JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 45, No. 2, Apr. 2010, pp. 293–309
COPYRIGHT 2010, MICHAEL G. FOSTER SCHOOL OF BUSINESS, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195
doi:10.1017/S0022109010000074

Factoring Information into Returns

David Easley, Soeren Hvidkjaer, and Maureen O'Hara*

Abstract

We examine the potential profits of trading on a measure of private information (PIN) in a stock. A zero-investment portfolio that is size-neutral but long in high PIN stocks and short in low PIN stocks earns a significant abnormal return. The Fama-French, momentum, and liquidity factors do not explain this return. However, significant covariation in returns exists among high PIN stocks and among low PIN stocks, suggesting that PIN might proxy for an underlying factor. We create a PIN factor as the monthly return on the zero-investment portfolio above and show that it is successful in explaining returns to independent PIN-size portfolios. We also show that it is robust to inclusion of the Pástor-Stambaugh liquidity factor and the Amihud illiquidity factor. We argue that information remains an important determinant of asset returns even in the presence of these additional factors.

I. Introduction

Despite what would seem to be a natural affinity between microstructure and asset pricing, the research in each area has remained largely disparate. Asset pricing research focuses on the role of market and other aggregate risks; microstructure has largely examined asset-specific or idiosyncratic adjustment of prices. This specialization has resulted in lacunae in each area: Asset pricing models have difficulty in explaining asset returns, while microstructure models have been viewed as irrelevant for all but the very short-run behavior of asset prices.

Recent research has attempted to bridge these fields by demonstrating that microstructure-related concepts such as liquidity and asymmetric information may play a role in explaining the cross section of asset returns. The liquidity-based research (see, e.g., Chordia, Roll, and Subrahmanyam (2000), Amihud (2002),

*Easley, dae3@cornell.edu, Department of Economics, Cornell University, 450 Uris Hall, Ithaca, NY 14853; Hvidkjaer, sh.fi@cbs.dk, Copenhagen Business School, Solbjerg Plads 3, 2000 Frederiksberg, Denmark; O'Hara, mo19@cornell.edu, Johnson Graduate School of Management, Cornell University, 447 Sage Hall, Ithaca, NY 14853. The authors thank Joel Hasbrouck, Ingrid Werner, participants at the Morgan Stanley Equity Market Microstructure Research Conference and European Finance Association Meetings (Maastricht), Yakov Amihud (the referee), and Stephen Brown (the editor) for helpful comments. We are grateful to Anchada Charoenrook and Jennifer Conrad for providing us with the Amihud factor. We thank Morgan Stanley for research support. The authors are solely responsible for the contents of this paper.

Pástor and Stambaugh (2003), or Acharya and Pedersen (2005)) has typically found that liquidity, measured in a variety of ways, plays an economically significant role in influencing asset returns. What is less clear from this liquidity research is exactly what causes this relationship to arise, but potential explanations include inventory constraints on market makers or more general limitations on dealers' risk-bearing capacity. Whatever the cause, research using the Pástor and Stambaugh factor (see Ang, Hodrick, Xing, and Zhang (2006)) or the Amihud factor (see Charoenrook and Conrad (2004)) supports the notion that liquidity may act as a factor in asset pricing.

In a related microstructure direction, researchers have shown that asymmetric information may play a role in the cross section of asset returns. Easley and O'Hara (2004) develop a theoretical model to show why equilibrium differences in asset returns will arise due to private information. The argument here is that asymmetric information creates a risk for uninformed traders because they are unable to optimally structure portfolios. While both informed and uninformed traders will optimally diversify, the uninformed do not know the correct (full information) weights to hold of each asset and so in equilibrium end up holding "too much" of the bad assets and "too little" of the good assets, setting the stage for information risk to affect asset returns. Gârleanu and Pedersen (2004) provide an alternative but related private information explanation for asset return effects.

Empirically, Easley, Hvidkjaer, and O'Hara (2002) show that the effect predicted by the Easley and O'Hara (2004) model is found in asset returns. Using a measure of private information (PIN) derived from a microstructure model, these authors find striking evidence that information risk is priced. Specifically, results from cross-sectional asset pricing regressions indicate that a 10% increase in PIN (approximately a two-standard-deviation move) gives rise to an increase in annual expected returns of 2.5%. This cross-sectional regression approach controls for the effect of other characteristics on returns, but it does not directly control for the sensitivity of returns to standard risk factors. While there is an unresolved debate as to whether characteristics or factor sensitivities are more important in explaining the cross section of returns (see, e.g., Daniel and Titman (1997), Davis, Fama, and French (2000), and Gebhardt, Hvidkjaer, and Swaminathan (2005)), the factor sensitivities are more appealing as controls for undiversifiable risk. Moreover, developing a factor approach to testing for information risk permits the comparison and inclusion of both an information factor and a liquidity factor into asset pricing investigations.

In this research, we develop a PIN factor and investigate its performance for explaining cross-sectional asset returns. Using the time-series regression approach of Fama and French (1993), we find intriguing evidence of success in explaining returns to independent PIN-size portfolios. In particular, inclusion of a PIN factor substantially reduces the intercepts relative to those of the Fama-French factors, momentum factor, plus liquidity factors model. Indeed, for 7 of 10 size decile portfolios, these intercepts become statistically insignificant, although they remain at least marginally significant for the three smallest portfolios. We interpret these results as supporting the role of PIN or information risk as a statistical factor in affecting asset returns.

These results also demonstrate that liquidity and information risk may both play distinctive and important roles in explaining asset returns. Some research (see Duarte and Young (2009)) argues that liquidity effects unrelated to information risk explain the relation between PIN and the cross section of returns. However, we find that the PIN factor remains significant even when both the Pástor-Stambaugh factor and the Amihud illiquidity factor are included. Therefore, PIN does not appear to be merely proxying for liquidity effects. Such a finding should not, however, be unexpected. Liquidity effects can arise for a variety of reasons that are unrelated to information, and measuring such liquidity effects is at best inexact.[1] Nonetheless, our results here suggest that factoring information and liquidity into returns can improve our understanding of both short- and long-run asset pricing behavior.

This paper is organized as follows. Section II provides a basic derivation of the PIN model. Section III then sets out the estimation technique and the data we use in this paper. Section IV examines the role of PIN in explaining cross-sectional returns by looking at whether portfolios based on sorts over PIN earn differential returns. Section V investigates the role of PIN as a factor. We create a PIN factor, and we investigate its performance relative to the Fama-French, momentum plus liquidity model. The paper's last section is a conclusion.

II. The PIN Model

The first step in our analysis is to estimate the probability of information-based trade (PIN) for each stock in our sample. As information-based trade is not directly observable, we use a structural microstructure model to make inferences about it. 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 the market maker's 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. We model the trading process and the market maker's learning process and use this structural model, along with trade data, to make an inference about the probability of information-based trade.

We follow Easley, Hvidkjaer, and O'Hara (2002) in modeling a market in which a competitive market maker trades a stock with informed and uninformed traders. Trade occurs over $t = 1, \dots, T$ discrete trading days, and within each trading day, trade occurs in continuous time. Information events occur between

[1]The Amihud measure reflects total price impacts and does not sort them into information-based and inventory-based components. So finding that the Amihud illiquidity factor is priced is not informative about information versus inventory price impacts. Whether this factor is actually priced or not, or more precisely whether it is reliably and stably priced, is still in debate. Charoenrook and Conrad (2004) find that ILLIQ, a factor-mimicking portfolio based on the Amihud illiquidity measure, is priced over the 1963–2003 sample period. Using returns on the Amihud factor provided by Charoenrook and Conrad, we find significant average returns in the period 1963–1983, but not in the 1984–2002 sample period of our analysis. Surprisingly, Duarte and Young (2009) find that the liquidity measure ILLIQ is priced in Fama-MacBeth regressions over their similar sample period.

trading days with probability α. When an information event occurs, it is either bad news, with probability δ, or good news, with probability $1-\delta$. Good news at date t is that the asset is worth $\bar{V}_t$, and bad news at date t is that it is worth $\underline{V}_t$.

During any trading day, orders arrive at the market according to Poisson processes. The market maker sets prices to buy or sell at each time during the day and then executes orders as they arrive. Traders informed of bad news sell, and those informed of good news buy. We assume that orders from the informed traders follow a Poisson process with daily arrival rate μ. Uninformed traders trade for liquidity reasons. We assume that buy and sell orders from uninformed traders arrive at the market according to independent Poisson processes with daily arrival rates ε_b for buy orders and ε_s for sell orders. If an order arrives at some time, 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.[2]

Suppose we now view this problem from the perspective of an econometrician. If we, like the market maker, observed 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, Kiefer, and O'Hara (1996), (1997a), (1997b), and Easley, Kiefer, O'Hara, and Paperman (1996) who demonstrate how to use structural models 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.

The first step in the estimation is to note that the total number of buys and sells per day is a sufficient statistic for the trade data. This occurs because of the assumption that arrivals follow independent Poisson processes. The likelihood function induced by this model for the total number of buys and sells on a single trading day is

$$(1) \quad \mathcal{L}((B,S)|\theta) = \alpha(1-\delta)e^{-(\mu+\varepsilon_b+\varepsilon_s)}\frac{(\mu+\varepsilon_b)^B(\varepsilon_s)^S}{B!S!} + \alpha\delta e^{-(\mu+\varepsilon_b+\varepsilon_s)}\frac{(\mu+\varepsilon_s)^S(\varepsilon_b)^B}{B!S!} + (1-\alpha)e^{-(\varepsilon_b+\varepsilon_s)}\frac{(\varepsilon_b)^B(\varepsilon_s)^S}{B!S!},$$

where (B,S) is the total number of buys and sells for the day and $\theta=(\mu,\epsilon_b,\epsilon_s,\alpha,\delta)$ is the parameter vector. This likelihood function is a mixture of three Poisson probabilities, weighted by the probability of having a "good news day" $\alpha(1-\delta)$, a "bad news day" $\alpha\delta$, and a "no news day" $(1-\alpha)$.

The model assumes that on each day, the arrivals of an information event and trades, conditional on information events, are drawn from identical and independent distributions. Thus the likelihood function for T days is a product of the

[2]A more extensive discussion of this structure can be found in Easley, Kiefer, and O'Hara (1996), (1997a), (1997b), and Easley, Kiefer, O'Hara, and Paperman (1996).

above likelihood over days. The log likelihood function, after dropping a constant term and rearranging, can be written as

$$
(2) \quad \mathcal{L}((B_t, S_t)_{t=1}^{T} | \theta) = \sum_{t=1}^{T} \left[-\varepsilon_b - \varepsilon_s + M_t(\ln x_b + \ln x_s) + B_t \ln(\mu + \varepsilon_b) + S_t \ln(\mu + \varepsilon_s) \right]
+ \sum_{t=1}^{T} \ln \left[\alpha(1-\delta) e^{-\mu} x_s^{S_t - M_t} x_b^{-M_t} + \alpha \delta e^{-\mu} x_b^{B_t - M_T} x_s^{-M_t} + (1-\alpha) x_s^{S_t - M_t} x_b^{B_t - M_t} \right],
$$

where $M_t = \min(B_t, S_t) + \max(B_t, S_t)/2$, $x_s = \varepsilon_s/(\mu + \varepsilon_s)$, and $x_b = \varepsilon_b/(\mu + \varepsilon_b)$. The factoring of $x_b^{M_t}$ and $x_s^{M_t}$ is done to increase the computing efficiency and reduce truncation error. This is important for stocks that have a large number of buys and sells, as otherwise we would need to compute small fractions (arrival rates) taken to large powers (numbers of trades).

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. As is standard in microstructure models, a market maker sets trading prices such that his expected losses to 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

$$
(3) \quad \text{PIN} = \frac{\alpha\mu}{\alpha\mu + \varepsilon_b + \varepsilon_s},
$$

where $\alpha\mu + \varepsilon_b + \varepsilon_s$ 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 ($\varepsilon_b = \varepsilon_s = \varepsilon$) and news is equally likely to be good or bad ($\delta = 0.5$), the percentage opening spread is

$$
(4) \quad \frac{\Sigma}{V_t^*} = \text{PIN} \frac{(\bar{V}_t - \underline{V}_t)}{V_t^*},
$$

where Σ is the spread, ask minus bid price, and V_t^* is the unconditional expected value of the asset given by $V_t^* = \delta \underline{V}_t + (1-\delta)\bar{V}_t$. The opening spread is therefore directly related to the probability of informed trading. Note that if PIN = 0, either because of the absence of new information ($\alpha = 0$) or traders informed of it ($\mu = 0$), the spread is also 0. This reflects the fact that only asymmetric information affects spreads when market makers are risk neutral.

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, Kiefer, and O'Hara (1996), (1997a), (1997b), Easley, Kiefer, O'Hara, and Paperman (1996), and Easley, O'Hara, and Paperman (1998) have used these measures of asymmetric information to show how spreads differ between 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 explore in an asset pricing context in Section IV of this paper.

III. Data and Estimation of PIN

We estimate our model for the sample of all ordinary common stocks listed on the New York Stock Exchange (NYSE) and the American Stock Exchange (AMEX) for the years 1983–2001. We focus on NYSE- and AMEX-listed stocks because the market microstructure of those venues most closely conforms to that of our structural model. We exclude real estate investment trusts (REITs), stocks of companies incorporated outside of the U.S, 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. Further, since we form portfolios based on year-end firm size, we exclude stocks for which this information is not available. In addition, we eliminate stocks with a year-end price below \$1. All other stocks are included. The final sample has between 1,705 and 2,281 stocks in the years 1983–2001.[3]

The likelihood function (2) depends upon the number of buys and sells each day for each stock in our sample. Transaction data give 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 transaction 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 market microstructure literature, and it essentially uses trade placement relative to the current bid and ask quotes to determine trade direction. Using this data, we maximize the likelihood function over the structural parameters, $\theta = (\alpha, \mu, \varepsilon_b, \varepsilon_s, \delta)$, 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. We then compute yearly PINs for each stock using the formula in equation (3).

[3]In previous research we estimated PINs up to 1998. The inclusion of the 3 additional years in this data set creates significant computational difficulties because of the increased number of trades in later years. But as we want to use PIN in asset pricing tests, it is important to have as long a time series as possible.

The maximum likelihood estimation converges for almost all stocks. Panel A of Table 1 lists the number of stocks for which we could estimate the likelihood function each year and which also had year-end price data in the Center for Research in Security Prices (CRSP). Of over 38,000 stock-years, we were able to obtain PIN estimates for all but 427. These failures were generally due to days of extremely high trading volume, which caused computational underflow in the optimization program. This issue occurs primarily in the last 6 years of the sample, coinciding with the advent of small day traders and the growth in automated trading, both of which have resulted in a marked increase in the number of trades. Furthermore, this occurs almost exclusively for the largest stocks rather than for smaller stocks. For instance, Table 1 shows that while only 45 of the 1,871 stocks (2.4%) in the 2001 sample do not obtain PIN estimates, these stocks account for 23.8% of the total market capitalization in our 2001 sample. This suggests that care is warranted in interpreting the asset pricing results for large stocks. We also suspect that the microstructure model provides a better description of the information environment among smaller stocks. For instance, the assumption that

TABLE 1

Summary Statistics

Panel A of Table 1 contains basic information on the yearly samples. N_EST is the number of stocks for which PIN estimates were obtained, while N_NOT_EST is the number of stocks for which estimates could not be obtained. FRAC_CAP is the total year-end market value of the stocks for which PIN estimates were not obtained divided by the total market value of the sample. Panel B presents summary statistics of PIN. Statistics were first computed based on each of the yearly samples, and the table reports the minimum, mean, and maximum across years of the statistics in the first column. P1, P25, P75, and P99 refer to percentiles of the yearly cross-sectional distribution; StdDev is the standard deviation; and CorrSz is the Spearman correlation between firm size and PIN.

Panel A. Summary Statistics of Yearly Samples

Year	N_EST	N_NOT_EST	FRAC_CAP
1983	2,080	2	0.001
1984	2,005	2	0.002
1985	1,943	3	0.000
1986	1,861	4	0.002
1987	1,913	5	0.001
1988	1,876	2	0.017
1989	1,818	9	0.008
1990	1,705	5	0.000
1991	1,808	14	0.003
1992	1,929	20	0.033
1993	2,081	9	0.013
1994	2,127	3	0.001
1995	2,142	17	0.008
1996	2,180	60	0.120
1997	2,239	62	0.237
1998	2,281	39	0.189
1999	2,145	63	0.256
2000	1,948	63	0.317
2001	1,826	45	0.238

Panel B. Summary Statistics of PIN

Statistics	Min	Mean	Max
P1	0.059	0.085	0.100
P25	0.115	0.153	0.173
Median	0.160	0.194	0.214
P75	0.219	0.246	0.268
P99	0.432	0.459	0.511
Mean	0.181	0.208	0.227
StdDev	0.067	0.078	0.103
CorrSz	–0.852	–0.661	–0.485

information events occur only once per day seems plausible for smaller stocks but less so for large stocks with thousands of trades per day.

Summary statistics for the PIN estimates are provided in Panel B of Table 1. We compute yearly statistics, and the table reports the minimum, mean, and maximum across years of these statistics. For instance, the average of the yearly cross-sectional median PINs is 0.194. The means of the yearly 25th and 75th cutoff points are 0.153 and 0.246, respectively. So on average, half of our stocks have PINs between these two levels. In prior work we found a strong negative correlation between PIN and firm size, defined as market value of equity at the beginning of the year. Panel B of Table 1 reports the cross-sectional Spearman correlation between PIN and firm size. The mean of these correlations is −0.661.

IV. Tests on PIN-Size Portfolios

Previous research has shown that size is an important determinant of excess returns (see Banz (1981), Barry and Brown (1984), and Fama and French (1992), (1993)). Size also plays an indirect role in our information-based explanation of why PIN matters for returns. If, as suggested by Barry and Brown (1984), size proxies for the total amount of information available about a stock, then controlling for size, PIN should proxy for how this total information is divided between public and private information.[4] Indeed, in the model of Easley and O'Hara (2004), both the amount of total information and the fraction of the information that is private affect expected returns. By first sorting stocks into portfolios according to size and then according to PIN, we hold the amount of total information approximately constant across PIN portfolios. Therefore, by analyzing the return differences across PIN portfolios within each size category, we investigate whether the split between public and private information affects expected returns, as proposed by Easley and O'Hara (2004).

At the beginning of each year, we sort stocks into size deciles based on market capitalization at the end of the prior year, and within each size decile, we sort into three portfolios based on the PINs estimated over the prior year, ranging from the lowest one-third of PINs to the highest one-third of PINs. This sequential sorting procedure produces 30 portfolios with an approximately equal number of stocks in each portfolio.[5] The results of the sorting procedure are reported in Table 2.

Panel A of Table 2 reports the average PIN for stocks in each portfolio. As expected within each PIN classification, average PIN decreases uniformly as we move from small to large stocks. Within size classifications, PIN increases as we move from low to high PIN stocks. This change is much larger for small size categories than it is for large size categories. For the smallest stocks, the average PIN changes from 0.20 for low PIN to 0.39 for large PIN; while for the largest stocks, it changes from 0.09 to 0.16. Most large stocks have small PINs, so

[4]Barry and Brown (1984) also use period of listing as a proxy for the total information available.

[5]Due to the strong negative correlation between size and PIN, independent sorts into size and PIN portfolios provide too few firms in the ⟨large size, high PIN⟩ and ⟨small size, low PIN⟩ cells to be useful.

TABLE 2

Characteristics and Returns of PIN Portfolios

At the beginning of each year during 1984–2002, stocks are sorted into size deciles based on market capitalization at the end of the prior year, and within each size decile, three portfolios are formed based on the PINs estimated over the prior year. Panel A of Table 2 contains the time-series average of the yearly value-weighted mean PIN for each portfolio, and DIF is the difference between high and low PIN portfolios. Panel B contains the time-series average of the average stock market value of the firms (in millions of dollars) in each portfolio, and DIF is the difference between high and low PIN portfolios. Panel C contains the time-series average of the monthly returns of each portfolio. Returns are weighted by the prior year-end market value. DIF is the average return difference between high and low PIN portfolios, and t(DIF) is the t-statistics of DIF. The last row provides return statistics on a portfolio, denoted PINF, that is equally invested in each of the individual portfolios in the 10 preceding rows.

Panel A. PIN

Size Portfolio	Low	2	High	DIF
Small	0.203	0.282	0.394	0.191
2	0.190	0.256	0.352	0.162
3	0.177	0.235	0.319	0.142
4	0.165	0.218	0.293	0.128
5	0.158	0.205	0.270	0.112
6	0.151	0.195	0.252	0.101
7	0.138	0.179	0.234	0.096
8	0.130	0.165	0.213	0.082
9	0.121	0.151	0.190	0.070
Large	0.094	0.124	0.161	0.068

Panel B. Stock Market Value of Firms

Size Portfolio	Low	2	High	DIF
Small	12.49	11.53	10.36	–2.130
2	34.56	33.90	32.56	–1.995
3	74.30	73.19	70.80	–3.498
4	141.0	141.9	137.3	–3.755
5	253.1	250.6	247.0	–6.069
6	439.3	434.1	421.2	–18.10
7	767.7	754.6	734.7	–32.96
8	1,393.5	1,361.4	1,337.4	–56.06
9	3,003.3	2,878.3	2,713.5	–289.8
Large	21,304.8	11,468.7	9,769.0	–11,535.8

Panel C. Returns

Size Portfolio	Low	2	High	DIF	t(DIF)
Small	0.428	0.552	1.164	0.736	2.63
2	0.197	0.468	0.970	0.773	3.14
3	0.339	0.383	0.911	0.573	2.72
4	0.825	0.939	0.974	0.149	0.71
5	0.724	0.888	0.933	0.209	1.18
6	0.901	1.083	1.056	0.155	1.06
7	1.105	1.045	0.992	–0.113	–0.78
8	1.081	1.136	1.116	0.035	0.28
9	1.059	1.161	1.167	0.109	0.69
Large	1.136	1.141	1.143	0.008	0.05
PINF	0.780	0.879	1.043	0.263	2.78

creating significant variation within this group is not possible. Because the change in PIN is small for large stocks, we do not expect to find significant variation in returns as we move across PIN categories for large stocks.

Panel B of Table 2 reports the average size for stocks in each portfolio. Again, as expected within each size classification, average size generally decreases as we move from low to high PIN. However, in percentage terms the changes are large only for the largest size classification, reflecting that the very largest firms tend to have very low PINs. To check on whether our results are due to a particular classification scheme or to a correlation within categories, we also considered a 5 × 5 size-PIN sort and a 10 × 5 size-PIN sort. All of the results are similar for each classification scheme, so we report only those for the 10 × 3 size-PIN sort.

Panel C of Table 2 reports value-weighted monthly returns for each portfolio.[6] These returns are the time-series average of the monthly returns on each portfolio. The differences of the returns on high and low PIN portfolios are positive and significant for the three smallest size deciles. They are 0.74, 0.77, and 0.57 with t-values of 2.63, 3.14, and 2.72 for the three smallest size deciles, respectively. The remaining differences vary in sign and are insignificantly different from 0. This suggests that the degree of asymmetric information, or at least our estimated PIN, only affects expected returns among small stocks.[7]

Table 2 also provides the mean return on a composite zero-investment portfolio (denoted PINF) formed by taking long (short) positions of equal size in the 10 high (low) PIN portfolios. The mean monthly return on this portfolio is 0.26%, and the corresponding t-value is 2.78. This abnormal return is thus both economically and statistically significant. Using the theory in Easley and O'Hara (2004), we interpret the return on this zero-investment portfolio as a return earned for bearing an information-based risk. To further examine this risk and to ask whether it can be explained by other standard risk factors, in the next section we conduct a factor analysis of this return.

V. Factor Tests

Fama and French (1993) find that the returns on three factor-mimicking portfolios, constructed as the overall stock market portfolio (r_M), a portfolio long in small firms and short in large firms (SMB), and a portfolio long in firms with high book-to-market equity (BE/ME) and short in low BE/ME firms (HML), contributes to explaining average returns on stock portfolios. The three-factor model, however, does not explain the returns to momentum portfolios (see Jegadeesh and Titman (1993), Fama and French (1996), and Grundy and Martin (2001)) or liquidity portfolios (see Pástor and Stambaugh (2003)). Carhart (1997) suggests adding a factor-mimicking portfolio based on momentum (UMD) (i.e., the returns on a diversified portfolio long in recent winners and short in recent losers) to the three-factor model. Pástor and Stambaugh (2003) suggest adding an additional factor related to liquidity risk. The liquidity measure essentially relies on signed volume and subsequent price reversals.[8] They then construct the Pástor-Stambaugh liquidity (PSLIQ) risk factor using the returns from a portfolio of long stocks with high return sensitivity to fluctuations in aggregate liquidity and short stocks with low sensitivities. An alternative approach to capturing liquidity

[6]The results with equal-weighted returns are similar, so we follow Fama and French (1993) in focusing on value-weighted returns.

[7]Since PIN is measured with error, it could be that PIN actually matters across all size categories, but either our measurements are more noisy for large stocks or the effects are smaller for large stocks and are overwhelmed by the noise.

[8]The Pástor and Stambaugh (2003) measure is akin to measuring the temporary price effects that arise due to illiquidity in a market. Microstructure research typically divides total price effects of trading into a permanent component (due to information) and a temporary component (due to liquidity). A price reversal is consistent with these temporary effects.

is suggested by Amihud (2002), who proposes an illiquidity measure defined as the daily price change divided by the daily volume in a stock.[9] Charoenrook and Conrad (2004) construct a factor-mimicking portfolio (ILLIQ) based on the Amihud illiquidity measure, and they show that it is priced over the 1963–2003 period. Both the Pástor and Stambaugh and the Amihud measures rely on daily rather than intradaily data. Therefore, while the measures are likely to be relatively coarse, they do provide the ability to investigate the effect of liquidity in long-horizon asset pricing studies.[10]

It is conceivable that PIN is correlated with beta, book-to-market, momentum, and liquidity, as well as with size, so that the returns being earned on PIN portfolios are just compensation for risk already captured by the Fama-French, momentum, and liquidity factor-mimicking portfolios. The time-series correlations between these factors and our composite portfolio, PINF, which is long in high PIN and short in low PIN stocks, are given in Table 3.[11] The highest correlation between the returns of PINF and another factor is the correlation of 0.55 with the momentum factor, UMD. Because UMD also has a high average return of 0.99, it appears to be the variable most likely to explain the returns to the PIN portfolios. The correlations with the remaining factors are relatively low, and none of the mean factor returns are significant at the 5% level.

The lack of significance for the means of the PSLIQ and ILLIQ factors might seem surprising, given the evidence in Pástor and Stambaugh (2003) and Charoenrook and Conrad (2004). However, while Pástor and Stambaugh obtain a positive intercept when regressing the PSLIQ factor against the Fama-French factors, the mean return is not significantly positive, even in the Pástor-Stambaugh sample period between 1966 and 1999. The ILLIQ factor mean return, by contrast, is very high in the period preceding our sample period, yielding a significantly positive mean return for the entire 1963–2003 sample period studied by Charoenrook and Conrad.[12]

[9]The Amihud measure is similar to a Kyle λ, which is simply the price change divided by the volume. In the Kyle model, the presence of informed traders causes λ to be positive, while if there is no asymmetric information, then λ is 0. The Amihud measure will thus pick up both information effects and other liquidity-related effects such as inventory pressures, tick size constraints, and the like.

[10]Indeed, Amihud ((2002), p. 32) stresses that the Amihud illiquidity measure was designed for situations where detailed microstructure data were not available, noting that "there are finer and better measures of illiquidity such as the bid ask spread (quoted or effective), transaction-by-transaction market impact, or the probability of information-based trading (i.e., PIN). These measures, however, require a lot of microstructure data that are not available in many stock markets."

[11]The Fama-French and the momentum factor returns were obtained from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/, which also details the procedure used to create the portfolio returns. The Pástor and Stambaugh (2003) liquidity factor returns were obtained from the Wharton Research Data Services. We are grateful to Anchada Charoenrook and Jennifer Conrad for supplying the Amihud illiquidity factor returns.

[12]The sample period seems to be very important for these results. The Pástor-Stambaugh value-weighted factor-mimicking portfolio has an average return of 0.26 (t-statistic = 1.20) in the period 1966–1999 and a higher return of 0.38 (t-statistic = 1.00) in our 1984–2002 sample period. The Amihud factor average return is 0.82 (t-statistic = 3.97) in the Charoenrook-Conrad sample period of 1963–2003. This is due to an even higher mean return of 1.28 (t-statistic = 3.74) in the subsample period prior to our study and an insignificant mean return of 0.20 (t-statistic = 0.86) in our 1984–2002 sample period.

TABLE 3

Factor Summary Statistics

In Table 3, all statistics are based on monthly returns in 1984–2002 of the Fama-French, momentum, liquidity, and PIN factor portfolios: market excess return (r_M), small stock returns minus large stock returns (SMB), high book-to-market stock returns minus low book-to-market stock returns (HML), past 1-year winner stock returns minus past loser stock returns (UMD), returns on high liquidity beta stocks minus returns on stocks with low liquidity betas (PSLIQ) as constructed by Pástor and Stambaugh (2003), returns on high Amihud illiquidity stocks minus returns on low illiquidity stocks as constructed by Charoenrook and Conrad (2004), and high PIN stock returns minus low PIN stock returns (PINF). The construction of the PINF portfolio is explained in Table 2.

				Autocorrelation for Lag				Correlations					
Variables	Mean	t(Mean)	StdDev	1	2	3	12	SMB	HML	UMD	PINF	PSLIQ	ILLIQ
r_M	0.55	1.80	4.61	0.02	–0.07	–0.04	–0.01	0.16	–0.52	–0.09	–0.21	–0.11	–0.09
SMB	–0.07	–0.32	3.50	–0.04	–0.00	–0.15	0.04		–0.44	0.10	–0.04	0.16	0.68
HML	0.36	1.60	3.37	0.09	0.05	0.07	–0.04			–0.07	–0.06	–0.21	–0.00
UMD	0.99	3.31	4.52	–0.07	–0.08	0.02	0.24				0.55	0.48	–0.12
PINF	0.26	2.78	1.43	0.07	–0.09	0.03	0.02					0.23	–0.18
PSLIQ	0.38	1.00	5.71	0.05	0.08	–0.16	–0.02						–0.14
ILLIQ	0.20	0.86	3.45	–0.01	0.07	–0.09	0.20						

To investigate whether the Fama-French, momentum, and liquidity factors jointly can explain the apparent effect of PIN on returns, we regress the time series of monthly returns for each of the 10 zero-investment PIN portfolios, R_i, on the Fama-French, momentum, and liquidity factors:

$$(5)\quad R_i \;=\; \alpha_i + \beta_i r_M + s_i\mathrm{SMB} + h_i\mathrm{HML} + m_i\mathrm{UMD} + l_i\mathrm{PSLIQ} + a_i\mathrm{ILLIQ} + \epsilon_i.$$

The results of this regression are reported in Table 4. The most important coefficients are the intercepts. If the factors explain the returns on our portfolios, then the intercepts should be indistinguishable from 0. For the three portfolios involving the smallest stocks (where we have significant variation in PIN and significant differences in raw returns) the intercepts are positive, large, and significantly different from 0. For these portfolios the intercepts are 0.83, 0.64, and 0.53, with t-values of 2.96, 2.78, and 2.68, respectively. For portfolios of larger stocks the intercepts are smaller, and none are significant. We also tested the null hypothesis that the intercepts are jointly equal to 0 using the F-test of Gibbons, Ross, and Shanken (GRS) (1989). The GRS statistic has a p-value of 0.015, so the null hypothesis is rejected.

The column labeled "All" in Table 4 also reports the results for regression (5) applied to the time series of monthly returns of the composite portfolio, PINF.[13] Here the intercept is 0.20 with a t-value of 2.43. Thus even after controlling for these six factors, there is a significant return of 0.20% per month on our zero-investment PINF portfolio that is unexplained.[14]

Since existing factors do not explain the returns to PIN-based portfolios, we ask whether a PIN factor can explain these returns. For PIN to play a role as a statistical factor, there must be common variation of returns within high PIN stocks

[13]The construction of this zero-investment portfolio, which is long in high PIN stocks and short in low PIN stocks, is described in Table 2.

[14]We have also checked that the effect of PIN is not a January effect. Excluding the month of January from regression (5) actually yields a higher intercept of 0.28 (t-statistic = 3.52) on the composite portfolio. The average monthly return on the composite portfolio outside January is 0.37% (t-statistic = 4.26).

TABLE 4

Fama-French Six-Factor Regressions

The time series of zero-investment portfolio returns, R_i, is created as explained in Table 2, and the following regression is performed for each portfolio i:

$$R_i = \alpha_i + \beta_i r_M + s_i \text{SMB} + h_i \text{HML} + m_i \text{UMD} + l_i \text{PSLIQ} + a_i \text{ILLIQ} + \epsilon_i. \tag{5}$$

Table 4 reports the point estimates for each of the coefficients and their t-statistics, along with the adjusted R^2. The sample period is 1984–2002. The last row reports the p-value from a Gibbons, Ross, and Shanken (1989) F-test of whether the 10 intercepts are jointly equal to 0.

	Size Portfolio										
Coefficients	Small	2	3	4	5	6	7	8	9	Large	All
α	0.83	0.64	0.53	0.03	0.22	0.01	–0.12	–0.07	0.03	–0.13	0.20
	(2.96)	(2.78)	(2.68)	(0.15)	(1.31)	(0.06)	(–0.81)	(–0.53)	(0.23)	(–0.93)	(2.43)
β_i	–0.25	–0.21	–0.23	–0.13	–0.21	–0.03	–0.05	0.01	0.06	0.11	–0.09
	(–3.58)	(–3.72)	(–4.64)	(–2.56)	(–4.99)	(–0.73)	(–1.30)	(0.32)	(1.75)	(3.07)	(–4.60)
s_i	0.08	–0.21	0.09	–0.12	–0.14	0.00	0.01	–0.11	–0.03	0.29	–0.01
	(0.62)	(–1.86)	(0.95)	(–1.17)	(–1.76)	(0.04)	(0.08)	(–1.80)	(–0.44)	(4.21)	(–0.35)
h_i	0.07	0.10	0.03	–0.04	–0.27	–0.13	–0.22	–0.16	–0.31	–0.01	–0.09
	(0.64)	(1.06)	(0.36)	(–0.54)	(–4.00)	(–2.40)	(–3.75)	(–3.28)	(–5.79)	(–0.14)	(–2.97)
m_i	0.10	0.24	0.25	0.24	0.21	0.19	0.12	0.16	0.13	0.10	0.17
	(1.53)	(4.26)	(5.41)	(4.82)	(5.05)	(5.87)	(3.41)	(5.35)	(4.00)	(2.86)	(8.96)
l_i	–0.02	–0.05	–0.07	–0.14	–0.05	0.02	–0.05	–0.08	0.03	0.10	–0.03
	(–0.40)	(–0.98)	(–1.80)	(–3.35)	(–1.52)	(0.80)	(–1.80)	(–3.05)	(1.14)	(3.36)	(–1.91)
a_i	–0.35	–0.06	–0.25	0.01	0.02	0.03	0.06	0.06	0.06	–0.15	–0.06
	(–2.82)	(–0.57)	(–2.90)	(0.14)	(0.26)	(0.55)	(0.87)	(1.12)	(1.03)	(–2.40)	(–1.56)
Adj. R^2	0.13	0.24	0.25	0.13	0.20	0.22	0.11	0.14	0.33	0.33	0.36

GRS-test p-value: 0.015

and within low PIN stocks. Such common variation would increase the volatility of the zero-investment PIN-based portfolios. Table 3 shows that the composite PIN portfolio, PINF, has a standard deviation of monthly returns of 1.43%. This is quite low compared to the standard deviation of the UMD portfolio of 4.52% or the PSLIQ portfolio of 5.71%. However, Table 5 shows that there is a strong correlation in returns across the PIN-based portfolios. Specifically, we find substantial correlation between returns to PIN portfolios for smaller firms, and substantial correlation between returns to PIN portfolios for larger firms. For example, the correlation in returns between size portfolios 1 and 2 is 0.32, and between portfolios 9 and 10 it is 0.46. However, there is little or even negative covariation in returns between small and large firm PIN portfolios. Perhaps small and large firm PIN portfolios are responding to information events that are different but correlated within size groups. Regardless of the interpretation, this covariation provides support for a PIN factor.[15]

We cannot use the composite PIN portfolio, PINF, to explain returns to the size portfolios, as PINF is composed of an equal weighting of all of the size portfolios. Instead, we form PIN factor-mimicking portfolios whose component stocks exclude the stocks in the portfolio whose return we are trying to explain.

[15]To give a perspective for the effect of the common variation within PIN portfolios, we replicated the sorting procedure by assigning a generated random number to each of the sample stocks each year and substituting this variable for PIN in the double-sorts. This procedure was repeated 50 times, yielding an average standard deviation of returns on the composite portfolio of 0.72%. The standard deviation across the 50 runs of the computed portfolio standard deviation was 0.033%.

TABLE 5

Return Correlations between PIN-Based Zero-Investment Portfolios

At the beginning of each year in 1984–2002, stocks are sorted into size deciles based on market capitalization at the end of the prior year, and within each size decile, three portfolios are formed based on the PINs estimated over the prior year. Returns are weighted by the prior year-end market value, and for each size decile, a zero-investment portfolio is created with a long position in the high PIN portfolio and a short position in the low PIN portfolio. The Pearson correlations between the 10 zero-investment portfolio returns are reported in Table 5, and the corresponding *p*-values are in parentheses.

	Size Portfolio								
Size Portfolio	2	3	4	5	6	7	8	9	10
1	0.319	0.250	0.207	0.233	0.036	0.027	0.023	–0.001	–0.026
2		0.400	0.384	0.332	0.085	0.067	0.049	–0.188	–0.146
3			0.294	0.310	0.209	0.150	0.083	–0.047	0.017
4				0.347	0.029	0.241	0.276	0.101	–0.061
5					0.215	0.297	0.275	0.242	0.016
6						0.253	0.124	0.305	0.303
7							0.286	0.316	0.166
8								0.411	0.078
9									0.459

Specifically, for size portfolio i, a PIN factor, PINF_{-i}, is formed as the composite zero-investment portfolio used in Tables 2–4, except that the zero-investment PIN portfolio for size portfolio i is excluded from PINF_{-i}. That is, for each of the 9 size portfolios (excluding size portfolio i), two value-weighted portfolios are created based on the stocks in the low and high PIN terciles, respectively. PINF_{-i} is then created as the average of the returns on the 9 high PIN portfolios minus the average of the returns on the 9 low PIN portfolios. This ensures that there is no overlap between the stocks in the factor-mimicking portfolio and the portfolio whose returns we wish to explain. This procedure is obviously more demanding than simply forming one PIN factor and using it for all size portfolios. Of course, one PIN factor, such as PINF, could be used to help explain returns to other portfolios.

Table 6 reports the results of the regressions

$$\begin{aligned} R_i \quad = \quad & \alpha_i + \beta_i r_M + s_i\text{SMB} + h_i\text{HML} + m_i\text{UMD} \\ & + l_i\text{PSLIQ} + a_i\text{ILLIQ} + p_i\text{PINF}_{-i} + \epsilon_i, \end{aligned} \tag{6}$$

where PINF_{-i} is the PIN factor for portfolio i. The results provide strong support for the PIN factor playing an important role in asset pricing. Looking first at the PIN factor coefficient, we find all coefficients to be positive, and they are significant at the 5% level for 8 of the 10 size portfolios. Thus, sensitivity to the PIN factor increases returns for stocks (or more precisely, portfolios) in accordance with our theory. That this effect is weaker for the largest stocks is consistent with findings in Aslan, Easley, Hvidkjaer, and O'Hara (2007) that information risk is more important for smaller stocks. Second, we find that inclusion of the PIN factor substantially reduces all of the intercepts relative to the results of Table 4. The intercepts of the three smallest size portfolios remain positive and significant at the 5% level. None of the other intercepts are positive and significantly different from 0. These results suggest that adding the PIN factor significantly reduces the unexplained portion of asset returns. Finally, the R^2 values are also substantially higher when including the PIN factor. The average R^2 across the 10 portfolios is

0.20 in the 6-factor regressions in Table 4, while it increases to 0.25 when including the PIN factor in Table 6.

TABLE 6

Fama-French Seven-Factor Regressions with PINF

The time series of zero-investment portfolio returns, R_i, is created as explained in Table 2, and the following regression is performed for each portfolio i:

$$R_i = \alpha_i + \beta_i r_M + s_i \text{SMB} + h_i \text{HML} + m_i \text{UMD} + l_i \text{PSLIQ} + a_i \text{ILLIQ} + p_i \text{PINF}_{-i} + \epsilon_i, \quad (6)$$

Table 6 reports the point estimates for each of the coefficients and their t-statistics, along with the adjusted R^2. The sample period is 1984–2002.

	Size Portfolio									
Coefficients	Small	2	3	4	5	6	7	8	9	Large
α	0.74	0.53	0.44	–0.18	0.07	–0.03	–0.25	–0.14	–0.06	–0.15
	(2.68)	(2.34)	(2.27)	(–0.89)	(0.45)	(–0.23)	(–1.76)	(–1.16)	(–0.42)	(–1.07)
β_i	–0.20	–0.15	–0.18	–0.05	–0.15	–0.01	0.01	0.05	0.11	0.12
	(–2.79)	(–2.65)	(–3.75)	(–0.94)	(–3.67)	(–0.19)	(0.20)	(1.42)	(2.98)	(3.19)
s_i	0.10	–0.21	0.10	–0.12	–0.14	0.01	0.01	–0.11	–0.02	0.29
	(0.75)	(–1.98)	(1.13)	(–1.23)	(–1.89)	(0.09)	(0.21)	(–1.83)	(–0.38)	(4.26)
h_i	0.15	0.19	0.09	0.05	–0.21	–0.11	–0.17	–0.13	–0.28	0.00
	(1.33)	(2.07)	(1.13)	(0.67)	(–3.32)	(–2.08)	(–3.07)	(–2.70)	(–5.37)	(0.04)
m_i	–0.02	0.10	0.16	0.08	0.07	0.16	0.02	0.10	0.06	0.08
	(–0.30)	(1.70)	(3.12)	(1.48)	(1.68)	(4.33)	(0.47)	(2.98)	(1.59)	(2.03)
l_i	–0.00	–0.02	–0.06	–0.12	–0.03	0.03	–0.04	–0.07	0.05	0.10
	(–0.01)	(–0.50)	(–1.46)	(–3.11)	(–0.92)	(1.04)	(–1.31)	(–2.75)	(1.74)	(3.46)
a_i	–0.33	–0.01	–0.23	0.07	0.07	0.05	0.10	0.09	0.09	–0.15
	(–2.73)	(–0.15)	(–2.75)	(0.86)	(0.99)	(0.74)	(1.53)	(1.57)	(1.53)	(–2.32)
p_i	0.69	0.78	0.55	0.96	0.77	0.18	0.56	0.34	0.41	0.10
	(2.94)	(4.15)	(3.51)	(5.81)	(5.92)	(1.70)	(5.01)	(3.57)	(3.85)	(0.89)
Adj. R^2	0.16	0.29	0.29	0.24	0.30	0.22	0.20	0.18	0.37	0.33

It is useful to compare our findings on the PIN factor with those of the Pástor-Stambaugh liquidity factor and the Amihud illiquidity factor. While the coefficients on the PIN factor are uniformly positive, the coefficients on both liquidity factors are not. The Amihud factor, for example, is negative for smaller stocks but generally positive for larger stocks, while the PSLIQ factor shows no consistent pattern. Similarly, only 3 of the 10 coefficient values for the Amihud measure are statistically significant in either Table 6 (with the PIN factor) or Table 4 (without the PIN factor). The PSLIQ factor shares a similar significance pattern. These results suggest that while liquidity may play a role in asset pricing, it clearly does not obviate the role played by information risk. Indeed, our strong findings on the PIN factor clearly underscore that information risk is a factor in asset pricing.

In summary, our findings are twofold. First, the Fama-French, momentum, and liquidity factors do not explain the differences in average returns across PIN-sorted portfolios. Second, a factor-mimicking portfolio based on PIN has good ability to capture those return differences, even when the factor-mimicking portfolio contains none of the stocks in the portfolios whose returns are explained. Still, the PIN factor fails to completely explain the returns to PIN-sorted portfolios among the smallest stocks.

VI. Conclusion

We show that zero-investment portfolios based on PIN sorts earn significant abnormal returns that cannot be explained by the Fama-French factors or by a momentum factor. Moreover, we find that significant covariation in returns exists among PIN portfolios, suggesting that PIN might proxy for an underlying factor in returns. Consequently, we form a factor-mimicking portfolio based on PIN and find that it helps to explain the returns to these portfolios, though it is unable to fully explain the return differences across PIN portfolios among the smallest stocks.

Our results show that PIN affects returns and that a PIN-based factor performs as well as other factors in a standard multifactor model designed to explain differences in returns. Our results thus join a growing literature showing that microstructure variables influence long-term asset returns. How best to measure such microstructure influences remains a source of discussion, but we would argue that our results here suggest that liquidity and information risk play distinctive roles in this price process. With regard to the PIN factor, a natural goal is to understand how such a factor role arises. Certainly, a first step is to provide a theoretical explanation of why this effect occurs. This cannot be done with the standard capital asset pricing model (CAPM) or the consumption-based CAPM (CCAPM), as PIN derives from an economy in which traders have differential information. If one builds a model without differential information, or one in which all traders in equilibrium have common information, then PIN cannot matter. We believe that asset pricing models in which differential information remains as an equilibrium phenomenon are more natural than those in which it never exists or disappears. Easley and O'Hara (2004) constructed such a model to explain how PIN as a characteristic could be priced in the cross section. We are currently investigating whether such a model can provide a factor interpretation for PIN.

References

Acharya, V. V., and L. H. Pedersen. "Asset Pricing with Liquidity Risk." *Journal of Financial Economics*, 77 (2005), 375–410.

Amihud, Y. "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects." *Journal of Financial Markets*, 5 (2002), 31–56.

Ang, A.; R. J. Hodrick; Y. Xing; and X. Zhang. "The Cross-Section of Volatility and Expected Returns." *Journal of Finance*, 61 (2006), 259–299.

Aslan, H.; D. Easley; S. Hvidkjaer; and M. O'Hara. "Firm Characteristics and Informed Trading: Implications for Asset Pricing." Working Paper, University of Houston, Cornell University, and INSEAD (2007).

Banz, R. W. "The Relationship between Return and Market Value of Common Stocks." *Journal of Financial Economics*, 9 (1981), 3–18.

Barry, C. B., and S. J. Brown. "Differential Information and the Small Firm Effect." *Journal of Financial Economics*, 13 (1984), 283–294.

Carhart, M. M. "On Persistence in Mutual Fund Performance." *Journal of Finance*, 52 (1997), 57–82.

Charoenrook, A., and J. Conrad. "Identifying Risk-Based Factors." Working Paper, University of North Carolina (2004).

Chordia, T.; R. Roll; and A. Subrahmanyam. "Commonality in Liquidity." *Journal of Financial Economics*, 56 (2000), 3–28.

Daniel, K., and S. Titman. "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns." *Journal of Finance*, 52 (1997), 1–33.

Davis, J. L.; E. F. Fama; and K. R. French. "Characteristics, Covariances, and Average Returns: 1929 to 1997." *Journal of Finance*, 55 (2000), 389–406.

Duarte, J., and L. Young. "Why is *PIN* Priced?" *Journal of Financial Economics*, 91 (2009), 119–138.
Easley, D.; S. Hvidkjaer; and M. O'Hara. "Is Information Risk a Determinant of Asset Returns?" *Journal of Finance*, 57 (2002), 2185–2221.
Easley, D.; N. M. Kiefer; and M. O'Hara. "Cream-Skimming or Profit-Sharing? The Curious Role of Purchased Order Flow." *Journal of Finance*, 51 (1996), 811–833.
Easley, D.; N. M. Kiefer; and M. O'Hara. "One Day in the Life of a Very Common Stock." *Review of Financial Studies*, 10 (1997a), 805–835.
Easley, D.; N. M. Kiefer; and M. O'Hara. "The Information Content of the Trading Process." *Journal of Empirical Finance*, 4 (1997b), 159–186.
Easley, D.; N. M. Kiefer; M. O'Hara; and J. B. Paperman. "Liquidity, Information, and Infrequently Traded Stocks." *Journal of Finance*, 51 (1996), 1405–1436.
Easley, D., and M. O'Hara. "Information and the Cost of Capital." *Journal of Finance*, 59 (2004), 1553–1583.
Easley, D.; M. O'Hara; and J. Paperman. "Financial Analysts and Information-Based Trade." *Journal of Financial Markets*, 1 (1998), 175–201.
Fama, E. F., and K. R. French. "The Cross-Section of Expected Stock Returns." *Journal of Finance*, 47 (1992), 427–465.
Fama, E. F., and K. R. French. "Common Risk Factors in the Returns on Stock and Bonds." *Journal of Financial Economics*, 33 (1993), 3–56.
Fama, E. F., and K. R. French. "Multifactor Explanations of Asset Pricing Anomalies." *Journal of Finance*, 51 (1996), 55–84.
Gârleanu, N., and L. H. Pedersen. "Adverse Selection and the Required Return." *Review of Financial Studies*, 17 (2004), 643–665.
Gebhardt, W. R.; S. Hvidkjaer; and B. Swaminathan. "The Cross-Section of Expected Corporate Bond Returns: Betas or Characteristics?" *Journal of Financial Economics*, 75 (2005), 85–114.
Gibbons, M. R.; S. A. Ross; and J. Shanken. "A Test of the Efficiency of a Given Portfolio." *Econometrica*, 57 (1989), 1121–1152.
Grundy, B. D., and J. S. Martin. "Understanding the Nature of the Risks and the Source of the Rewards to Momentum Investing." *Review of Financial Studies*, 14 (2001), 29–78.
Jegadeesh, N., and S. Titman. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance*, 48 (1993), 65–91.
Lee, C. M., and M. J. Ready. "Inferring Trade Direction from Intraday Data." *Journal of Finance*, 46 (1991), 733–746.
Pástor, L., and R. F. Stambaugh. "Liquidity Risk and Expected Stock Return." *Journal of Political Economy*, 111 (2003), 642–685.