Journal of Financial Markets 1 (1998) 175–201

Journal of
FINANCIAL
MARKETS

# Financial analysts and information-based trade

David Easley[a], Maureen O'Hara[b,*], Joseph Paperman[c]

[a] *Department of Economics, Cornell University, 517 Malott Hall, Ithaca, NY 14853, USA*
[b] *Johnson Graduate School of Management, Cornell University, 517 Malott Hall, Ithaca, NY 14853, USA*
[c] *Department of Accounting, University of Washington, DC, USA*

Received 1 April 1998; received in revised form 6 April 1998

**Abstract**

In this research, we investigate the informational role of financial analysts. Using a trade-based empirical technique, we estimate the probability of information-based trading for a sample of NYSE stocks that differ in analyst coverage. We determine how this probability differs across stocks followed by many analysts, and we investigate whether analysts increase or create the flow of information. We also determine the 'normal' level of noise trading in each sample stock, thereby giving us the ability to assess the depth of the market for stocks with differing analysts followings. Our most important empirical result is that the number of financial analysts is not a good proxy for information-based trading. © 1998 Elsevier Science B.V. All rights reserved.

*JEL classification:* G14; G24

*Keywords:* Microstructure; Financial analysts; Trade

## 1. Introduction

Few issues in finance are more puzzling than the role played by financial analysts. Although ubiquitous participants in the financial markets, how exactly they affect the market equilibrium is unclear. Under one view, financial analysts provide the crucial function of uncovering and disseminating information to the market. It is their activities, reflected in the trades of their clients, that allow markets to become efficient. Yet, whether analysts actually gather information, or serve merely to promote a variety of special interests, is debatable. The

* Corresponding author. Tel.: 607-255-3645; fax: 607-254-4590; e-mail: mo19@cornell.edu.


*PII* S1386-4181(98)00002-0

compensation structure of analysts provides strong incentives to provide, at best, censored versions of their findings, and it is generally acknowledged that analysts' recommendations are biased towards good news.[1] This suggests that analysts may be, at best, noisy signals of information, and, at worst, merely another marketing mechanism to sell stocks. Moreover, there remains the conundrum that if markets are already efficient, then it is hard to justify either the information-gathering activities of analysts or its expense. Viewed from this perspective, analysts are superfluous, albeit expensive, market participants.

Recently, researchers have directed greater attention to the role of financial analysts. A number of researchers (see Shores, 1990; Bhushan, 1989, Brennan and Hughes, 1991; Brennan et al., 1993; Chung et al., 1995) employ financial analysts as a direct proxy for informed trading. This perspective is supported by research by Womack (1996), who argues that analysts forecasts (particularly sell recommendations) are informative, and by Skinner (1990) and Brennan et al. (1993), who find faster adjustment to new information for stocks with wider analysts following. Brennan and Subrahmanyam (1995) investigate this informed trader link further by estimating the effects of financial analysts on adverse selection costs as measured by the Kyle $\lambda$. If financial analysts are 'informed' traders, then one might expect to find greater adverse selection costs, reflecting the increased risk markets makers face in trading with the informed. Brennan and Subrahmanyam, however, find the opposite: firms with many analysts face a smaller adverse selection cost than do firms with fewer analysts. This result could be explained by enhanced competition between informed agents resulting in faster incorporation of information, but it is also consistent with the notion that financial analysts are uninformed, or even misinformed, traders. Certainly, the puzzle remains as to what role financial analysts play in the market.

In this research, we develop a new approach for investigating the informational role of financial analysts. Using a technique developed in Easley et al. (1997a) and Easley et al. (1996), we estimate the risk of information-based trading for a sample of NYSE stocks that differ in analyst coverage. Our analysis uses the information in trade data to estimate the probability of informed trade. This allows us to determine not only whether the probability of informed trading differs across stocks followed by many analysts, but also how the components of informed trading differ. For example, we can determine how frequently new private information occurs, and whether the probability of these information events differs with analyst coverage. This allows us to investigate

[1] See, for example, Incredible 'Buys': Many Companies Press Analysts to Steer Clear of Negative Ratings', *Wall Street Journal*, 19 July 1995, where it is argued that analysts reports are so biased that only unsophisticated investors pay heed. A similar argument is found in 'Who is Pulling the Strings', Euromoney, April 1996.

whether analysts create private information. We can also determine the 'normal' level of noise trading across volume categories, thereby giving us the ability to assess the depth of the market for stocks with differing analysts followings.

Our most important empirical result is that the number of financial analysts is not a positive proxy for information-based trading. We find that the probability of information-based trade is lower for stocks with many analysts, a result consistent with the analysis of Brennan and Subrahmanyam (1995). What may be of more interest are our results on why this occurs. We show that stocks with more analysts do have more informed trade, but that they have even greater rates of uninformed trade. This suggests that while analysts' clients may be trading on the basis of information, analysts attract even more uninformed following to a stock, and it is this greater depth that reduces the overall risk of information-based trading. Such an effect is consistent with Merton's (1987) analysis in which traders transact only in stocks with which they are familiar, and it supports the view that analysts serve to increase trading volume by showcasing stocks to uninformed traders.

We also show that financial analysts do not appear to create new private information; the probability of private information events is the same across stocks with many and few analysts, as are the probabilities of good and bad information events. This is consistent with Womack's (1996) finding that only 24 of 694 added-to-buy-list recommendations discuss facts deemed private or new. Our results thus confirm the view that analysts recommendations are generally based on public, rather than private, information.

Because our empirical technique provides a direct estimate of the probability of information-based trade for individual stocks, we are able to confirm the predictions of our model using spread data. Using regression analysis, we also show that financial analysts provide little explanatory power beyond that contained in volume, price level, and the probability of informed trade. We interpret these results as showing that the influence of analysts on the market equilibrium lies primarily in their effects on the composition of trade, rather than on the production or provision of new information.

Finally, we note that these regression results also provide reassuring evidence on the reasonableness of our trade-based estimating technique. Our empirical technique relies on the information conveyed by the frequency and imbalance of trade. Although conceptually simple, this approach has been shown to work quite well in a variety of applications. Our approach can be extended to include other market statistics, for example, trade size effects, an issue we explicitly deal with in Easley et al. (1997b). We do not include trade size as a variable in this analysis because there is no evidence in the literature that analysts affect the average trade size. We believe that our results would hold under more complex specifications, but that remains an empirical question for future research.

This paper is organized as follows. In the next section, we specify a continuous time sequential trade model, and we develop the likelihood function that we use in our estimation. Section 3 discusses the data and our sample selection technique. Because the number of analysts may be affected by market capitalization and industry effects, we use a matched pairs technique to select a sample of few and many analysts firms. In Section 4, we estimate the model and calculate the probability of information-based trading for each stock in our sample. We then examine how these parameters differ across stocks with many and few analysts. In Section 5, we investigate the robustness of our results using regressions of stock spreads on the number of analysts, price level, trading volume and the probability of information-based trading. In Section 6, we summarize our results, and discuss their implications for the role of financial analysts.

## 2. The model

### 2.1. Trade process

In this section, we set out a mixed discrete-and-continuous time, sequential trade model of market making.[2] The number of financial analysts does not directly enter this model of trade and price formation. Instead, we view the number of analysts as potentially affecting the parameters of this process. As each parameter is introduced we discuss how it may be affected by analysts. After deriving the trade process model, we estimate its parameters, and we then determine to what extent the parameter values vary with analyst following.

Informed and uninformed individuals trade a single risky asset and money with a market maker over $i = 1, \dots, I$ discrete trading days. Within any trading day, time is continuous, and it is indexed by $t \in [0, T]$. The market maker stands ready to buy or sell one unit of the asset at his posted bid and ask prices at any time. Because he is competitive and risk neutral, these prices are the expected value of the asset conditional on his information at the time of trade.

One aspect of this trade process that may be affected by financial analysts is the probability of private information events. We define a private information event as the occurrence of a signal that is not publicly observable about the future value of the asset where the signal may be good news or bad news. We view information events as occurring, if at all, prior to the beginning of any trading day. Of course, both public and private information events may occur and analysts may affect the probability of both types of events. Public information events may affect prices, but we view them as not directly affecting

[2] This model is similar to that used in our investigation of the effects of liquidity on the probability of informed trade, see Easley et al. (1996).

trade.[3] Our focus here is on private information events. Private information events are independently distributed across days and occur with probability $\alpha$.

If financial analysts discover information, which otherwise would never have become private information, and make it possible for their clients to trade on this information, then they raise the probability of private information events, $\alpha$. If analysts discover information that would eventually have been private information anyhow then they do not affect $\alpha$.[4] Alternatively, analysts could lower $\alpha$ by making what would otherwise have been private information so widely available that it effectively becomes public. We do not assume that any of these effects occur; our model allows for any effect, and so whether financial analysts create private information becomes an empirical question that we investigate.

These information events are good news with probability $1-\delta$, or bad news with probability $\delta$. Analysts tend to make only positive or neutral recommendations so it is possible that they affect the value of $\delta$. Of course, rational traders may be able to correctly invert any signal that analysts send in which case $\delta$ may be unaffected. It is also possible (but perhaps not plausible) that analysts only follow firms with predominantly good news, generating a correlation between analyst following and $\delta$.

After the end of trading on any day, and before nature moves again, the full information value of the asset is realized. Let $(V_i)_{i=1}^{I}$ be the random variables giving the value of the asset at the end of trading days $i=1,\dots,I$. These values will naturally be correlated. We do not make any specific assumptions about the correlations as they are not needed for our analysis. We let the value of the asset conditional on good news on day $i$ be $\bar{V}_i$; similarly it is $V_i$ conditional on bad news on day $i$.[5] The value of the asset if no news occurs on day $i$ is denoted $V_i^*$, where we define $V_i^* = \partial V_i + (1-\partial)\bar{V}_i$.[6]

Trade arises from both informed traders (those who have seen any signal) and uninformed traders. On any day, arrivals of uninformed buyers and uninformed sellers are random variables determined by independent Poisson processes with

[3] In effect, we define information events as public if they do not affect trading. Such events may cause price changes, but little or no trade should be generated by a truly public information event. To the extent that seemingly public information events affect trade, they have a private component (such as understanding how to use this particular information) and we classify them as private information events. The motivation for this distinction is that our empirical technique extracts information from trade data and so we cannot identify the events that we call public.

[4] If this occurs then analysts may change the identity of the informed traders and the timing of private information. They do not create new private information.

[5] In our empirical work we look at several stocks at once. We assume that the random variables giving asset values are independent across firms. For this reason, the analysis in the text is done for a single firm.

[6] This assumption is not required for our analysis (only that $\bar{V} \geqslant V \geqslant *\underline{V}$ is needed), however the assumption makes the spread properties used in later sections more tractable.

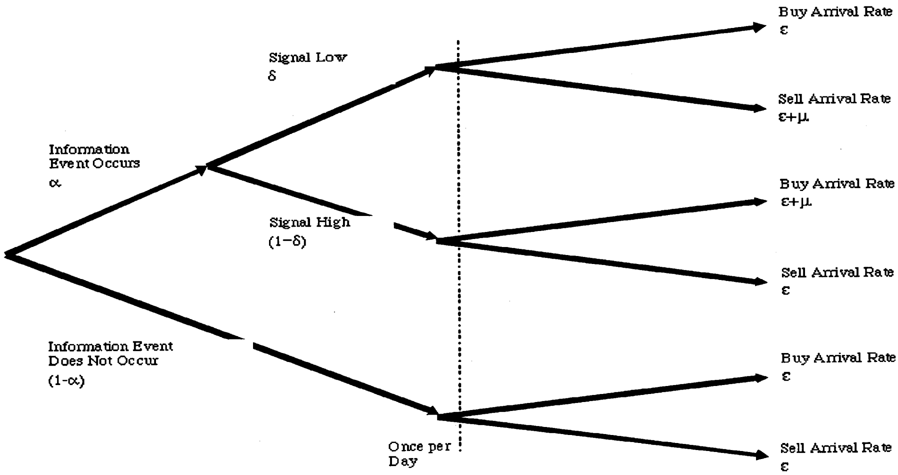


Fig. 1. Tree Diagram of the Trading Process. $\alpha$ is the probability of an information event, $\delta$ is the probability of a low signal, $\mu$ is the rate of informed trade arrival, and $\varepsilon$ is the rate of uninformed buy and sell trade arrivals. Nodes to the left of the dotted line occur once per day.

arrival rate $\varepsilon$ per day. The existence of analysts following a stock may attract uninformed traders, and, conversely, the existence of depth in a stock (a high value for $\varepsilon$) may attract analysts. In either case, it seems reasonable to expect $\varepsilon$ and the number of analysts in a stock to be positively correlated.[7]

On days for which information events have occurred, informed traders also arrive. These informed traders may include analysts' clients, as well as other traders who obtain information directly. If financial analysts provide private information to their clients (whether they affect $\alpha$ or not) we would expect to find a positive correlation between the number of analysts in a stock and the rate of informed trade. We assume that all informed traders are risk neutral and competitive. If a trader observes a good signal, then the profit maximizing trade is to buy the stock; conversely, he will sell if he observes a bad signal. We assume that the arrival of news to one trader at a time, and his subsequent arrival at the market, also follows a Poisson process. The arrival rate for this process is $\mu$ per day. All of these arrival processes are assumed to be independent.

The tree given in Fig. 1 describes this trading process. At the first node of the tree, nature selects whether an information event occurs. If an event occurs, nature then determines if it is good news or bad news. The three nodes (no event, good news, and bad news) before the dotted line in Fig. 1 occur only once per

[7] Of course, this hypothesized effect of analysts on the rate of uninformed trade (and all of the other effects of analysts on parameters discussed in the text) is an empirical question. Our theoretical model imposes no restrictions on these effects. We determine them empirically.

day. Then, given the node selected for the day, traders arrive according to the relevant Poisson process. That is, on good event days, the arrival rates are $\varepsilon + \mu$ for buy orders and $\varepsilon$ for sell orders. On bad event days, the arrival rates are $\varepsilon$ for buys and $\varepsilon + \mu$ for sells. Finally, on non-event days, only uninformed traders arrive and the arrival rate of both buys and sells is $\varepsilon$.

### 2.2. *Trades and prices*

Each day nature selects one of the three branches of the tree. The market maker knows the probability attached to each branch, and he knows the order arrival process for each of the branches. He does not know, however, which of the three branches has been selected by nature. We assume that the market maker is a Bayesian who uses the arrival of trades, and the rate of trading, to update his beliefs about the occurrence of information events. Because days are independent, we can analyze the evolution of his beliefs separately on each day. Let $P(t) = (P_n(t), P_b(t), P_g(t))$ be the market maker's prior belief about the events 'no news' (n), 'bad news' (b), and 'good news' (g) at time $t$. So his prior belief at time 0 is $P(0) = (1 - \alpha, \alpha\delta, \alpha(1 - \delta))$.

To determine quotes at time $t$, the market maker updates his prior conditional on the arrival of an order of the relevant type. The time $t$ prior expected value of the asset, conditional on the history of trade prior to time $t$, is

$$\mathrm{E}[V_i|t] = P_n(t)V_i^* + P_{\mathrm{b}}(t)\underline{V}_i + P_{\mathrm{g}}(t)\bar{V}_i. \tag{1}$$

For a given trade at time $t$, it is straightforward to derive bid and ask prices:

$$b(t) = \mathrm{E}[V_i|t] - \frac{\mu P_{\mathrm{b}}(t)}{\varepsilon + \mu P_{\mathrm{b}}(t)}(E[V_i|t] - \underline{V}_i) \tag{2}$$

and

$$a(t) = \mathrm{E}[V_i|t] + \frac{\mu P_{\mathrm{g}}(t)}{\varepsilon + \mu P_{\mathrm{g}}(t)}(\bar{V}_i - E[V_i|t]). \tag{3}$$

These equations demonstrate the explicit role played by arrivals of informed and uninformed traders in affecting trading prices. If there are no informed traders ($\mu = 0$), then trade carries no information, and so the bid and ask are both equal to the prior expected value of the asset. Alternatively, if there are no uninformed traders ($\varepsilon = 0$), then $b(t) = \underline{V}_i$ and $a(t) = \bar{V}_i$ for all t. At these prices no informed traders will trade either, and the market, in effect, shuts down. Generally, both informed and uninformed traders will be in the market, and so the bid is below $\mathrm{E}[V_i|t]$ and the ask is above $\mathrm{E}[V_i|t]$.

The factors influencing the spread are easier to identify if we write the spread explicitly. Let $\Sigma(t) = a(t) - b(t)$ be the spread at time $t$. Calculation shows that

$$\sum(t) = \frac{\mu P_{\mathrm{g}}(t)}{\varepsilon + \mu P_{\mathrm{g}}(t)}(\bar{V}_i - \mathrm{E}[V_i|t]) + \frac{\mu P_{\mathrm{b}}(t)}{\varepsilon + \mu P_{\mathrm{b}}(t)}(\mathrm{E}[V_i|t] - \underline{V}_i). \tag{4}$$

The spread at time t can be viewed in two parts. The first term is the probability that a buy is information-based times the expected loss to an informed buyer, and the second is a symmetric term for sells. The probability that any trade that occurs at time $t$ is information-based is the probability a buy is information-based times the probability of any buy occurring plus the probability a sell is information-based sell times the probability of a sell occurring, explicitly,

$$\mathrm{PI}(t) = \frac{\mu(1 - P_{\mathrm{n}}(t))}{\mu(1 - P_{\mathrm{n}}(t)) + 2\varepsilon}. \tag{5}$$

This probability depends on the rates of informed and uninformed trading, as well as on the market maker's beliefs regarding the occurrence and composition of information events. So if there is no possibility of news ($P_{\mathrm{n}}(t) = 1$) or if no one trades on private information ($\mu = 0$), then $\mathrm{PI}(t) = 0$ and there is no spread. Alternatively, if all trades are information based ($\varepsilon = 0$), then $\mathrm{PI}(t) = 1$ and the spread is wide enough ($\bar{V}_i - \underline{V}_i$) to prevent anyone from profiting on private information.

The spread for the opening quotes has a particularly simple form in the natural case in which good and bad events are equally likely. That is, if $\delta = 1 - \delta$ then[8]

$$\sum(0) = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon}[\bar{V}_i - \underline{V}_i]. \tag{6}$$

The first term in this equation is the probability that a trade at the opening of the trading day is information-based. This probability of trading with an informed trader is clearly a crucial factor influencing the size of spreads. If the number of financial analysts affects this probability through its effect on the trade parameters $\alpha$, $\mu$ and $\varepsilon$, then analysts affect spreads. Our model predicts how initial spreads will differ with these parameters, and thus provides insights into how spreads will differ with the number of analysts.

If, like the market maker, we knew the parameters of the problem, $\theta = (\alpha, \delta, \varepsilon, \mu)$, and observed the order arrival process, then we could compute the stochastic process of bids and asks. Although we can observe the order arrival process, we do not know the parameters. These parameters can be estimated, however, from the data on order arrivals.

[8] In our empirical work we find that $\delta = 0.5$ is a good approximation.

### 2.3. The likelihood function

Estimating the parameter vector $\theta = (\alpha, \delta, \varepsilon, \mu)$ is more complex than just estimating arrival rates from independent Poisson processes. The difficulty arises because we cannot directly observe the arrival of any information events or trades governed by these parameters. Elsewhere we have demonstrated how our model provides the structure necessary to extract information on the parameters from the observable variables, buys and sells.

In our model, buys and sells follow one of three Poisson processes on each day. We do not know which process is operating on any day, but we do know that the data reflect the underlying information structure, with more buys expected on days with good events, and more sells on days with bad events. Similarly, on no-event days, there are no informed traders in the market, and so fewer trades arrive. These rates and probabilities are determined by a mixture model in which the weights on the three possible components (i.e., the three branches of the tree reflecting no news, good news, and bad news) reflect their probability of occurrence in the data. The probabilities of a no-event day, a bad-event day, and a good-event day are, respectively, given by $1-\alpha$, $\alpha\delta$, and $\alpha(1-\delta)$, so the likelihood of observing $B$ buys and $S$ sells on a day of unknown type is

$$L((B,S)|\theta) = (1-\alpha)*\mathrm{e}^{-\varepsilon}\frac{(\varepsilon)^B}{B!}\,\mathrm{e}^{-\varepsilon}\frac{(\varepsilon)^S}{S!} + \alpha\delta*\mathrm{e}^{-\varepsilon}\frac{(\varepsilon)^B}{B!}\,\mathrm{e}^{-(\mu+\varepsilon)}\frac{[(\mu+\varepsilon)]^S}{S!}$$

$$+\,\alpha(1-\delta)*\mathrm{e}^{-(\mu+\varepsilon)}\frac{[(\mu+\varepsilon)]^B}{B!}\,\mathrm{e}^{-\varepsilon}\frac{(\varepsilon)^S}{S!}. \tag{7}$$

For any given day, the maximum likelihood estimator of the information event parameters $\alpha$ and $\delta$ will be either 0 or 1, reflecting that information events occur only once a day. Over multiple days, however, these parameters can be estimated from the daily numbers of buys and sells. Thus, intra-day data allows us to estimate the trader selection probabilities in our model, and inter-day data to estimate the information event parameters.[9] Because days are independent, the likelihood of observing the data $M = (B_i, S_i)_{i=1}^{I}$ over $I$ days is just the product of the daily likelihoods,[10]

$$L(M|\theta) = \prod_{i=1}^{I} L(\theta|B_i, S_i). \tag{8}$$

[9] This is just intended to provide some intuition about how the estimation works. Of course, we actually use the entire data set to determine the joint parameter vector.

[10] Easley et al. (1997a) tested the independence of information events across days, and found that they could not reject the independence assumption.

To estimate the parameter vector $\theta$ from any data set $M$, we maximize the likelihood defined in Eq. (8). This provides direct estimates of the rate of informed and uninformed trading in a particular stock, as well as of the information event structure surrounding that stock.

## 3. The data

We now turn to determining how financial analysts affect the structure of trading in a stock. Using the methodology developed in the previous section, we can estimate directly from the order flow how the composition of trading differs between stocks and investigate the incidence and direction of private information events. By examining these estimates for stocks followed by many analysts and for stocks followed by few analysts, we can determine the correlation between the number of analysts and the probability of information-based trading in a stock.

The number of analysts used in our estimation is taken from the Institutional Brokers Estimation System (I/B/E/S) database. I/B/E/S compiles analysts forecasts from over 400 brokerage firms.[11] The trade data are drawn from the ISSM transactions data base. The ISSM database provides time-stamped bid and ask quotes, depths and transactions. Our estimation technique requires knowing the number of buys and sells for each stock in our sample. This data can be derived from the ISSM data with two adjustments. First, large orders sometimes have multiple participants on one side of the trade. Reporting conventions may treat such transactions as multiple trades, when in fact only one trade has actually occurred. Hasbrouck (1988) suggests combining trades occurring within 5 s at the same price, and we follow this procedure in the paper. Second, while the ISSM data identifies trades, it does not indicate whether any trade was initiated by a buyer or a seller. This classification can be done using a technique developed by Lee and Ready (1991). Trades above the midpoint of the spread are classified as buys; those below the midpoint are called sells. Midpoint trades are classified by comparison to the previous trade price. Trades executed at higher prices than the previous trades are called buys, and those at lower prices are called sells. This technique undoubtedly misclassifies some trades, but it is standard, and it has been shown to work reasonably well.[12]

The time period for our study is 1 October 1991 through 22 December 1991. Our choice of sample period was restricted by several considerations. Because

[11] The IBES database thus provides a complete depiction of sell-side market analysts. There are also buy-side analysts, but their numbers are not generally available.

[12] Following Lee and Ready, we also use a 5 s lag on quotes to adjust for differences in reporting times between quotes and trades.

the number of analysts following a stock may change over time, we use a single time period for all firms in our sample. The number of analysts forecasting earnings tends to increase throughout the year, stabilizing approximately 2 months before the earnings announcement. For this reason, we chose stocks that have the same fiscal year end month so that the fixed sample time period falls at the same point in all firms fiscal years. To satisfy these considerations, firms with December fiscal year ends for whom I/B/E/S reported the number of analysts forecasting fiscal 1991 annual earning in October 1991 are selected. Fiscal 1991 is chosen because it is the most recent year which satisfied data availability requirements. A total of 2136 firms satisfy these data requirements.

The remainder of the sample selection process is driven by the testing procedures to be used. We are interested in determining whether the number of analysts following a stock is related in any systematic way with the information structure for that stock. Two alternative methods of testing for these effects can be used. The first method, which is more commonly seen in the finance literature, is to select a broad sample and then use regressions to test for significant partial correlations between variables of interest. The effect of 'nuisance' variables, such as size or industry, is eliminated by including them as independent variables in the regressions. This method is appropriate when the expected functional form of the relationship between the dependent variable and the independent variable is known. For the case of financial analysts, however, this is problematic. As several researchers have shown (see Bhushan, 1989; O'Brien and Bhushan, 1990), there are many factors influencing the number of analysts and some of these factors are endogenous. Of particular concern is that trading volume may both be affected by the number of analysts and be the cause of analyst following. This introduces a difficult estimation problem in determining the relevant impact of financial analysts on trading behavior.

In this paper, we chose the alternative testing methodology of constructing a matched sample of firms. The matched sample approach allows for elimination of 'nuisance' variables by explicitly controlling for these variables in the construction of the sample of firms. The matching criteria we use are listed exchange, industry, and firm size. Exchange listing is relevant because listing requirements can introduce differences in firm groups. Since our model is most appropriate for a stock specialist, we select only those firms listed on the NYSE for whom I/B/E/S data was available during our sample period. This restricts the sample to 1168 firms.

Numerous researchers have also shown significant industry effects on the number of analysts following a firm. To incorporate these effects, each firm is assigned to an industry based on its three digit SIC code obtained from the CRSP data files. We then analyze pairs of firms where both firms in a pair are in the same industry. Another matching criterion is firm size, reflecting the importance that researchers have demonstrated of firm size on analyst following. An important issue is what measure of size to use. One possible measure is to look at trading volume, taking the view that analysts are more likely to follow heavily traded stocks. The difficulty with this measure, however, is that analysts in turn affect trading volume, so that determining which is endogenous (volume or analysts) is problematic. We proxy firm size by the total market capitalization for each firm, where the market capitalization is taken from the CRSP data files. One advantage of using market capitalization is that the endogenity problems noted above do not arise. That is, unlike size measures such as volume or turnover, it is not immediately obvious how financial analysts can cause market capitalization. Within each industry, therefore, all possible pairings of firms by market capitalization are calculated. We then select the pairs that have the closest market capitalizations.[13]

The final step in our sample selection was to ensure that in each pair of firms the number of analysts differs as much as possible. Pairs are included in the sample if the firm with the greater analyst following had at least the lesser of double or 10 more analysts than the firm with fewer analyst following.[14] These procedures resulted in a sample of 60 matched pairs, or 120 firms, which are listed in Table 3 in the appendix. These firms are matched with respect to listing, industry and market capitalization, but differ with respect to the number of analysts following the firm. It is this difference that we now exploit to determine whether financial analysts are related in any systematic way to the information structure for a stock.

## 4. Estimation

The next step in our analysis is to estimate the parameters of the structural model. Recall that the trade process depends on four parameters: $\alpha$, the probability of private information events; $\delta$, the probability that the information is bad news; $\mu$, the arrival rate of traders who know the private information if it exists (i.e. the informed traders); and $\varepsilon$ the arrival rate of uninformed traders. These parameters, in turn, determine the probability of information-based trading in a stock. It is this probability and its relation to the number of financial analysts that we wish to investigate.

[13] Some SIC codes as reported on CRSP did not seem consistent with the actual industry of firms in the sample. We manually checked the firms in the Standard and Poors Manual to ensure that firms were properly matched by industry. This removed 78 pairs from our potential sample.

[14] Firms were included in only one pair. Where firms could be assigned to multiple pairs (i.e. there were multiple firms within one industry that met our selection criteria), we chose pairings so as to maximize the total number of pairs in the sample.

### 4.1. Parameter estimates

We estimate the parameters of the trade process for each stock in our sample by maximizing the likelihood function conditional on the stock's trade data as described in the previous section. The two probability parameters $\alpha$ and $\delta$ were restricted to [0, 1] by a logit transform of unrestricted parameters, and the two rate parameters $\varepsilon$ and $\mu$ were restricted to $[0, \infty]$ by a logarithmic transform. We then maximized over the unrestricted parameters using the quadratic hill-climbing algorithm GRADX from the GQOPT package.[15] Standard errors for the economic parameter estimates were calculated from the asymptotic distribution of the transformed parameters using the delta method.[16] The standard errors show that the model can be estimated quite precisely. The arrival rate variables are estimated with great accuracy, reflecting the precision that arises with the large number of transactions in our data set. The information parameters $\alpha$ and $\delta$ have larger standard errors, but are still estimated reasonably precisely.[17] Of particular importance is the estimation of our composite probability of information-based trading. In all but four cases, this variable is significantly different from zero. These results suggest that our procedure is able to accurately estimate our underlying variables.

Panel A of Table 1 provides the means of our estimated parameters for the high analysts and low analysts groups. To test whether the individual parameter estimates tend to be lower or higher for the high versus low analyst stocks, we use non-parametric statistics, specifically the Wilcoxon Signed Ranks test. The Wilcoxon test allows us to compare two samples with a magnitude and a directional hypothesis. That is, we can test whether the values for one sample tend to be higher or lower than for the second sample. These test statistics are also given in Panel A of Table 1.

We first consider the estimates of the information event parameter, $\alpha$, which is the probability of a private information event occurring before the start of a trading day. One possible role for analysts is the creation of new private information, suggesting that the probability of private information events should be higher for stocks with more analysts. Table 1 shows that this is not the case: the mean $\alpha$ is 0.368 for high analysts stocks and it is 0.387 for low analysts stocks. This ordering of the means for $\alpha$ is consistent with the alternative explanation that analysts turn private information into public information. However, the Wilcoxon test statistic is 1.29, which is less than the 0.05 significance level of 1.96. Thus, the likelihood of private information events is not

---

[15] A grid of starting points is used to insure that a global maximum is found.

[16] For discussion of the delta method see Goldberger (1991).

[17] Intuitively, we have one observation of nature's choice per day. So we have only 60 observations to use in estimating $\alpha$ and $\delta$. However, we have many observations of trades per day, so the Poisson parameters $\varepsilon$ and $\mu$ can be estimated more precisely.

Table 1
Parameter estimates by analyst following

This table gives mean parameter estimates for sample firms with high analyst following (denoted HIGH) and low analyst following (denoted LOW). The top panel is for our total 60 pair (120 firm) sample. The lower panel is for the 41 pair sample in which market capitalization differences between the pairs are less than 25%. We report means and standard deviations for parameters $\mu$, the rate of informed trading, $\varepsilon$ the rate of uninformed trading, $\alpha$ the probability of information events, $\delta$ the probability of bad news, and PINF the composite probability of information-based traded. Analysts is the mean number of financial analysts following firms in each group, and market capitalization is the mean market capitalization in thousands of dollars for 1991 as reported on CRSP. The Wilcoxon sign rank statistic is a test for differences in the two populations, where the critical level for 0.05 significance is 1.96.

| | $\mu$ | $\varepsilon$ | $\alpha$ | $\delta$ | PINF | Analysts | Mkt. Cap. |
|---|---|---|---|---|---|---|---|
| *Panel A* | | | | | | | |
| | | | *All stocks* | | | | |
| High analysts group: | | | | | | | |
| Mean | 34.05 | 24.75 | 0.368 | 0.383 | 0.182 | 16.65 | 1371394 |
| STD | 40.14 | 27.52 | 0.196 | 0.242 | 0.064 | 8.25 | 1987102 |
| Low analysts group: | | | | | | | |
| Mean | 19.27 | 13.15 | 0.387 | 0.383 | 0.212 | 7.02 | 1324772 |
| STD | 17.13 | 12.26 | 0.183 | 0.280 | 0.082 | 5.17 | 1968578 |
| Wilcoxon test | 3.58 | 4.73 | − 1.29 | 0.31 | − 2.43 | | |
| *Panel B* | | | | | | | |
| | | | 41 *stocks* | | | | |
| High analysts group: | | | | | | | |
| Mean | 36.51 | 25.91 | 0.356 | 0.403 | 0.180 | 16.98 | 1366467 |
| STD | 46.27 | 29.72 | 0.173 | 0.246 | 0.066 | 8.88 | 1971766 |
| Low analysts group: | | | | | | | |
| Mean | 15.71 | 11.63 | 0.401 | 0.419 | 0.214 | 7.29 | 1391480 |
| STD | 14.56 | 12.06 | 0.166 | 0.284 | 0.073 | 5.45 | 2095574 |
| Wilcoxon test | 4.05 | 4.89 | − 1.76 | − 0.16 | − 2.55 | | |

significantly different for stocks followed by many financial analysts versus those followed by few analysts.[18]

[18] This result complements our finding in earlier research (Easley et al., 1996) that the probability of information events did not differ between high and medium volume stocks. Interestingly, however, low volume stocks do have significantly fewer information events. Since such stocks typically have no analyst following they would not be included in the sample considered here.

The second information parameter in our model is $\delta$, which is the probability that new information is bad news. There is no compelling reason to expect the probability of bad news to be correlated with the number of analysts following a stock, and hence the estimation of this parameter provides a simple check on the reasonableness of our model. Table 1 shows that the means for the two groups are the same, being 0.383 for both the high analysts group and the low analysts group. The Wilcoxon test statistic is 0.31. Thus, as expected, there is no significant evidence of a difference in the direction of information between our groups of stocks.

We now consider the parameters relating to the arrival rates of uninformed and informed traders. The arrival rate of uninformed traders is our estimated parameter $\varepsilon$, while the informed arrival rate is our estimated parameter $\mu$. If financial analysts' clients trade on the basis of private information, then we would expect to find a higher $\mu$ for the many analysts stocks. If financial analysts simply generate trade unrelated to information, then this should be reflected in a higher estimate of $\varepsilon$. Note that such a higher $\varepsilon$ could also reflect the effect of analyst following on the population of investors trading a stock. In particular, Merton (1987) argued that investors trade securities with which they are familiar. If analysts play a 'certification' role of essentially advertising a stock, then this could induce more investors to follow a stock, thereby generating more uninformed trade.

Table 1 shows dramatic differences in both informed and uninformed arrival rates across the two groups. For uninformed traders, the rate falls from 24.75 for the high analysts group to 13.15 for the low analysts group. The Wilcoxon test statistic of 4.73 confirms that this difference is significant at the 0.05 level. Thus, high analyst stocks tend to have more uninformed trade than do low analyst stocks. The behavior of informed order arrivals exhibits similar behavior. The mean value of $\mu$ is 34.05 for the high analysts group and it is 19.27 for the low analysts group. The Wilcoxon test confirms that this difference is also significant with a test statistic of 3.58 Thus, higher analysts stocks also tend to have higher arrival rates of informed traders.

These results demonstrate that although high and low analysts stocks do not significantly differ in their information structure ($\alpha$ and $\delta$), they do differ in the level and composition of activity in the stock. High analyst stocks have significantly higher levels of both informed and uninformed trade than do low analyst stocks. This suggests that while financial analysts do not create new private information, they do act, at least to some extent, as informed traders in their own and their client's trading activities. This result is consistent with the hypothesis that the clients of financial analysts receive early notification of events that would otherwise have become private information to other traders. The higher level of uninformed trading is consistent with a number of possibilities. One is simply that the clients of some analysts act as uninformed traders. Such a finding corresponds with there being differences in skill levels between analysts, with those less skilled essentially having no more insight than uninformed traders. A second possibility is the certification role suggested by Merton whereby sufficient analyst following is a condition for at least some traders to consider following a stock.

What is not directly apparent from the individual parameter estimates, however, is whether either group has a greater probability of information-based trade. The greater uninformed activity in high analyst stocks would tend to reduce the probability that any trade is information-based, but the higher level of informed activity has the opposite effect. From the point of view of price-setting agents, what matters is not the rates of such trading, but rather the probability of information-based trading given that a trade occurs. Having estimated the parameter values, we can now determine this probability of informed trade for the stocks in our sample, and examine how it differs between high and low analyst stocks.

### 4.2. The probability of information-based trade

The probability of information-based trade is a composite variable reflecting the various parameters characterizing the trade process. This probability is given by our model in Eq. (11), where we showed that the probability of informed trade is

$$\mathrm{PI} = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} \tag{9}$$

for the market maker's initial beliefs. As is apparent from Eq. (9), the probability of information-based trading depends on the arrival rates of traders (both informed and uninformed) and on the probability that new information exists. Consequently, it is the interaction of our estimated parameters that matters for determining the effect of information-based trading on spreads.[19]

We calculated this probability for each stock in our sample. The mean values are reported in Table 1 (individual estimates are given in the last column of Table 3 in the appendix). The data reveal an intriguing result: the probability of informed trade is clearly *lower* for the high analyst stocks. The mean results show that the probability of information-based trades is approximately 0.182 for the stocks in the high analyst sample, but it rises to 0.212 for the low analyst group. Pairwise testing using the Wilcoxon signed ranks test reveals that we can

[19] Because the PI variable is composed of estimated parameters, there is a potential errors in variables problem in the estimation. There is no immediate way to determine the scale of this problem, but the underlying precision of the estimated variables suggests that it is likely to be small. Our regression testing in the next section provides one check on the reasonableness of the estimates.

reject the hypothesis that the two groups are not different at the 0.05 significance level. The test statistics are given in Table 1.[20]

These results demonstrate that the number of analysts is not a good (positive) proxy for information-based trade.[21] Instead we find, as did Brennan and Subrahmanyam (1995), that the number of analysts is negatively related to the probability of information-based trade. This occurs because of the correlations of the number of financial analysts with $\alpha$, $\varepsilon$ and $\mu$. We know that the probability of private information events, $\alpha$, is uncorrelated with the number of analysts. So the effect of the number of analysts on the probability of information based trade is contained its effect on $\varepsilon$ and $\mu$. Analyst following is positively correlated with the rate of informed trade as embodied in our parameter $\mu$. However, the rate of uninformed trade, $\varepsilon$, increases even more with analyst following. Thus, while the market maker's total expected losses to informed traders increases with analyst following, because he is spreading these costs over a wider base of uninformed trades, the adverse selection cost per trade actually declines.

## 5. Spreads and regression testing

The analysis thus far provides estimates the level of information-based trading for the stocks in our sample, and these estimates, in turn, form the basis for our analysis of the role of financial analysts. Because our analysis involves estimation of unobservable variables, however, there is always the concern that our estimates are not actually capturing what we claim. In this section of the paper, we investigate this issue by testing how well our estimated model predicts bid–ask spreads. This provides an independent check on the validity of our approach, and it allows us to test for any cross-sectional influences of financial analysts not captured by our analysis.

Our testing approach makes use of the fact that we have used only trade data to estimate adverse selection costs for the stocks in our sample. Typically

[20] One concern that often arises in empirical work is whether the results could be due to specification issues. In particular, our matched sample technique, while designed to minimize the confounding effects of other variables, may inadvertently introduce biases into our analysis. For example, market capitalization does tend to be lower for our low analysts firms, reflecting the fact that bigger firms have more analysts. To check whether market capitalization differences are actually driving our results, we selected the 41 pairs in our sample in which market capitalization differences are less than 25%. Panel B of Table 1 provides the mean parameter estimates and the Wilcoxon test statistics for this subsample. These statistics show that our results are not driven by size differences. The sub-sample results are essentially the same as those for the full sample, with that of the small difference pairs perhaps even stronger.

[21] Several studies have used analyst following as a proxy for the level of private information being produced about firms (see for example Skinner, 1990 and Brennan and Hughes, 1991).

researchers have used spreads or some decomposition of spreads to measure adverse selection cost, with larger spreads reflecting greater such costs. Because price data is not used in our estimation of the trade process, we can use spreads to provide an independent verification of our results. This will allow us to ensure that the adverse selection cost measure we have developed is consistent with those of other studies that use spreads as a measure of adverse selection costs. We can also determine whether analyst following has any explanatory power in the cross sectional variation of spreads that is not captured in our measure. If analyst following has such incremental explanatory power, then it would indicate that our model has not fully captured the effect of financial analysts on the adverse selection problem faced by market makers.

### 5.1. Predicted spread relationship

According to our theoretical model, as embodied in Eq. (4), the adverse selection component of a stock's spread is determined by the probability of information–based trades and the range of eventual values for the stock. Within our model the number of analysts can influence spreads only through differences in the parameters of our model. If our model is correct, and the parameter estimation procedure is effective, then the estimated probability of informed trade should affect spreads in a predictable way. More importantly, the number of analysts following the firm should have no incremental explanatory power beyond our measure of adverse selection costs.[22]

To investigate the relationship between spreads and our estimate of the probability of informed trade, we return to the opening spread derived in Eq. (6). Substituting in Eq. (9) yields the opening spread on day $i$ as[23]

$$\sum = [\bar{V}_i - \underline{V}_i]\mathrm{PI}. \tag{10}$$

The first term in this equation reflects the expected loss to an informed trader when one does arrive, while the second term is the probability a trade is information-based. While we have an estimate for the probability of informed trade, PI, we do not have a direct observation of the range of possible values, $[\bar{V}_i - \underline{V}_i]$. To create a proxy for this range, we need to estimate the sum of the

[22] These incremental explanatory power tests are somewhat problematic. The number of analysts reported by I/B/E/S can be thought of as a proxy for the 'true' unobservable amount of analyst following for the firm. In this case we are measuring the variable with error and our sample selection process would tend to maximize this error. The coefficients estimated for this dependent variable would then be biased towards zero.

[23] All of the variables in this equation are stock specific. Asset values and estimated parameters differ across stocks, and it is this variability that we are exploring. While these values should all be viewed as indexed by firm in this equation and the following regressions, we have not included the index in the text in order to keep the notation simple.

expected value increase on good days, $\Delta\bar{V}$, and the expected value decrease on bad days, $\Delta\underline{V}$.

To construct the proxy for the range in possible values, we first identify good and bad days. Using our parameter estimates and the daily trade data from ISSM, for each day in our sample period we can calculate the end of the day posterior probability that the day was a good news, bad news, or no-event day. A day is classified as a good news day when the posterior probability of a good event is greater than or equal to 0.8. Days are classified as a bad news day when the posterior probability of a bad event is greater than or equal to 0.8. For days classified as good news days we calculate $\Delta\bar{V}$ as the closing quote midpoint minus the opening quote midpoint. Similarly on bad news days we calculate $\Delta\underline{V}$ as the opening quote midpoint minus the closing quote midpoint. We then average $\Delta\bar{V}$ over all good news days and $\Delta\underline{V}$ over all bad news days. Our proxy for $[\bar{V}_i - \underline{V}_i]$, denoted $\Delta V$, is then calculated as $\Delta\bar{V} + \Delta\underline{V}$.[24] It is possible that analysts could affect adverse selection costs through this range of values as well as the probability of informed trade. The correlation between analyst following and $\Delta V$ is relatively low, 0.139. Thus, it does not appear that the number of analysts following a stock is correlated the magnitude of the impact of information.

Because our measure of price changes may not be exactly equal to the possible range of prices we allow our proxy to be a linear function of $\Delta V$. Using our proxy for the range of values the spread equation can be re-written as

$$\sum = \beta * \Delta V * \mathrm{PI} \tag{11}$$

where $\beta$ is the constant in our linear relationship. Eq. (11) reflects the intuitive relationship that spreads are influenced positively by the probability of informed trade.

### 5.2. Regression equations

One reason for using our methodology was to avoid confounding factors that affect spreads. In examining spreads, therefore, it is important that we control for these confounding factors not included in our model. If inventory effects matter, then average daily dollar volume of trade may be important to the market maker in setting his spreads. Based on the inventory cost literature, we expect that the market maker's inventory costs should be positively related to the difference between his current inventory and his optimal inventory, and negatively related to trading volume. The positive relationship with inventory position arises because a risk averse market maker will demand increasing

[24] For those firms which did not have at least one good news day and one bad news day, we use the average measure of the proxy for firms in the same one digit SIC code.

premiums the further he is from his optimal position. The negative relation with volume arises because higher trading volume allows the market maker to adjust his inventory position more quickly. Unfortunately, we have no data on the actual or optimal inventory position of market makers. We include in our regressions a term for trading volume.

Other factors that may affect spreads are the market maker's monopoly power and operating costs. These components would include any fixed costs of making a market, the return on the market maker's investment, and the return on the market maker's time. We expect that operating costs would tend to be fixed, while the amount invested would depend somewhat on the price of the stock. Therefore, we incorporate this component of the spread in our regressions as a constant term plus a linear function of stock price. The resulting regression equation is

$$\sum = \beta_0 + \beta_1 * V + \beta_2 * \Delta V * \mathrm{PI} + \beta_3 * \log(\mathrm{Volume}) + \eta. \tag{12}$$

As discussed above, we expect $\beta_0$ and $\beta_1$ to be positive and $\beta_3$ to be negative. If our model accurately estimates the probability of informed trade, we would expect $\beta_2$ to be significantly positive.[25]

### 5.3. Results

Regression results are presented in Table 2. The stock price, $V$, is calculated as the CRSP end-of-day price averaged over the sample period. The dollar volume, Volume, is calculated as the daily dollar volume of trades from ISSM averaged over the sample period. The PI and $\Delta V$ variables are obtained as outlined in the previous section. The opening spread, $\Sigma$, is taken as the daily opening spread from ISSM averaged over the sample period. The first column in Table 2 shows the results of estimating regression Eq. (12) for the 120 stocks in our sample. All of the coefficients have the expected sign and are statistically significant.[26] The regression explains a significant portion of the cross sectional variation in spreads with an adjusted $R^2$ of 75.11% and an $F$-statistic of 120.71. We interpret these results as evidence that our approach is reasonably successful in capturing the adverse selection cost component of spreads.

We next ask whether our probability of informed trade parameter captures all the information contained in analyst following. To answer this question we

[25] The regression was also run with dollar volume and share volume instead of log (Volume). The results for these regressions were similar with slightly lower F values.

[26] We caution that the specific coefficient estimates should be interpreted with care because the regressors are stochastic. The direction of this bias is indeterminate depending on the correlation of errors for the independent variables.

Table 2
The effect of PI and analysts on spreads

This table presents the results from estimating the linear regression given by

$$\sum = \beta_0 + \beta_1 \cdot V + \beta_2 \cdot \Delta V \cdot \mathrm{PI} + \beta_3 \cdot \log(\mathrm{Vol}) + \beta_4 \cdot \log(\mathrm{FA}) + \beta_5 \cdot \log(\mathrm{FA1}) + \beta_6 \cdot \log(\mathrm{FA2}) + \beta_7 \cdot \log(\mathrm{FA3}+5) + \beta_8 \cdot \log(\mathrm{FA6}) + \eta.$$

In Regression 1 the coefficients on the financial analyst variables log(FA), log(FA1), log(FA2), log(FA3 + 5) and log(FA6) are restricted to zero. In Regression 2 all coefficients are unconstrained. The dependent variable $\Sigma$ is the average quoted opening spread calculated from information on the ISSM database for the period 1 October 1991–22 December, 1991. The average price, $V$, is obtained by averaging the CRSP daily prices for this same time period. The variable Vol is the average daily dollar volume. The probability of informed trade, PI, is reported in the Appendix and $\Delta V$ is our proxy for the range of prices. The number of analysts in each stock and the stock's SIC code are reported in Table 3. The regressions use OLS, $t$-statistics are reported in parenthesis.

| Variable | Regression 1 | Regression 2 |
|---|---|---|
| Intercept | 0.3849 (8.94) | 0.3623 (7.15) |
| Price | 0.0051 (13.65) | 0.0052 (13.45) |
| $\Delta V$ times PI | 0.2096 (3.98) | 0.2205 (4.07) |
| log(Vol) | − 0.0282 ( − 5.68) | − 0.0247 ( − 3.44) |
| log(FA) | | − 0.0117 ( − 1.01) |
| log(FA1) | | 0.0080 (0.88) |
| log(FA2) | | 0.0034 ( − 0.39) |
| log(FA3 + 5) | | 0.0184 (2.11) |
| log(FA6) | | 0.0067 (0.84) |
| Adjusted $R^2$ | 0.7511 | 0.7562 |
| $F$-statistic | F(3, 116) = 120.71 | F(8, 111) = 47.12 |

include the number of analysts following the firm in regression Eq. (12) to determine if it has any incremental explanatory power. Because the number of analysts following a firm tends to be industry-driven, we use the number of analysts relative to the average for the industry.[27] Each firm in our sample is classified into one of seven industry groups by the first digit of its SIC code, and the number of analysts in each group is then denoted by FA1, FA2, …, FA7. Of the seven groups formed, SIC codes 5 and 7 had fewer than five firms each, so we merged these industry groups with SIC codes 3 and 4, respectively, for a total of five industry groups.[28] We treated SIC code group 4 + 7 as the base group, and then added variables for the other four SIC code groups 1, 2, 3 + 5, and 6 to our

[27] We also tried (FA) and log(FA) alone as well as non-logarithmic industry relative analyst following with similar results.

[28] The SIC code group to merge with was selected based on the nearest mean number of analyst following. In both cases the range of analyst following in the receiving SIC code group was unchanged by this procedure.

Table 3
Sample stocks

This table gives the sample of stock used in our analysis. Ticker gives the company's ticker symbol, market capitalization is for 1991 and is taken from the CRSP tapes, SIC is the companies' SIC code, and analysts is the number of analysts following the stock as of October 1991 as given by the I/B/E/S tapes.

| Name | Ticker | Mkt. Cap. | SIC | Analysts |
|---|---|---|---|---|
| NEWMONT GOLD CO | NGC | 4,168,781 | 1041 | 20 |
| AMERN BARRICK | ABX | 3,807,223 | 1041 | 31 |
| NUEVO ENERGY CO | NEV | 103,075 | 1311 | 5 |
| WAINOCO OIL | WOL | 89,402 | 1311 | 11 |
| BERRY PETROLEUM | BRY | 229,442 | 1311 | 3 |
| PLAINS PETE | PLP | 278,816 | 1311 | 10 |
| WESTN GAS RES | WGR | 490,714 | 1311 | 3 |
| NOBLE AFFILIATES | NBL | 601,912 | 1311 | 18 |
| RANGER OIL LTD | RGO | 740,830 | 1311 | 16 |
| SANTA FE ENG RES | SFR | 574,470 | 1311 | 8 |
| ENRON OIL & GAS | EOG | 1,480,050 | 1311 | 19 |
| ANADARKO PETE CO | APC | 1,321,488 | 1311 | 29 |
| SANTA FE ENERGY | SFP | 155,303 | 1381 | 3 |
| POGO PRODUCG | PPP | 154,446 | 1382 | 7 |
| READING & BATES | RB | 401,503 | 1381 | 4 |
| ROWAN COS | RDC | 418,646 | 1381 | 24 |
| WRIGLEY, WM JR | WWY | 3,160,474 | 2067 | 9 |
| HERSHEY FOODS | HSY | 3,324,131 | 2066 | 24 |
| LOEWS CP | LTR | 7,380,848 | 2111 | 7 |
| AMERN BRANDS | AMB | 9,247,860 | 2111 | 24 |
| ST JOE PAPER | SJP | 922,595 | 2631 | 2 |
| FED PAPER BD | FBO | 1,158,753 | 2631 | 16 |
| CENT NEWSPAPERS | ECP | 441,693 | 2711 | 3 |
| AFFIL PUBNS INC | AFP | 403,038 | 2711 | 10 |
| SCRIPPS E.W. CO. | SSP | 1,311,146 | 2711 | 5 |
| WASH POST | WPO | 1,946,556 | 2711 | 14 |
| MARVEL ENTMT GRO | MRV | 549,000 | 2731 | 2 |
| HOUGHTON MIFFLIN | HTN | 406,952 | 2731 | 7 |
| GRACE W R | GRA | 3,490,457 | 2819 | 15 |
| UNION CARBIDE | UK | 2,568,591 | 2819 | 27 |
| RHONE-POUL RORER | RPR | 8,926,370 | 2834 | 12 |
| UPJOHN CO | UPJ | 7,228,710 | 2834 | 39 |
| MARION MERR DOW | MKC | 10,050,522 | 2834 | 23 |
| WARNER-LAMBERT | WLA | 10,441,960 | 2834 | 39 |
| LAWTER INTL INC | LAW | 586,265 | 2893 | 3 |
| FERRO CP | FOE | 721,392 | 2899 | 6 |

Table 3. Continued.

| Name | Ticker | Mkt. Cap. | SIC | Analysts |
|---|---|---|---|---|
| TOSCO CP | TOS | 765,829 | 2911 | 5 |
| DIAM SHAM | DRM | 545,795 | 2911 | 12 |
| MURPHY OIL CP | MUR | 1,528,776 | 2911 | 10 |
| KERR-MCGEE | KMG | 1,862,807 | 2911 | 24 |
| WOLVERINE WW | WWW | 74,552 | 3149 | 1 |
| TIMBERLAND CO | TBL | 67,653 | 3143 | 5 |
| CRANE CO | CR | 743,006 | 3494 | 6 |
| KEYSTONE INTL | KII | 932,663 | 3494 | 17 |
| VARCO INTL INC | VRC | 193,446 | 3533 | 6 |
| SMITH INTL INC | SII | 280,837 | 3533 | 13 |
| AYDIN CP | AYD | 112,705 | 3662 | 2 |
| WATKINS JHNSN | WJ | 82,918 | 3662 | 10 |
| TELEDYNE INC | TDY | 1,087,480 | 3662 | 9 |
| E SYS INC | ESY | 1,226,809 | 3662 | 20 |
| FRUEHAUF TRAILER | FTC | 138,309 | 3711 | 5 |
| ALLEN GROUP | ALN | 184,612 | 3714 | 10 |
| TEXTRON | TXT | 3,403,590 | 3728 | 11 |
| LOCKHEED CP | LK | 2,849,040 | 3721 | 23 |
| ACUSON CP | ACN | 1,135,876 | 3845 | 13 |
| BARD C R INC | BCR | 1,626,157 | 3840 | 27 |
| AMSCO INTL INC | ASZ | 648,554 | 3842 | 2 |
| DIAGNOSTIC PRODS | DP | 448,203 | 3842 | 10 |
| GREEN MTN PWR | GMP | 126,939 | 4911 | 1 |
| TUCSON ELEC PWR | TEP | 115,722 | 4911 | 9 |
| EMPIRE DIST ELEC | EDE | 308,370 | 4911 | 3 |
| BLACK HILLS CP | BKH | 377,216 | 4911 | 7 |
| INTST PWR CO | IPW | 313,808 | 4911 | 3 |
| EASTERN UTIL | EUA | 344,912 | 4911 | 12 |
| IOWA-ILL G & E | IWG | 704,524 | 4911 | 7 |
| PORTLAND GEN CP | PGN | 736,473 | 4911 | 17 |
| CENT MAINE PWR | CTP | 689,425 | 4911 | 6 |
| PSI RESOURCES | PIN | 956,620 | 4911 | 18 |
| HAWAIIAN ELEC | HE | 871,122 | 4911 | 7 |
| IDAHO POWER CO | 1DA | 976,839 | 4911 | 17 |
| LG & E ENERGY CP | LGE | 1,022,295 | 4911 | 10 |
| GULF STS UTILS | GSU | 1,107,564 | 4911 | 20 |
| SEAGULL ENERGY C | SGO | 315,835 | 4922 | 12 |
| KN ENERGY INC | KNE | 258,378 | 4923 | 5 |
| MAPCO INC | MDA | 1,826,767 | 4925 | 10 |
| PANHANDLE EASTN | PEL | 1,649,738 | 4922 | 21 |

Table 3. Continued.

| Name | Ticker | Mkt. Cap. | SIC | Analysts |
|---|---|---|---|---|
| WESTCOAST ENERGY | W | 1,019,592 | 4923 | 10 |
| ENSERCH CP | ENS | 888,268 | 4924 | 20 |
| CMNWLTH ENERGY | CES | 392,696 | 4931 | 5 |
| SOUTHN IND G & E | SIG | 531,663 | 4931 | 11 |
| CENT HUDSON G & E | CNH | 451,663 | 4931 | 7 |
| PUB SVC N MEX | PNM | 407,297 | 4931 | 14 |
| CIPSCO INC | CIP | 952,545 | 4931 | 8 |
| DELMARVA P & L | DEW | 1,114,754 | 4931 | 16 |
| UTILICORP UTD | UCU | 961,505 | 4931 | 7 |
| BOSTON EDISON | BSE | 1,038,114 | 4931 | 17 |
| ARROW ELECTRS | ARW | 312,937 | 5065 | 5 |
| ANTHEM ELECTRS | ATM | 455,965 | 5065 | 14 |
| WEIS MKTS INC | WMK | 1,127,995 | 5411 | 2 |
| VONS COMPANIES | VON | 1,027,093 | 5411 | 10 |
| BALTIMORE BCP | BBB | 66,927 | 6025 | 2 |
| HIBERNIA CP LA | HIB | 77,715 | 6021 | 6 |
| STD FED BK MICH | SFB | 563,706 | 6022 | 7 |
| SIGNET BKG CP VA | SBK | 619,505 | 6025 | 14 |
| SYNOVUS FINL COR | SNV | 678,258 | 6021 | 4 |
| UJB FINL CP | UJB | 669,650 | 6022 | 10 |
| KEYCORP | KEY | 2,104,895 | 6021 | 16 |
| CHASE MANHATTAN | CMB | 2,345,603 | 6025 | 27 |
| BANK OF NEW YORK | BK | 2,168,598 | 6022 | 19 |
| WELLS FARGO CA | WFC | 3,013,100 | 6025 | 32 |
| NBB BANCORP INC | NBB | 128,400 | 6036 | 1 |
| COAST SVGS FINL | CSA | 119,648 | 6035 | 8 |
| DIME BANCORP INC | DME | 78,064 | 6036 | 6 |
| CALIF FEDL BANK | CAL | 57,654 | 6035 | 14 |
| DOWNEY SLA CA | DSL | 193,932 | 6123 | 1 |
| METROPOL FINL ND | MFC | 198,187 | 6122 | 7 |
| RELIANCE GROUP H | REL | 307,519 | 6311 | 3 |
| NWNL COMPANIES | NWN | 370,512 | 6311 | 10 |
| OLD REP INTL | ORI | 827,718 | 6311 | 3 |
| 20TH CENTY INDS | TW | 1,072,808 | 6311 | 11 |
| MUTUAL RISK MGMT | MM | 261,787 | 6331 | 5 |
| ZENITH NATL INS | ZNT | 320,439 | 6331 | 11 |
| AHMANSON H F CA | AHM | 2,015,083 | 6711 | 15 |
| 1ST INTST BCP CA | I | 1,882,350 | 6711 | 28 |
| REPUBLIC NY CP | RNB | 2,440,078 | 6711 | 10 |
| BARNETT BKS FLA | BBI | 2,330,011 | 6711 | 28 |

Table 3. Continued.

| Name | Ticker | Mkt. Cap. | SIC | Analysts |
|---|---|---|---|---|
| CITY NATL CP CA | CYN | 382,482 | 6712 | 5 |
| MNC FINL INC MD | MNC | 425,700 | 6712 | 12 |
| COMERICA INC MI | CMA | 1,673,506 | 6712 | 12 |
| MELLON BK CP PA | MEL | 1,768,511 | 6711 | 23 |

regression. The resulting regression equation is

$$\sum = \beta_0 + \beta_1 * V + \beta_2 * \Delta V * \mathrm{PI} + \beta_3 * \log(\mathrm{Volume}) + \beta_4 * \log(\mathrm{FA}) + \beta_5 * \log(\mathrm{FA1}) + \beta_6 * \log(\mathrm{FA2}) + \beta_7 * \log(\mathrm{FA3} + 5) + \beta_8 * \log(\mathrm{FA6}) + \eta. \tag{13}$$

The results of estimating Eq. (13) are reported in the second column of Table 2. The additional cross-sectional variation explained by analyst following is trivial (a mere 0.51%). The $F$-statistic for the test of deleting the analyst variables is $F(5, 111) = 1.48$ which is not statistically significant at the $\alpha = 10\%$ level. To determine whether collinearity is a problem, we calculated variance inflation factors (VIFs) for the dependent variables.[29] The largest VIF factor is 2.84, which along with the fact that the regression coefficients were fairly stable when independent variables are added or subtracted from the regression equation, suggests that collinearity is not a serious problem for our independent variables.

Finally, because our matching on firm size is not perfect, we performed a Chow test to determine if our results could be driven by differences in market capitalization. The 60 pairs were ranked based on percentage difference in market capitalization and divided into two groups of 30 pairs each. Eq. (13) was estimated for each sub-sample, and we then tested the null hypothesis of no structural change between the sub-samples. The test statistic for this test is $F(4, 112) = 1.27$, which is not significant at the $\alpha = 5\%$ level. We conclude from this that our results are not driven by differences in market capitalization.

[29] The VIF for a variable is the inverse of one minus the $R^2$ of the regression of that variable on all of the other independent variables. A value at or above 10.0 is usually regarded as an indication of serious collinearity (see Neter et al., 1983).

## 6. Conclusions

How financial analysts affect the structure of information-based trade in stocks is a subject of widespread interest. In this paper, we investigated this issue by estimating the structure of trade for a sample of stocks with high and low analyst following. Our analysis revealed a number of important insights. We showed that analysts do not appear to create private information; the probability of private information events is the same for stocks with high or low analyst following. Similarly, despite the widely reported bias in analysts' reports, they do not introduce any difference into the probability that information on a stock is good or bad news. Where analyst coverage does matter is in the composition of trade. We found that stocks with high analyst following have both higher rates of informed and uninformed trading. Overall, however, high analysts stocks face a lower probability of information-based trading.

We also showed that financial analysts are not a good positive proxy for information-based trading. Indeed, financial analysts are actually associated with a *reduced* level of information-based trading. This dictates that studies using analysts as positive proxies for informed trading cannot hope to capture the true relationship. A better approach would be to directly estimate the extent of information-based trading using an approach such as that developed here. As our regression results show, our estimated probability of informed trade has significant explanatory power with respect to actual spread behavior.

While our findings provide new insights into the role of financial analysts, much remains to be learned. Our results relate analyst following to the overall level of information-based trading in a stock. Yet, it may be that the effect of analysts is more time-specific. In particular, perhaps it is at specific times such as before earnings announcements that analyst behavior is informed. In that case, analysts' trading might have the effect of increasing the speed with which new information is impounded into prices, a result consistent with extant empirical evidence. Similarly, it may be that analysts affect the composition of trades in a more complex fashion than is tested for here. In particular, if analysts skew the distribution of trade size towards larger orders, then order size effects may arise. Finally, it may be that analysts coverage differentially affects a stock's reaction to public information, an issue which cannot be addressed with our private information-based model. It would also be interesting to know whether analysts following can explain any of the very puzzling intraday seasonal effects found in security trading. We look forward to addressing some of these issues in future research.

## Acknowledgements

We would like to thank an anonymous referee, the editor Avanidhar Subrahmanyam, Paul Irvine, and seminar participants at Boston University, Columbia

University, Cornell University, the University of North Carolina, Notre Dame, Rice University and the Western Finance Association meetings for helpful comments. We also thank Gideon Saar for programming assistance and I/B/E/S for providing access to analysts data. This research was supported by the National Science Foundation (Grant SBR-9631583).

## References

Bhushan, R., 1989. Firm characteristics and analyst following. Journal of Accounting and Economics 11, 255–274.

Brennan, M., Hughes, P., 1991. Stock prices and the supply of information. Journal of Finance 46, 1665–1691.

Brennan, M., Jegadeesh, N., Swaminathan, B., 1993. Investment analysis and the adjustment of stock prices to common information. Review of Financial Studies 6, 799–824.

Brennan, M., Subrahmanyam, A., 1995. Investment analysis and price formation is securities markets. Journal of Financial Economics 38, 361–381.

Chung, K.H., McInish, T., Wood, R., Wyhowski, D.J., 1995. Production of information, information asymmetry and the bid-ask spread: empirical evidence from analysts forecasts. Journal of Banking and Finance 19, 1025–1046.

Easley, D., Kiefer, N., O'Hara, M., 1997a. One day in the life of a very common stock. Review of Financial Studies 10, 805–835.

Easley, D., Kiefer, N., O'Hara, M., 1997b. The information content of the trading process. Journal of Empirical Finance 4, 159–186.

Easley, D., Kiefer, N., O'Hara, M., Paperman, J., 1996. Liquidity, information, and infrequently traded stocks. Journal of Finance, 1405–1436.

Goldberger, A., 1991. A course in econometrics. Harvard University Press, Cambridge.

Hasbrouck, J., 1988. Trades, quotes, inventory, and information. Journal of Financial Economics 22, 229–252.

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

Merton, R., 1987. A simple model of capital market equilibrium with incomplete information. Journal of Finance 42, 483–510.

Neter, J., Wasserman, W., Kunter, M., 1983. Applied linear regression models. Richard D. Irwin, Homewood, IL.

O'Brien, P., Bhushan, R., 1990. Analyst following and institutional ownership. Journal of Accounting Research 28, 55–76.

Shores, D., 1990. The association between interim information and security returns surrounding earnings announcements. Journal of Accounting Research 23 (1).

Skinner, D., 1990. Option markets and the information content of accounting earnings releases. Journal of Accounting and Economics 13, 191–211.

Womack, K.L, 1996. Do brokerage analysts recommendations have investment value? Journal of Finance 51, 137–167