

## Is Information Risk a Determinant of Asset Returns?

DAVID EASLEY, SOEREN HVIDKJAER, and MAUREEN O'HARA\*

### ABSTRACT

We investigate the role of information-based trading in affecting asset returns. We show in a rational expectation example how private information affects equilibrium asset returns. Using a market microstructure model, we derive a measure of the probability of information-based trading, and we estimate this measure using data for individual NYSE-listed stocks for 1983 to 1998. We then incorporate our estimates into a Fama and French (1992) asset-pricing framework. Our main result is that information does affect asset prices. A difference of 10 percentage points in the probability of information-based trading between two stocks leads to a difference in their expected returns of 2.5 percent per year.

ASSET PRICING IS FUNDAMENTAL to our understanding of the wealth dynamics of an economy. This central importance has resulted in an extensive literature on asset pricing, much of it focusing on the economic factors that influence asset prices. Despite the fact that virtually all assets trade in markets, one set of factors not typically considered in asset-pricing models are the features of the markets in which the assets trade. Instead, the literature on asset pricing abstracts from the mechanics of asset price evolution, leaving unsettled the underlying question of how equilibrium prices are actually attained.

Market microstructure, conversely, focuses on how the mechanics of the trading process affect the evolution of trading prices. A major focus of this extensive literature is on the process by which information is incorporated into prices. The microstructure literature provides structural models of how prices become efficient, as well as models of volatility, both issues clearly of importance for asset pricing. But of perhaps more importance, microstructure models pro-

\* Cornell University, University of Maryland, and Cornell University. The authors would like to thank Anat Admati; Yakov Amihud; Kerry Back; Patrick Bolton; Douglas Diamond; Ken French; William Gebhardt; Gordon Gemmill; Mark Grinblatt; Campbell Harvey; David Hirshleifer; Schmuël Kandell; Charles Lee; Bhaskaran Swaminathan; Zhenyu Wang; Ingrid Werner; two referees; the editor (René Stulz); and seminar participants at Cornell University, DePaul University, the Federal Reserve Bank of Chicago, Georgia State University, the Hong Kong University of Science and Technology, INSEAD, the Massachusetts Institute of Technology, Princeton University, the Red Sea Finance Conference, the Stockholm School of Economics, Washington University, the University of Chicago, the University of Essex, the Western Finance Association meetings, and the European Finance Association meetings for helpful comments. We are also grateful to Marc Lipson for providing us with data compaction programs.

vide explicit estimates of the extent of private information. The microstructure literature has demonstrated the important link between this private information and an asset's bid and ask trading prices, but it has yet to be demonstrated that such information actually affects asset-pricing fundamentals.

If a stock has a higher probability of private information-based trading, should that have an effect on its required return? In traditional asset-pricing models, the answer is no. These models rely on the notion that if assets are priced "efficiently," then information is already incorporated and hence need not be considered. But this view of efficiency is static, not dynamic. If asset prices are continually revised to reflect new information, then efficiency is a process, and how asset prices become efficient cannot be separated from asset returns at any point in time.

That stock returns might depend upon features of the trading process has been addressed in various ways in the literature. Perhaps the most straightforward approach is that of Amihud and Mendelson (1986), who argue that liquidity should be priced. Their argument is that investors maximize expected returns net of transactions (or liquidity) costs, where a simple measure of these costs is the bid-ask spread. In equilibrium, traders will require higher returns to hold stocks with larger spreads. Amihud and Mendelson (1986, 1989) and Eleswarapu (1997) present empirical evidence consistent with this liquidity hypothesis. Supporting evidence using other measures of liquidity is provided by Brennan and Subrahmanyam (1996), Amihud, Mendelson, and Lauterbach (1997), Brennan, Chordia, and Subrahmanyam (1998), Datar, Naik, and Radcliffe (1998), and Amihud (2000). But the overall research on this issue is mixed, with Eleswarapu and Reinganum (1993), Chen and Kan (1996), and Chalmers and Kadlec (1998) concluding that liquidity is not priced. Certainly, part of the difficulty in resolving this issue is simply that transactions costs, however measured, are often quite small. While identifying illiquidity with transactions costs seems sensible, finding these liquidity effects among the noise in asset returns need not be an easy task.

Our focus in this paper is on showing empirically that a different aspect of the trading and price discovery process—private information—affects cross-sectional asset returns. We first present a simple example to provide the intuition for why the existence of private information should affect stock returns.<sup>1</sup> The importance of this information effect is then an empirical question. However, this question is difficult to answer because the extent of private information is not directly observable. To deal with this problem, we use a structural market microstructure model to generate a measure of the probability of information-based trading (PIN) in individual stocks. We then estimate this measure using high-frequency data for NYSE-listed stocks for the period 1983 to 1998. The resulting estimates are a time series of individual stock probabilities of information-based trading for a very large cross

<sup>1</sup> The theoretical case for why information affects asset returns is developed more fully in Easley and O'Hara (2000). We present in this paper a brief example of this cross-sectional effect.

section of stocks. We investigate whether these information probabilities affect cross-sectional asset returns by incorporating our estimates into a Fama and French (1992) asset pricing framework.

Our main result is that information does affect asset prices: Stocks with higher probabilities of information-based trading have higher rates of return. Indeed, we find that a difference of 10 percentage points in PIN between two stocks leads to a difference in their expected returns of 2.5 percent per year. The magnitude, and statistical significance, of this effect provides strong support for the premise that information affects asset-pricing fundamentals.

Our focus on the role of information in asset pricing is related to several recent papers. In a companion theoretical paper, Easley and O'Hara (2000) develop a multiasset rational expectations equilibrium model in which stocks have differing levels of public and private information. In equilibrium, uninformed traders require compensation to hold stocks with greater private information, resulting in cross-sectional differences in returns. Admati (1985) generalized Grossman and Stiglitz's (1980) analysis of partially revealing rational expectations equilibrium to multiple assets and showed how individuals face differing risk-return trade-offs when differential information is not fully revealed in equilibrium. Wang (1993) provides an intertemporal asset-pricing model in which traders can invest in a riskless asset and a risky asset. In this model, the presence of traders with superior information induces an adverse selection problem, as uninformed traders demand a premium for the risk of trading with informed traders. However, trading by the informed investors also makes prices more informative, thereby reducing uncertainty. These two effects go in opposite directions, and their overall effect on asset returns is ambiguous. Because this model allows only one risky asset, it is not clear how, if at all, information would affect cross-sectional returns. Jones and Slezak (1999) also develop a theoretical model allowing for asymmetric information to affect asset returns. Their model relies on changes in the variance of news and liquidity shocks over time to differentially affect agents' portfolio holdings, thereby influencing asset returns. These theoretical papers suggest that asymmetric information can affect asset returns, the issue of interest in this paper.

An alternative stream of the literature considers the effects of incomplete, but symmetric, information on asset prices. Building from Merton (1987), a number of authors (see, e.g., Basak and Cuoco (1998) and Shapiro (2002, p. 49)) analyze asset pricing when traders can be unaware of the existence of some assets. In this setting, cross-sectional differences in returns can emerge simply because traders cannot hold assets they do not know about; the lack of demand for these unknown assets results in their commanding a higher return in equilibrium.<sup>2</sup> The symmetric information structure here is an im-

<sup>2</sup> A variant of this model is that traders face some exogenous participation constraints. For example, international investors may be unable to purchase assets that are available to domestic investors, and this can create cross-sectional differences in returns. A model that does allow asymmetric information in this setting is Brennan and Cao (1997).

portant difference between these models and our analysis. There is no risk of trading with traders who have better information because everyone knows the same information. Some traders may face participation constraints on holding some assets, but these constraints are not related to information on future asset performance. Indeed, as Merton pointed out, a problem with the incomplete information model is that any higher asset premium can be arbitrated away by traders who do not face such constraints. That is not the case with our asymmetric information risk; the risk remains in equilibrium even though all traders know it is there.

A related literature on estimation risk also examines the effect of differential information on expected returns (see Barry and Brown (1984, 1985), Barry and Jennings (1992), or Coles, Loewenstein, and Suay (1995) for examples). This literature asks how differences in investor confidence about return distributions affects expected returns. The basic conclusion is that securities for which there is little information will have higher expected returns. These securities are riskier for investors than securities about which they have more information, but this risk is different from that measured by  $\beta$ , and thus it affects measured excess returns. The difference between this approach and ours is that we focus on differences of information across investors, each of whom knows the structure of returns, and we ask whether having private versus public information affects security returns.

Finally, two recent empirical papers related to our analysis are Brennan and Subrahmanyam (1996) and Amihud (2000). These authors investigate how the slope of the relationship between trade volume and price changes affects asset returns. This measure of illiquidity relies on the price impact of trade, and it seems reasonable to believe that stocks with a large illiquidity measure are less attractive to investors. Brennan and Subrahmanyam find support for this notion using two years of transactions data to estimate the slope coefficient  $\lambda$ , while Amihud establishes a similar finding using daily data. What economic factors underlie this result is not clear. Because  $\lambda$  is derived from price changes, factors such as the impact of price volatility on daily returns, or inventory concerns by the market maker could influence this variable, as could adverse selection.<sup>3</sup> Neither analysis addresses whether their illiquidity measure is proxying for spreads, or for the more fundamental information risk we address. Our analysis here focuses directly on private information by deriving a trade-based measure of information risk. This

<sup>3</sup> The Kyle (1985)  $\lambda$  has not been tested as to its actual linkage with private information. While it seems reasonable to us that such a theoretical linkage would exist, there are a number of reasons why this empirical measure is problematic. For example, the actual Kyle model assumes a call market structure in which orders are aggregated and it is only the net imbalance that affects the price. Actual markets do not have this structure, so in practice  $\lambda$  is estimated on a trade-by-trade basis (as in Brennan and Subrahmanyam), or is a time-series change in price per volume over some interval (as in Amihud). Either approach may introduce noise in the specification. Moreover, because the  $\lambda$  calculation also involves both price and the quantity of the trade, its actual value may be affected by factors such as the size of the book, tick size consideration, and market maker inventory.

PIN measure has been shown in previous work (see Easley, Kiefer, & O'Hara (1996a, 1997a, 1997b)), Easley et al. (1996b), and Easley, O'Hara & Paperman (1998)) to explain a number of information-based regularities, providing the link to private information we need to investigate cross-sectional asset-pricing returns.

The PIN variable is correlated with other variables that we do not include in our return estimation. In particular, as would be expected with an information measure, PIN is correlated with spreads. It is also correlated with the variability of returns and with volume or turnover. One might suspect that the probability of information-based trade only seems to be priced because it serves as a proxy for these omitted variables. We show, however, that this is not the case. We show that over our sample period, spreads do not affect asset returns but PIN does. When spreads or the variability of returns are included with PIN in the return regressions, the probability of information-based trade remains highly significant, and its effect on returns is changed only slightly. Volume remains a factor in asset pricing, but it does not remove the influence of PIN. We view these results as strong evidence that the probability of information-based trade is priced in asset returns.

The paper is organized as follows. Section I provides the theoretical intuition for our analysis. We construct a partially revealing rational expectation example in which private information affects asset prices because it skews the portfolio holdings of informed and uninformed traders in equilibrium. We then turn in Section II to the empirical testing methodology. We set out a basic microstructure model and we demonstrate how the probability of information-based trading is derived for a particular stock. Estimation of the model involves maximum likelihood, and we show how to derive these estimating equations. In Section III, we present our estimates. We examine the cross-sectional distribution of our estimated parameters, and we examine their temporal stability. Section IV then puts our estimates into an asset-pricing framework. We use the cross-sectional approach of Fama and French (1992) to investigate expected asset returns. In this section, we present our results, and we investigate their robustness. We also investigate the differential ability of spreads, variability of returns, turnover, and our information measure to affect returns. Section V summarizes our results and discusses their implications for asset-pricing research.

## **I. Information and Asset Prices**

To show that whether information is private or public affects asset returns, we construct a simple rational expectations equilibrium asset-pricing example. We use this example only to motivate our empirical search for information effects, so we keep the exposition here as simple as possible. A more general analysis of the effects of public and private information on asset prices is found in a companion theoretical paper by Easley and O'Hara (2000).

In standard capital-asset-pricing models (CAPM), individuals have common beliefs, and assets are priced according to these beliefs. Market risk must be held, and in equilibrium, individuals are compensated with greater expected returns for holding it. But no one must hold idiosyncratic risk, and so there is no market compensation for doing so. We show here that when there is differential information that is not fully revealed in equilibrium, and individuals thus have differing beliefs, the story is more complex.

We begin with a simple example that we make more complex, and more realistic, in three stages. First, we consider a market in which individuals have, for whatever reason, arbitrarily differing beliefs. These individuals perceive differing risk-return trade-offs in the market, and, most importantly, they typically hold differing portfolios. Thus, they will rationally choose to hold idiosyncratic risk. They believe that they are being compensated for holding this risk because they believe that assets are mispriced. Whether they are in fact compensated for holding idiosyncratic risk depends on how their beliefs relate to the truth. A trader with correct beliefs, in a world in which some others are incorrect, correctly perceives an expected excess return to holding some assets and he overweights his portfolio, relative to the market, in these assets.

We then drop the arbitrary beliefs assumption, and we generate differences in beliefs from common priors and private information that is not fully revealed by equilibrium asset prices. In this second part of our example, there is nothing irrational about any of the individuals: They have common priors, they receive differing information, they make correct inferences from their information and market prices, but they end up with differing beliefs as long as all information is not fully revealed. Individuals with private information have better beliefs, that is, they correctly perceive expected excess returns on some assets, and they hold better portfolios than do those without private information. In this case, the existence of private information generates expected (average) excess returns to some assets.

Third, we ask what happens if this private information is made public. Here the traders hold a common portfolio. They hold only market risk, not any idiosyncratic risk. We show that this reduces the expected excess returns to assets about which information was previously private. In this case, individuals are compensated only for holding market-wide risk whereas before they were compensated for the extra risk induced by the private information. We conclude that private, rather than public, information can yield an increase in required expected returns.

For our example, we consider a two-period, dates  $t$  and  $t + 1$ , consumption-based asset-pricing model. There are two possible states of the world,  $s = 1, 2$ , at date  $t + 1$ , and there are two assets with state contingent payoffs. Asset one pays three units of the single consumption good in state one and zero units in the other state; asset two pays three units of the good in state two and zero units in state one. Since each asset is identified with the state in which it pays off, we index both states and assets by  $s$ . Initially we con-



sider an economy in which there is one unit of each asset available. The price of the consumption good is one at each date and asset prices at date  $t$  are  $p = (p_1, p_2)$ .

There are two traders indexed by  $i = 1, 2$ . Each trader is endowed with one unit of the consumption good and with one-half of the available assets. Traders have logarithmic utility of consumption at dates  $t$  and  $t + 1$  and they discount future utility with discount factor  $0 < \rho < 1$ . Trader  $i$ 's belief about the distribution of states tomorrow is  $q^i = (q_1^i, q_2^i)$ . Trader  $i$  chooses current consumption and asset purchases,  $(c_t^i, \alpha_1^i, \alpha_2^i)$ , to maximize his expected utility of current and future state contingent consumption,  $(c_t^i, c_{t+1}^i(1), c_{t+1}^i(2))$ ,

$$\ln(c_t^i) + \rho[q_1^i \ln(c_{t+1}^i(1)) + q_2^i \ln(c_{t+1}^i(2))] \quad (1)$$

subject to his budget constraints

$$c_t^i = w_t^i - p_1 \alpha_1^i - p_2 \alpha_2^i, c_{t+1}^i(1) = 3\alpha_1^i, c_{t+1}^i(2) = 3\alpha_2^i, \quad (2)$$

where  $w_t^i = 1 + (p_1 + p_2)/2$  is date  $t$  wealth.

If the traders have common beliefs,  $q$ , then it is easy to show that equilibrium asset prices are  $p_s^* = 2q_s\rho$ , and that, in equilibrium, each trader holds one-half a unit of each asset. Thus, the traders hold risk-free portfolios. There is no single risk-free asset in this economy, but traders can create one by buying equal amounts of the two risky assets. The shadow risk-free rate (of return) this implies is  $3/2\rho^{-1}$ . So the price of asset  $s$  is its expected payoff,  $3q_s$ , divided by the risk-free rate,  $3/2\rho^{-1}$ . As there is no market risk in the economy, this is just a simple example of the consumption-based CAPM.

If the traders have differing beliefs, similar calculations show that equilibrium prices are  $p_s = 2q_s\rho$  where  $\bar{q}_s = 1/2(q_s^1 + q_s^2)$  is the average belief about the probability of state  $s$ . So again, the price of each asset is an expected return divided by the risk-free rate (the risk-free rate is still  $3/2\rho^{-1}$ ), but now the expectation is a market or average expectation—not any individual's expectation. This is an example of Lintner's (1969) generalization of the CAPM to heterogeneous beliefs. Although traders can still create a risk-free asset, they choose to hold idiosyncratic risk. This calculation shows that equilibrium date  $t + 1$  consumptions for trader  $i$  in states one and two are  $(3q_1^i/2\bar{q}_1, 3q_2^i/2\bar{q}_2)$ . Unless beliefs are identical, traders have random date  $t + 1$  consumptions even though there is no market risk in this economy. This occurs because each trader believes the assets are mispriced, and so each trader is willing to accept some risk in order to take advantage of the perceived mispricing.

Suppose that trader one's beliefs are, in fact, correct. Then the expected return on asset  $s$  is  $3q_s^1$ . So asset prices are not (correct) expected returns divided by the risk-free rate. There is a positive excess return (expected return divided by price minus the risk-free rate) on one asset and a negative

excess return on the other asset. This occurs even though there is no market-wide risk: There will be three units of the good tomorrow for sure. The trader with better beliefs holds a better portfolio; he overweights his portfolio in the asset with a positive excess return.

This simple example shows that if traders have differing beliefs they perceive differing risk-return trade-offs and they may choose to hold idiosyncratic risk. Each trader believes that he is being compensated for the risk he holds. In fact, at least one of these traders has incorrect beliefs. So his perceptions about the risk-return trade-off are also incorrect. Next, we introduce private information into the economy so that neither trader has incorrect beliefs—they have a common prior and more or less information.<sup>4</sup>

The common prior on states is  $(\frac{1}{2}, \frac{1}{2})$ . Trader one, the informed trader, receives a private signal  $y \in \{1, 2\}$  with probability  $\frac{1}{2}$  on each signal. If  $y = 1$ , then the conditional probability of state one is  $\frac{3}{4}$ , and if  $y = 2$ , the conditional probability of state one is  $\frac{1}{4}$ . Trader two is uninformed, but he knows this structure and he uses this knowledge, along with equilibrium prices, to make any inferences that he can about the informed trader's information. Unless we introduce further randomness into the economy, prices will reveal the informed trader's signal and, in equilibrium, traders will have common beliefs. To prevent this uninteresting case, we use the standard device of noisy supply.<sup>5</sup> The aggregate supply of assets one and two is given by the random variable  $x \in \{(3/5, 1), (1, 3/5)\}$  with probability  $\frac{1}{2}$  on each supply vector. This random supply is equally divided across the traders to form their initial endowments of assets.<sup>6</sup> We assume that  $x$  and  $y$  are uncorrelated. So there are four states of the world at time  $t$ ,  $z \in Z = \{z_1, z_2, z_3, z_4\} = \{(1, (3/5, 1)), (1, (1, 3/5)), (2, (3/5, 1)), (2, (1, 3/5))\}$ , each of which is equally likely.

Calculation shows that rational expectations equilibrium prices and shadow risk-free rates are as shown in Table I. Equilibrium prices in states  $z_2$  and  $z_3$  are equal so the date  $t$  equilibrium is nonrevealing in these date  $t$  states. Prices in each of states  $z_1$  and  $z_4$  differ from all others, so if the date  $t$  state is one of these, the equilibrium is revealing. Thus, equilibrium beliefs of the uninformed trader are  $(\frac{3}{4}, \frac{1}{4})$  in state  $z_1$ ,  $(\frac{1}{2}, \frac{1}{2})$  in states  $z_2$  and  $z_3$ , and

<sup>4</sup> We do not believe that traders' differing beliefs necessarily come from this type of common prior structure. Disagreement about probabilities seems far more natural than does a common prior. When traders disagree, market prices provide information about others beliefs, but without some further structure it is not clear how or if traders should use this to change their own beliefs. We use the standard common beliefs and information structure to analyze the effects of private versus public information. The analysis can be done without common priors.

<sup>5</sup> An alternative that works equally well is to introduce noisy traders. In our analysis, this is easily done by having some traders whose beliefs are random and who do not learn from prices.

<sup>6</sup> We assume that traders do not make an inference about the state of the world from their endowment. Alternatively, we could assume that the uninformed trader has a constant endowment and that only the informed trader's endowment varies. We do not do this only because it complicates the calculations. Another standard alternative is to allow a random exogenous supply of the assets. We do not do this only because then we would have a partial equilibrium model, which is more difficult to compare to the usual consumption-based asset-pricing structure.



**Table I**  
**Equilibrium Prices and Shadow Risk-free Rates**

Date $t$ State	Price of Asset 1	Price of Asset 2	Risk-free Rate
$z_1$	$5/2\rho$	$1/2\rho$	$\rho^{-1}$
$z_2$	$5/4\rho$	$5/4\rho$	$6/5\rho^{-1}$
$z_3$	$5/4\rho$	$5/4\rho$	$6/5\rho^{-1}$
$z_4$	$1/2\rho$	$5/2\rho$	$\rho^{-1}$

$(\frac{1}{4}, \frac{3}{4})$  in state  $z_4$ . The informed traders beliefs are  $(\frac{3}{4}, \frac{1}{4})$  in states  $z_1$  and  $z_2$  and  $(\frac{1}{4}, \frac{3}{4})$  in states  $z_3$  and  $z_4$ . So we have endogenously generated rational differences in beliefs.

Because of the differing beliefs that the traders have in states  $z_2$  and  $z_3$ , they hold differing portfolios in these states and they accept risk that, in aggregate, they do not have to hold. As before with exogenous differences in beliefs, the traders perceive differing risk-return trade-offs and they believe that they are being compensated for the risk that they hold. The interesting question is what is the market compensation for risk? The expected excess return on an asset is typically defined to be the expected return on the asset minus the risk-free rate. In this economy, because of the differing beliefs, the expected return cannot be computed uniformly across traders, so instead we compute it from the point of view of an outside observer. This is the expected return that would be measured by empirical averages of returns. The outside observer sees three possible date  $t$  states  $z_1$ ,  $z_2 = z_3$ , and  $z_4$  with probabilities  $\frac{1}{4}$ ,  $\frac{1}{2}$ , and  $\frac{1}{4}$ . Computing the average excess return in each state, and then averaging over states, yields an average excess return, for each risky asset, of  $0.1 \rho^{-1}$ .

In this economy, there is a random supply of assets, so there is market risk, and traders must be compensated for holding that risk. But the traders have different information, so they hold different amounts of the market risk. They do not each hold half of the market. Does this require more compensation for risk than an economy with public information? To answer this question, we compute the expected excess return on risky assets when all information is public. That is, we assume that in each state of the world, at date  $t$  both traders receive the signal,  $y$ . We then compute for each state the market clearing prices and the risk-free rate of return. These prices differ across all four possible date  $t$  states, and for each state we compute the average excess return on asset holding and then average this over the date  $t$  states. The result is an average excess return, for each risky asset, of  $0.057 \rho^{-1}$ . This return is less than the return in the private information economy.

In this economy, when information about the payoff on risky assets is private rather than public, the market requires a greater expected excess return. This occurs because when information is private rather than public and uninformed investors cannot perfectly infer it from prices, they view the

asset as being more risky. Uninformed investors could avoid this risk, but they choose not to do so. To completely avoid this risk, uninformed traders would have to hold only the risk-free asset, but this is not optimal; they receive higher expected utility by holding some of the risky, private information assets. They are rational, so they hold an optimally diversified portfolio, but no matter how they diversify, uninformed traders are taken advantage of by informed traders who have learned which assets to hold. Although the example has only one trading period, it is easy to see that uninformed investors also would not choose to avoid this risk by buying and holding a fixed portfolio over time. In each trading period in an intertemporal model, uninformed investors reevaluate their portfolios. As prices change, they optimally change their holdings.

In our example, information about the payoff on an asset provides information about the return to holding the entire market of risky assets. As long as information is useful, this feature cannot generally be avoided in a finite asset state-space framework, as the payoff to holding the market is the sum of payoffs to holding all of the individual assets. So the example does not allow us to isolate the effect of asset-specific information versus information about a common component in security returns. In our companion theoretical paper, we consider an arbitrary number of assets and continuous states spaces and show that similar asset-specific effects on required excess returns emerge.

The example suggests that the existence of private information, either about a common component of asset returns or about a single asset in a finite asset economy, should affect asset prices. The significance of this effect is an empirical question. The natural approach to this empirical question would be to measure the extent of differential information asset by asset, look for any common components, and ask whether either the asset-specific measures or the common components are priced. But because private information is not directly observable, it cannot be measured directly; its presence can only be inferred from market data. Fortunately, the market microstructure literature provides a way to do this. The probability of information-based trade (PIN) from Easley et al. (1997b) measures the prevalence of private information in a microstructure setting. This measure is derived in a market microstructure model, not in the Walrasian setting used in the example, but it has been shown to have predictive power as a measure of private information.

In the next section we derive the PIN measure and show how to estimate it.<sup>7</sup> We then use PIN as a proxy for the private information in the theoretical example and ask whether it is priced in a Fama–French asset-pricing regression.

<sup>7</sup> If a stock has more private information and an unchanged amount of public information, its equilibrium expected return falls. This occurs because risk is reduced. Here we keep the underlying information structure fixed and vary the split of this information between public and private.

## II. Microstructure and Asset Prices

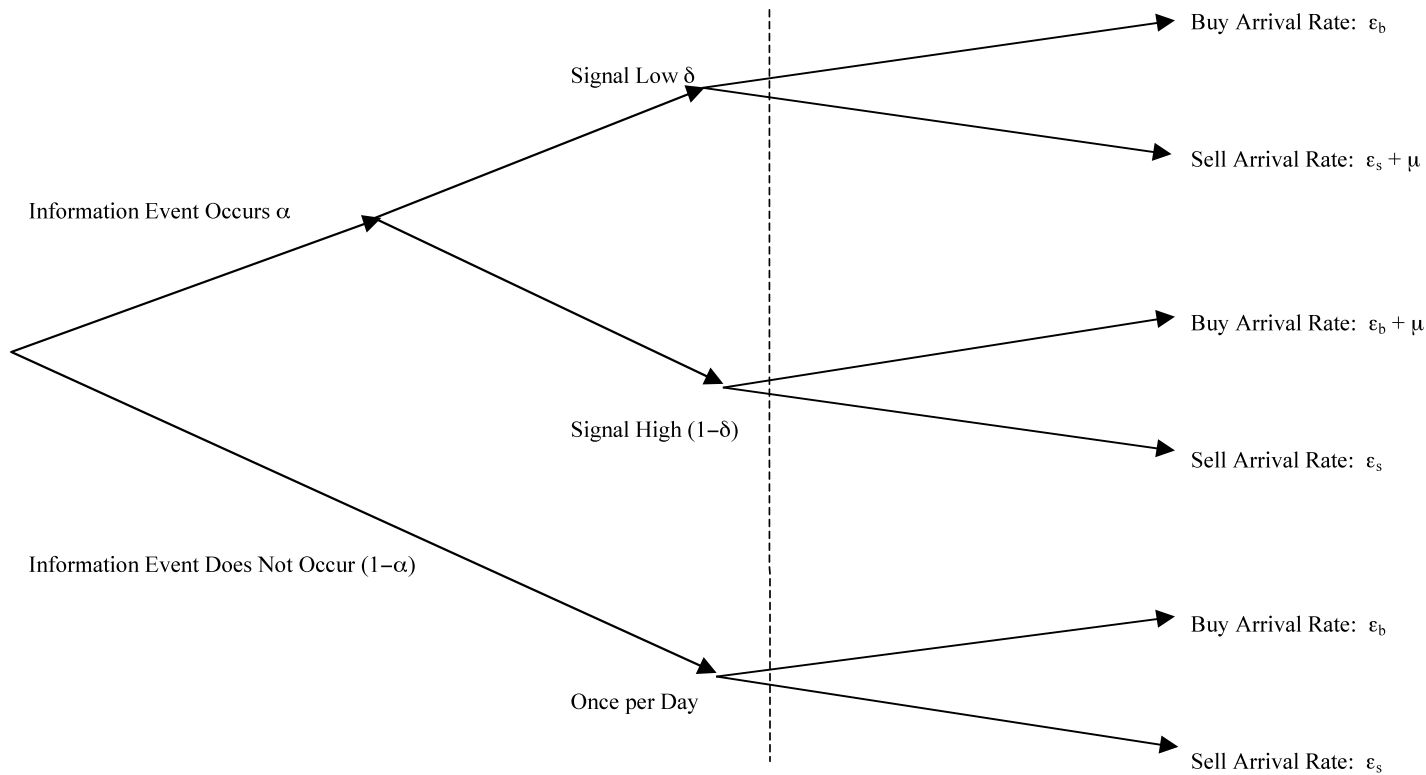
Consider what we know from the microstructure literature (see O'Hara (1995) for a discussion and derivation of microstructure models). Microstructure models can be viewed as learning models in which market makers watch market data and draw inferences about the underlying true value of an asset. Crucial to this inference problem is their estimate of the probability of trade based on private information about the stock. Market makers watch trades, update their beliefs about this private information, and set trading prices. Over time, the process of trading, and learning from trading, results in prices converging to full information levels.

As an example, consider the simple sequential trade tree diagram given in Figure 1. Microstructure models depict trading as a game between the market maker and traders that is repeated over trading days  $i = 1, \dots, I$ . First, nature chooses whether there is new information at the beginning of the trading day, these events occur with probability  $\alpha$ . The new information is a signal regarding the underlying asset value, where good news is that the asset is worth  $\bar{V}_i$ , and bad news is that it is worth  $\underline{V}_i$ . Good news occurs with probability  $(1 - \delta)$  and bad news occurs with the remaining probability,  $\delta$ . Trading for day  $i$  then begins with traders arriving according to Poisson processes throughout the day. The market maker sets prices to buy or sell at each time  $t$  in  $[0, T]$  during the day, and then executes orders as they arrive. Orders from informed traders arrive at rate  $\mu$  (on information event days), orders from uninformed buyers arrive at rate  $\epsilon_b$ , and orders from uninformed sellers arrive at rate  $\epsilon_s$ . Informed traders buy if they have seen good news and sell if they have seen bad news. If an order arrives at time  $t$ , the market maker observes the trade (either a buy or a sale), and he uses this information to update his beliefs. New prices are set, trades evolve, and the price process moves in response to the market maker's changing beliefs. This process is captured in Figure 1.

Suppose we now view this problem from the perspective of an econometrician. If we, like the market maker, observe a particular sequence of trades, what could we discover about the underlying structural parameters and how would we expect prices to evolve? This is the intuition behind a series of papers by Easley et al. (1996a, 1997a, 1997b), who demonstrate how to use a structural model to work backwards to provide specific estimates of the risks of information-based trading in a stock. They show that these structural models can be estimated via maximum likelihood, providing a method for determining the probability of information-based trading in a given stock.

Is this probability a good proxy for the information risk described in the previous section? We would argue that it is.<sup>8</sup> The information risk we have

<sup>8</sup> The theoretical example raises the possibility that there is a common component in private information across stocks. We estimate PIN stock by stock, so our empirical work does not explicitly take this into account. It would be interesting to investigate a more complex microstructure model with both stock-specific and common information and ask how these factors are separately priced.



**Figure 1. Tree diagram of the trading process.**  $\alpha$  is the probability of an information event,  $\delta$  is the probability of a low signal,  $\mu$  is the rate of informed trade arrival,  $\epsilon_b$  is the arrival rate of uninformed buy orders, and  $\epsilon_s$  is the arrival rate of uninformed sell orders. Nodes to the left of the dotted line occur once per day.

modeled here (and in more detail in Easley and O'Hara (2000)) is viewed from the perspective of an uninformed trader, but it should be remembered that the market maker is similarly uninformed. The risk is due to private, not public, information, and this too is a feature of the microstructure setting above. The information risk is greatest when there are more frequent information events (captured here by  $\alpha$ ), and/or more informed traders getting the information (captured by  $\mu$ ), and it is mitigated by the willingness of other traders to hold the stock (captured by the  $\epsilon_s$ ). There are, of course, differences between the rational expectations framework and microstructure models, and some of these may be important. But the underlying information variable is what matters for our analysis, and the tractability of the microstructure variable provides a coherent way to estimate this. To the extent that this proxy does not capture the information risk we seek, then we would not expect to find any significant effects of this variable on asset returns.

Returning to the structural model, the likelihood function induced by this simple model of the trade process for a single trading day is

$$\begin{aligned} L(\theta|B,S) = & (1 - \alpha)e^{-\epsilon_b} \frac{\epsilon_b^B}{B!} e^{-\epsilon_s} \frac{\epsilon_s^S}{S!} \\ & + \alpha\delta e^{-\epsilon_b} \frac{\epsilon_b^B}{B!} e^{-(\mu+\epsilon_s)} \frac{(\mu + \epsilon_s)^S}{S!} \\ & + \alpha(1 - \delta)e^{-(\mu+\epsilon_b)} \frac{(\mu + \epsilon_b)^B}{B!} e^{-\epsilon_s} \frac{\epsilon_s^S}{S!}, \end{aligned} \quad (3)$$

where  $B$  and  $S$  represent total buy trades and sell trades for the day respectively, and  $\theta = (\alpha, \mu, \epsilon_b, \epsilon_s, \gamma)$  is the parameter vector. This likelihood is a mixture of distributions where the trade outcomes are weighted by the probability of it being a “good news day”  $\alpha(1 - \delta)$ , a “bad-news day”  $(\alpha\delta)$ , and a “no-news day”  $(1 - \alpha)$ .

Imposing sufficient independence conditions across trading days gives the likelihood function across  $I$  days

$$V = L(\theta|M) = \prod_{i=1}^I L(\theta|B_i, S_i), \quad (4)$$

where  $(B_i, S_i)$  is trade data for day  $i = 1, \dots, I$  and  $M = ((B_1, S_1), \dots, (B_I, S_I))$  is the data set.<sup>9</sup> Maximizing (4) over  $\theta$  given the data  $M$  thus provides a way to determine estimates for the underlying structural parameters of the model (i.e.,  $\alpha, \mu, \epsilon_b, \epsilon_s, \delta$ ).

<sup>9</sup> The independence assumptions essentially require that information events are independent across days. Easley et al. (1997b) do extensive testing of this assumption and are unable to reject the independence of days.

This model allows us to use observable data on the number of buys and sells per day to make inferences about unobservable information events and the division of trade between the informed and uninformed. In effect, the model interprets the normal level of buys and sells in a stock as uninformed trade, and it uses this data to identify  $\epsilon_b$  and  $\epsilon_s$ . Abnormal buy or sell volume is interpreted as information-based trade, and it is used to identify  $\mu$ . The number of days in which there is abnormal buy or sell volume is used to identify  $\alpha$  and  $\delta$ . Of course, the maximum likelihood actually does all of this simultaneously. For example, consider a stock that always has 40 buys and 40 sells per day. For this stock,  $\epsilon_b$  and  $\epsilon_s$  would be identified as 40 (where the parameters are daily arrival rates),  $\alpha$  would be identified as zero, and  $\delta$  and  $\mu$  would be unidentified. Suppose, instead, that on 20 percent of the days there are 90 buys and 40 sells; and, on 20 percent of the days there are 40 buys and 90 sells. The remaining 60 percent of the days continue to have 40 buys and 40 sells. The parameters in this example would be identified as  $\epsilon_b = \epsilon_s = 40$ ,  $\mu = 50$ ,  $\alpha = 0.4$ , and  $\delta = 0.5$ .

One might conjecture that this trade imbalance statistic is too simplistic to capture the actual influence of informed trading. In particular, because trading volume naturally fluctuates, perhaps these trade imbalance deviations are merely natural artifacts of random market influences and are not linked to information-based trade as argued here. However, it is possible to test for this alternative by restricting the weights on the mixture of distributions to be the same across all days. This “random volume” model is soundly rejected in favor of the information-mixture model derived above (see Easley et al. (1997b) for procedure and estimation results). A second concern is that the model uses only patterns in the number of trades, and not patterns in volume, to identify the structural parameters.<sup>10</sup> It is possible to add trade size to the underlying approach, in which case the sufficient statistic for the trade process is the four-tuple (#large buys, #large sells, #small buys, and #small sales). This greatly increases the computational complexity, but as shown in Easley et al. (1997a), there appears to be little gain in doing so, as the trade size variables do not generally reveal differential information content. Given the extensive estimation required in this project, we have chosen to use the simple model derived above; to the extent that this omits important factors, we would expect the ability of our estimates to predict asset returns to be reduced.

We now turn to the economic use of our structural parameters. The estimates of the model's structural parameters can be used to construct the theoretical opening bid and ask prices.<sup>11</sup> As is standard in microstructure models, a market maker sets trading prices such that his expected losses to

<sup>10</sup> This number of trades approach is consistent with the findings of Jones, Kaul, and Lipson (1994), who argue that volume does not provide information beyond number of trades.

<sup>11</sup> Given any history of trades, we can also construct the theoretical bid and ask prices at any time during the trading day. But in our empirical work we focus on opening prices so we provide here only the derivation for the opening spread.



informed traders just offset his expected gains from trading with uninformed traders. This balancing of gains and losses is what gives rise to the “spread” between bid and ask prices. The opening spread is easiest to interpret if we express it explicitly in terms of this information-based trading. It is straightforward to show that the probability that the opening trade is information-based, PIN, is

$$PIN = \frac{\alpha\mu}{\alpha\mu + \epsilon_s + \epsilon_b}, \quad (5)$$

where  $\alpha\mu + \epsilon_s + \epsilon_b$  is the arrival rate for all orders and  $\alpha\mu$  is the arrival rate for information-based orders. The ratio is thus the fraction of orders that arise from informed traders or the probability that the opening trade is information based. In the case where the uninformed are equally likely to buy and sell ( $\epsilon_b = \epsilon_s = \epsilon$ ) and news is equally likely to be good or bad ( $\delta = 0.5$ ), the percentage opening spread is<sup>12</sup>

$$\frac{\Sigma}{V_i^*} = (PIN) \frac{(\bar{V}_i - V_i)}{V_i^*}, \quad (6)$$

where  $V_i^*$  is the unconditional expected value of the asset given by  $V_i^* = \delta\bar{V}_i + (1 - \delta)V_i$ . The opening spread is therefore directly related to the probability of informed trading. Note that if PIN equals zero, either because of the absence of new information ( $\alpha$ ) or traders informed of it ( $\mu$ ), the spread is also zero. This reflects the fact that only asymmetric information affects spreads when market makers are risk neutral.

Returning to our example of a stock that has trade resulting in estimated parameters of  $\epsilon_b = \epsilon_s = 40$ ,  $\mu = 50$ ,  $\alpha = 0.4$ , and  $\delta = 0.5$ , we see that PIN for this stock would be 0.2. This means that for this stock the market maker believes that 20 percent of the trades come from informed traders. This risk of information-based trade results in a spread, but the size of this spread also depends on the variability of the value of the stock. If this stock typically has a range of true values of \$4 around an expected value on day  $i$  of \$50, then its opening spread,  $\Sigma$ , would be predicted to be \$0.80, resulting in an opening percentage spread around \$50 of 1.6 percent.

Neither the estimated measure of information-based trading nor the predicted spread is related to market maker inventory because these factors do not enter into the model. Instead, these estimates represent a pure measure of the risk of private information. More complex models can also be estimated, allowing for greater complexity in the trading and information processes. Easley et al. (1996a, 1996b, 1997a, 1997b, 1998) have used these measures of asymmetric information to show how spreads differ between

<sup>12</sup> The predicted spread can also be derived for the general case where  $\epsilon_b \neq \epsilon_s$  and  $\delta \neq 0.5$ ; see Easley, O'Hara, and Saar (2001).

frequently and infrequently traded stocks, to investigate how informed trading differs between market venues, to analyze the information content of trade size, and to determine if financial analysts are informed traders.

Whether asymmetric information also affects required asset returns is the issue of interest in this paper. The model and estimating procedure detailed above provide a mechanism for determining the probability of information-based trading, and it is this PIN variable that we will explore in an asset-pricing context in Section IV of this paper. Asset-pricing considerations, however, are inherently dynamic, focusing as they do on the return that traders require over time to hold a particular asset. This dictates that any information-linked return must also be dynamic, and hence we need to focus on the time-series properties of our estimated information measure. Prefatory to this, however, is the more fundamental problem of estimating PIN when the underlying structural variables can be time-varying.

In the next section, we address these estimation issues. Using time-series data for a cross section of stocks, we maximize the likelihood functions given by our structural model. We use our estimates of the structural parameters to calculate PIN, and we investigate the temporal stability of these estimates. Having established the statistical properties of our estimates, we then address the link between information and asset pricing in the following section.

### **III. The Estimation of Information-Based Trading**

#### *A. Data and Methodology*

We estimate our model for a sample of all ordinary common stocks listed on the New York Stock Exchange (NYSE) for the years 1983 to 1998. We focus on NYSE-listed stocks because the market microstructure of that venue most closely conforms to that of our structural model. We exclude real estate investment trusts, stocks of companies incorporated outside of the United States, and closed-end funds. We also exclude a stock in any year in which it did not have at least 60 days with quotes or trades, as we cannot estimate our trade model reliably for such stocks. The final sample has between 1,311 and 1,846 stocks which we analyze each year.

The likelihood function given in equation (4) depends on the number of buys and sells each day for each stock in our sample. Transactions data gives us the daily trades for each of our stocks, but we need to classify these trades as buys or sells. To construct this data, we first retrieve transactions data from the Institute for the Study of Security Markets (ISSM) and Trade And Quote (TAQ) data sets. We then classify trades as buys or sells according to the Lee–Ready algorithm (see Lee and Ready (1991)). This algorithm is standard in the literature and it essentially uses trade placement relative to the current bid and ask quotes to determine trade direction.<sup>13</sup> Using this

<sup>13</sup> See Ellis, Michaely, and O'Hara (2000) for an analysis of alternative trade classification algorithms and their accuracy.

data, we maximize the likelihood function over the structural parameters,  $\theta = (\alpha, \mu, \epsilon_B, \epsilon_S, \delta)$ , for each stock separately for each year in the sample period. This gives us one yearly estimate per stock for each of the underlying parameters.<sup>14</sup>

The underlying model involves two parameters relating to the daily information structure ( $\alpha$ , the probability of new information, and  $\delta$ , the probability that new information is bad news) and three parameters relating to trader composition ( $\mu$ , the arrival rate of informed traders, and  $\epsilon_s$  and  $\epsilon_b$ , the arrival rates of uninformed sellers and buyers). Information on  $\mu$ ,  $\epsilon_s$ , and  $\epsilon_b$  accumulates at a rate approximately equal to the square root of the number of trade outcomes, while information on  $\alpha$  and  $\delta$  accumulates at a rate approximately equal to the square root of the number of trading days. The difference in information accumulation rates dictates that the precision of our  $\mu$  and  $\epsilon$  estimates will exceed that of our  $\alpha$  and  $\delta$  estimates, but the length of our time series is more than sufficient to provide precise estimates of each variable.

The maximum likelihood estimation converges for almost all stocks. Of more than 20,000 time series, only 716 did not converge. These failures were generally due to series with days of such extremely high trading volume compared to normal levels that convergence was not possible. Further, the estimation yielded only 456 corner solutions in  $\delta$ , the probability of an information event being bad news. Such corner solutions arise because a sustained imbalance of trading (e.g., more buys than sells) will result in the estimates of the probability of bad news being driven to one or zero. There are only six corner solutions found for  $\alpha$ , the probability of any day being an information day.<sup>15</sup> This finding is reassuring as it suggests the economically reasonable result that private information is a factor in the trading of every stock.

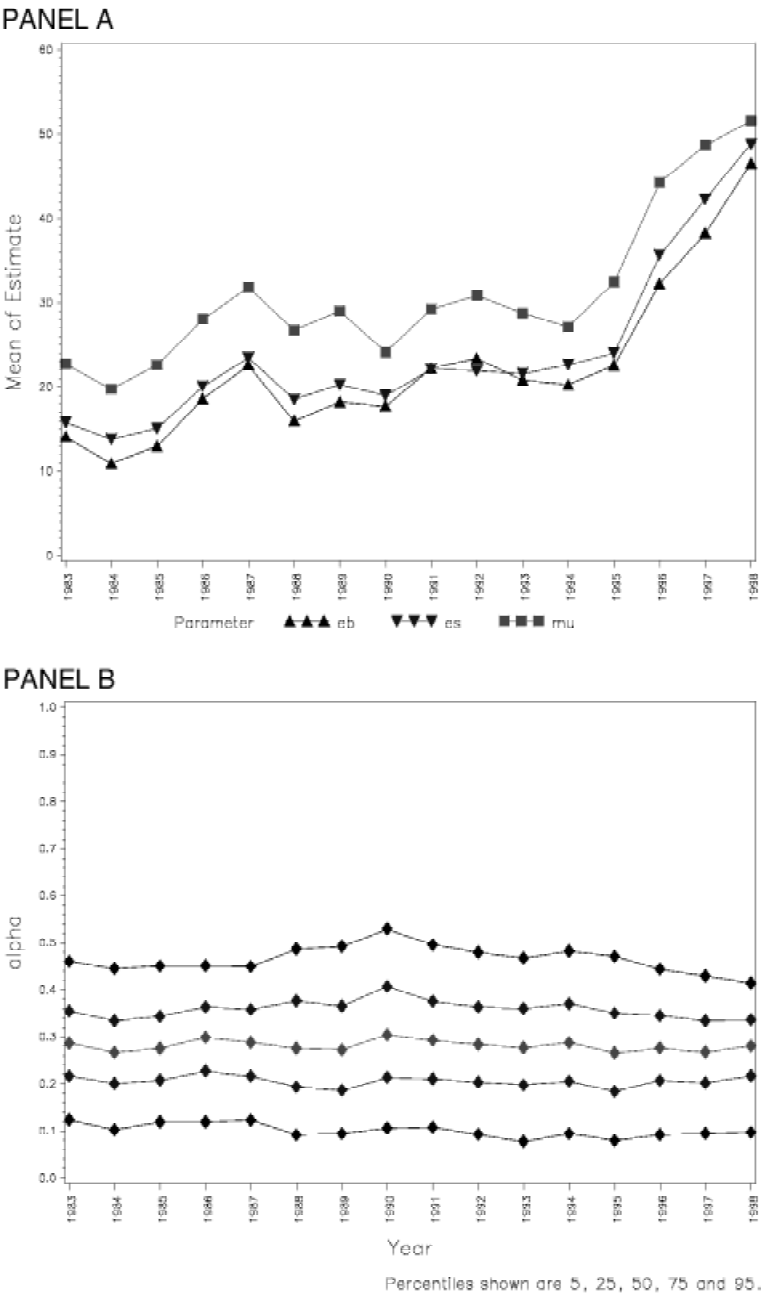
### B. Distribution of Parameter Estimates

The time-series patterns of the cross-sectional distribution of the individual parameter estimates are shown in Figure 2. The parameter estimates generally exhibit reasonable economic behavior. The estimates of  $\mu$ ,  $\epsilon_s$ , and  $\epsilon_b$  are related to trading frequency, and hence show an upward trend as trading volume increases on the NYSE over our sample period.<sup>16</sup> On the other hand, the estimates of  $\alpha$  and  $\delta$  are stable across years, and so, as expected, they do not trend.

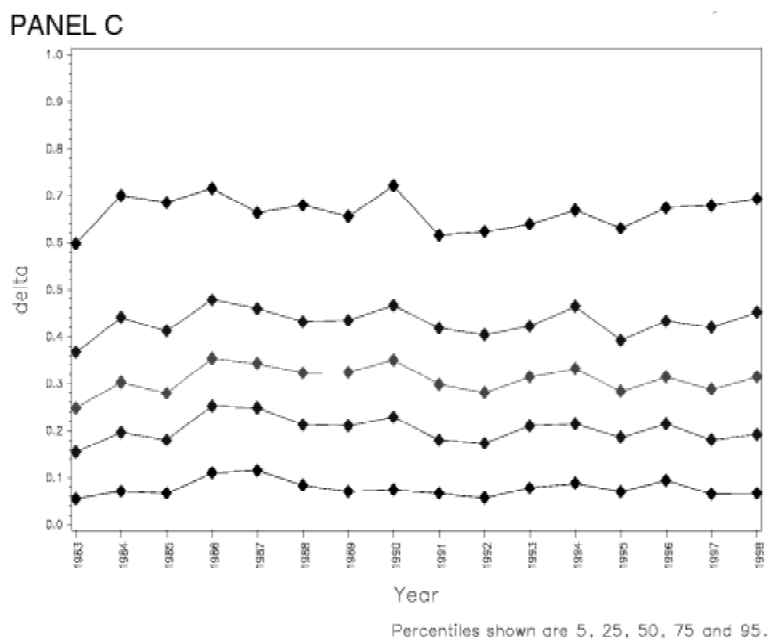
<sup>14</sup> We chose an annual estimation period because of the need to estimate the time series of the large number of stocks in our sample. The model can be estimated using as little as 60 trading days of data provided there is sufficient trading activity. We estimated our parameters over rolling 60-day windows for a subsample of stocks, but found little difference with the annual estimates.

<sup>15</sup> The better performance of  $\alpha$  over  $\delta$  is not surprising, as only the fraction of days that have information events is used for the estimation of  $\delta$ , while the algorithm uses the whole sample in estimating  $\alpha$ . Indeed, corner solutions to  $\delta$  are mainly found in stocks with low  $\alpha$  estimates.

<sup>16</sup> These estimates also show a peak at the time of the 1987 market crash, and a fall-off in the low volume years following the crash.



**Figure 2. Parameter distributions.** The figure shows the cross-sectional distribution over time of the estimated parameters in the microstructure model given by the likelihood function in equation (4). Panel A gives the annual cross-sectional mean of the trading frequency parameters,  $\epsilon_b$ ,  $\epsilon_s$ , and  $\mu$ . Panel B shows the 5th, 25th, 50th, 75th, and 95th percentiles each year in the sample period for the cross-sectional distribution of  $\alpha$ , the probability that an information event has occurred. (Figure continues on next page.)

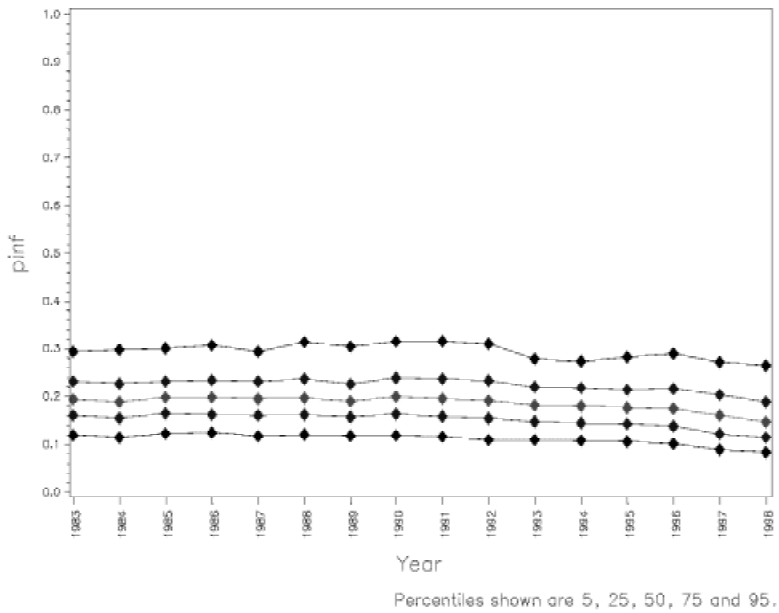


**Figure 2—Continued. Parameter distributions.** Panel C shows the same percentiles for  $\delta$ , the probability of an information day containing bad news.

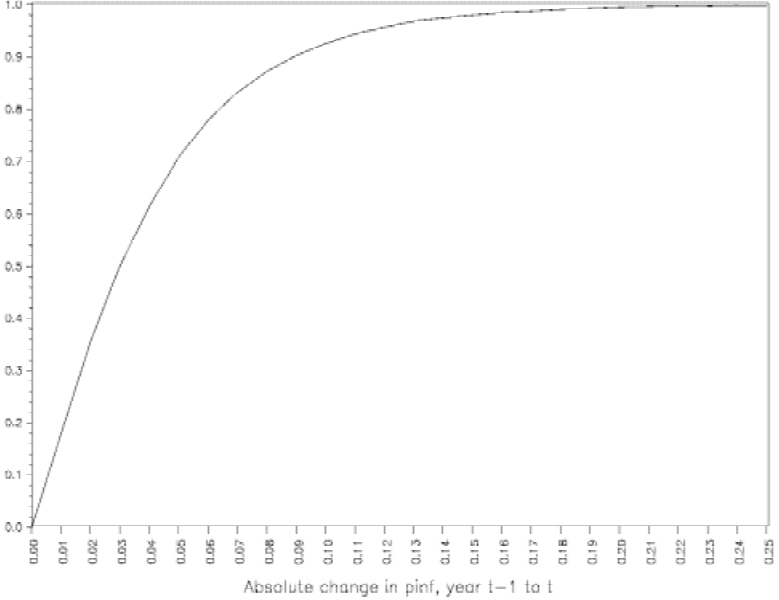
Our particular interest is in the composite variable PIN, the probability of information-based trading. PIN is computed from equation (5) using the yearly estimates of  $\alpha$ ,  $\delta$ ,  $\mu$ ,  $\epsilon_S$ , and  $\epsilon_b$ ; thus we obtain one estimate of PIN for each stock each year. The estimated PIN is very stable across years, both individually and cross-sectionally. Panel A of Figure 3 shows the cross-sectional pattern of PIN. Not only is the median almost constant around 0.19, but the individual percentiles also appear to be stable across years. On an individual stock level, absolute changes between years are relatively small. Panel B of Figure 3 shows the cumulative distribution of year-to-year absolute changes in individual stock PINs. We find that 50 percent of absolute changes are within three percentage points (out of a median of 19 percentage points), while 95 percent are within 11 percentage points. Thus, individual stocks exhibit relatively low variability in the probability of information-based trading across years.

An interesting question is how these PIN estimates relate to the underlying trading volume in the stock. We calculated the cross-sectional correlations between PIN and the logarithm of average daily trading volume for each stock for each year of our sample. The average correlation over the 16 years in our sample is  $-0.58$ , with a range of  $-0.38$  to  $-0.71$ . Hence, we find that across stocks within the same year, PIN is negatively correlated with trading volume. This is consistent with previous empirical work (see Easley et al. (1996b)) showing that actively traded stocks face a lower ad-

PANEL A



PANEL B



**Figure 3. PIN distributions.** The figure shows the cross-sectional distribution of the estimated probability of information based trading, PIN, given by equation (5). Panel A shows the 5th, 25th, 50th, 75th, and 95th percentiles each year in the sample period for the cross-sectional distribution of PIN. Panel B shows the cumulative distribution of absolute price changes from year  $t - 1$  to year  $t$  of individual stock PIN estimates.



verse selection problem due to informed trading. Note, then, that across stocks within each year, PIN is negatively correlated with trading volume, while across time, PIN estimates remain constant, even though trading volume increases. These are exactly the patterns we would expect if PIN captures the underlying information structure.

Given that the parameter estimates are stable across years, we pool the years to further illustrate the distribution of the parameters across stocks. Figure 4 shows these pooled distributions for our estimated parameters, and Table II presents summary statistics. It is clear from the figure that the composite parameter PIN is rather tightly distributed around the mode 0.18, while  $\alpha$  and, in particular,  $\delta$ , are more dispersed over the parameter space. The skewness of  $\delta$  is consistent with the generally rising stock prices over this period; stocks typically did well, so the probability of bad news was generally lower than that of good news. We have aggregated the uninformed trading variables to depict the balance between uninformed buying and selling. Over our time interval, uninformed traders were marginally more likely to sell, while informed traders were more likely to buy. This, too, is consistent with the economic conditions of our sample, as informed traders were better able to capture the benefits of good news and thus rising stock prices.

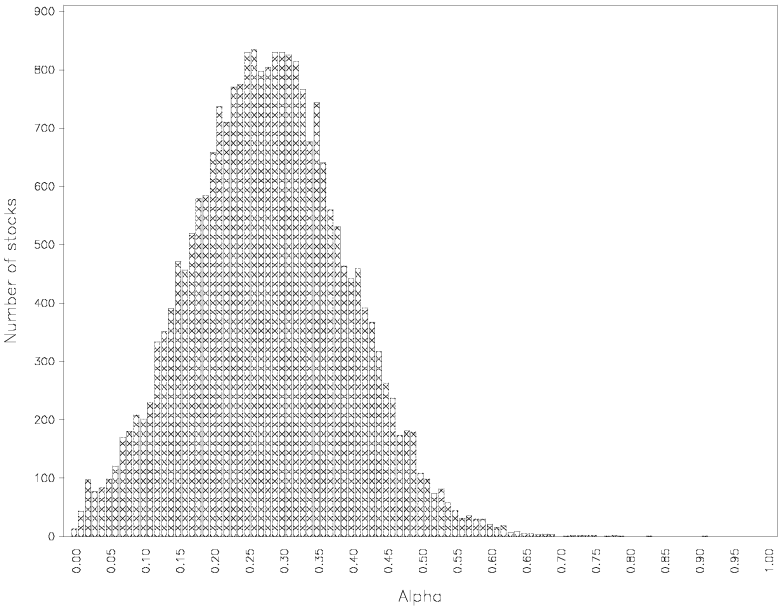
In summary, we have been able to estimate our structural model for a cross section of stocks. The individual parameter estimates appear economically reasonable, and the small standard errors of our estimates indicate strong statistical significance. The time series of our estimates indicate a remarkable stability, with very little year-to-year movement in our estimated parameters. Our contention is that the estimated variables measure the components of information-based trading, and their combination into our PIN variable provides a concrete measure of this risk for each stock.

#### **IV. Asset-Pricing Tests**

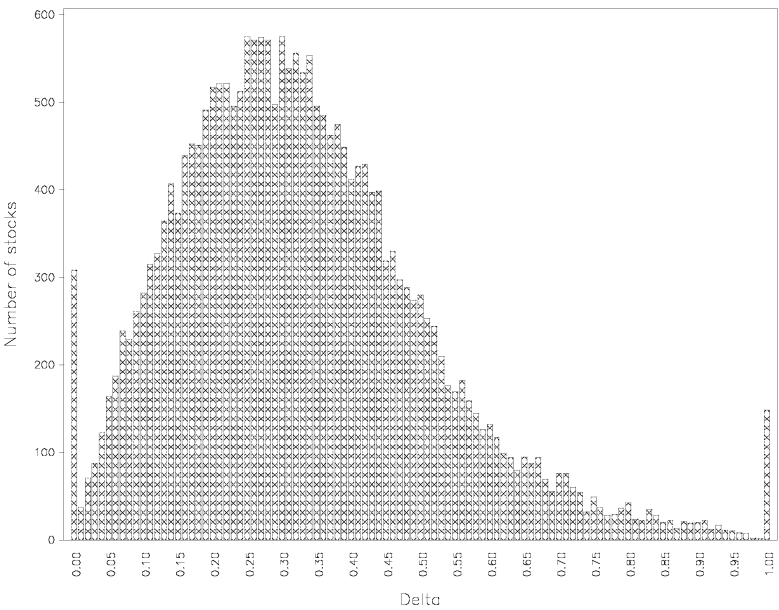
##### *A. Data and Methodology*

For the asset-pricing tests, we need to use additional data on firm characteristics and returns. These data are available from the monthly CRSP files and the annual COMPUSTAT files. Data are not available for all of our listed firms, so the sample used in our asset-pricing tests is drawn from the intersection of the NYSE-listed firms on the CRSP and COMPUSTAT files. The monthly samples contain between 997 and 1,316 stocks for the period 1984 to 1998, yielding 180 monthly observations to aggregate over time. One concern we note at the outset is the length of our sample period. Asset-pricing tests typically employ long sample periods, but transaction data, which we need to calculate our PIN variable, are not available prior to 1983, and since we employ lagged PIN estimates, the asset-pricing tests begin in 1984. Longer sample periods enhance the ability to find statistically significant factors influencing returns, so our limited sample period imposes a particularly stringent constraint on our testing approach.

PANEL A

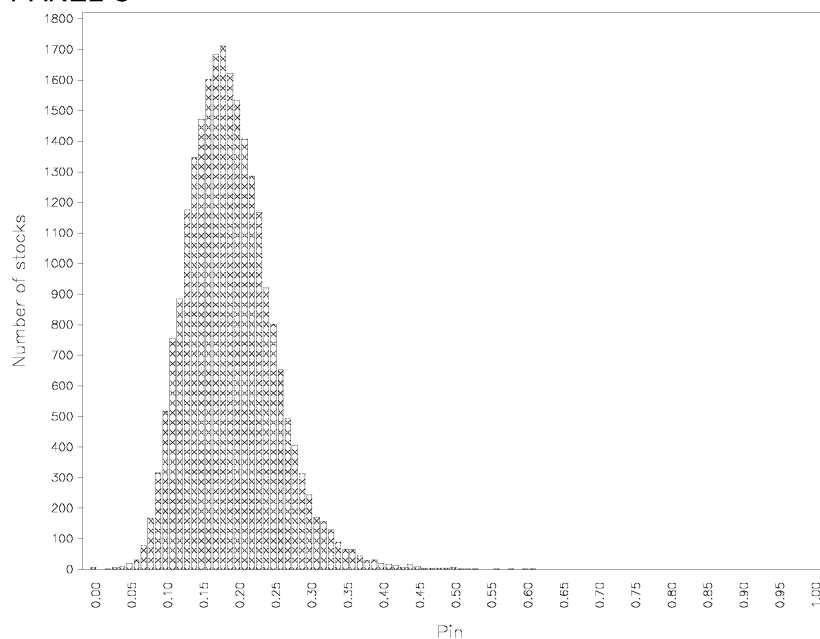


PANEL B

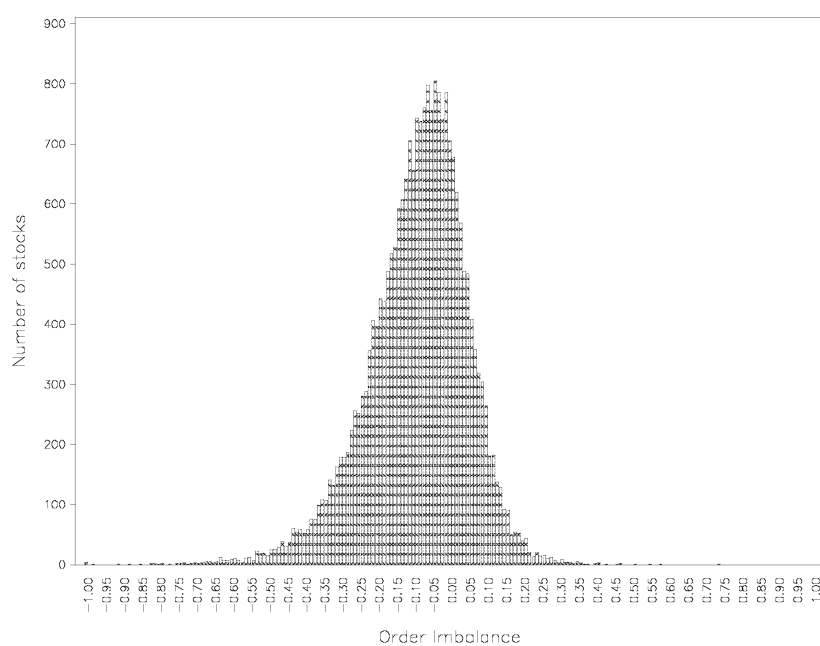


**Figure 4. Parameter distributions with pooled data.** The figure shows the empirical distribution of the microstructure model parameters with all stocks and all years pooled. Panel A gives the distribution of  $\alpha$ , the probability that an information event has occurred. Panel B shows the distribution for  $\delta$ , the probability of an information day containing bad news. (Figure continues on next page.)

PANEL C



PANEL D



**Figure 4—continued.** Parameter distributions with pooled data. Panel C contains the distribution of PIN. Panel D shows the distribution of the uninformed order flow imbalance,  $(\epsilon_b - \epsilon_s)/(\epsilon_b + \epsilon_s)$ .

**Table II**  
**Parameter Summary Statistics**

The table contains time-series averages across years 1983 to 1998 of cross-sectional means, medians, standard deviations, and the median of parameter standard errors from the likelihood estimation. The following are variable descriptions:  $\alpha$  is the probability of an information event,  $\delta$  is the probability of a low signal,  $\mu$  is the rate of informed trade arrival,  $\epsilon_b$  is the arrival rate of uninformed buy orders, and  $\epsilon_s$  is the arrival rate of uninformed sell orders. PIN is the probability of informed trading as given by equation (5).

Variable	Mean	Median	Std. Dev.	Median Std. Err.
$\alpha$	0.283	0.281	0.111	0.035
$\delta$	0.331	0.309	0.181	0.066
$\mu$	31.075	21.303	32.076	0.996
$\epsilon_b$	22.304	11.437	31.519	0.324
$\epsilon_s$	24.046	13.095	31.427	0.299
PIN	0.191	0.185	0.057	0.019

As a first check on the reasonableness of our hypothesis that PIN is related to returns, we compute excess returns for portfolios of stocks sorted independently according to PIN and size.<sup>17</sup> Returns for each stock are taken from the CRSP monthly return files, using the CRSP delisting return in the month of possible delisting. All returns are in excess of the one-month T-bill rates.  $PIN_{it}$  is the probability of information-based trading in year  $t$  whose estimation is described in the previous section.  $SIZE_{it}$  is the logarithm of market value of equity in firm  $i$  at the end of year  $t$ . Panel A of Table III provides average excess returns for portfolios of stocks sorted each year into three groups of PIN and five groups of size. Information about the effectiveness of our sorts is provided in Panels C and D while the number of stocks in each portfolio is provided in Panel B.

The data reveal an important link between PIN and excess returns. With one exception, excess return increases as we move from low to high PIN within each size group. The exception occurs for the large size-high PIN group, but this group has only 8.8 stocks on average, and so inferences here are likely to be unreliable. It is also interesting to note that the difference between high and low PIN excess returns becomes smaller in absolute and relative value as we move from small to large stocks. This could be occurring because PIN only measures the presence of informed traders and not the impact of their information. It is plausible that private information tends to have a greater impact on price for small stocks than for large stocks. It is this convolution of PIN and price impact that matters for excess returns just as they matter for spreads in equation (6). Other factors, such as beta, are

<sup>17</sup> We know from previous work that size is related to both PIN and returns. We sort on it here, and include it in our regressions, in an attempt to control for the total amount of information available about firms. If within each size group the total amount of information about firms is constant, then an increase in PIN implies more private information and less public information.

**Table III**  
**Portfolio Excess Returns**

The table contains results for portfolios of stocks sorted independently by size and PIN, where size is the market value of equity measured at the end of year  $t - 1$  and PIN is the probability of informed trading as given by equation (5) estimated over year  $t - 1$ . For each year, stocks are sorted into three PIN groups, ranging from low to high, and five size groups, ranging from small to large. The reported results are averages of the relevant variables over the sample period 1984 to 1998. Panel A reports average excess returns, Panel B reports the average number of stocks in each portfolio, Panel C reports the average size in \$ millions of the stocks in each portfolio, and Panel D reports the average PIN of stocks in each portfolio.

Size/PIN	Low	Medium	High	Size/PIN	Low	Medium	High
Panel A: Excess Returns				Panel B: Number of Stocks			
Small	0.148	0.202	0.474	Small	21.9	60.1	142.5
2	0.462	0.556	0.743	2	30.5	79.0	115.7
3	0.647	0.695	0.892	3	54.4	95.5	75.1
4	0.873	0.837	0.928	4	93.4	99.0	32.7
Large	0.953	1.000	0.643	Large	174.4	41.6	8.8
Panel C: Size				Panel D: PIN			
Small	75.3	79.4	71.2	Small	0.134	0.186	0.257
2	257.6	257.3	244.6	2	0.138	0.186	0.247
3	665.2	641.9	602.3	3	0.137	0.183	0.237
4	1,719.1	1,593.6	1,473.9	4	0.138	0.180	0.231
Large	9,845.7	4,439.9	4,814.3	Large	0.127	0.175	0.233

not controlled for in our portfolios, but these results do show that PIN is related to returns. Whether this relationship occurs because of the covariance of PIN and other factors is the question we address next by including PIN in a standard asset-pricing regression.

To allow for comparability with previous work, our methodology follows that of Fama and French (1992). Fama and French explored the determinants of the cross-sectional variation in returns and found that beta, size, and book-to-market (i.e., the ratio of the book value of equity to the market value of equity) all influenced returns. Consequently, we include these variables, as well as our estimated PIN variable, in our analysis of asset-pricing returns. We also explore whether the effect of PIN can be captured by previously suggested proxies for liquidity, namely bid-ask spreads and share turnover, or by return variation, by including these variables in the asset-pricing regressions.

We calculate betas using the following approach. *Preranking portfolio betas* are estimated for individual stocks using monthly returns from at least two years to, when possible, five years, before the test year. Thus, for each stock, we use at least 24 monthly return observations in the estimation. We regress these stock returns on the contemporaneous and lagged value-weighted CRSP NYSE/AMEX index. Preranking portfolio betas are then given

as the sum of the two coefficients (this approach, suggested by Dimson (1979), is intended to correct for biases arising from nonsynchronous trading). Next, 40 portfolios are sorted every January on the basis of the estimated betas, and monthly portfolio returns are calculated as equal-weighted averages of individual stock returns. *Postranking portfolio betas* are estimated from the full sample period, such that one beta estimate is obtained for each of the 40 portfolios. Portfolio returns are regressed on contemporaneous and lagged values of CRSP index returns. The portfolio beta,  $\hat{\beta}_p$ , is then the sum of the two coefficients. We use individual stocks in the cross-sectional regressions, so individual stock betas are taken as the beta of the portfolio to which they belong. Because the portfolio compositions change each year, individual stock betas vary over time.

Book value of common equity is obtained from the annual COMPUSTAT files (item 60). Following Fama–French, we exclude firms with negative book values, and we set BE/ME values outside the 0.005 and 0.995 fractiles equal to these fractiles, respectively. We take logs, such that the explanatory variable,  $BM_{it-1}$ , is  $\ln(BE_{t-1}/ME_{t-1})$  for firm  $i$ .

For each month in the sample period 1984 to 1998, we ran the following cross-sectional regression:

$$R_{it} = \gamma_{0t} + \gamma_{1t}\hat{\beta}_p + \gamma_{2t}PIN_{it-1} + \gamma_{3t}SIZE_{it-1} + \gamma_{4t}BM_{it-1} + \eta_{it}, \quad (7)$$

where  $R_{it}$  is the excess return of stock  $i$  in month  $l$  of year  $t$  (monthly subscripts omitted),  $\gamma_{jt}$ ,  $j = 1, \dots, 5$ , are the estimated coefficients, and  $\eta_{it}$  is the mean-zero error term. The coefficients from the cross-sectional regressions are averaged through time, using the standard Fama and MacBeth (1973) methodology. Because this procedure is inefficient under time-varying volatility, we also use the correction technique suggested by Litzenberger and Ramaswamy (1979). This correction weights the coefficients by their precisions when summing across the cross-sectional regressions, and is essentially a weighted least-square methodology. We report both the unadjusted and the Litzenberger–Ramaswamy adjusted coefficients.

A problem with almost all variables provided as alternatives to beta as the explanatory variable of the cross section of returns (e.g., size, book-to-market, earnings-to-price, turnover, etc.) is that these variables depend on the security price. Miller and Scholes (1982) noted that the inverse of price may be a good measure of the conditional beta, and therefore regression analysis may be capturing mismeasurement of beta, rather than some alternative priced factor. Berk (1995) makes a related point. Because the estimation of PIN involves only trades, we avoid this potential critique in our inclusion of PIN.

Our primary interest lies in the time-series average of  $\gamma_{2t}$ , namely the coefficient for PIN. Our hypothesis is that a higher risk of information-based trading for a stock translates into a higher required return for that stock, so we expect a significantly positive average coefficient on PIN.



**Table IV**  
**Summary Statistics**

The table contains means, medians, minimum value, and maximum values on the variables included in the asset-pricing tests. All statistics are calculated from the full sample, that is, pooling all months. *RETURN* is the percentage monthly return in excess of the one-month T-bill rate. *BETAs* are portfolio betas estimated from the full period using 40 portfolios. *PIN* is the probability of informed trading given by equation (5) and estimated yearly for each stock. *PPIN* is the portfolio PIN calculated by first sorting all stocks into 40 portfolios according to PIN each year, and then taking the average within each portfolio of the individual stock PIN in the following year. *SIZE* is the logarithm of year-end market value of equity. *BM* is the logarithm of book value of equity divided by market value of equity. *SPREAD* is the yearly average of the daily opening spreads of stock *i*. *STD* is the daily return standard deviation for stock *i* in year *t*. *TURNOVER* is the logarithm of the average monthly turnover year  $t - 3$  to  $t - 1$ , and *CVTURN* is the logarithm of the coefficient of variation of the monthly turnover year  $t - 3$  to  $t - 1$ .

Variable	Mean	Median	Min.	Max.
<i>RETURN</i>	0.73	0.47	-100.64	339.69
<i>BETA</i>	1.09	1.09	0.52	1.64
<i>PIN</i>	0.19	0.18	0.00	0.53
<i>PPIN</i>	0.19	0.19	0.10	0.31
<i>SIZE</i>	13.29	13.30	6.65	18.62
<i>BM</i>	-0.52	-0.47	-3.35	2.39
<i>SPREAD</i>	1.52	1.14	0.14	15.07
<i>STD</i>	2.10	1.88	0.46	14.92
<i>TURNOVER</i>	1.56	1.59	-2.33	4.37
<i>CVTURN</i>	-0.68	-0.69	-2.07	1.30

## B. Results

Summary statistics on the variables in the asset-pricing regression are provided in Table IV. The procedure on beta sorting portfolios results in a reasonably broad variation in beta, with beta ranging between 0.52 and 1.64. As noted in the previous section, our estimated PIN variable has a mean of 0.19, while ranging from 0 to 0.53. The means of the size and book-to-market variables are also consistent with prior work on this sample period.

We first investigate the interrelationships of the explanatory variables, and in particular how PIN correlates with each variable. Table V presents time-series means of the monthly bivariate correlations of the variables in the asset-pricing tests. One of the largest absolute correlations is between size and PIN, with an average correlation of -0.58. This finding confirms results from earlier research that larger firms tend to have lower probabilities of informed trading. One might expect that stocks with greater private information have higher systematic volatility, and this appears to be the case: PIN is positively correlated with beta, with a correlation of 0.163. We had weaker priors on the relationship between PIN and BM, but note a positive correlation (0.168). The correlation between return and PIN is rather low, but the correlation between return and the other explanatory variables is similarly low.

Table V  
Simple Correlations

The table contains the time-series means of monthly bivariate correlations of the variables in the asset-pricing tests. *RETURN* is the percentage monthly return in excess of the one-month T-bill rate. *BETAs* are portfolio betas estimated from the full period using 40 portfolios. *PIN* is the probability of informed trading given by equation (5) and estimated yearly for each stock. *PPIN* is the portfolio PIN calculated by first sorting all stocks into 40 portfolios according to PIN each year, and then taking the average within each portfolio of the individual stock PIN in the following year. *SIZE* is the logarithm of year-end market value of equity. *BM* is the logarithm of book value of equity divided by market value of equity. *SPREAD* is the yearly average of the daily opening spreads of stock *i*. *STD* is the daily return standard deviation for stock *i* in year *t*. *TURNOVER* is the logarithm of the average monthly turnover year *t* – 3 to *t* – 1, and *CVTURN* is the logarithm of the coefficient of variation of the monthly turnover year *t* – 3 to *t* – 1.

	BETA	PIN	SIZE	BM	SPREAD	STD	TURNOVER	CVTURN	PPIN
RETURN	–0.015	–0.006	0.023	–0.005	–0.022	–0.038	–0.021	–0.019	–0.008
BETA		0.163	–0.207	0.009	0.222	0.434	0.294	0.089	0.171
PIN			–0.576	0.168	0.353	0.239	–0.187	0.412	0.572
SIZE				–0.384	–0.708	–0.493	0.123	–0.547	–0.568
BM					0.273	0.112	–0.030	0.160	0.162
SPREAD						0.748	–0.116	0.397	0.328
STD							0.294	0.331	0.236
TURNOVER								–0.006	–0.169
CVTURN									0.397

Return and beta are negatively correlated, but, as discussed below, this is in line with prior findings in this sample period. Likewise, the positive correlation between return and size is opposite of that reported in earlier periods, but it is consistent with findings from our sample period. Finally, the low correlation between return and *BM* is not unexpected given Loughran’s (1997) finding that book-to-market arises primarily in Nasdaq stocks, and our sample uses only NYSE firms.

The results from the asset-pricing tests are provided in Table VI. The results give striking evidence that the risk of informed trading as captured by PIN is priced in the required returns of stocks. Looking at the weighted least-squares results, we find a positive and significant coefficient on PIN (*t*-value 4.362). The point estimate of the PIN coefficient has the natural interpretation that a difference of 10 percentage points in PIN between two stocks translates into a difference in required return of 0.21 percent per month. This is an economically meaningful and significant difference. We also find a significant and positive coefficient on size (*t*-value 9.994), and a significant, but negative, coefficient on beta (*t*-value –6.22). This latter finding, while inconsistent with standard asset-pricing theory, is consistent with the findings of Fama and French (1992), Chalmers and Kadlec (1998), and Datar et al. (1998), who investigate similar sample periods. Book-to-market is not significant, a finding not unexpected given our earlier discussion.

**Table VI**  
**Asset-Pricing Tests**

The table contains time-series averages of the coefficients in cross-sectional asset-pricing tests using standard Fama and MacBeth (1973) methodology and Litzenberger and Ramaswamy (L-R; 1979) precision-weighted means (weighted least-squares (WLS)). The dependent variable is the percentage monthly return in excess of the one-month T-bill rate. *BETAs* are portfolio betas calculated from the full period using 40 portfolios. *PIN* is the probability of information-based trading in stock *i* of year  $t - 1$ . *SIZE* is the logarithm of market value of equity (*ME*) in firm *i* at the end of year  $t - 1$ , and *BM* is the logarithm of the ratio of book value of common equity to market value of equity for firm *i* in year  $t - 1$ . *T*-values are given in parentheses.

	<i>Beta</i>	<i>PIN</i>	<i>SIZE</i>	<i>BM</i>
Fama-MacBeth	-0.175 (-0.481)	1.800 (2.496)	0.161 (2.808)	0.051 (0.480)
L-R WLS	-0.482 (-6.22)	2.086 (4.362)	0.168 (9.994)	0.042 (1.120)

That PIN affects asset returns is consistent with the economic analysis motivating our work. We believe that the PIN variable captures aspects of the dynamic efficiency of stock prices. These dynamic effects arise because information-based trading affects not only the spread, but the evolution of prices as well. Our results are consistent with this dynamic efficiency influencing the required returns for stocks.

### *C. Potential Errors in Variables for PIN*

PIN, just like beta, is an estimated variable, and so it is potentially subject to an errors-in-variables (EIV) bias. This difficulty has long been recognized in beta estimation, and the portfolio approach of Fama and French is designed to correct for this problem. In effect, the solution is to create an instrumental variable to use in place of the variable in question. Assuming that the EIV problem has been corrected for beta, the only variable in equation (7) measured with potential errors is PIN. In this case, the bias on the coefficient on PIN is unambiguously downward; hence the bias is against finding any effect of PIN. However, we have done a portfolio construction for PIN similar to that of beta. Thus, a simple test for the importance of this problem is to see if our estimated PIN coefficients change when we use this portfolio approach.

Forty portfolios are sorted each year,  $t - 2$ , on the basis of the estimated PINs. The portfolio PIN is then the average of the estimated PINs for the year  $t - 1$ , and this variable, PPIN, is then assigned to each stock in the portfolio. This technique is meant to ensure that the error in PPIN is uncorrelated with the error term. The instrument should also be highly correlated with the original variable. Table V shows that the average cross-sectional correlation between PIN and PPIN is 0.572. We then run the same cross-sectional regression, now using these portfolio PINs. The results are pro-

**Table VII**  
**Errors in Variables**

The table contains time-series averages of the coefficients in cross-sectional asset-pricing tests using standard Fama and MacBeth (1973) methodology and Litzenberger and Ramaswamy (L-R; 1979) precision-weighted means. The dependent variable is the percentage monthly return in excess of the one-month T-bill rate. Betas are portfolio betas calculated from the full period using 40 portfolios. *PIN* is the probability of information-based trading in stock *i* of year  $t - 1$ . *PPIN* is the portfolio pin calculated by first sorting all stocks into 40 portfolios according to *PIN* each year, and then taking the average within each portfolio of the individual stock *PIN* in the following year. *SIZE* is the logarithm of market value of equity (*ME*) in firm *i* at the end of year  $t - 1$ , and *BM* is the logarithm of the ratio of book value of common equity to market value of equity for firm *i* in year  $t - 1$ . *T*-values are given in parentheses.

	<i>Beta</i>	<i>PIN</i>	<i>PPIN</i>	<i>SIZE</i>	<i>BM</i>
Fama-MacBeth	0.022 (0.058)		1.882 (1.435)	0.143 (2.399)	0.010 (0.096)
L-R WLS	-0.290 (-3.52)		2.400 (2.717)	0.152 (8.629)	-0.001 (-0.016)
Fama-MacBeth	0.013 (0.035)	1.758 (2.362)	0.813 (0.647)	0.167 (2.703)	0.013 (0.121)
L-R WLS	-0.298 (-3.60)	1.940 (3.561)	1.174 (1.241)	0.179 (9.394)	0.001 (0.019)

vided in the upper half of Table VII. The coefficients on *PPIN* are very close to the coefficients on *PIN* in Table VI, but the *t*-values are substantially reduced. *PPIN* remains significant only in the weighted least-squares regression where the coefficient on *PPIN* is 2.40 with a *t*-value of 2.717.

The insignificant change in the coefficient on *PIN* and the reduced *t*-values suggest that, though it may be correcting for an EIV problem, the portfolio approach primarily adds noise to the *PIN* variable. In a single regression framework, a test for EIV is to include both the instrument and the original variable in the regression. If there is an EIV problem, and the instrument fixes it, then the coefficient on the original variable should be zero, while if there is not an EIV problem, then the coefficient on the instrument would be zero. Therefore, we included both *PPIN* and *PIN* in the regressions, and the results are reported in the bottom half of Table VII. In this case, the coefficient on *PIN* is not significantly different from the result of the regression excluding *PPIN*. The coefficient on *PPIN*, however, is not significantly different from zero. This strongly suggests that the portfolio approach applied to *PIN* does not correct for any EIV problem. Thus we focus our discussion on regressions using *PIN* rather than the instrument *PPIN*.

#### *D. Alternative Explanations*

It is natural to ask whether *PIN* works in our asset-pricing regressions because it is a fundamental priced variable, or because it is serving as a proxy for some omitted variable. There are three obvious candidates, and

innumerable less obvious candidates, for the omitted variable designation. The most obvious candidate is spreads. Earlier researchers (e.g., Amihud and Mendelson (1986)) found a positive relationship between spread and returns, so it could be that PIN is serving as a proxy for spread. Second, a stock with a high PIN is one with substantial imbalances in trades, and thus is a stock whose price is likely to be highly variable.<sup>18</sup> So it could be that PIN is serving as a proxy for the variability of returns on the stock. Of course, to the extent that this risk can be diversified away, it should not be priced, and any nondiversifiable component of the risk should be picked up by  $\beta$ . But we know that CAPM does not work well over this time period, so this risk could be positively related to observed average excess returns.<sup>19</sup> Finally, there has been substantial interest in the role of volume, or turnover, in explaining asset-price behavior. Therefore, we test whether PIN is merely serving as a proxy for these measures. We consider each of these variables in turn, and while we show that they are correlated with PIN as expected, we also show that when they are included in the returns regression they do not eliminate the direct effect of PIN. Results of the regressions are reported in Table VIII.

We first consider the bid-ask spread. We define the variable  $SPREAD_{it-1}$  to be the average of the daily opening percentage spreads for firm  $i$  in year  $t - 1$ . We know from Table V that  $PIN$  and  $SPREAD$  are positively, but not perfectly, correlated. Indeed, the correlation is a relatively low 0.353, reflecting that spreads can be affected by many factors other than information. Is  $PIN$  or  $SPREAD$  the better predictor of returns? We test this by first including  $SPREAD$  in place of  $PIN$  in the asset-pricing regression and then by including both  $SPREAD$  and  $PIN$  in the regression. When  $SPREAD$  is included, and  $PIN$  is excluded,  $SPREAD$  is insignificant in the Fama-MacBeth regression, while it is significantly negative in the weighted least-squares regression. This result is not what would be expected from liquidity-based theories of asset returns. When  $SPREAD$  and  $PIN$  are both included,  $PIN$  remains highly significant ( $t$ -value = 3.88) with the correct positive sign. The inclusion of spread reduces the coefficient on PIN only slightly from 2.09 to 1.87.

That it is  $PIN$ , and not  $SPREAD$ , that affects asset-pricing returns is consistent with the economic analysis motivating our work. While traders undoubtedly care about spreads, they are more concerned with the risk of holding the stock, and this is affected by the extent of private information. Information-based trading does give rise to spreads, but spreads in actual markets can be

<sup>18</sup> This relationship is not completely straightforward because although trade moves prices, public information events, which in our model do not generate trade, also move prices. So there could be stocks with low PIN, that is little information-based trade, and highly variable prices caused by the release of public information. This is consistent with our analysis in that public events can move the range of true values of the stock.

<sup>19</sup> Note that our theoretical model (equation (6)) includes essentially a standard deviation term as it measures the losses due to the risk of informed trading. Thus, it seems reasonable that PIN and standard deviation could both play a role due to risk. We thank the referee for raising this point.

Table VIII  
Alternative Explanations

The table contains time-series averages of the coefficients in cross-sectional asset-pricing tests using standard Fama and MacBeth (1973) methodology and Litzenberger and Ramaswamy (L-R; 1979) precision-weighted means. The dependent variable is the percentage monthly return in excess of the one-month T-bill rate. Betas are portfolio betas calculated from the full period using 40 portfolios. *PIN* is the probability of information-based trading in stock *i* of year *t* - 1. *SIZE* is the logarithm of market value of equity (ME) in firm *i* at the end of year *t* - 1, and *BM* is the logarithm of the ratio of book value of common equity to market value of equity for firm *i* in year *t* - 1. *SPREAD* is the average daily opening percentage spread for firm *i* in year *t* - 1. *STD* is the standard deviation of daily returns for firm *i* in year *t* - 1. *TURNOVER* is the logarithm of the average monthly turnover year *t* - 3 to *t* - 1, and *CVTURN* is the logarithm of the coefficient of variation of the monthly turnover year *t* - 3 to *t* - 1. *T*-values are given in parentheses.

	<i>Beta</i>	<i>PIN</i>	<i>SIZE</i>	<i>BM</i>	<i>SPREAD</i>	<i>STD</i>	<i>TURNOVER</i>	<i>CVTURN</i>
Fama-MacBeth	-0.146 (-0.407)		0.106 (2.122)	0.037 (0.353)	-0.060 (-0.745)			
L-R WLS	-0.425 (-5.47)		0.081 (4.333)	0.035 (0.930)	-0.060 (-2.66)			
Fama-MacBeth	-0.164 (-0.458)	1.694 (2.483)	0.144 (2.849)	0.044 (0.419)	-0.052 (-0.650)			
L-R WLS	-0.441 (-5.66)	1.866 (3.882)	0.124 (5.746)	0.041 (1.100)	-0.051 (-2.27)			
Fama-MacBeth	0.159 (0.486)		0.053 (1.138)	-0.009 (-0.088)		-0.318 (-2.73)		
L-R WLS	0.008 (0.102)		0.034 (2.172)	-0.016 (-0.416)		-0.354 (-13.0)		
Fama-MacBeth	0.141 (0.435)	1.411 (2.104)	0.084 (1.748)	-0.002 (-0.024)		-0.311 (-2.67)		
L-R WLS	-0.008 (-0.100)	1.405 (2.932)	0.064 (3.458)	-0.010 (-0.271)		-0.345 (-12.6)		
Fama-MacBeth	0.047 (0.144)		0.119 (2.166)	0.051 (0.486)			-0.305 (-3.27)	-0.210 (-2.44)
L-R WLS	-0.237 (-2.88)		0.134 (7.762)	0.046 (1.222)			-0.336 (-9.95)	-0.178 (-2.73)
Fama-MacBeth	0.025 (0.077)	1.328 (1.718)	0.139 (2.444)	0.054 (0.519)			-0.288 (-2.99)	-0.242 (-2.86)
L-R WLS	-0.260 (-3.14)	1.454 (2.956)	0.156 (8.306)	0.048 (1.276)			-0.315 (-9.17)	-0.214 (-3.23)
Fama-MacBeth	0.276 (0.939)	1.415 (2.005)	0.133 (2.694)	-0.019 (-0.190)	0.203 (2.179)	-0.466 (-3.38)	-0.072 (-0.927)	-0.193 (-2.15)
L-R WLS	0.107 (1.225)	1.350 (2.723)	0.121 (5.274)	-0.018 (-0.485)	0.198 (5.724)	-0.489 (-11.2)	-0.090 (-2.19)	-0.159 (-2.39)

affected by many other factors such as minimum tick sizes, specialist continuity rules, and even market power. These factors dictate that spreads will be, at best, a noisy proxy for the risk of informed trading. *PIN* is also a proxy for the risk of informed trading, and as it is estimated from trade data, it too could be influenced by the factors listed above. But our results



show that it is priced. We believe our results provide strong evidence that information plays a deeper role, one beyond that captured, however imperfectly, in spreads.

We next consider whether *PIN* or the variability of returns on a stock is the better predictor of excess returns. We define  $STD_{it-1}$  to be the standard deviation of daily returns on stock  $i$  in year  $t - 1$ . As expected, Table V shows that *STD* and *PIN* are positively correlated, with a correlation coefficient of 0.24. When *STD* is included in the pricing regression in place of *PIN*, it is highly significant with a negative coefficient. This also changes the coefficient on  $\beta$  from negative and significant to positive and insignificant. This occurs in part because the  $\beta$  used in our regressions is the portfolio  $\beta$  and not the individual stock's  $\beta$ . Of more importance for us is that when both *STD* and *PIN* are included, the coefficient on *PIN* remains positive and significant ( $t$ -statistic = 2.93). The effect of *STD* indicates the weakness of CAPM, or at least our standard implementation of it, over this period. But it has little effect on the pricing of the probability of information-based trade.

Next, we consider whether *PIN* can be interpreted as a proxy for volume effects. Volume can be measured in many ways, but previous research on asset-pricing effects has typically used turnover, or daily volume divided by shares outstanding. This measure avoids any of the price-beta concerns noted earlier, and also allows for greater comparability across stocks. Datar et al. (1998) present evidence that there is a negative relationship between turnover and returns. Further, Chordia, Subrahmanyam, and Anshuman (2001) argue that both turnover and the volatility of turnover should affect price behavior, and they find that both variables negatively affect returns. We calculate share turnover in each stock for each month in year  $t - 3$  to  $t - 1$ , as the number of shares traded divided by the number of shares outstanding. The natural logarithm of the average turnover is then used in the asset-pricing regressions for year  $t$ . As a proxy for the variability of turnover, we follow Chordia et al. and employ the natural logarithm of the coefficient of variation (*CV*) of the monthly turnover in years  $t - 3$  to  $t - 1$ . Including *Turnover* and *CV turnover* for each stock in the estimating equation (without *PIN*) reveals a strong negative effect on returns of both variables, similar to the findings in Chordia et al. When *PIN* is included along with these variables, the coefficient on *PIN* remains positive and significant ( $t$ -statistic = 2.956).<sup>20</sup> Thus, it appears that the influence of *PIN* on returns is not proxying for the effects of volume.

Finally, to be sure that *PIN* is not serving as a proxy for the joint effects of these alternative variables, we also ran the regression including all of the alternative variables, spread, standard deviation of daily returns, turnover, and volatility of turnover. The coefficient on *PIN* in the weighted least-

<sup>20</sup> Including only turnover we find similar effects. *Turnover* alone enters negatively and significantly, while *Turnover* and *PIN* together are both significant ( $t$ -statistics of -9.88 and 2.42, respectively) and retain their predicted signs.

squares (WLS) regression is 1.35 with a  $t$ -statistic of 2.723. Thus *PIN* is not proxying for the joint effect of these alternative variables. It is also worth noting from the regression that spread now is positive and significant. The inclusion of the standard deviation of returns is important for this result, suggesting that the underlying asset pricing model may be mis-specified.

In summary, the positive relationship between expected return and the probability of informed trading seems to be robust to the inclusion of different explanatory variables in the cross-sectional regressions. Thus, there is evidence that the risk of informed trading is, indeed, an important determinant of required stock returns.

## V. Conclusions

We have investigated the role of information-based trading in affecting asset returns. Our premise is that in a dynamic market, asset prices are continually adjusting to new information. This evolution dictates that the process by which asset prices become informationally efficient cannot be separated from the process generating asset returns. Our theoretical model suggests that private information influences this price evolution, and in so doing, affects the risk of holding the asset. We investigate this link between asset prices and private information by using the structure of a sequential trade market microstructure model to derive an explicit measure of the probability of information-based trading for an individual stock. We then estimate this probability for a large sample of NYSE-listed stocks. Incorporating our probability estimates into a standard asset-pricing framework revealed strong support for our hypothesis: Information-based trading has a large and significantly positive effect on asset returns. Indeed, our estimated information variable and firm size are the predominant factors explaining returns.

That the risk of information-based trading affects asset returns raises a host of important questions regarding asset pricing in general, and asset-pricing models in particular. Brevity precludes addressing all of these, but we do feel it useful to consider three general issues. These involve the theoretical basis for our result, the empirical properties of *PIN*, and the implications of our results for future research.

Of particular importance is why this can occur in a seemingly efficient capital market. A natural objection to all candidates put forward to explain asset returns is that, with the exception of systematic risk, the actions of arbitrageurs should remove any such proposed influence on the market. While this may be accurate for some factors, we do not believe that it is accurate with respect to asymmetric information. In a world with asymmetric information, an uninformed investor is always at a disadvantage relative to traders with better information. In bad times, this disadvantage can result in the uninformed trader's portfolio holding too much of the stock; in good times, the trader's portfolio has too little of the stock. Holding many stocks cannot

remove this effect because the uninformed do not know the proper weights of each asset to hold. In this sense, asymmetric information risk is systematic because, like market risk, it cannot be diversified away.<sup>21</sup>

In our empirical work we found that PIN, the probability of information-based trading variable, actually dominated all other variables, including  $\beta$ , in explaining returns. Given our argument that information has a systematic component, this should not be unexpected. We caution, however, that our results do not mean that only private information matters in asset pricing. What our results do suggest is that the effects of information may be more pervasive, and important, than our simple theories, and asset-pricing models, have thus far considered.

The success of our PIN variable naturally leads to questions regarding its empirical properties. A very useful exercise would be to examine the cross-sectional determinants of PIN, and in particular how PIN relates to variables such as industry or accounting measures. Not surprisingly, this is a large endeavor and one we hope to address in future work. One benefit of such a project could be to determine a set of "sufficient statistics" for PIN that involves accounting data. As is clear from this paper, the actual calculation of PIN requires a tremendous amount of computation. Replicating PIN with more easily available data would make it easier to apply, and would have the added benefit of explaining why it is that some accounting data appears to be informative for asset pricing.

Finally, our results here suggest a number of directions for future research. Our results suggest that a firm could lower its cost of capital if it could reduce its PIN. Public disclosure of information that would otherwise be private could do this. Botosan (1997) provides support for this idea by showing that for a sample of firms with low analyst following greater disclosure of information reduces the cost of capital by an average of 28 basis points. There is now a substantial body of work suggesting that volume, and volume-linked variables, play an important role in asset pricing. We have shown here that PIN is not a volume effect, but there remains the intriguing question of whether volume effects may not be proxying for some of the underlying components of PIN, such as the rate of uninformed trade or the probability of new information. Investigating the role of the components of PIN would provide insight into this issue. An equally intriguing issue is momentum. There is now widespread, if in some cases grudging, acceptance of the fact that momentum affects asset prices. These momentum effects appear to arise over relatively short time intervals (months 3–12), and they pose a challenge for virtually all asset-pricing theories. One possible explanation is that momentum is somehow linked to the underlying information

<sup>21</sup> It is also not the case in our model that the informed traders will simply trade the effect away because they too face risk in holding the asset. Informed traders are also risk averse, and so there will always be a premium to hold the risky asset. However, because the stock is relatively less risky for the informed, in equilibrium, their expected holdings of the asset exceed that of the uninformed (see Easley and O'Hara (2000) for more analysis and discussion).

structure of the stock. Testing for such effects using our approach would require finer estimates (i.e., monthly) of our PIN variable, as well as potentially a longer time frame. We hope to consider this in future research.

## REFERENCES

- Admati, Anat, 1985, A noisy rational expectations equilibrium for multi-asset securities markets, *Econometrica* 53, 629–658.
- Amihud, Yakov, 2000, Illiquidity and stock returns: Cross-sectional and time series effects, Working paper, New York University.
- Amihud, Yakov, and Haim Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics* 17, 223–249.
- Amihud, Yakov, and Haim Mendelson, 1989, The effects of beta, bid-ask spread, residual risk, and size on stock returns, *Journal of Finance* 44, 479–486.
- Amihud, Yakov, Haim Mendelson, and B. Lauterbach, 1997, Market microstructure and securities values: Evidence from the Tel Aviv Stock Exchange, *Journal of Financial Economics* 45, 365–390.
- Barry, Christopher B., and Stephen J. Brown, 1984, Differential information and the small firm effect, *Journal of Financial Economics* 13, 283–294.
- Barry, Christopher B., and Stephen J. Brown, 1985, Differential information and security market equilibrium, *Journal of Financial Quantitative Analysis* 20, 407–422.
- Barry, Christopher B., and Robert H. Jennings, 1992, Information and diversity of analyst opinion, *Journal of Financial and Quantitative Analysis* 27, 169–183.
- Basak, Suleyman, and Domenico Cuoco, 1998, An equilibrium model with restricted stock market participation, *Review of Financial Studies* 11, 309–341.
- Berk, Jonathan B., 1995, A critique of size-related anomalies, *Review of Financial Studies* 8, 275–286.
- Botosan, Caval A., 1997, Disclosure level and the cost of equity capital, *The Accounting Review* 72, 323–349.
- Brennan, Michael J., and Henry H. Cao, 1997, International portfolio investment flows, *Journal of Finance* 52, 1851–1880.
- Brennan, Michael J., Tarun Chordia, and Avanidhar Subrahmanyam, 1998, Alternative factor specifications, security characteristics, and the cross-section of expected stock returns, *Journal of Financial Economics* 49, 345–373.
- Brennan, Michael J., and Avanidhar Subrahmanyam, 1996, Market microstructure and asset pricing: On the compensation for illiquidity in stock returns, *Journal of Financial Economics* 41, 441–464.
- Chalmers, J.M., and Gregory B. Kadlec, 1998, An empirical examination of the amortized spread, *Journal of Financial Economics* 48, 159–188.
- Chen, Nai-fu F., and Raymond Kan, 1996, Expected return and the bid-ask spread, in K.S.S. Saitou and K. Kubota, eds.: *Modern Portfolio Theory and Applications* (Gakujutsu Shuppan Center, Osaka, Japan).
- Chordia, Tarun, Avanidhar Subrahmanyam, and V.R. Anshuman, 2001, Trading activity and expected stock returns, *Journal of Financial Economics* 59, 3–32.
- Coles, Jeffrey L., Uri Loewenstein, and J. Suay, 1995, On equilibrium pricing under parameter uncertainty, *Journal of Financial and Quantitative Analysis* 30, 347–364.
- Datar, V., Narayan Naik, and R. Radcliffe, 1998, Liquidity and stock returns: An alternative test, *Journal of Financial Markets* 1, 203–219.
- Dimson, Elroy, 1979, Risk measurement when shares are subject to infrequent trading, *Journal of Financial Economics* 7, 197–226.
- Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, 1996a, Cream-skimming or profit-sharing? The curious role of purchased order flow, *Journal of Finance* 51, 811–833.
- Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, 1997a, The information content of the trading process, *Journal of Empirical Finance* 4, 159–186.

- Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, 1997b, One day in the life of a very common stock, *Review of Financial Studies* 10, 805–835.
- Easley, David, Nicholas M. Kiefer, Maureen O'Hara, and Joseph Paperman, 1996b, Liquidity, information, and less-frequently traded stocks, *Journal of Finance* 51, 1405–1436.
- Easley, David, and Maureen O'Hara, 2000, Information and the cost of capital, Working paper, Johnson Graduate School of Management, Cornell University.
- Easley, David, Maureen O'Hara, and Joseph Paperman, 1998, Financial analysts and information-based trade, *Journal of Financial Markets* 1, 175–201.
- Easley, David, Maureen O'Hara, and Gideon Saar, 2001, How stock splits affect trading: A microstructure approach, *Journal of Financial and Quantitative Analysis* 36, 25–51.
- Eleswarapu, V.R., 1997, Cost of transacting and expected returns in the Nasdaq market, *Journal of Finance* 52, 2113–2127.
- Eleswarapu, V.R., and Marc R. Reinganum, 1993, The seasonal behavior of liquidity premium in asset pricing, *Journal of Financial Economics* 34, 373–386.
- Ellis, Katrina, Roni Michaely, and Maureen O'Hara, 2000, The accuracy of trade classification rules: Evidence from Nasdaq, *Journal of Financial and Quantitative Analysis* 35, 529–551.
- Fama, Eugene F., and Kenneth R. French, 1992, The cross section of expected stock returns, *Journal of Finance* 47, 427–465.
- Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.
- Grossman, Sanford, and Joseph Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393–408.
- Jones, Charles M., Gautam Kaul, Marc L. Lipson, 1994, Transactions, volume, and volatility, *Review of Financial Studies* 7, 631–651.
- Jones, Charles M., and Steven L. Slezak, 1999, The theoretical implications of asymmetric information on the dynamic and cross-sectional characteristics of asset returns, Working paper, University of North Carolina, Chapel Hill.
- Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1335.
- Lee, Charles M.C., and Mark J. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733–746.
- Lintner, John, 1969, The aggregation of investor's diverse judgments and preferences in purely competitive security markets, *Journal of Financial and Quantitative Analysis* 4, 347–400.
- Litzenberger, Robert, and Krishna Ramaswamy, 1979, The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence, *Journal of Financial Economics* 7, 163–196.
- Loughran, Timothy, 1997, Book-to-market across firm size, exchange, and seasonality: Is there an effect? *Journal of Financial and Quantitative Analysis* 32, 249–268.
- Merton, Robert C., 1986, A simple model of capital market equilibrium with incomplete information, *Journal of Finance* 42, 483–510.
- Miller, Merton H., and Myron S. Scholes, 1982, Dividends and taxes: Some empirical evidence, *Journal of Political Economy* 90, 1118–1141.
- O'Hara, Maureen, 1995, *Market Microstructure Theory* (Blackwell Publishers, Cambridge, MA).
- Shapiro, Alexander, 2001, The investor recognition hypothesis in a dynamic general equilibrium: Theory and evidence, *Review of Financial Studies* 15, 97–145.
- Wang, Jiang, 1993, A model of intertemporal asset prices under asymmetric information, *Review of Economic Studies* 60, 249–282.