The Probability and Magnitude of Information Events

Elizabeth R. Odders-White[†]   Mark J. Ready[‡]

Department of Finance
School of Business
University of Wisconsin – Madison
Madison, WI 53706

April 2006

Abstract

Models of adverse selection risk generally assume that market makers offset expected losses to informed traders with expected gains from the uninformed. We recognize that the expected loss captures a combination of two effects: 1) the *probability* that some traders have private information, and 2) the likely *magnitude* of that information. We use a maximum-likelihood approach to separately estimate the probability and the magnitude of private information and test our procedure on a simulated data set. We then estimate the parameters for NYSE-listed stocks from 1993 through 2003, and show that our estimates can be used to predict future extreme returns. Finally, we examine the time-series and cross-sectional properties of the probability and magnitude of information and find a decline in the frequency of private information events following the introduction of Regulation FD. Our results shed light on the price discovery process and have implications for many areas of finance.

1. Introduction

Liquidity suppliers in securities markets are always aware that other traders may have better information. This idea, which was first formalized in seminal work by Kyle (1985) and Glosten and Milgrom (1985), has spurred an extensive theoretical and empirical literature devoted to quantifying adverse selection risk. Kyle (1985) modeled the behavior of a single market maker who sets a "break-even" price based on the net combined order flow of informed and uninformed traders. The equilibrium price allows the market maker to offset expected losses to the informed trader with expected gains from the uninformed. Thus, the price impact is a function of the degree of asymmetric information in the market and provides an indication of the expected loss to the informed trader. Glosten and Milgrom (1985) consider a slightly different setting but rely on the same basic tradeoff between gains from uninformed and losses to better informed traders.

Liquidity suppliers' expected loss to informed traders has been measured empirically in a variety of ways. The most easily observable metrics are quoted and effective spreads. (The effective spread is a simple variant of the quoted spread that accounts for the fact that trades may occur at prices inside or outside the quotes.) Although the effective spread stems purely from asymmetric information in the Glosten and Milgrom (1985) model, empirical studies have shown that spreads may also reflect other types of costs, like order processing and inventory risk. Several authors have developed approaches to estimating the information component of the effective spread by isolating the permanent impact of each trade.[1] An alternative method of separating the transitory effects from the expected loss to informed traders is to simply measure

---

[1] See, for example, Roll (1984), Glosten and Harris (1988), Stoll (1989), George, Kaul, and Nimalendran (1991), Hasbrouck (1991), Lin, Sanger, and Booth (1995), and Huang and Stoll (1997) for models that decompose effective spreads.

the aggregate price impact over time intervals that encompass several trades, as in Breen,

Hodrick, and Korajczyk (2002), among others.

We recognize that each of these measures of expected loss captures a combination of two

effects: 1) the *probability* that some traders have private information, and 2) the likely *magnitude*

of that information. The extant literature has focused primarily on the combined effect, rather

than the probability and the magnitude separately. The notable exception is a model developed

by Easley, Kiefer, O'Hara, and Paperman (1996).[2] Although their model is most often used to

estimate the probability that a particular trader is informed (PIN), PIN is simply a function of the

underlying parameters in their model, which include the probability that some traders receive a

private signal (i.e., the probability of a private information event). Easley et al. (1996) do not

estimate the magnitude of private information, nor do they consider the relation between the

probability of an information event and the expected loss to informed traders.

Separate consideration of the probability and the magnitude of private information leads

to a variety of insights regarding adverse selection risk and the way information is incorporated

into prices. For example, suppose Firm A has frequent private information events with little

value impact, whereas Firm B has *in*frequent private information events with *large* value impact.

Although the two firms may have similar expected losses to informed traders, differences

between the probability of an information event and the magnitude of that event lead to different

predictions regarding the return patterns of Firms A and B. In particular, firms with a low

probability of information events will have more return reversals, while firms with a high

probability of information events will have more continuations.

---

[2] Bernhardt and Hughson (2002) provide a model where information events arrive *before each trade* with probability γ, and are publicly revealed following each trade. In their empirical estimation they assume $\gamma=1$, but they also assume informed traders randomly choose to be informed, so there are some individual trades that have no information component.

We propose a simple model that formalizes this idea. We then develop a method for separately estimating the components of the expected loss to informed traders. We test our estimation procedure using a simulated data set, and then estimate the probability and the magnitude of private information for NYSE-listed stocks using rolling one-year windows from 1993 through 2003. We demonstrate that our parameter estimates can be used to predict the probability of future extreme overnight returns. This not only offers additional evidence of the validity of our approach, but also has implications for option pricing, risk management, and corporate finance.

Finally, we examine the time series and cross-sectional properties of our estimates of the probability and magnitude of information events. We find a positive relation between firm size and adverse selection costs in our sample that is driven by the higher frequency of information events for larger firms. We also find that adverse selection risk is related to past volatility – not because private information events are more frequent for firms with more volatile overall stock returns, but because events tend to be larger in magnitude for these firms. We further show that the probability of information events declined dramatically in late 2000, the period corresponding to the implementation of the SEC's Regulation FD (Fair Disclosure). We also find that the positive relation between the probability of private information events and firm size is attenuated in the period following Regulation FD. Together, these results suggest that Regulation FD had the intended result of reducing the flow of private information, especially for large firms.[3]

---

[3] This finding is consistent with Eleswarapu, Thompson, and Venkataraman, (2004) but inconsistent with Sidhu, Smith and Whaley (2005).

2. Model

We begin the discussion of our model with an example. Consider a simple setting in which informed and uninformed traders trade each stock at prices set by a single competitive market maker. The market maker knows that Firm A is likely to have a private information event, but does not know with certainty whether an event has actually occurred. Thus, he moves the price in response to order flow to reflect the expected value of the information event. This expectation includes the possibility of no event, so if the market maker subsequently learns that a private information event did in fact take place (which is generally the case for Firm A), he will incorporate the full impact of the private information, usually resulting in a small price continuation. On the few occasions that an event does not occur, the market maker will subsequently learn that he was reacting to order flow that was entirely composed of liquidity traders, so he will completely reverse the previous price response.

In contrast, Firm B has high information risk but little realized informed trading, so the price impact serves as protection against large events that rarely occur. Consequently, in most cases, the market maker will reverse the original price impact after learning that the prior period's trading was not information-motivated. In the rare cases when information events do occur, the initial price response to the order flow will tend to be too small, because it was based on the market maker's low ex-ante probability. Thus, there will tend to be relatively large price continuations in the following period.

We now formalize the intuition in this illustrative example. Each period in our model resembles the single-period Kyle (1985) model, in which a market maker sets a price upon observing the net combined order flow from a single informed trader and uninformed traders. We assume that the single trading period in the Kyle model corresponds to one trading day

because this is the interval we will use when estimating the model's parameters in Section 4.

(Our model does not require any specific assumption about the length of the period.)[4] We

further assume that the true value of the private signal is revealed before trading on the following

day. Our setting differs from the Kyle model in that the informed trader receives a private signal

on day t with probability α, and only trades if such a private information event has occurred.

Thus, the market maker faces uncertainty as to whether the total order imbalance originated from

uninformed traders alone or from both the uninformed and informed traders. In the Kyle (1985)

model, $\alpha = 1$, so the net order always reflects the orders of both the informed and uninformed.

In our model, the informed trader's private signal is not the only channel of information

about the value of the stock. We also allow for public news arrival, both during the day (denoted

$r_{pd,t}$) and after the close of trading (the "overnight" public return, denoted $r_{po,t}$). These returns are

normally distributed with mean zero and variance, σ_{pd}^2 and σ_{po}^2, and they are independent of all

other random variables in the model. Public news is simultaneously observed by all market

participants; therefore, public information is not impounded into prices through order flow and

has no effect on the informed trader's strategy. The public information shocks do not add any

new insights to our model, but we include them because they are important for the estimation of

the parameters. Without the publicly-observed shocks, our maximum likelihood procedure

would be forced to attribute all of the price volatility to private information.

Our model can be described as follows. As in the Kyle model, the net order from the

uninformed traders, u_t, is normally distributed with mean zero and variance σ_u^2, and is

independent of all other random variables in the model, including the net order from the

informed trader, x_t. The risk-neutral market maker's response to the total order imbalance, y_t, is

---

[4] The impact of different timing assumptions on our estimates is examined in Section 3.3.

measured as a return, $r_{y,t}$, and is assumed to be linear with slope λ (i.e., $r_{y,t} = \lambda y_t$). Note that Kyle (1985) defines λ as the *dollar* change in share price (per share of order imbalance), while we measure λ as the *percentage* change in share price (per share of order imbalance). Here again, our motivation is empirical rather than theoretical. We will be estimating our parameters over one year, and we believe that public and private information shocks are more likely to be stationary in percentage terms than in dollar terms. We measure both intra-day and overnight returns as price changes (plus distributions) divided by the quote midpoint at the start of the day, so the full-day return between two successive opening quotes is just the sum of the intra-day and overnight returns.

As in Kyle (1985), the informed trader chooses x_t to maximize expected profit given the private signal and given the market maker's assumed response. Each day, with probability α the informed trader receives a private signal, $r_{i,t}$, that is normally distributed with mean zero and variance σ_i^2; $r_{i,t} = 0$ with probability $(1-\alpha)$. $r_{i,t}$ is independent of both the uninformed volume on day t and any public information. Given the market maker's response, the informed trader's optimal strategy is to choose $x_t = r_{i,t}/(2\lambda)$; thus, if an event occurs, $x_t \sim N(0, \sigma_i^2/(4\lambda^2))$, as in Kyle (1985). On non-event days, $x_t = 0$.

We assume that the market maker chooses λ so that his expected profit is zero. The market maker's expected profit is $p_t E[y_t(r_{y,t} - r_{i,t})]$, where $r_{y,t} = \lambda y_t$, p_t is the pre-trade price, and $r_{i,t}$ represents the informed trader's private signal. In the appendix, we show that this expected profit equals zero when

$$\lambda = (1/2)\alpha^{1/2}(\sigma_i/\sigma_u) \tag{1}$$

Note that when events occur every day, $\alpha = 1$ and expression (1) reduces to that given in Kyle (1985), with one modification. Because of our different definition of λ, our expression includes

σ_i, the standard deviation of the *return* (proportional price impact) associated with the private information event. In contrast, Kyle's expression for λ contains $\Sigma_0^{1/2}$, which is the standard deviation of the *dollar* per share impact of the private information event.

When events do not occur every day, our setup differs from Kyle's in that the linear pricing rule shown in equation (1) does not provide zero profits conditioned on the order flow. Specifically, smaller values of y_t will be more common on days with no private information event, so the adjustment given in (1) will be too large (the market maker will have positive conditional expected profit). Similarly, on days with large y_t, it is more likely that an information event has occurred so the market maker's conditional expected profit will be negative.[5] We believe, however, that our assumed linear pricing rule is reasonable given the NYSE specialist's affirmative obligation to maintain a "fair and orderly market." This obligation may impose costs on the specialist during times of large order imbalances that may be recouped on days with less extreme imbalances.[6]

Also note that rearranging equation (1) yields

$$\lambda\sigma_u = (1/2)\alpha^{1/2}(\sigma_i)$$

The left hand side of this equation is the price impact associated with one standard deviation of uninformed order flow. This price impact is the cost borne by the uninformed traders. The right hand side expresses this cost as a function of the two components that are the focus of this paper: the probability of information events (α) and their likely magnitude (σ_i). One of our empirical questions, addressed later in the paper, is whether there are substantial differences in the relative importance of these components across firms.

---

[5] See Spiegel and Subramanyam (2000).
[6] Our assumption is consistent with Hansch, Naik, and Viswanathan's (1999) finding that market makers on the London Stock Exchange make money on small trades and lose money on larger trades.

The daily sequence of events in our model can be summarized as follows:

1. Public (intraday) news, $r_{pd,t}$, arrives.

2. The informed trader receives a private signal, $r_{i,t}$, with probability α.

3. Uninformed traders submit a net order, u_t. The informed trader submits an order, x_t.

    (If $r_{i,t} = 0$, then $x_t = 0$.) The market maker observes the net combined order flow, $y_t =$

    $u_t + x_t$, and determines the price change using $r_{y,t} = \lambda y_t$. Trade occurs. The return

    from the opening quote midpoint to the closing quote midpoint is $r_{d,t} = r_{pd,t} + r_{y,t}$.

4. Public (overnight) news, $r_{po,t}$, arrives.

5. The private signal observed at the start of the trading day, $r_{i,t}$, becomes public. The

    return from the closing quote midpoint on day t to the opening quote midpoint on

    day $t+1$ is $r_{o,t} = r_{po,t} + r_{i,t} - r_{y,t}$, which includes the difference between the true signal

    on day t and the market maker's response to the order imbalance that day.

The total return from the opening quote midpoint on day t to the opening quote midpoint
on day $t+1$ is:

$$r_t = r_{d,t} + r_{o,t} = (r_{pd,t} + r_{y,t}) + (r_{po,t} + r_{i,t} - r_{y,t}) = r_{i,t} + r_{pd,t} + r_{po,t} \qquad (2)$$

3. The frequency of reversals

It is clear that our model will generate returns that will be a mixture of distributions,
mixing days with events and days without events. What may be less clear is that our model also
generates particular patterns in adjacent intraday and overnight returns, with sign reversals being
far more common for firms with low levels of α. The maximum likelihood procedure uses all of
these features of the return distributions. Before describing our procedure in detail, we discuss
the intuition behind the relation between α and the frequency of reversals, where reversals are

defined as cases where the trade imbalance, y_t, and the subsequent overnight return, $r_{o,t}$, have

opposite signs.

In our simple model, there are no costs of order processing and no compensation for

dealer inventory, so the price-impact return, $r_{y,t} = \lambda y_t$, stems solely from adverse selection risk.

On $1-\alpha$ of the days the market maker learns that no event occurred (i.e., $r_{i,t} = 0$), so he fully

reverses the initial price impact. In these cases the overnight return, $r_{o,t}$, is more likely to have

the opposite sign of the imbalance, y_t. (When no event occurs, $r_{a,t}$ and y_t will have the opposite

signs unless the public overnight return, $r_{po,t}$, is large and has the same sign as $r_{y,t}$.) In contrast,

on α of the days the market maker learns that a private information event did occur, and in these

cases he will generally move the next day's opening price in the same direction as y_t because he

does not fully incorporate the effect of the private information on the first day.[7] Thus, firms with

high probabilities of information events (α) will have fewer reversals than firms with low α's.[8]

To illustrate more concretely the relation between the frequency of private information

events, α, and the frequency of reversals, we simulate a data set and examine its properties. In

the simulations, we let α assume ten different values (0.05, 0.10, 0.15, 0.20, 0.25, 0.35, 0.45,

0.55, 0.65, and 0.85), and we let the standard deviation of the informed trader's signal, σ_i, range

from 0.02 to 0.10 in increments of 0.02. The volatility of the intraday public information, σ_{pd}, is

fixed at 0.02, and the volatility of overnight public information is set to 0.01, both of which are

approximately equal to the average estimated values from the NYSE data described in Section

---

[7] When an event occurs, $r_{o,t}$ will have the same sign as y_t unless (a) the overnight public return is large and has the opposite sign of the imbalance, or (b) u_t (the uninformed part of the order imbalance) is large and has the same sign as x_t (the informed order).
[8] In our model, as in the single-period Kyle (1985) model, approximately half of the private information is incorporated into the security price as a result of the informed trading, and the remainder is incorporated when the event becomes public. If the informed trader were allowed to trade multiple times throughout the day, and the market maker could update his estimate of the probability that an information event took place by observing the order flow in each trading interval, the reversal effect would be reduced. In Section 3.3, we simulate a model with two rounds of trade to investigate the impact of this alternative assumption on our estimates.

4.3. These parameter values yield 50 different combinations of α and σ_i. Each combination is viewed as a separate "firm." For each firm, we generate 500 years of daily data (252 observations/year). Total daily returns are determined according to equation (2), using the simulated private signals to generate informed order flow.[9]

The results in Table 1 demonstrate the relation between reversal frequencies and α and σ_i. Reversal frequencies are computed as the fraction of observations for which the imbalance and the overnight return have opposite signs. The table demonstrates that reversal frequencies are generally decreasing in α, as expected. Not surprisingly, the pattern is most pronounced for high values of σ_i because when the magnitude of private information is small, its effect on returns is more easily swamped by the public component. For a given α, reversal frequencies are increasing in σ_i for exactly the same reason. The relation is not quite monotone in α, because for very small values of α, the price response to order flow is very small, so although it is nearly always reversed (for roughly 1-α of the observations), these small reversals are swamped by the overnight public information.

Llorente et al. (2002) also look at probabilities of return reversals, but in a different context. They model risk-averse investors who trade either for hedging or speculative purposes. The relative importance of these two motives determines whether returns will exhibit reversals (because hedging trades dominate) or continuations (because speculative trades dominate). In contrast, in our model investors are risk-neutral, so there is no hedging motive for trade. Instead, the uncertainty regarding the occurrence of information events, coupled with the

---

[9] The parameter σ_u, which is the standard deviation of uniformed order flow, does not affect the return pattern, because the informed trader's intensity is adjusted to match it. As a result, the informed trader's impact on prices is only a function of σ_i and α, and the pattern in returns is only a function of σ_i, α, σ_{pd}, and σ_{po}.

specialist's positive obligation to maintain fair and orderly markets, leads to underreaction

(continuations) or overreaction to order flow (reversals).

4. Estimating α, σ_i, σ_{pd}, and σ_{po}

Our maximum likelihood approach uses the information in the full sequence of returns

and order imbalances to estimate α and σ_i, as well as σ_{pd}, σ_{po}, and σ_u. Intuitively, two features of

the data allow separate identification of α and σ_i:

- Reversals between order flow and overnight returns (as illustrated in Table 1)

- Kurtosis in the return distribution (specifically, days with information events have

    higher return variance)

The maximum likelihood approach uses the fact that conditioned on whether there has

been a private information event, the three variables y_t, $r_{d,t}$, and $r_{o,t}$ are jointly normally

distributed. In the appendix, we derive conditional variance/covariance matrices as functions of

the parameters α, σ_i, σ_{pd}, σ_{po}, and σ_u.[10] The unconditional density for y_t, $r_{d,t}$, and $r_{o,t}$ is obtained

by weighting the two conditional densities by $(1-\alpha)$ and α, respectively. The parameter

estimates are obtained by maximizing the following log-likelihood function:

$$L(\alpha,\sigma_u,\sigma_i,\sigma_{p,d},\sigma_{p,o}) = \sum_{t=1}^{N} \ln \left[\begin{array}{c} (1-\alpha) f_n \left(y_t, r_{d,t}, r_{o,t}; \alpha, \sigma_u, \sigma_i, \sigma_{p,d}, \sigma_{p,o} \right) \\ + \alpha f_e \left(y_t, r_{d,t}, r_{o,t}; \alpha, \sigma_u, \sigma_i, \sigma_{p,d}, \sigma_{p,o} \right) \end{array} \right] \qquad (3)$$

where f_n and f_e are the conditional, multivariate normal densities on day t assuming no event

occurs and an event occurs, respectively.

---

[10] Although our model shares some features with that of Bernhardt and Hughson (2002), our empirical estimation procedure is quite different. One obvious difference is that they estimate their model trade-by-trade, whereas we use aggregate intraday and overnight periods. More importantly, they use unsigned trade volumes. In contrast, our use of signed trade volumes allows us to exploit the conditional covariances between trade volumes and price changes illustrated in Table 1.

4.1. Estimates from simulated data

Before proceeding with our analysis, we test the accuracy of our estimation procedure on the simulated data set described in Section 3. We maximize the likelihood function in equation (3) for each firm year, restricting the estimates of α, σ_i, σ_{pd}, and σ_{po} to lie between 0.00001 and 1. Although the standard deviations could theoretically exceed 1, it would imply that one-day returns often exceed $\pm 100\%$. Furthermore, the true σ_i values range from 0.02 to 0.10. The maximum likelihood estimation procedure converges to a boundary for 0.4% of the firm years. Results for the remaining observations are summarized in Tables 2 and 3.[11]

Table 2 compares the α estimates to the true values. The estimates appear to be unbiased, and the precision (true value/standard deviation of estimates) is increasing in both α and σ_i. The results for the σ_i estimates in Table 3 are similar. These estimates are unbiased, and the precision is increasing in σ_i and increasing in α for all but the smallest value of σ_i. The fact that the precision is tied to the values of α and σ_i is not surprising. Attempting to separately estimate the probability and the magnitude of private information has little value when virtually no information arrives through order flow (i.e., when α and σ_i are small).

4.2. The effect of different timing

Our model and estimation procedures assume that private information events arrive each day, and that the private information becomes public at the end of each trading day. Furthermore, the model assumes that the day's trading is well-approximated by the one-round version of Kyle (1985). In this subsection, we investigate the sensitivity of the estimates to these

---

[11] The α estimates equal 1 for another 0.2% of the observations. We do not omit these observations, as $\alpha = 1$ is simply the standard Kyle (1988) model.

two assumptions by simulating data from alternative models with *different timing assumptions* and then applying our estimation procedure to the simulated data. Thus, in this subsection, the assumptions underlying the estimates are inconsistent with the process that generates the data. Our purpose is to investigate the sensitivity of our estimates to these alternative timing assumptions.

In our first alternative model, we assume that private information events become public after two days, giving the informed investor two opportunities to trade based on his or her private signal. In this alternative model there can be two private information events "active" during a day – one observed at the start of the current day and another observed at the start of the previous day. We assume that different informed traders observe each of these signals. The quote midpoint is set to the expected value based on public information at the start of each day, and informed traders are risk neutral, so each can ignore the possible presence of the other informed trader when determining their optimal trade. In our second alternative model, information events last only one day, but there are two rounds of trading during each day.

In developing our alternative models, we are interested in isolating the importance of the particular timing assumption, so we try to maintain as much similarity as possible to our original model. Information events still arrive at the start of each day with probability α and their magnitude is normally distributed with standard deviation σ_i. We continue to assume that the market maker follows a linear pricing rule, and in each round of trading we assume that the informed trader uses a trading strategy that is linear in the difference between the current quote midpoint (which is equal to the expected value based on all public information) and his or her updated expected value.

In both alternative models the informed traders trade twice based on their signal. The second time they trade, their decision is the same as in Kyle (1985): their optimal order is equal to the difference between their updated estimate of the value and the quote midpoint multiplied by $1/(2\lambda)$. In the first round of trade, we limit the informed trader to a linear strategy for analytical tractability, and we solve numerically for the optimal trade intensity parameter.

In the first alternative model, we assume that the market maker updates the quote midpoint at the start of each day to reflect the new expected value of the asset based on the observed total order flow. The market maker observes the public revelation of the private information event two days prior (if an event occurred), so he can calculate the part of the past day's volume due to that earlier event. The remainder of yesterday's volume must be from the uninformed traders and from potential informed trading associated with an event that occurred yesterday that is not yet public. In the second alternative model, the market maker faces a similar updating problem in the middle of each day, after the first round of trading.

The potential equilibria in our alternative models are characterized by three parameters: the market maker's price response (λ) and the informed traders' optimal trading intensity in each round (β_1 and β_2). We define a combination of these three parameters to be an equilibrium if the following conditions are satisfied:

- The market maker's expected profit is zero
- $\beta_2 = 1/(2\lambda)$
- The market maker conjectures that the informed traders use β_1 as their trading intensity in the first period and updates quote midpoints based on this conjecture. Given this, it is in fact optimal for the informed traders to use this same intensity in the first period (i.e., the conjecture is borne out).

For each combination of α and σ_i, we solve numerically for the equilibrium values of λ, β_1 and β_2. In order to estimate the necessary expected values, we simulate 500,000 trading days (approximately 2,000 years) of information events and uninformed volumes, and we consider various combinations of the parameters, using the average values of market maker and informed trader profits as proxies for expected values. We start by choosing a trial value of λ. For each given λ, we select an initial estimate for the market maker's conjectured value of β_1 and then find the actual value that maximizes the average informed trader profit. The more sensitive the market maker to order flow, the more that aggressive trading by the informed trader in the first period reduces the opportunity for profitable trade in the second period. Accordingly, the higher the market maker's conjectured value of β_1, the lower the informed trader's optimal choice, and there is a point where the two values coincide. For each given λ, we find this equilibrium value of β_1 (at which the conjectured value and optimal choice coincide), and then we search over λ to find the level that drives the average market maker profit to zero.

For each combination of α and σ_i used in Tables 1 through 3, we find the equilibrium values for λ, β_1 and β_2, and simulate 500 years of daily data using these values. We then apply our estimation procedure to these simulated data. Table 4 shows the result of this exercise using the alternative model with information events that last two days. Not surprisingly, the estimated values of α are higher than the true values, because the impact of each information event is spread across two days. Importantly, the biases in the estimates of α are not strongly related to the level of σ_i. Table 5 shows the results using the alternative model with two rounds of trading during the day. In this case, the estimates are reasonably close to the true values, with the exception of the lowest values of α. Again, the biases are not strongly related to the level of σ_i.

In summary, the above simulation results suggest that even if the true timing of information events and rounds of trading differ from the one-day assumption in our model, one can still reasonably compare the estimates of α and σ_i across firms. Of course, this conclusion is subject to the caveat that the timing be similar across the firms being compared. For example, suppose firms of types X and Y have similar levels of α, but firms of type X tend to have two days between the private and public observation of information events, whereas firms of type Y have just one day. In this case, firms of type X will have higher *estimated* α's than firms of type Y.

4.3. Estimates from actual data

We now apply our estimation method to a sample of NYSE-listed stocks. We consider 40 overlapping one-year time windows, starting each January, April, July, and October. The first window covers the period from January-December 1993, and the final window covers the period from October 2002-September 2003.[12] In each one-year time window, we include all common stocks from the CRSP database that are listed on the NYSE over the entire period, with no symbol or CUSIP changes, no stock splits or other unusual distributions, and no cash dividends in excess of 10% of the ex-dividend stock price. We also eliminate any stocks where the company has multiple issues included in CRSP.

For each stock we collect intraday trade and quote data from the TAQ database. We sign each trade using the Lee and Ready (1991) algorithm, aligning trades with the most recent NYSE-quote that has been in effect for at least 5 seconds. We measure the share imbalance using all NYSE trades reported between the open and the close, except those greater than 10,000

---

[12] We exclude the one-year window ending in December 2003 because in Section 4 we use our estimates to predict extreme returns for the next quarter.

shares.[13] Each day, we also measure quote midpoint returns from the open to the close and from

the close to the next day's open, adjusting overnight returns for dividends using data from CRSP.

Our model focuses on firm-specific news, so we purge market-wide components from

both intraday and overnight returns by running daily cross-sectional regressions of each return

measure on a constant, historical beta (based on the previous 5 years of monthly returns), the

natural logarithm of market capitalization, and the natural logarithm of the book-to-market ratio.

These independent variables are motivated by Fama and French (1992, 1993), and all are

measured as of the start of the one-year estimation window.[14] The residuals from these

regressions become our estimates of $r_{d,t}$ and $r_{o,t}$.

Similarly, the price impact measure in our model captures each stock's price change in

response to *unexpected* trade imbalance. Imbalance has been shown to be persistent [Chordia

and Subrahmanyam (2004)], and subject to market-wide factors, so after scaling imbalance by

the number of shares outstanding, we run daily cross-sectional regressions on a constant, the

previous day's imbalance, and market capitalization [see Corwin and Lipson (2005)]. The

residuals from these regressions become our estimates of y_t.

As in the simulations, we omit firm years that converge to a boundary (3% of the

sample)[15]. We also omit firm years with fewer than 200 days with non-missing data or whose

numerical estimates failed to converge (0.2% of the sample). Mean α and σ_i estimates across the

---

[13] We eliminate trades of over 10,000 shares to reduce the impact of very large transactions that are likely to be uninformed, liquidity trades.

[14] Prior to taking logs, we impose minimum and maximum values for the book-to-market ratio of 0.017 and 3.13, respectively. These cutoffs are approximately equal to the cutoffs used by Fama and French (1992) for the last year of their sample. They impose these cutoffs (particularly the lower bound) to avoid extreme values after taking logs. Unlike Fama and French (1992), we do not discard observations with negative book values. Instead, we set the book-to-market ratio equal to one (so the natural logarithm is zero) and include a separate dummy variable in each regression that identifies these firms.

[15] 50% of these are cases where σ_{po} converged to the lower boundary (0.00001), and 41% are cases where α and/or σ_i converged to the lower boundary (also 0.00001), so the estimated adverse selection is essentially zero. The results of our empirical tests are similar when all boundary observations are included.

remaining 50,770 firm-year observations are 0.44 and 0.02, respectively. The mean intraday public information parameter, σ_{pd}, is 0.02, more than double the average value of the overnight public information parameter, σ_{po}, of 0.008. Additional summary statistics on the estimates of α and σ_i are presented in Table 6. The table displays mean α estimates for each decile of the α distribution. Average α's range from 0.038 for the lowest decile to 0.964 for the highest decile. For each α decile, Table 6 also shows average σ_i estimates and quintiles of the σ_i distribution. α and σ_i appear to be negatively related, but there is substantial dispersion in the σ_i estimates within each α decile. The results suggest that the way information is incorporated into prices varies across firms – for example, via frequent small events (high α and low σ_i) versus rare large events (low α and high σ_i) – indicating that the decomposition of the expected loss to informed traders into the probability and the magnitude of private information is a useful exercise. In Section 6, we further examine time-series and cross-sectional properties of the parameter estimates.

5. Predictions for future extreme returns

We have tested the validity of our estimation procedure using simulations, but it would also be helpful to confirm that the empirical estimates behave as expected. One way to do so is to examine the relation between the estimates of α, σ_i, σ_{pd}, and σ_{po} and future extreme returns. Intuitively, the probability of an extreme return will be concave in α. When α is very low, information events almost never occur, so returns are approximately normally distributed. Likewise, for very high levels of α, events occur nearly every day, so the distribution of the daily return is also approximately normal. For intermediate levels of α, daily returns will follow a truer mixture of normal distributions, increasing the likelihood of extreme outcomes.

In addition, firms with high values of σ_i (relative to σ_{pd} and σ_{po}) are more likely to experience extreme returns. In our model, returns due to public information are normally distributed and, therefore, will rarely be extreme. In contrast, the infrequent arrival of private information events can generate extreme returns. The larger the magnitude of this "lumpy" private information relative to the "smooth" public information, the more extreme the event.

We test whether extreme returns in the next quarter can be predicted using estimates of α, σ_i, σ_{pd} and σ_{po} from the current year. This analysis not only validates our empirical estimates, but also has implications for areas of finance like risk management, option pricing, and corporate finance, in which the ability to predict extreme outcomes is valuable.

Let r_{qmax} represent the maximum absolute daily excess return ($r_t = r_{d,t} + r_{o,t}$) over the quarter following the current year, and let σ_r equal the robust standard deviation of daily excess returns over the current year. Robust standard deviations are computed as the 68.27[th] percentile of the return distribution.

We define three different levels of extreme events: $e_n=1$ if $r_{qmax} > n\sigma_r$ for $n \in \{4,5,6\}$. We let P_n represent the true probability that $e_n=1$ and $P_n^{est}(\alpha,\sigma_i,\sigma_{pd},\sigma_{po})$ represent the predicted probability that $e_n=1$ given our estimates of α, σ_i, σ_{pd}, and σ_{po}. The probability that the *maximum* absolute return will be above $n\sigma_r$ is equal to one minus the probability that *all* of the observations will be *below* $n\sigma_r$. The probability that a single day's return will be below $n\sigma_r$ depends on whether an event has occurred. If Q is the number of trading days in the quarter, then

$$P_n^{est}(\alpha,\sigma_i,\sigma_{pd},\sigma_{po}) = 1 - \Pr\{|r_t| \le n\sigma_r\}^Q \tag{4}$$

where

$$\Pr\{|r_t| \le n\sigma_r\} = \alpha\left(1 - 2\Phi\left(-n\sigma_r \Big/ \sqrt{\sigma_i^2 + \sigma_{pd}^2 + \sigma_{po}^2}\right)\right) + (1-\alpha)\left(1 - 2\Phi\left(-n\sigma_r \Big/ \sqrt{\sigma_{pd}^2 + \sigma_{po}^2}\right)\right)$$

and $\Phi(\bullet)$ is the standard normal cumulative distribution function.

We estimate probit regressions to test the ability of our model to predict which firms will experience extreme returns over the coming quarter. We start with the 48,326 observations used in Table 6 and delete 69 observations for which there are no good returns in the following quarter. We are particularly interested in the ability of the model to capture the possibility of extreme returns when they have not yet occurred (exploiting the price sensitivity to order flow and subsequent reversals), so we exclude 6,375 observations where the maximum absolute daily return during the estimation year was more than 10 times the robust standard deviation.

We estimate three different specifications. The first specification includes only an intercept and the predicted probability:

$$P_n = \Phi\left[a + b\Phi^{-1}\left(P_n^{est}\left(\alpha, \sigma_i, \sigma_{pd}, \sigma_{po}\right)\right)\right] \tag{5}$$

Note that if the predicted probability of an extreme return using our estimates is equal to the actual probability, then the intercept, a, will be zero and the slope, b, will be one. The second and third specifications add control variables to equation (5). The first control variable, which is used in both the second and third specifications, is an estimate of the probability of extreme returns that uses only the historical open-to-open return distribution. Open-to-open returns are assumed to come from a mixture of two normal distributions with different standard deviations; thus, the likelihood function for open-to-open returns has the following form:

$$L(\alpha, \sigma_1, \sigma_2) = \sum_{t=1}^{N} \ln\left[\alpha\Phi\left(r_t; \sigma_1\right) + (1 - \alpha)\Phi\left(r_t; \sigma_2\right)\right] \tag{6}$$

where $\Phi(r; \sigma)$ is the normal CDF, with standard deviation σ, evaluated at r. This mixture-of-normals assumption is consistent with our model, in which open-to-open returns are normally distributed with variance equal to $\sigma_1 = \sigma_i^2 + \sigma_{pd}^2 + \sigma_{po}^2$ on days with information events (with

probability α) and variance equal to $\sigma_2 = \sigma_{pd}^2 + \sigma_{po}^2$ on days with no information events. In

contrast to our model-based estimates, this mixture-of-distributions approach ignores the

additional information contained in the order imbalance data and it ignores the conditional

covariances between intraday and overnight returns. The mixture-of-distributions approach is

more flexible than our model-based approach for predicting extreme returns, because it

implicitly allows some of the information events to be publicly observed, whereas our model

assumes all heteroskedasticity in daily returns results from private information events. We

maximize the likelihood function in equation (6) to produce estimates for the mixing parameter

(which corresponds to α in our model) and the estimates for the two standard deviations. We

then use an approach analogous to equation (4) to calculate the probabilities of extreme events,

P_n^{mix}, and add $\Phi^{-1}(P_n^{mix})$ to the probit specification shown in equation (5).

The third specification for the probit model adds several more control variables that

reflect the shape of the tails of the return distribution during the estimation period. Specifically,

we include the ratio of the sample standard deviation to the robust standard deviation, as well as

the fraction of days during the estimation period in which the absolute return exceeded three,

four, five, six, or seven robust standard deviations. These six variables are designed to test the

degree to which the predicted probabilities, based either on our model or on the mixture-of-

distributions approach, act as proxies for particular features of the return distribution from the

estimation period.

We are also concerned that both our model-based estimates and the mixture of

distribution estimates may be picking up heteroskedasticity in the return distribution that is due

to time variation in the level of volatility. Firms that had an increase in volatility late in the

sample may be more likely to have relatively large returns (compared to the full year standard

deviation) over the coming quarter. Thus, we include a control variable equal to the ratio of the

sample standard deviation of daily returns for the final month of the estimation interval to the

sample standard deviation for the full estimation period.

The results are reported in Table 7. The coefficients on our model-based predicted

probability are positive and statistically significant for all specifications, meaning an increase in

the predicted probability of an extreme event using our estimated parameters is associated with a

higher frequency of realized extreme events in the coming quarter. The fact that these

coefficients are different from one (and the intercept terms are non-zero) suggests that our

predicted probabilities do not match the observed probabilities. The coefficients on the

probability estimates based on the mixture-of-distributions model are also positive and

significant for 5- and 6-σ events, but are somewhat sensitive to the inclusion of the remaining

control variables. The fact that the mixture-of-distributions approach provides information

beyond our model-based estimates may indicate that there is also heteroskedasticity in the

publicly-observed components of returns, as this is not captured by our model. The shape of the

return distribution over the estimation period also matters to some extent, and the significant

coefficient on the ratio of recent volatility to total volatility for the year is evidence of persistence

in the volatility of returns.

We test the robustness of the results in Table 7 in several ways. We describe the results

of each of these tests, focusing on the coefficients on the predicted probabilities. (To conserve

space, we do not include separate tables.) First, we estimate the three specifications without

transforming the predicted probabilities using the inverse of the Normal CDF. These tests yield

positive and statistically significant coefficients on our model-based probability in all cases. We

also use the sample standard deviation instead of our robust standard deviation to define the

extreme return cutoffs for the both the dependent and control variables, and the results are qualitatively similar to those shown in Table 7, but the coefficients on the predicted probabilities tend to have smaller magnitudes than those reported in the table. Finally, we run the tests including the observations with an absolute return greater than ten times the robust standard deviation during the estimation year. Again, the results are similar to those in Table 7, with slightly smaller coefficients on our model-based probabilities.

The results in Table 7 demonstrate that our estimates of α, σ_i, σ_{po}, and σ_{po} yield a statistic that can be used to predict future extreme returns. Together with the simulation results, this confirms the validity of our estimation procedure. In addition, the ability to predict extreme outcomes is valuable in option pricing and risk management. The question as to whether the predicted probability from our model has power above and beyond other possible measures (e.g., the implied volatilities from traded options) remains an open question for future research.

6. Time series and cross-sectional variation in α and σ_i

Having demonstrated that our estimates of the probability (α) and magnitude (σ_i) of private information are reasonable, we now analyze how the estimates vary over time and across firms. Knowledge of the types of firms or time periods for which α's or σ_i's are high or low will deepen our understanding of the way in which information is incorporated into prices. This knowledge can then be applied in many different areas of finance. For example, changes in a given firm's α and/or σ_i may be useful in predicting corporate events such as takeovers. Moreover, α and σ_i may have an impact on the type of market structure that best suits particular firms.

To examine the time-series properties of our estimates, we run panel data regressions of our estimates of α and σ_i on time dummies, allowing for firm-level random effects. The estimated coefficients from these regressions are plotted in Figure 1.[16] As shown in Panel A, α was fairly volatile in the mid- to late-1990's, and then experienced a dramatic drop-off beginning in 2000. This drastic decrease in the latter half of the sample may be related to Regulation FD, enacted by the Securities and Exchange Commission in October 2000. Reg FD was designed to reduce the selective disclosure of information to certain individuals and mandates that firms publicly disclose material information. Introduction of the rule may have reduced the degree of information asymmetry, particularly the kind of signals that would be received one day by an informed trader and publicly revealed the next.[17] Although existing studies have documented changes in market maker and trader behavior surrounding reductions in tick size, we see no consistent patterns in the α estimates surrounding the decrease to sixteenths in June 1997 or to decimals in January 2001.[18]

Panel B of Figure 1 illustrates the time trend in σ_i. There was a marked increase in the magnitude of private information from 1995 through 2000, perhaps simply due to an increase in total volatility over the period. Like α, σ_i declined following Reg FD, which suggests that the regulatory change may have reduced both the frequency and the magnitude of private information events. The fact that σ_i remains higher in the post-FD period than during the 1990's could again reflect higher *total* volatility after 2000.

---

[16] Intervals are centered around the observation date. For example, 1998 includes the four one-year windows beginning between July 1997 and June 1998.

[17] Eleswarapu, Thompson, and Venkataraman (2004) find that information asymmetry costs declined following the implementation of Reg FD. Our results suggest this decline was related to a decline in the frequency of information events.

[18] See, for example, Goldstein and Kavajecz (2000), Jones and Lipson (2001), and Bacidore, Battalio, and Jennings (2003).

The fact that both α and σ_i are likely related to total volatility makes interpreting the graphs in Panels A and B of Figure 1 more difficult. As the uncertainty about the value of a firm increases, one would expect increases in the volatilities of both publicly- and privately-observed signals. This suggests that the fraction of total volatility that stems from private information events, measured as $\sqrt{\alpha}\sigma_i \big/ \sigma$, may provide a way to examine differences in the information environment while controlling for the overall level of firm volatility. The sharp decrease in the fraction of total volatility that stems from private information events surrounding Reg FD (Panel C), to levels below those in the 1990's, suggests that there was indeed a shift of information from private to public channels during this period.

Before concluding the discussion of the changes in estimates surrounding Reg FD, it is important to point out that there were other events occurring in the same time frame that probably resulted in substantial changes to the information environment. This period saw a dramatic decline in valuations of technology stocks and a dramatic decline in the frequency and size of acquisitions. There were also other important legal and regulatory events during this period, including prosecution of several major investment banks resulting from the actions of their stock analysts, and the U.S. SEC's passage of new auditor independence rules. In light of these events, it may be inappropriate to attribute changes in our estimates during this period solely to Reg FD.

Next, we analyze how the probability (α) and magnitude (σ_i) of private information relate to underlying firm characteristics. Our cross-sectional analysis proceeds in three steps. First, we use economic reasoning to identify candidate explanatory variables. Note that, in general, we do not have explicit predictions as to how α and σ_i vary cross-sectionally. Accordingly, the results

in this section do not constitute tests of a specific theory, but rather, provide the first indication of how these new measures of private information relate to observable firm characteristics.

Second, for each explanatory variable, we group the observations by decile and examine plots of the averages of the α's and σ_i's relative to the average firm characteristic for each decile. In some cases, we transform the explanatory variable so that the relation between the explanatory variable and the average estimates is approximately linear. For example, we use a logarithmic transformation for market capitalization. Finally, guided by the patterns in these plots, we perform multivariate panel data regressions.

A number of factors have the potential to impact firms' α's and σ_i's. Larger firms may be more stable and predictable, and information may become public quickly for these firms, suggesting that both the probability and magnitude of private information events will be lower for larger firms. Likewise, smaller, less established firms may have higher α's and σ_i's because these companies are changing rapidly. On the other hand, information production may be limited for smaller firms, reducing the frequency with which information is incorporated into prices and lowering α. (The effect on σ_i is unclear.) Finally, because decision making is more dispersed in larger firms, it may be more difficult to control information leakage, suggesting more frequent private information events.

These hypotheses suggest that analyst following, industry group, and spending on research and development may also have an impact on α and σ_i. Firms with greater analyst following may have increased information production, potentially increasing α, but the fact that this information will be made public more quickly suggests that both α and σ_i may be reduced. Innovative firms' efforts to keep new technologies private could lead to large, infrequent private information events. If these firms tend to spend more on research and development, we would

expect a negative relation between R&D and α, and a positive relation between R&D and σ_i. Furthermore, if these firms are concentrated within particular industries, we may also see variation in α and σ_i across industry groups.

Financial strength may also affect the magnitude of information events. Unhealthy firms may face a greater risk of extreme corporate events like bankruptcies, takeover attempts, or restructurings, and firms that are performing extremely well may be more likely to experience extreme events like acquisitions or announcements of major new product lines. This suggests that α and σ_i may be related to measures of solvency and past performance like debt-to-equity ratio, past return, and book-to-market ratio. α and σ_i may also be positively related to trading activity, as higher trading volume may signal greater divergence of opinion.

Because in many cases we do not have clear priors as to the functional form, or even the direction, of the relation between α and σ_i and the firm characteristics discussed above, we begin by creating plots to determine how the estimates vary with each factor. For each one-year window, we divide the sample into deciles based on the value of each firm characteristic independently. Then we pool the sample (across all years), and for each decile, we graph the average values of α and σ_i against the average value of the given firm attribute. Although they are not shown to conserve space, we also create plots for $\sqrt{\alpha}\sigma_i \big/ \sigma$, and we comment on how they compare to the graphs for α and σ_i.

Firm characteristics are calculated using data from the Center for Research in Securities Prices (CRSP), COMPUSTAT, and I/B/E/S databases. Firm size is computed as the natural logarithm of the product of stock price and shares outstanding (in millions) as of the end of the one-year estimation window. Book-to-market and debt-equity ratios are computed using book

values from the most recently reported quarter (also as of the end of the estimation interval). We

also obtain R&D from the most recent quarterly report and scale it by firm size. Unfortunately,

R&D is missing for over half of the observations in our sample and is reported as zero for

another fifteen percent of the observations. We use the first two digits of the COMPUSTAT SIC

code to sort firms into ten industry groups: durables, nondurables, utilities, energy, construction,

business equipment, manufacturing, transportation, financial, and business services. Past

performance is captured by the excess return (relative to the S&P 500) over the prior six months,

and trading activity is measured as the average monthly trading volume over the past twelve

months scaled by the total shares outstanding. Volatility is measured using both the monthly

standard deviation computed over the prior 60 months and the equity beta (estimated from a

monthly market model regression over the same period) to examine differences in systematic and

total risk. Analyst following is computed as the number of analysts reporting annual earnings

estimates as of the most recent summary date in I/B/E/S prior to the end of the estimation

interval. To control for the strong positive correlation between analyst following and firm size,

we regress the natural logarithm of (1+number of analysts) on the natural logarithm of market

capitalization and use the residuals as our measure. Adding one to the number of analysts allows

us to compute the measure for firms with no analyst following.

Graphs of α and σ_i are presented in Figure 2. Panel A shows that α is clearly increasing

in firm size. This is consistent with the hypothesis that larger firms have greater production of

private information and/or with the hypothesis that it is more difficult for large firms to control

information leakage. In contrast, σ_i is decreasing in firm size, suggesting that large firms

experience smaller events. The fraction of total volatility due to private information events (not

reported) is increasing across firm size deciles, indicating that the higher σ_i for small firms may simply be a reflection of the higher total variability of these firms.

Panel B shows that σ_i (like $\sqrt{\alpha}\sigma_i / \sigma$, which is not reported) is generally increasing in analyst following. In contrast, α is slightly lower for firms with unusually low or unusually high analyst following compared to other firms their size, but α is fairly constant across the middle deciles. The absence of a strong trend is perhaps not surprising given the complexity of the relation between the number of analysts providing earnings estimates and firms' information environment. A firm may be followed by many analysts because it is a highly visible firm and coverage is expected or because public information is readily available and easy to analyze. Firms that provide analysts with inside information also may have a large number of analysts, either because the analysts are attracted to such a firm or because a large number of analysts cause increasing pressure for such disclosure. On the other hand, such firms may give preferential access to a small number of analysts, which might ultimately decrease the number of analysts as others recognize their efforts would be at a serious disadvantage. Finally, as pointed out by Sidhu, Smith, and Whaley (2005), simultaneously disclosing private information to a large number of analysts may have the same effect as public disclosure because when many traders share private information they trade very aggressively and the price impact is nearly immediate.

Not surprisingly, σ_i is increasing in both historical volatility and beta. In contrast, the relations between α (and $\sqrt{\alpha}\sigma_i / \sigma$, not reported) and these measures are not consistently positive, and there is a clear *negative* relation between α and historical volatility. As expected, firms with extreme book-to-market ratios (Panel F) and firms that experienced either very low or

very high returns over the past six months (Panel H) appear to have more extreme events.

Furthermore, these events are less frequent for firms with extreme past returns. (The graph of

$\sqrt{\alpha}\sigma_i \big/ \sigma$ versus book-to-market is similar to that for α, while the graph versus past return

resembles that for σ_i .) The magnitude is generally larger for more actively-traded firms (Panel

G). There is no clear relation between the parameter estimates and debt-to-equity ratio (Panel

E). Likewise, no clear patterns emerge in the plots of α and σ_i versus R&D, which are not

reported due to the large fraction of missing values. Finally, there is little variation across

industries (also not reported), but utilities have a higher average α and lower average σ_i than

firms in other industries.

The plots in Figure 2 have the advantage that they do not require ex-ante specification of

the functional form, but their disadvantage is that they do not allow simultaneous consideration

of multiple explanatory variables. In our multivariate regressions, we use the plots in Figure 2 to

guide our choice of variables and functional forms.

We include logged firm size, adjusted analyst following, volatility, beta, debt-to-equity

ratio, and turnover as explanatory variables in our regressions. Due to non-linearities apparent in

the graphs, we use two dummy variables to capture book-to-market ratio – one for the lowest

book-to-market decile and one for the highest decile – and we take the absolute value of the

return over the past six months before including it as a regressor. Because R&D is missing for

many observations in our sample and did not show any clear patterns (perhaps due to the small

sample size), we omit this variable from the regressions. Finally, we add industry dummies,

along with time dummies for each of the forty periods in our sample.

We run separate panel data regressions for the probability, α, and magnitude, σ_i, of

private information events, as well as for the fraction of total volatility that is driven by private

information events, $\sqrt{\alpha}\sigma_i\big/\sigma$. We allow for correlation in the error terms for each firm over time. Because our observation intervals are overlapping (one-year windows rolling quarter-by-quarter), we model the error structure to include an additional covariance term for observations for the same firm that cover periods beginning within three quarters of one another.

The regression results are presented in Table 8. The results for $\sqrt{\alpha}\sigma_i\big/\sigma$ demonstrate that firms' return variability tends to be driven less by private information in some industries than in others. The fraction of total volatility that stems from private information events is significantly lower for durables, energy, and financial firms relative to the other firms in the sample. This is in spite of the higher probability of information events for energy firms, so the lower magnitude of private information has a more dominant effect. Larger firms are more likely to have private information events, and these events tend to make up a greater portion of total volatility. Likewise, higher analyst following indicates a greater frequency of private information events.

In contrast to the univariate plots in Figure 2, both α and σ_i increase with historical volatility, and α also increases with beta. Firms with book-to-market ratios in the highest decile tend to have both more frequent and larger private information events that make up a greater fraction of total volatility. High book-to-market ratios may indicate depressed stock prices, perhaps related to financial distress. As in Figure 2, if a firm's total return over the past six months has been either very low or very high, the firm is less likely to experience a private information event in the current period but, if it does, it will tend to be larger in magnitude. Finally, more actively traded firms tend to have larger but less frequent private information events.

Collectively, the results in Figures 1 and 2 and Table 8 reveal important insights about asymmetric information that would be unobservable when focusing only on measures of expected loss. For example, we find a positive relation between firm size and the fraction of volatility due to private information events that is driven by a higher frequency of information events for larger firms. This result is consistent with the hypothesis that it is more difficult for larger firms to control information leakage. If Reg FD made firms more willing to spend resources to control information leakage, then this could be one reason for the decrease in α in the latter part of the sample (Figure 1), and one might expect that the effect would be stronger for the larger firms in the sample. We examine this hypothesis in Figure 3, which reproduces the graphs for market capitalization in Figure 2 (Panel A) separately for the pre- and post-FD periods. One-year intervals ending before October 2000 fall into the pre-FD period, and intervals beginning after October 2000 are labeled post-FD. (Intervals that span both periods are omitted from the graphs in Figure 3.)

The results in Panel A of Figure 3 demonstrate that α declined across all firm sizes following Reg FD and that this decrease is, in fact, less pronounced for the smallest deciles. Interestingly, σ_i increased in the post-FD period for many firm-size deciles (Panel B), but as we saw in Figure 2, this is due to an increase in total volatility in the latter part of the sample period. The results in Panel C of Figure 3 confirm that the fraction of volatility stemming from private information events decreased substantially after Reg FD, especially for large firms. Panel D shows that while estimated total volatility increased in the post-FD period, the proportional increase was similar across the firm-size deciles. The patterns in Figure 3 are consistent with the hypothesis that larger firms had more information leakage in the pre-FD period, and because Reg

FD increased the scrutiny on this leakage, these firms show a larger decline in the fraction of total volatility stemming from private information events.

7. Conclusion

In the existing literature, adverse selection risk is typically measured on a single dimension, based on the anticipated loss to informed traders. Many studies examine the determinants of this expected loss – as captured empirically by spreads, price impact measures, and adverse selection components – and find that these measures vary across firms and over time. While this work is clearly instructive, we recognize that measures of the expected loss capture a combination of two effects: 1) the *probability* of a private information event, and 2) the likely *magnitude* of the information. We develop a method of separately estimating the probability and the magnitude of private information using returns and trade imbalances.

We validate our estimation procedure using a simulated data set, and then estimate these parameters for NYSE-listed stocks from 1993 through 2003. We show that our parameter estimates can be used to predict future extreme returns, which not only offers additional evidence of the validity of our estimates, but also has implications for option pricing, risk management, and corporate finance. Finally, we examine the time series properties of the probability and magnitude of information.

Our work suggests that focusing only on the expected loss to informed traders provides an incomplete picture, as firms with similar expected losses can have markedly different probabilities and magnitudes of private information events. For example, we find a positive relation between firm size and adverse selection cost in our sample that is driven by the higher frequency of information events for larger firms. In contrast, we find that past volatility is

related to adverse selection risk primarily because private information events tend to be larger for firms with more volatile overall stock returns, not because events are more frequent for these firms.

We also show that the probability and magnitude of information events declined dramatically in late 2000, which corresponds to the implementation of the SEC's Regulation FD. We find that the positive relation between firm size and the fraction of total return volatility that stems from private information is attenuated in the period following Regulation FD. Together, these results suggest that Regulation FD had the intended result of reducing the flow of private information from firms to analysts.

In summary, we believe that the ability to separately estimate the probability and the magnitude of private information events will yield many other applications that are important to investors, regulators, and researchers. For example, changes in α and σ_i may be useful in predicting corporate events such as takeovers. More generally, analyzing α and σ_i independently helps to clarify the distinction between the *risk* of informed trading and the *degree of realized* informed trading, deepening our understanding of the price discovery process.

References

Bacidore, J., Battalio, R., Jennings, R., 2003. Order submission strategies, liquidity supply, and trading in pennies on the New York Stock Exchange. Journal of Financial Markets 6, 337-362.

Breen, W., Hodrick, L., Korajczyk, R., 2002. Predicting equity liquidity. Management Science 48, 470-483.

Bernhardt, D., Hughson, E., 2002. Intraday trade in dealership markets. European Economic Review 46, 1697-1732.

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

Corwin, S., Lipson, M., 2005. Order flow, Trader type, and the determinants of commonality in prices and liquidity. University of Notre Dame working paper.

Chordia, T., Subrahmanyam, A., 2004. Order imbalance and individual stock returns:Theory and evidence. Journal of Financial Economics 72, 485-518.Easley, D., Kiefer, N., O'Hara, M., Paperman, J., 1996. Liquidity, information, and infrequently traded stocks. Journal of Finance 51, 1405-1436.

Eleswarapu, V., Thompson, R., Venkataraman, K., 2004. The impact of Regulation Fair Disclosure: trading costs and information asymmetry. Journal of Financial and Quantitative Analysis 39, 209-225.

Fama, E., French, K., 1992. The cross-section of expected stock returns. Journal of Finance 47, 3-56.

Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3-56.

George, T., Kaul, G., Nimalendran, M., 1991. Estimation of the bid-ask spread and its components: A new approach. Review of Financial Studies 4, 623-656.

Glosten, L., Harris, L., 1988. Estimating the components of the bid-ask spread. Journal of Financial Economics 21, 123-142.

Glosten, L., Milgrom, P., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71-100.

Goldstein, M., Kavajecz, K., 2000. Eights, sixteenths and market depth: changes in tick size and liquidity provision on the NYSE. Journal of Financial Economics 56, 125-149.

Gomes, A., Gorton, G., Madureira, L., 2004. SEC Regulation Fair Disclosure, Information, and the Cost of Capital. University of Pennsylvania working paper.

Hansch, O., Naik, N., Viswanathan, S., 1999. Preferencing, internalization, best execution, and dealer profits. Journal of Finance 54, 1799-1828.

Hasbrouck, J., 1991. Measuring the information content of stock trades. Journal of Finance 46, 179-207.

Huang, R., Stoll, H., 1997. The components of the bid-ask spread: A general approach. Review of Financial Studies 10, 995-1034.

Jones, C., Lipson, M., 2001. Sixteenths: Direct evidence of institutional trading costs," Journal of Financial Economics 59, 253-278.

Kyle, A., 1985. Continuous auctions and insider trading. Econometrica 53, 1315-1336.

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

Lin, J., Sanger, G., Booth, G., 1995. Trade size and components of the bid-ask spread, Review of Financial Studies 8, 1153-1183.

Llorente, G., Michaely, R., Sarr, G., Wang, J., 2002. Dynamic volume-return relation of individual stocks. Review of Financial Studies 15, 1005-1047.

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

Sidhu, B., Smith, T., Whaley, R., 2005. Regulation Fair Disclosure and the cost of adverse selection. Duke University working paper.

Spiegel, M., Subrahmanyam, A., 2000. Asymmetric information and news disclosure rules. Journal of Financial Intermediation 9, 363-403.

Stoll, H., 1989. Inferring the components of the bid-ask spread: Theory and empirical tests. Journal of Finance 44, 115-134.

Figure 1: Time Trends in Estimates of α, σ_i, and $\sqrt{\alpha}\sigma_i/\sigma$

Panels A through C plot coefficients from panel data regressions of α, σ_i, and $\sqrt{\alpha}\sigma_i/\sigma$, respectively, on time dummies, allowing for firm-level random effects.

Panel A: Estimates of the Frequency of Private Information Events (α)
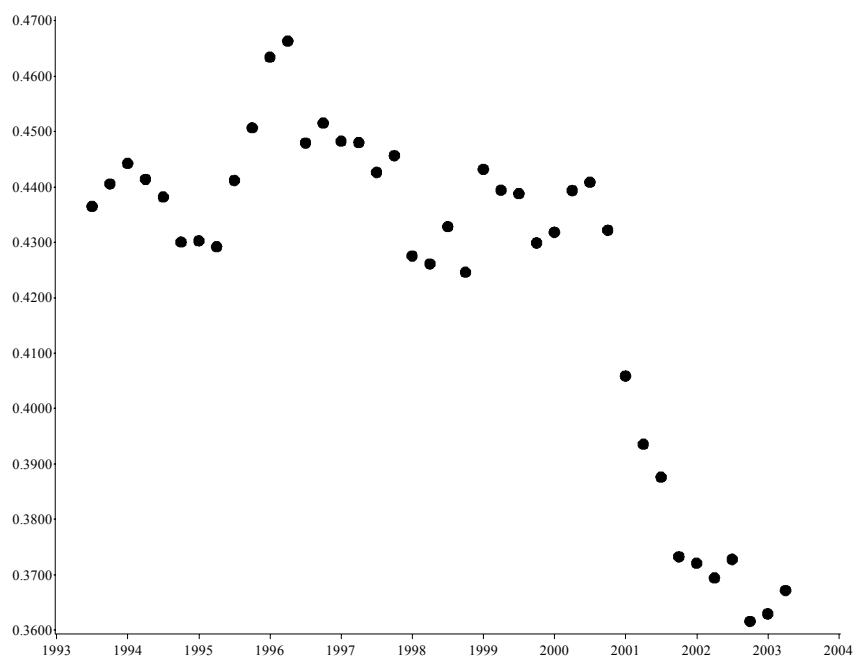
Panel B: Estimates of the Magnitude of Private Information Events (σ_i)
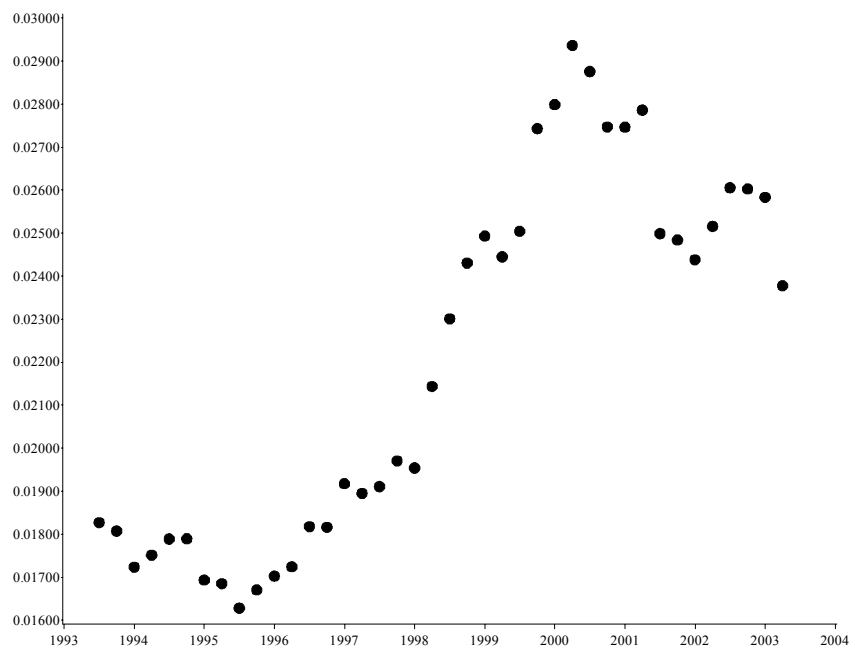
Panel C: The Fraction of Estimated Volatility Due to Private Information Events

$$\left(\sqrt{\alpha}\sigma_i \Big/ \sigma \right)$$
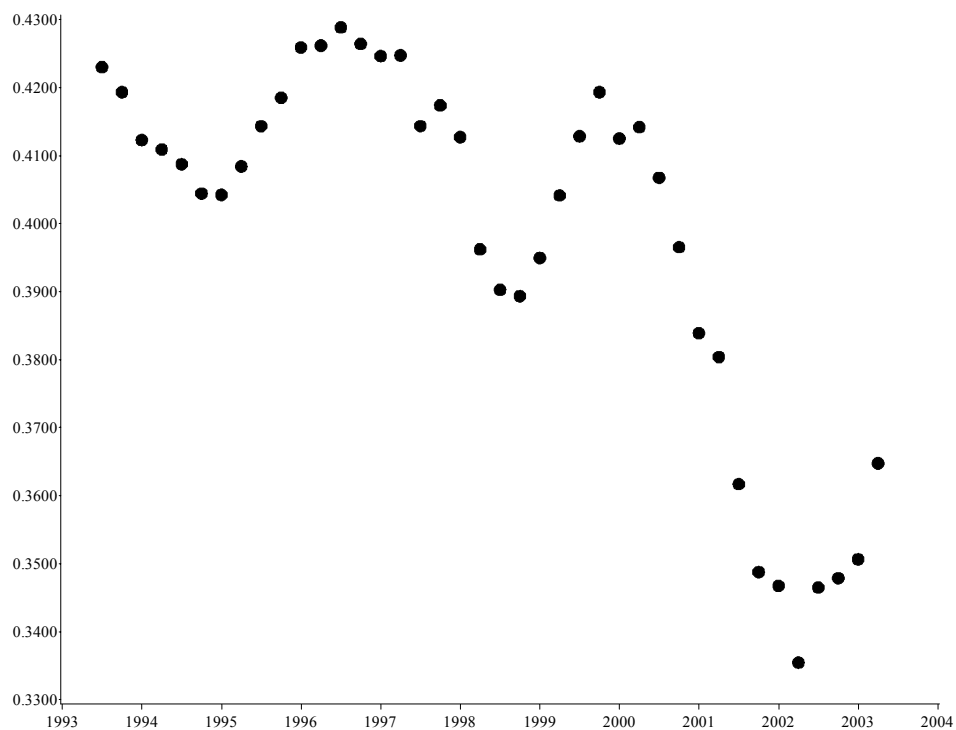
Figure 2: Parameter Estimates and Firm Characteristics

For each one-year estimation window, we assign each observation to a decile based on a particular characteristic. We then pool the sample and plot the average α and σ_i estimates for each decile versus the average value of the given characteristic. In some cases, we transform the characteristic so that its relation with the average estimates is approximately linear.

Panel A: Average Estimates by Firm Size Decile
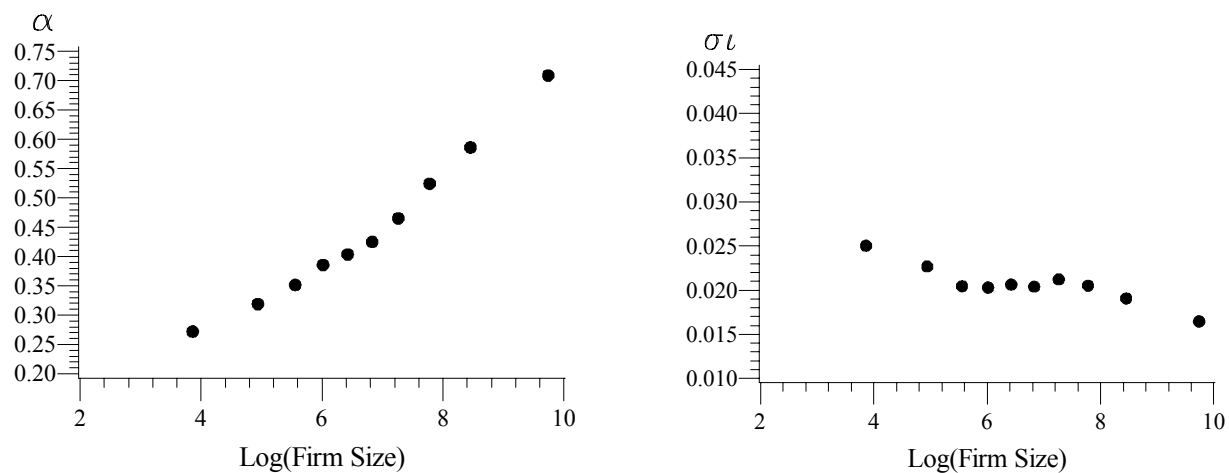
Panel B: Average Estimates by Adjusted Analyst Following Decile
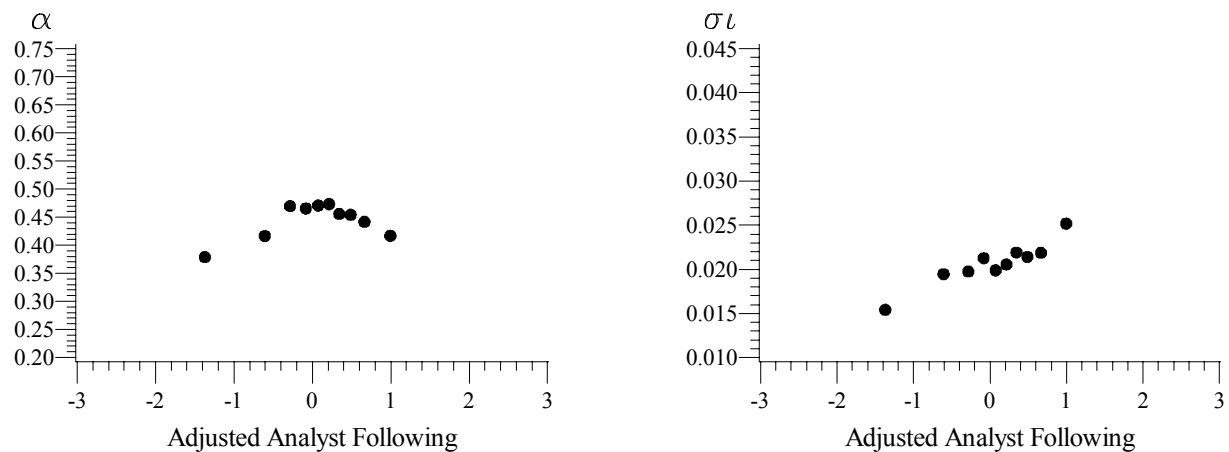
Figure 2 (continued)

Panel C: Average Estimates by Volatility Decile
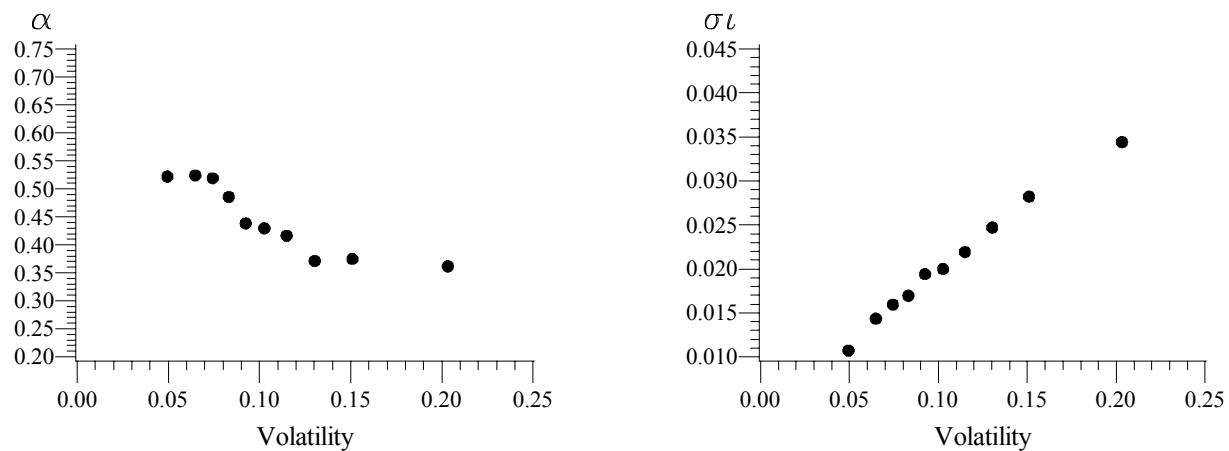
Panel D: Average Estimates by Equity Beta Decile
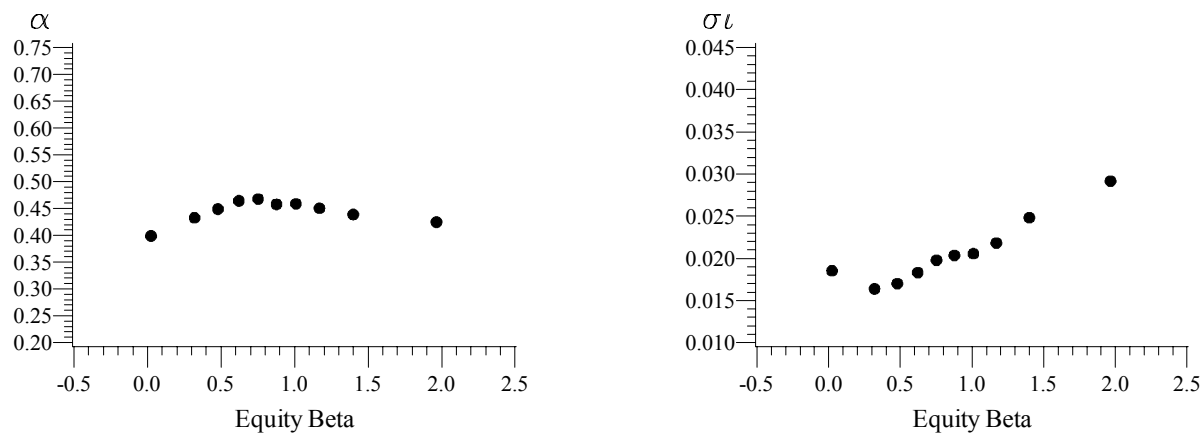
Panel E: Average Estimates by Debt-to-Equity Ratio Decile
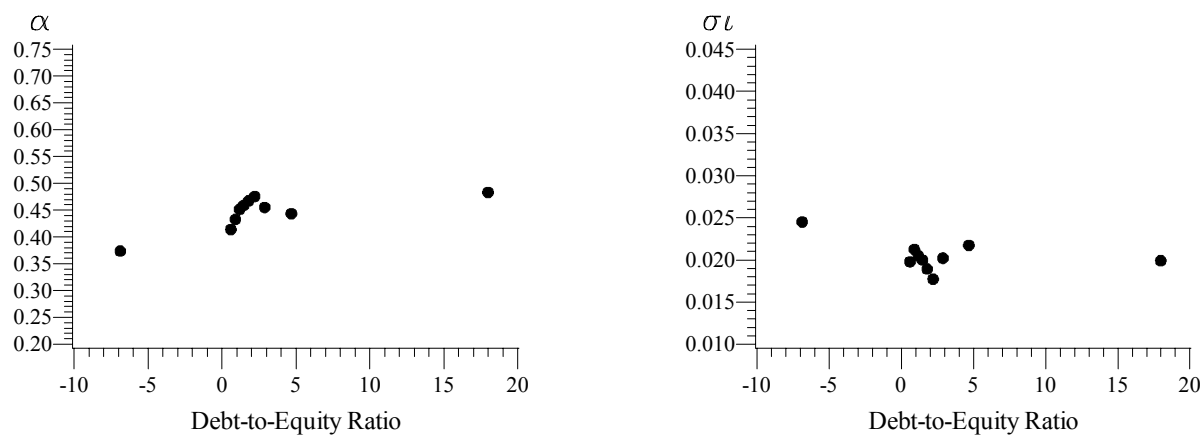
Figure 2 (continued)

Panel F: Average Estimates by Book-to-Market Decile
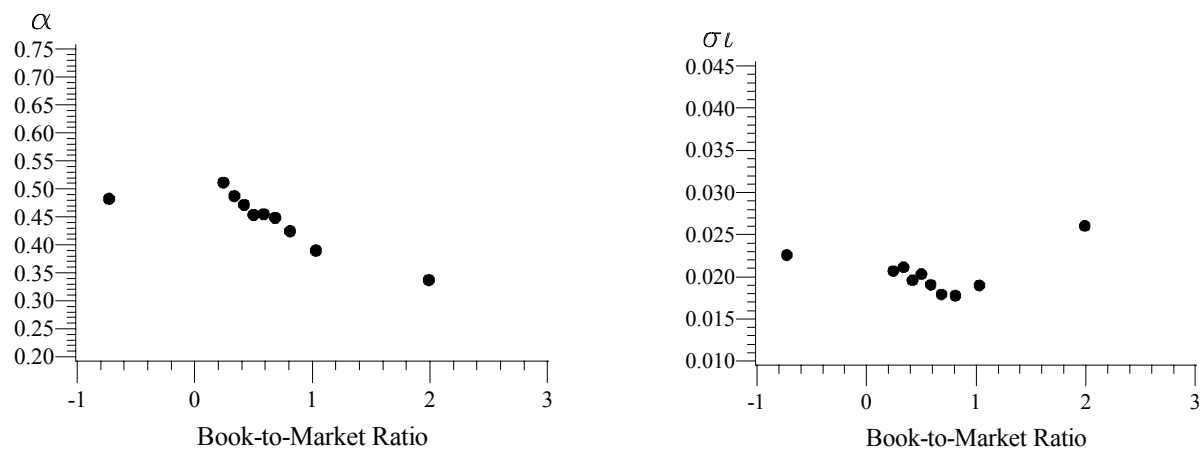
Panel G: Average Estimates by Turnover Decile
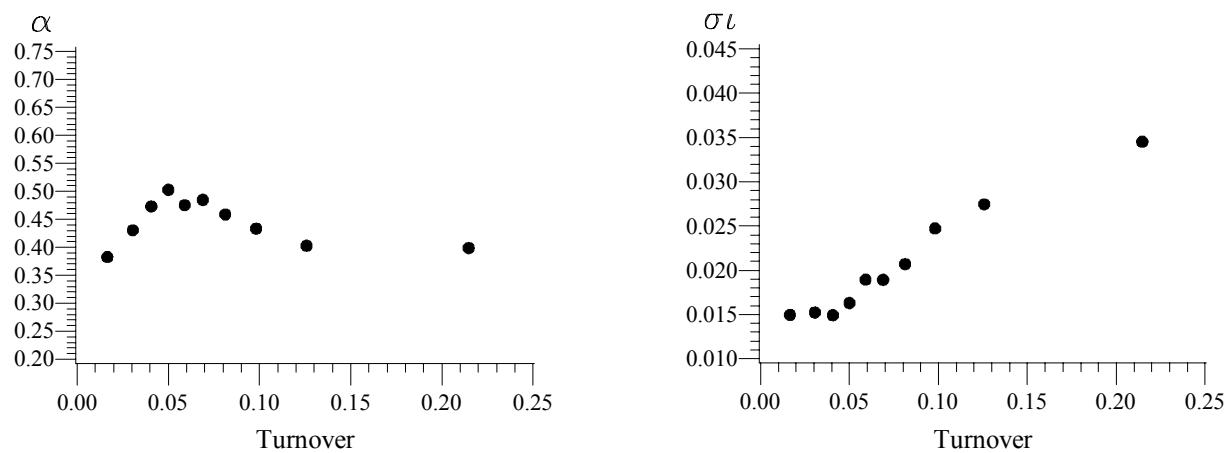
Panel H: Average Estimates by Past 6-Month Return Decile
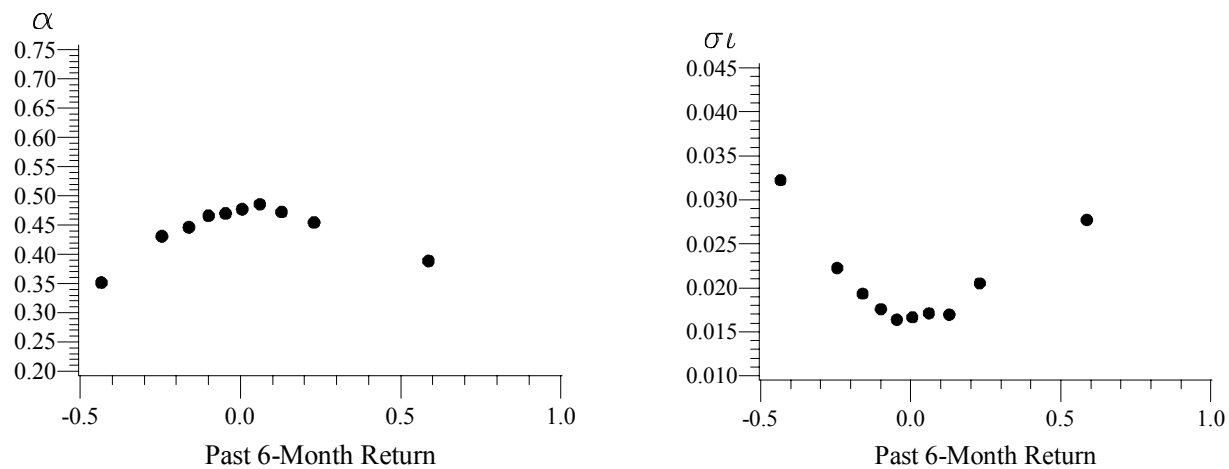
Figure 3: Average Estimates by Market Capitalization, Pre- and Post-Reg FD

For each one-year estimation window, we assign each observation to a decile based on market capitalization. We then average the estimates for each decile in the pre-FD sample and in the post-FD sample. We plot these averages against each decile's average of the natural logarithm of market capitalization (in millions) across the full sample period. O's represent pre-FD averages, and +'s denote post-FD averages. α is the fraction of days with an information event, σ_i is the standard deviation of the information events, the volatility ratio is $\sqrt{\alpha}\sigma_i\big/\sigma$, and σ is the total standard deviation including public signals.

Panel A

Panel B

Panel C

Panel D
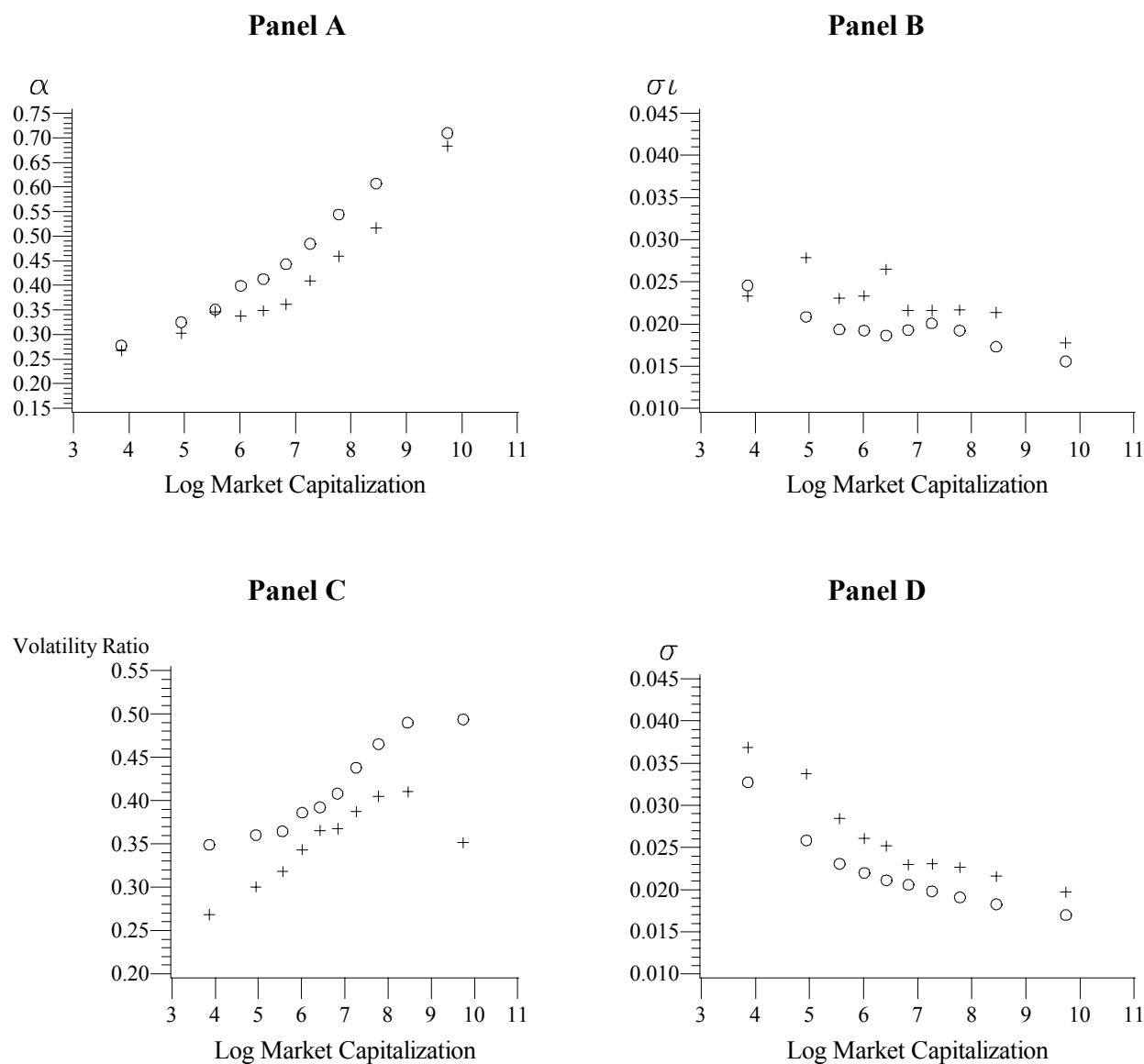
Table 1: α and the Frequency of Reversals

We simulate 500 years of daily data (500 x 252 = 126,000 trading days) for 50 different combinations of α and σ_t parameters. For each set of parameter values, the reversal frequency is the fraction of days where y_t and $r_{o,t}$ have opposite signs.

True α values	True σ_i values				
	0.020	0.040	0.060	0.080	0.100
0.05	0.555	0.614	0.660	0.704	0.735
0.10	0.571	0.634	0.689	0.728	0.761
0.15	0.574	0.642	0.695	0.730	0.758
0.20	0.575	0.644	0.694	0.725	0.749
0.25	0.574	0.639	0.687	0.713	0.735
0.35	0.571	0.629	0.666	0.686	0.701
0.45	0.564	0.608	0.641	0.658	0.670
0.55	0.551	0.593	0.612	0.630	0.639
0.65	0.543	0.570	0.589	0.599	0.603
0.85	0.519	0.529	0.539	0.542	0.544

Table 2: α Estimates from Simulated Data

We simulate 500 years of daily data (500 x 252 = 126,000 trading days) for 50 different combinations of α and σ_t parameters. For each year of data, α and σ_i are estimated by maximizing the likelihood function in equation (3) subject to the constraints that the estimates lie between 0.00001 and 1. For each pair of true α and σ_i values, the table reports the average of the α estimates across the 500 years. The standard deviation of the α estimates across the 500 years is shown in parentheses.

True α values	True σ_i values				
	0.020	0.040	0.060	0.080	0.100
0.05	0.053	0.052	0.051	0.051	0.051
	(0.029)	(0.020)	(0.017)	(0.016)	(0.015)
0.10	0.105	0.102	0.101	0.101	0.099
	(0.041)	(0.030)	(0.025)	(0.022)	(0.020)
0.15	0.156	0.151	0.150	0.150	0.151
	(0.053)	(0.035)	(0.029)	(0.025)	(0.025)
0.20	0.212	0.200	0.203	0.199	0.200
	(0.058)	(0.040)	(0.034)	(0.029)	(0.027)
0.25	0.263	0.254	0.247	0.249	0.249
	(0.070)	(0.047)	(0.037)	(0.032)	(0.030)
0.35	0.350	0.353	0.347	0.349	0.348
	(0.082)	(0.053)	(0.043)	(0.037)	(0.034)
0.45	0.454	0.451	0.449	0.450	0.451
	(0.088)	(0.059)	(0.047)	(0.041)	(0.037)
0.55	0.561	0.548	0.547	0.543	0.546
	(0.096)	(0.059)	(0.048)	(0.045)	(0.044)
0.65	0.649	0.651	0.642	0.644	0.650
	(0.103)	(0.068)	(0.051)	(0.044)	(0.041)
0.85	0.835	0.841	0.840	0.842	0.844
	(0.119)	(0.069)	(0.053)	(0.051)	(0.040)

Table 3: σ_i Estimates from Simulated Data

We simulate 500 years of daily data (500 x 252 = 126,000 trading days) for 50 different combinations of α and σ_i parameters. For each year of data, α and σ_i are estimated by maximizing the likelihood function in equation (3) subject to the constraints that the estimates lie between 0.00001 and 1. For each pair of true α and σ_i values, the table reports the average of the σ_i estimates across the 500 years. The standard deviation of the σ_i estimates across the 500 years is shown in parentheses.

True α values	True σ_i values				
	0.020	0.040	0.060	0.080	0.100
0.05	0.021	0.042	0.063	0.083	0.103
	(0.007)	(0.010)	(0.014)	(0.015)	(0.018)
0.10	0.021	0.041	0.061	0.081	0.102
	(0.006)	(0.007)	(0.008)	(0.010)	(0.012)
0.15	0.021	0.040	0.061	0.081	0.101
	(0.007)	(0.005)	(0.007)	(0.008)	(0.009)
0.20	0.020	0.040	0.060	0.081	0.101
	(0.008)	(0.005)	(0.006)	(0.007)	(0.008)
0.25	0.021	0.040	0.060	0.081	0.100
	(0.008)	(0.004)	(0.005)	(0.006)	(0.007)
0.35	0.021	0.040	0.060	0.080	0.100
	(0.009)	(0.004)	(0.005)	(0.005)	(0.006)
0.45	0.021	0.040	0.060	0.080	0.100
	(0.010)	(0.004)	(0.005)	(0.005)	(0.006)
0.55	0.022	0.040	0.060	0.080	0.100
	(0.011)	(0.004)	(0.004)	(0.005)	(0.006)
0.65	0.022	0.040	0.060	0.080	0.100
	(0.012)	(0.004)	(0.004)	(0.005)	(0.005)
0.85	0.023	0.040	0.060	0.080	0.099
	(0.014)	(0.003)	(0.004)	(0.005)	(0.005)

Table 4: α and σ_i Estimates from Simulated Data in Which Information Events Last Two Days

We use a model in which the information events lasts two days, and we simulate 500 years of daily data (500 x 252 = 126,000 trading days) for 50 different combinations of α and σ_i parameters. We then use our estimation procedure, which assumes the information events last one day, to estimate the parameters. For each year of data, α and σ_i are estimated by maximizing the likelihood function in equation (3) subject to the constraints that the estimates lie between 0.00001 and 1. For each pair of true α and σ_i values, Panel A reports the average of the yearly α estimates and Panel B reports the average of the yearly σ_i estimates.

True α values	True σ_i values				
	0.020	0.040	0.060	0.080	0.100
	Panel A: Average of α Estimates				
0.05	0.071	0.067	0.065	0.068	0.068
0.10	0.143	0.139	0.149	0.156	0.161
0.15	0.222	0.234	0.248	0.255	0.262
0.20	0.304	0.327	0.347	0.357	0.369
0.25	0.398	0.451	0.453	0.472	0.481
0.35	0.584	0.672	0.692	0.694	0.696
0.45	0.742	0.858	0.870	0.861	0.867
0.55	0.851	0.929	0.943	0.940	0.940
0.65	0.895	0.957	0.962	0.966	0.964
0.85	0.956	0.976	0.979	0.979	0.980
	Panel B: Average of σ_i Estimates				
0.05	0.018	0.036	0.051	0.067	0.082
0.10	0.017	0.031	0.044	0.058	0.071
0.15	0.016	0.028	0.041	0.054	0.066
0.20	0.015	0.027	0.039	0.051	0.063
0.25	0.017	0.025	0.038	0.050	0.062
0.35	0.015	0.025	0.037	0.050	0.062
0.45	0.022	0.026	0.038	0.052	0.065
0.55	0.021	0.028	0.041	0.056	0.070
0.65	0.025	0.030	0.045	0.061	0.076
0.85	0.020	0.034	0.052	0.071	0.089

Table 5: α and σ_i Estimates from Simulated Data in Which There Are Two Rounds of Trading Each Day

We use a model with two rounds of informed trading during the day and simulate 500 years of daily data (500 x 252 = 126,000 trading days) for 50 different combinations of α and σ_i parameters. We then use our estimation procedure, which assumes a single round of trading, to estimate the parameters. For each year of data, α and σ_i are estimated by maximizing the likelihood function in equation (3) subject to the constraints that the estimates lie between 0.00001 and 1. For each pair of true α and σ_i values, Panel A reports the average of the yearly a estimates and Panel B reports the average of the yearly σ_i estimates.

	True σ_i values				
True α values	0.020	0.040	0.060	0.080	0.100
	Panel A: Average of α Estimates				
0.05	0.079	0.082	0.084	0.085	0.086
0.10	0.111	0.110	0.110	0.105	0.108
0.15	0.174	0.164	0.165	0.163	0.165
0.20	0.238	0.227	0.222	0.219	0.220
0.25	0.306	0.287	0.278	0.276	0.276
0.35	0.426	0.399	0.389	0.387	0.383
0.45	0.535	0.499	0.491	0.484	0.486
0.55	0.641	0.603	0.589	0.586	0.580
0.65	0.759	0.706	0.682	0.677	0.677
0.85	0.907	0.895	0.876	0.867	0.858
	Panel B: Average of σ_i Estimates				
0.05	0.011	0.023	0.036	0.050	0.062
0.10	0.019	0.036	0.054	0.073	0.091
0.15	0.019	0.037	0.056	0.075	0.094
0.20	0.020	0.038	0.057	0.077	0.095
0.25	0.020	0.038	0.058	0.078	0.098
0.35	0.021	0.039	0.060	0.080	0.101
0.45	0.021	0.040	0.060	0.082	0.102
0.55	0.022	0.040	0.061	0.082	0.104
0.65	0.023	0.040	0.061	0.083	0.104
0.85	0.024	0.040	0.060	0.082	0.105

Table 6: α and σ_i Estimates from NYSE Data

α and σ_i are estimated by maximizing the likelihood function in equation (3) using rolling one-year windows (beginning in January, April, July, and October) from 1993 through 2003, subject to the constraints that the estimates lie between 0.00001 and 1. For each decile of α estimates, the table presents mean α and σ_i estimates, along with quintiles of the σ_i distribution.

| | | σ_i estimates | | | | |
| α estimates | | 20th | 40th | | 60th | 80th |
Decile	Mean	percentile	percentile	Mean	percentile	percentile
1	0.038	0.007	0.023	0.065	0.067	0.112
2	0.110	0.005	0.011	0.028	0.025	0.049
3	0.183	0.006	0.011	0.021	0.020	0.034
4	0.265	0.007	0.013	0.019	0.019	0.028
5	0.357	0.008	0.012	0.017	0.017	0.023
6	0.458	0.008	0.012	0.015	0.015	0.021
7	0.567	0.008	0.010	0.013	0.013	0.017
8	0.685	0.007	0.009	0.012	0.012	0.015
9	0.819	0.006	0.008	0.010	0.010	0.013
10	0.964	0.005	0.006	0.008	0.008	0.011

Table 7: Probit Regressions of Actual Extreme Events on Predicted Frequencies

Actual extreme events (e_n=1) occur when $r_{qmax} > n\sigma_r$ for n∈{4,5,6}, where r_{qmax} is the maximum absolute daily market-adjusted return over the quarter following the current year, and σ_o is the robust standard deviation of daily market-adjusted returns over the current year. P_n^{est} is the predicted probability that e_n=1 from our model given the estimated parameter values ($\alpha,\sigma_i,\sigma_{pd},\sigma_{po}$). P_n^{mix} is the predicted probability that e_n=1 from the mixture of distributions model. $\Phi^{-1}(\bullet)$, is the inverse of the standard normal cumulative distribution function. Marginal values for the coefficients on the variables related to P_n^{est} and P_n^{mix} are shown in square brackets and p-values for all coefficients are shown in parentheses.

Explanatory Variable	$r_{qmax} > 4\sigma_o$			$r_{qmax} > 5\sigma_o$			$r_{qmax} > 6\sigma_o$		
Intercept	0.137	0.142	0.124	-0.023	-0.056	-0.248	-0.172	-0.250	-1.062
	(0.000)	(0.000)	(0.589)	(0.308)	(0.026)	(0.286)	(0.000)	(0.000)	(0.000)
$\Phi^{-1}(P_n^{est})$	0.267	0.271	0.297	0.219	0.192	0.198	0.203	0.132	0.122
	[0.549]	[0.558]	[0.612]	[1.495]	[1.309]	[1.345]	[3.319]	[2.135]	[1.927]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\Phi^{-1}(P_n^{mix})$		-0.005	-0.002		0.031	0.015		0.081	0.052
		[-0.005]	[-0.002]		[0.041]	[0.020]		[0.171]	[0.107]
		(0.674)	(0.910)		(0.001)	(0.258)		(0.000)	(0.000)
Last month std. dev. /full year std. dev.			0.379			0.348			0.329
			(0.000)			(0.000)			(0.000)
Sample standard dev./robust standard dev.			-0.275			-0.162			0.315
			(0.173)			(0.424)			(0.126)
Fraction of days in estimation year with:									
$e_3 = 1$			0.014			0.248			0.111
			(0.990)			(0.821)			(0.925)
$e_4 = 1$			3.827			4.661			4.401
			(0.022)			(0.008)			(0.019)
$e_5 = 1$			1.958			2.592			-0.854
			(0.449)			(0.338)			(0.768)
$e_6 = 1$			-7.651			-1.551			-5.302
			(0.044)			(0.694)			(0.208)
$e_7 = 1$			-0.793			-6.068			-3.319
			(0.863)			(0.204)			(0.520)

Table 8: Regression Results

Coefficients estimates and p-values are reported for the panel data regressions described in Section 6. Coefficient estimates have been multiplied by 100 to improve readability. Quarterly time dummy estimates are not reported to conserve space.

| | Dependent Variable | | | | | |
| | α | | σ_i | | $\sqrt{\alpha}\sigma_i/\sigma$ | |
Explanatory Variable	Estimate	p-value	Estimate	p-value	Estimate	p-value
Durables	-1.64	0.219	-0.15	0.301	-2.49	0.006
Nondurables	-1.93	0.083	0.25	0.038	0.50	0.494
Utilities	2.40	0.023	0.12	0.235	-0.76	0.250
Energy	5.43	0.000	-0.49	0.000	-3.16	0.000
Construction	2.53	0.189	-0.44	0.026	-1.58	0.145
Business Equipment	-2.20	0.013	-0.11	0.224	-0.58	0.360
Transportation	1.29	0.542	-0.15	0.418	-0.75	0.518
Financial	-3.13	0.000	-0.16	0.025	-2.79	0.000
Business Services	-4.67	0.000	0.32	0.002	-0.65	0.283
Log(Firm Size)	7.70	0.000	-0.11	0.000	2.38	0.000
Adjusted Analyst Following	1.84	0.000	0.03	0.481	2.03	0.000
Volatility	26.97	0.001	8.54	0.000	7.22	0.210
Equity Beta	1.45	0.005	-0.09	0.146	-0.03	0.941
Debt/Equity	0.00	0.418	0.00	0.162	0.00	0.680
Low Book-to-Market	-0.50	0.506	-0.11	0.172	-1.59	0.002
High Book-to-Market	3.37	0.000	0.22	0.011	1.00	0.032
Turnover	-76.13	0.000	8.65	0.000	39.98	0.000
Abs(Past 6 Month Return)	-4.48	0.000	0.60	0.000	0.54	0.177

Appendix

Derivation of λ

We assume that the market maker chooses λ so that the average profit across all trading days is zero. Let p_t denote the share price after incorporating the intraday public news ($r_{pd,t}$). In dollars per share, the market maker's new price is $p_t + p_t\lambda y_t$, and the informed trader knows the true value is $p_t + p_t r_{i,t}$. The expected profit is:

$$E[y_t[(p_t + p_t\lambda y_t)-(p_t + p_t r_{i,t})]] = p_t\, E[y_t(\lambda y_t -r_{i,t})]$$

Conditioning on whether an event occurs (and ignoring p_t because we will be setting the profit equal to zero) yields

$$E[y_t(\lambda y_t -r_{i,t})] = \alpha E[y_t(\lambda y_t -r_{i,t})|\text{event}] + (1-\alpha)E[y_t(\lambda y_t -r_{i,t})|\text{no event}]$$

where

$$E[y_t(\lambda y_t -r_{i,t})|\text{event}] = E[(x_t+u_t)(\lambda(x_t+u_t) -2\lambda x_t)|\text{event}]$$

$$= \lambda E[u_t^2 - x_t^2\,|\text{event}]$$

$$= \lambda\sigma_u^2 - \sigma_i^2/(4\lambda)$$

and

$$E[y_t(\lambda y_t -r_{i,t})|\text{ no event}] = E[u_t(\lambda u_t -0)|\text{no event}] = \lambda\sigma_u^2$$

Therefore,

$$E[y_t(\lambda y_t -r_{i,t})] = \alpha[\lambda\sigma_u^2 - \sigma_i^2/(4\lambda)] + (1-\alpha)\lambda\sigma_u^2$$

And setting this equal to zero yields expression (1) in the text:

$$\lambda = (1/2)\alpha^{1/2}(\sigma_i/\sigma_u)$$

Derivation of Conditional Variances and Covariances of (y_t, $r_{y,t}$, $r_{o,t}$)

The sample consists of a sequence of triples (y_t, $r_{d,t}$, $r_{o,t}$) that are jointly distributed but

independent of all other variables. (The parameters to be estimated are $\alpha, \sigma_i, \sigma_u, \sigma_{pd}$, and σ_{po}.)

If there is <u>no</u> event on day t, then $y_t = u_t$, $r_{d,t} = r_{pd,t} + \lambda u_t$, and $r_{o,t} = r_{po,t} - \lambda u_t$.

Thus, (y_t, $r_{d,t}$, $r_{o,t}$) are jointly normally distributed with mean zero and

$$\mathrm{Var}(y_t) = \mathrm{Var}(u_t) = \sigma_u^2$$

$$\mathrm{Var}(r_{d,t}) = \mathrm{Var}(r_{pd,t} + \lambda u_t) = \sigma_{pd}^2 + \lambda^2 \sigma_u^2$$

$$= \sigma_{pd}^2 + \alpha \sigma_i^2/4 \quad [\text{using } \lambda = (1/2)\alpha^{1/2}(\sigma_i/\sigma_u))]$$

$$\mathrm{Var}(r_{o,t}) = \sigma_{po}^2 + \alpha \sigma_i^2/4$$

$$\mathrm{Cov}(r_{d,t}, r_{o,t}) = -\lambda^2 \sigma_u^2 = -\alpha \sigma_i^2/4$$

$$\mathrm{Cov}(r_{d,t}, y_t) = \lambda \sigma_u^2 = \alpha^{1/2} \sigma_i \sigma_u/2$$

$$\mathrm{Cov}(r_{o,t}, y_t) = -\lambda \sigma_u^2 = -\alpha^{1/2} \sigma_i \sigma_u/2$$

If there <u>is</u> an event on day t, then $x_t = r_{i,t}/(2\lambda)$, and $y_t = x_t + u_t = r_{i,t}/(2\lambda) + u_t$,

$r_{d,t} = r_{pd,t} + \lambda(x_t + u_t) = r_{pd,t} + r_{i,t}/2 + \lambda u_t$, and $r_{o,t} = r_{po,t} + r_{i,t} - \lambda(x_t + u_t) = r_{po,t} + r_{i,t}/2 - \lambda u_t$.

Thus, (y_t, $r_{d,t}$, $r_{o,t}$) are jointly normally distributed with mean zero and

$$\mathrm{Var}(y_t) = \mathrm{Var}(x_t + u_t) = \sigma_i^2/4\lambda^2 + \sigma_u^2 = (1+1/\alpha)\sigma_u^2 \quad [\text{again using } \lambda = (1/2)\alpha^{1/2}(\sigma_i/\sigma_u)]$$

$$\mathrm{Var}(r_{d,t}) = \sigma_{pd}^2 + \sigma_i^2/4 + \lambda^2 \sigma_u^2 = \sigma_{pd}^2 + (1+\alpha)\sigma_i^2/4$$

$$\mathrm{Var}(r_{o,t}) = \sigma_{po}^2 + \sigma_i^2/4 + \alpha \sigma_i^2/4 = \sigma_{po}^2 + (1+\alpha)\sigma_i^2/4$$

$$\mathrm{Cov}(r_{d,t}, r_{o,t}) = \sigma_i^2/4 - \lambda^2 \sigma_u^2 = (1-\alpha)\sigma_i^2/4$$

$$\mathrm{Cov}(r_{d,t}, y_t) = \sigma_i^2/4\lambda + \lambda \sigma_u^2 = \alpha^{-1/2} \sigma_i \sigma_u/2 + \alpha^{1/2} \sigma_i \sigma_u/2$$

$$\mathrm{Cov}(r_{o,t}, y_t) = \sigma_i^2/4\lambda - \lambda \sigma_u^2 = \alpha^{-1/2} \sigma_i \sigma_u/2 - \alpha^{1/2} \sigma_i \sigma_u/2$$

Although the *un*conditional distribution is not normal, the unconditional variances and

covariances can be expressed using the parameters as follows

$$\text{Var}(y_t) = \alpha(1+1/\alpha)\sigma_u^2 + (1-\alpha)\sigma_u^2 = 2\sigma_u^2$$

$$\text{Var}(r_{d,t}) = \alpha(\sigma_{pd}^2 + (1+\alpha)\sigma_i^2/4) + (1-\alpha)(\sigma_{pd}^2 + \alpha\sigma_i^2/4) = \sigma_{pd}^2 + \alpha\sigma_i^2/2$$

$$\text{Var}(r_{o,t}) = \alpha(\sigma_{po}^2 + (1+\alpha)\sigma_i^2/4) + (1-\alpha)(\sigma_{po}^2 + \alpha\sigma_i^2/4) = \sigma_{po}^2 + \alpha\sigma_i^2/2$$

$$\text{Cov}(r_{d,t}, r_{o,t}) = \alpha(1-\alpha)\sigma_i^2/4 + (1-\alpha)-\alpha\sigma_i^2/4 = 0$$

$$\text{Cov}(r_{d,t}, y_t) = \alpha(\alpha^{-1/2}\sigma_i\sigma_u/2 + \alpha^{1/2}\sigma_i\sigma_u/2) + (1-\alpha)(\alpha^{1/2}\sigma_i\sigma_u/2) = \alpha^{1/2}\sigma_i\sigma_u$$

$$\text{Cov}(r_{o,t}, y_t) = \alpha(\alpha^{-1/2}\sigma_i\sigma_u/2 - \alpha^{1/2}\sigma_i\sigma_u/2) + (1-\alpha)(-\alpha^{1/2}\sigma_i\sigma_u/2) = 0$$