated component of the efficient price-change variance; and R^2_w is the proportion of variance in efficient price changes that is explained by trades.

are the correlations between the spread and other variables. These correlations are generally positive, but far from perfect. In particular, the correlation between the spread and R^2_w is only .136. As an empirical matter, therefore, the two quantities appear to be measuring different things and cannot be used interchangeably.

As noted in Section 3, cross-sectional analyses are subject to biases related to differing number of observations. To investigate the possibility that this bias is accounting for the cross-sectional findings, bias-corrected statistics were constructed using the bootstrap procedure discussed in Section 3. The average values of these statistics are presented in Table 3.

By comparing the values in Table 3 with their counterparts in Table 1, it is clear that the inverse relation between capitalization and R^2_w is partially due to bias in the estimates. Taking the total sample estimates for R^2_w, the estimated bias is $.343 - .330 = .013$. Furthermore, the estimated bias is higher for the low-capitalization issues: for subsamples 1–4, the average estimated biases in R^2_w are .031, .014, .009, and .007, respectively. Nevertheless, the patterns of the bias-corrected estimates across market-value subsamples are similar to those reported for the uncorrected statistics. The correlations for the bias-corrected statistics (not reported) are also close to those reported in Table 2, although slightly weaker. Since the bias correction itself introduces additional noise, this attenuation may simply reflect the increased variance of these estimates.

It was noted in the data description section that nearly contemporaneous trades and quotes were resequenced to place the trade before the quote. If this resequencing is not performed, the implied R_w^2's are much lower (the average over the total sample is .15), and although the lowest capitalization subsample still has the highest value of R_w^2, the correlation between R_w^2 and market value is not significantly different from zero. The capitalization correlations with Impulse, σ_x/hour, and $\sigma_{w,x}$/hour maintain, however, the same signs and statistical significance as with the resequenced data.

7. Intraday Patterns

Intraday patterns in various market characteristics have attracted significant interest. Transactions volume and return variances have typically been found to exhibit "U-shaped" patterns, with concentrations at the beginning and end of the trading day. These findings are consistent with the theoretical predictions of Admati and Pfleiderer (1988), whose analysis takes into account trading concentration due to the strategic behavior of informed and certain uninformed traders.

The Admati and Pfleiderer model also predicts that when volume is high, the market impact of a trade is lower (their hypothesis 1). In an analysis of transaction prices and (unsigned) volume, however, Foster and Viswanathan (1990b) find larger trade impacts at the beginning and (to a lesser extent) at the end of the day. Their econometric model assumes serial independence of trades, no feedback from prices to trade sizes, and a market impact that is linear in current trade size. The techniques presented here may be used to address the market impact question in a more general econometric framework.

The present intraday analysis divides the trading data into the first half-hour of trading (9:30:00 A.M. to 10:00:00 A.M.), the middle day (10:00:01 A.M. to 3:30:00 P.M.), and the last half-hour of trading (3:30:01 P.M. to 4:00:00 P.M.). To ensure sufficient numbers of observations, the analysis was restricted to those firms from the original 177-firm sample that possessed at least 500 observations in each time period. This left 46 firms in the sample (of which 40 were from the highest market-value subsample, and six were from the second-highest market-value subsample).

To establish a connection to earlier findings in this area, consider first the hourly standard deviations of the efficient price changes (σ_w/hour). These are indeed higher in the beginning and ending periods, confirming the results found for ordinary returns. The pattern is statistically significant for the beginning period, but not for the ending period.[5] The absence of an end-of-day upturn is consistent with the findings of Foster and Viswanathan for large-capitalization issues. As summarized by the σ_x figures, trading volumes are also slightly higher in the beginning and ending half-hours (not statistically significant). The absence of a pronounced opening effect may seem surprising. It should be recalled, however, that these figures ignore opening trades, which account for a substantial portion of the trading volume in large firms.

Intraday patterns in the spread have been examined by McInish and Wood (1989), who find evidence for larger spreads at the beginning, but not the end, of trading. The average spreads reported for the present analysis in Table 4 confirm this. Spread behavior is closely related to transaction returns. Since a large component of intraday transaction return variance is due to the spread, higher beginning-of-day spreads would lead to higher beginning-of-day return variances, even holding constant the variance of returns in efficient prices. The estimates of σ_w/hour constructed using the present analysis should be less affected by this spread-related bias and therefore constitute stronger evidence of intraday return patterns.

Consider next the trade informativeness measures discussed in this article. As in the preceding section, the Impulse statistic measures the persistent log change in the quote midpoint due to a 1000-share buy order. This is (statistically significantly) higher if the order occurs in the first half-hour of the day. Of course, since both Impulse and σ_x/hour are higher in the first half-hour, it is not surprising that the absolute measure of trade informativeness $\sigma_{w,x}$ is also higher in this period. In an absolute sense, therefore, trades move prices more in the first half-hour of trading than in the rest of the day. The proportional measure R_w^2 is actually lower in the first half-hour, however. (The statistical significance of this is marginal: 12 out of 46 by the sign test.) Another way of stating this finding is that although the information revealed by trades ($\sigma_{w,x}$) is higher in early trading, this is more than offset by a higher intensity of all other public information (which may include trading data on other securities).

The findings for Impulse and $\sigma_{w,x}$ support the Foster and Viswanathan findings of a higher market impact (higher trade informativeness) at the beginning of the trading day. While these findings contradict the predictions of the Admati and Pfleiderer model, several possible sources of the discrepancy deserve mention. Trading in their model proceeds as a sequence of call markets, rather than the dealer

[5] Statistical significance in this section is assessed using sign tests. In the case of σ_w/hour, for example, an indicator variable for a given firm is set to +1 if the value of σ_w/hour for the morning is above the value for midday and to 0 if below. To reject the null hypothesis of equal values for all 46 firms at the .01 level, the number of +1's or the number of 0's needs to be below 13.

Table 4
Summary intraday statistics

	9:30 A.M.–10:00 A.M.	10:00 A.M.–3:30 P.M.	3:30 P.M.–4:00 P.M.
No. obs.	1202.	8781.	1129.
	[514.]	[4235.]	[495.]
Impulse × 100	.112*	.074	.065
	[.057]	[.037]	[.037]
σ_x/hour	335.952	314.963	309.547
	[218.024]	[193.009]	[189.580]
σ_w/hour (×100)	1.029*	.522	.605*
	[.367]	[.233]	[.274]
$\sigma_{w,x}$/hour (×100)	.569*	.312	.350
	[.254]	[.129]	[.161]
R_w^2	.318	.370	.347
	[.128]	[.085]	[.108]
Spread (×100)	.853*	.773	.778
	[.488]	[.474]	[.471]

Values are averages for the total sample of 46 firms. All values refer to the 62 trading days in the first quarter of 1989 during the indicated times. The spread is the time-weighted average spread. Impulse, σ_x/hour, σ_w/hour, $\sigma_{w,x}$/hour, and R_w^2 are based on the quadratic VAR model described in Section 6. Impulse is the implied impact on the log quote-midpoint of a 1000-share buy order; σ_x/hour is the standard deviation of the innovation in signed trade volume, a measure of trading intensity; σ_w/hour is the standard deviation of the change in the efficient price; $\sigma_{w,x}$/hour is the square root of the trade-correlated component of the efficient price-change variance; and R_w^2 is the proportion of variance in efficient price changes that is explained by trades.

* The associated morning or afternoon value differs from the corresponding midday value using the sign test discussed in Section 7 at the .01 level.

market assumed here. Also, the point at which the market does in fact function as a call (the open) is excluded from the analysis.

8. Summary

This article has proposed measuring trade informativeness in a security market by decomposing the efficient price variance into trade-correlated and -uncorrelated components. The trade-correlated component ($\sigma_{w,x}^2$) reflects the persistent impact of trades on the efficient price, an impact that by implication arises from the market's perception of the private information in the trade. This decomposition also leads to a relative measure of the relative trade informativeness (R_w^2), which is the coefficient of determination in a regression of the changes in the efficient price on trades. The quantities are computed from a vector autoregressive model of trades and quote revisions, in which the efficient price is the implicit random-walk component of the quote.

Applied to a sample of NYSE companies, two findings are of particular importance. First, in both absolute and relative senses, trade informativeness appears to be more large for firms with lower market capitalizations. Second, in an absolute sense, trades are more informative in the beginning of the trading day than at other times. Relative to all public information, however, trades are slightly less informative at the beginning of the trading day.

By way of proposing directions for further research, it is found that on average, approximately 34 percent of the variance in efficient prices is explained by trades. Since the prevalence and precision of the private information that presumably gives rise to this figure are not known, interpretation of this magnitude is an open issue. If this proportion is not the result of a sample or methodological artifact, it may imply the presence of a very sizable quantity of asymmetric information. The nature, sources, and production costs of this information stand as worthwhile objects of further study.

Appendix A: Proof of Proposition 1

The proof is a modification of that found in Watson (1986). The vector moving average (VMA) representation for the trade/quote revision system is given by (5) in the text. The r_t equation may be written more compactly as

$$r_t = a^*(L)v_{1,t} + b^*(L)v_{2,t}, \tag{A1}$$

where a^* and b^* are the lag polynomial operators $[a^*(L) = a_0^* + a_1^*L + a_2^*L^2 + \cdots, b^*(L) = b_0^* + b_1^*L + b_2^*L^2 + \cdots]$ and, in conformance with (5), a_0^* is normalized to unity. Alternatively from (1),

$$r_t = (1 - L)q_t = (1 - L)m_t + (1 - L)s_t = w_t + (1 - L)s_t. \tag{A2}$$

The autocovariance generating function for a time series $\{x_t\}$ is defined as $g_x(z) = \Sigma c_x(j)z^j$, where $c_x(j) = Ex_tx_{t-j}$. [See Sargent (1979).] Similarly, the cross-covariance generating function for two series $\{x_t\}$ and $\{y_t\}$ is defined as $g_{xy}(z) = \Sigma c_{xy}z^j$, where $c_{xy}(j) = Ex_ty_{t-j}$. From (A1) the autocovariance generating function for r_t may be written

$$\begin{aligned} g_r(z) &= E[a^*(z)v_{1,t} + b^*(z)v_{2,t}][a^*(z^{-1})v_{1,t} + b^*(z^{-1})v_{2,t}] \\ &= a^*(z)a^*(z^{-1})\sigma_1^2 + b^*(z)\Omega b^*(z^{-1})'. \end{aligned} \tag{A3}$$

Alternatively from (A2) this may be written as

$$\begin{aligned} g_r(z) = g_w(z) &+ (1 - z^{-1})g_{ws}(z) + (1 - z)g_{sw}(z) \\ &+ (1 - z)(1 - z^{-1})g_s(z). \end{aligned} \tag{A4}$$

Equating (A3) and (A4), and letting $z = 1$, shows that

$$g_r(1) = g_w(1) = \sigma_w^2 = [a^*(1)]^2\sigma_1^2 + [b^*(1)]\Omega[b^*(1)]'. \tag{A5}$$

Since the lag polynomials evaluated at unity are equal to their respective coefficient sums, this is equivalent to Proposition 1.

Appendix B: Computational Details for the Model in Section 3

To eliminate the unobservable variable, m_t (the efficient price), it is useful to first work out the innovations representation for $(q_t - m_t)$:

$$\begin{aligned}(q_t - m_t) &= aL(q_t - m_t) + bx_t \\ &= aL(q_t - m_t) + b[-cL(q_t - m_t) + v_{1,t}] \\ &= (a - bc)L(q_t - m_t) + bv_{2,t},\end{aligned} \tag{B1}$$

or, by rearranging, $(q_t - m_t) = [b/(1 - (a - bc)L)]v_{2,t}$. An expansion about $L = 0$ gives the innovations (moving average) representation for $(q_t - m_t)$ in terms of current and lagged values of $v_{2,t}$. The moving average representation for x_t is obtained as an expansion of the last term in

$$\begin{aligned}x_t &= -cL(q_t - m_t) + v_{2,t} = \frac{-bcL}{1 - (a - bc)L}v_{2,t} + v_{2,t} \\ &= \left(\frac{1 - aL}{1 - (a - bc)L}\right)v_{2,t}.\end{aligned} \tag{B2}$$

An expansion of this gives the x_t equation from Equations (13). If $a < 1$, the above may be inverted to obtain the autoregressive representation

$$x_t = \left(\frac{-bcL}{1 - aL}\right)x_t + v_{2,t}. \tag{B3}$$

An expansion of this gives the x_t equation from Equations (12).

Turning now to the quote process, Equation (9) may be written as

$$(1 - aL)q_t = (1 - aL)m_t + bx_t, \tag{B4}$$

which may be rearranged to give $q_t = m_t + (b/(1 - aL))x_t$. Taking the first difference of this to compute the quote revision gives

$$r_t = (1 - L)q_t = (1 - L)m_t + \frac{b(1 - L)}{1 - aL}x_t. \tag{B5}$$

Substituting in for $(1 - L)m_t$ and x_t gives

$$\begin{aligned}r_t &= zv_{2,t} + v_{1,t} + \left[\frac{b(1 - L)}{1 - aL}\right]\left[\frac{1 - aL}{1 - (a - bc)L}v_{2,t}\right] \\ &= \left[\frac{z(1 - (a - bc)L) + b(1 - L)}{1 - (a - bc)L}\right]v_{2,t} + v_{1,t}.\end{aligned} \tag{B6}$$

This is the moving average representation for quote revisions, and an expansion gives the r_t equation from Equations (13). To obtain the autoregressive form, substitute in from (B2) for $v_{2,t}$:

$$r_t = \left[\frac{z(1-(a-bc)L)+b(1-L)}{1-(a-bc)L}\right]\left[\frac{1-aL}{1-(a-bc)L}\right]^{-1} x_t + v_{1,t}$$

$$= \frac{z(1-(a-bc)L)+b(1-L)}{1-aL} x_t + v_{1,t}. \tag{B7}$$

This may be expanded to give the r_t equation from Equations (12). Finally, to get the sum of the $v_{2,t}$ coefficients in the moving average representation for r_t, let $L = 1$ in (B6), yielding z.

References

Admati, A., and P. Pfleiderer, 1988, "A Theory of Intraday Patterns: Volume and Price Variability," *Review of Financial Studies,* 1, 3–40.

Bagehot, W. (pseudonym), 1971, "The Only Game in Town," *Financial Analysts Journal,* 8, 31–53.

Campbell, J. Y., and N. G. Mankiw, 1987, "Permanent and Transitory Components in Macroeconomic Fluctuations," *American Economic Review,* 77, 111–117.

Chiang, R., and P. C. Venkatesh, 1988, "Insider Holdings and Perceptions of Information Asymmetry: A Note," *Journal of Finance,* 43, 1041–1048.

Christie, A., 1982, "The Stochastic Behavior of Common Stock Variances," *Journal of Financial Economics,* 10, 407–432.

Copeland, T., and D. Galai, 1983, "Information Effects on the Bid-Ask Spread," *Journal of Finance,* 38, 1457–1469.

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, 1990, "The Process of Price Adjustment in Security Markets," working paper, Johnson Graduate School of Management, Cornell University.

Efron, B., and G. Gong, 1983, "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation," *American Statistician,* 37, 36–48.

Foster, D. F., and S. Viswanathan, 1990a, "A Theory of the Interday Variations in Volume, Variance, and Trading Costs in Securities Markets," *Review of Financial Studies,* 3, 593–624.

Foster, D. F., and S. Viswanathan, 1990b, "Variations in Volumes, Variances and Trading Costs," working paper, Futures and Options Research Center, Duke University.

Foster, D. F., and S. Viswanathan, 1990c, "Trading Costs of Target Firms and Corporate Takeovers," working paper, Futures and Options Research Center, Duke University.

Glosten, L. R., 1987, "Components of the Bid-Ask Spread and the Statistical Properties of Transaction Prices," *Journal of Finance,* 42, 1293–1307.

Glosten, L. R., 1989, "Insider Trading, Liquidity and the Role of the Monopolistic Specialist," *Journal of Business,* 62, 211–235.

Glosten, L. R., and L. E. Harris, 1988, "Estimating the Components of the Bid/Ask Spread," *Journal of Financial Economics,* 21, 123–142.

Glosten, L. R., and P. R. Milgrom, 1985, "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics,* 14, 71–100.

Hasbrouck, J., 1988, "Trades, Quotes, Inventories and Information," *Journal of Financial Economics,* 22, 229–252.

Hasbrouck, J., 1990, "Assessing the Quality of a Security Market: A New Approach to Transaction Cost Measurement," working paper, Department of Finance, Stern School, New York University.

Hasbrouck, J., 1991, "Measuring the Information Content of Stock Trades," *Journal of Finance,* 46, 179–207.

Judge, G. G., et al., 1985, *The Theory and Practice of Econometrics* (2d ed.), John Wiley & Sons, New York.

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

Lee, C. M., and M. J. Ready, 1991, "Inferring Trade Direction from Intraday Data," *Journal of Finance,* 46, 733–746.

Madhavan, A., and S. Smidt, 1990, "A Bayesian Model of Intraday Specialist Pricing," working paper, Wharton Department of Finance; forthcoming in *Journal of Financial Economics.*

McInish, T., and R. Wood, 1989, "An Analysis of Intraday Patterns in Bid/Ask Spreads for NYSE Stocks," working paper, Fogelman College of Business and Economics, Memphis State University.

Nelson, C., and S. Beveridge, 1981, "A New Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to the Measurements of the 'Business Cycle,'" *Journal of Monetary Economics,* 10, 139–162.

Newbold, P., 1983, "Model Checking in Time Series Analysis," in Arnold Zellner (ed.), *Applied Time Series Analysis of Economic Data,* U.S. Department of Commerce, Bureau of the Census, pp. 133–149.

Sargent, T. J., 1979, *Macroeconomic Theory,* Academic Press, New York.

Stock, J. H., and M. W. Watson, 1988, "Variable Trends in Economic Time Series," *Journal of Economic Perspectives,* 2, 147–174.

Stoll, H. R., 1989, "Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests," *Journal of Finance,* 44, 115–134.

Watson, M. W., 1986, "Univariate Detrending Methods with Stochastic Trends," *Journal of Monetary Economics,* 18, 49–75.