Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Asymmetric Effects of Informed Trading on the Cost of Equity Capital

Michael J. Brennan, Sahn-Wook Huh, Avanidhar Subrahmanyam

MANAGEMENT SCIENCE
Vol. 62, No. 9, September 2016, pp. 2460–2480
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2015.2250

Asymmetric Effects of Informed Trading on the Cost of Equity Capital

Michael J. Brennan
Anderson School of Management, University of California, Los Angeles, Los Angeles, California 90095; and Manchester Business School, University of Manchester, Manchester M13 9PL, United Kingdom, mbrennan@anderson.ucla.edu

Sahn-Wook Huh
School of Management, University (SUNY) at Buffalo, Buffalo, New York 14260, swhuh@buffalo.edu

Avanidhar Subrahmanyam
Anderson School of Management, University of California, Los Angeles, Los Angeles, California 90095, subra@anderson.ucla.edu

We decompose *PIN*, the probability of informed trading, into good-news (*PIN_G*) and bad-news (*PIN_B*) components, which we estimate at a quarterly frequency. We first assess the validity of *PIN* as a measure of informed trading by calculating its association with measures of the adverse-selection component of the cost of trading. We then provide new evidence that *PIN_G* and *PIN_B* capture informed trading around earnings announcements by showing that they predict positive and negative earnings surprises, respectively. Conjecturing that investors who take long positions will be more concerned about informed selling than about informed buying since the former depresses the sale price whereas the latter raises it, we then investigate asymmetry in the pricing of private information. We find strong evidence of such asymmetry in that the effect of *PIN_B* on the cost of equity capital is large and highly significant, whereas the effect of *PIN_G* is small and statistically insignificant.

Data, as supplemental material, are available at http://dx.doi.org/10.1287/mnsc.2015.2250.

Keywords: information asymmetry; decomposition of *PIN*; *PIN_G*; *PIN_B*; *AdjPIN*; *PSOS*; adverse-selection and noninformation components of trading costs; earnings announcements; earnings surprises; *CAR*; *SUE*; cost of equity capital

History: Received August 26, 2014; accepted May 9, 2015, by Neng Wang, finance. Published online in *Articles in Advance* December 10, 2015.

1. Introduction

There is now extensive evidence that the cost of equity capital or required return depends on the (il)liquidity of the markets in which the securities are traded; see, for example, Amihud and Mendelson (1986), Brennan and Subrahmanyam (1996), and Amihud (2002). More recently, Pástor and Stambaugh (2003), as well as Acharya and Pedersen (2005), relate systematic liquidity risk to expected returns. In papers such as Kyle (1985) and Glosten and Milgrom (1985), illiquidity arises purely from adverse selection due to informed trading. Easley et al. (1996) use a model that is close in spirit to that of Glosten and Milgrom (1985) to estimate the probability of informed trading, which they denote by *PIN*. Easley et al. (2002) show first that, given a security's fundamentals, its bid–ask spread is proportional to *PIN* and, second, that their estimate of *PIN* is a significant determinant of stock returns.

The relevance of *PIN*, both as a measure of informed trading and as a determinant of stock returns, has proved to be a topic of debate. Thus, Brown and Hillegeist (2007) find a negative relation between *PIN* and the quality of a firm's annual report, which they attribute to the crowding out of private information-based trading by public disclosure of information. Chen et al. (2007) report that the sensitivity of corporate investment to stock market prices is higher for firms with high values of *PIN*, and they argue that this is evidence that the stock prices of these firms reflect more private information through informed trading. Also consistent with the private information story, Ellul and Pagano (2006) find that IPO stocks with higher levels of *PIN* in secondary market trading have higher levels of IPO underpricing, which they attribute to compensation for expected adverse selection in the after-market. Vega (2006) finds that higher levels of preannouncement informed trading as measured by *PIN* reduce postannouncement earnings drift, and Ellul and Panayides (2013) document that *PIN* increases after analyst coverage termination, which is consistent with the role of analysts in reducing information asymmetry. On the other hand, Benos and Jochec (2007) find no evidence that *PIN* captures informed trading around earnings announcement dates, and Aktas et al. (2007) find no evidence that it captures information asymmetry before merger announcements.

Turning to the pricing of *PIN*, Mohanram and Rajgopal (2009) report that, although it is priced for NYSE/AMEX stocks in their sample period, the pricing evidence is not robust to alternative specifications and time periods. Duarte and Young (2009) use a more elaborate model to decompose *PIN* into two components: one that is related to "pure" informed trading (*AdjPIN*) and the other to liquidity shocks (*PSOS*). Their asset-pricing tests, which use raw returns and annual estimates of *PIN* and its components, find that only *PSOS* is priced. From this they conclude that *PIN* does not measure informed trading and that it is *PSOS*, a measure of illiquidity that is unrelated to information asymmetry, that drives the relation between *PIN* and the cross section of stock returns.

Most studies of *PIN* rely on estimates that are measured at an annual frequency (e.g., Aktas et al. 2007, Mohanram and Rajgopal 2009, Duarte and Young 2009, Lai et al. 2014). To the extent that *PIN* varies over time, and this is the assumption of studies that attempt to document time variation around information events, annual estimates reduce the power of tests. Therefore, we estimate *PIN* and its components on a quarterly basis, which allows us to explore variation in the probabilities of informed trading more efficiently and yields more powerful tests of the informed trading hypothesis. We examine the robustness of our results to alternative assumptions.

Considering the doubts that have been raised about the validity of *PIN*, we first assess its validity as a measure of informed trading by measuring its association with measures of the adverse-selection and noninformation components of trading costs estimated from the Glosten and Harris (1988) and Foster and Viswanathan (1993) models, and with the illiquidity measure of Amihud (2002). We also compare *PIN* with the quarterly *AdjPIN* and *PSOS* measures of Duarte and Young (2009). We find that *PIN* is highly correlated (0.26–0.27) with the adverse-selection component of trading costs, and only weakly correlated (<0.05) with the noninformation component. In contrast, *AdjPIN*, which is intended to isolate the component of *PIN* that is related to informed trading, has a lower correlation with the adverse-selection component (0.18) than does *PIN*. *PSOS*, which is intended to capture the component of *PIN* that is unrelated to informed trading, is highly correlated (0.23) with the adverse-selection component, and its correlation with the noninformation component is close to 0 (0.03–0.04). We conclude from this that *PIN* is the superior measure of the probability of informed trading and thus concentrate our empirical analyses on *PIN* and its components.

We provide further evidence that *PIN* is associated with informed trading by examining its behavior before quarterly earnings announcements. In the quarter before an earnings announcement, the probability of informed trading, *PIN*, is an increasing function of the absolute value of the earnings surprise. When we decompose *PIN* into the probability of informed buying on good news (*PIN_G*) and the probability of informed selling on bad news (*PIN_B*), we find that *PIN_G* is an increasing function of the earnings surprise, and *PIN_B* is a decreasing function of the surprise. This is consistent with the three *PIN* variables capturing informed trading on both good and bad news. We also find strong evidence that our quarterly *PIN* estimates are priced for NYSE/AMEX stocks over the full sample period, 1983–2010, as well as over subperiods that reflect different trading regimes, 1983–1997 and 2001–2010, and different data sources, 1993–2010 and 1983–2006.

Our primary concern is whether the two components of *PIN* have similar effects on the cost of equity capital. We conjecture that, since investors are more likely to need to sell stocks quickly to meet cash needs than to need to buy stocks quickly, they will be more concerned about informed traders with negative information, whose actions could depress the sale price in the event of a forced liquidation, than about traders with positive information whose actions could only raise the sale price. This leads to the hypothesis that it is *PIN_B*, the probability of informed trading on bad news, that is responsible for the pricing of *PIN*. Our asset-pricing tests are consistent with this and show that only *PIN_B* is priced; there is no evidence that investors demand a return premium for holding stocks with a high probability of informed trading on good news, *PIN_G*. Our results are robust to a series of experiments, including the test that controls for the *AdjPIN* and *PSOS* measures of Duarte and Young (2009).

Our findings are generally consistent with recent accounting studies, such as Bhattacharya et al. (2012), who examine the relations between earnings quality and the cost of capital, providing the evidence that information asymmetry indirectly affects the cost of equity capital, and Levi and Zhang (2015), who find that temporary increases in information asymmetry before earnings announcements lead to a higher cost of equity capital. The asymmetric effect of good- and bad-news informed trading on the cost of capital is also related to the findings in Brennan et al. (2012), who show that expected returns are significantly positively related to the sensitivity of the price change to sell orders but are not related to the sensitivity of prices to buy orders, and to Hong et al. (2000) and Kothari et al. (2008), who suggest that bad news is released more slowly than good news, since managers are reluctant to disclose bad news.

This paper is organized as follows. In §2, we describe the *PIN* measure and its decomposition and then discuss data and estimation of the *PIN* measures. In §3, we state the basic hypothesis, describe the test methodology, and define control variables. In §4, we report correlations between *PIN* and the adverse-selection and noninformation components of the cost of transacting and the behavior of the *PIN* statistics before quarterly earnings announcements. In §5, we examine the effects of *PIN* and its two components on the cost of equity capital. In §6, we report robustness tests. In §7, we conclude the paper. The online appendix is available as supplemental material at http://dx.doi.org/10.1287/mnsc.2015.2250.

2. The *PIN* Measure and Its Decomposition

2.1. Decomposition Into *PIN_G* and *PIN_B*

The probability of informed trading measure, *PIN*, developed by Easley et al. (1996 and 2002) is derived from the following simplified model of trading. Each day a private information event is assumed to occur with probability α, while no information event occurs with probability $1-\alpha$. If the information event occurs, it contains bad news with probability δ and good news with probability $1-\delta$. Orders from uninformed buyers (sellers) arrive randomly at the rate of ϵ_b (ϵ_s) each day, while orders from informed traders arrive randomly at the rate μ, but only if the information event has occurred. Informed traders buy if there is good news and sell if there is bad news that day. Then the likelihood of observing B_j buy orders and S_j sell orders on trading day j is given by

$$\begin{aligned} L(B_j, S_j \mid \theta) = {} & \alpha(1-\delta)e^{-(\mu+\epsilon_b)}\frac{(\mu+\epsilon_b)^{B_j}}{B_j!}e^{-\epsilon_s}\frac{\epsilon_s^{S_j}}{S_j!} \\ & + \alpha\delta e^{-\epsilon_b}\frac{\epsilon_b^{B_j}}{B_j!}e^{-(\mu+\epsilon_s)}\frac{(\mu+\epsilon_s)^{S_j}}{S_j!} \\ & + (1-\alpha)e^{-\epsilon_b}\frac{\epsilon_b^{B_j}}{B_j!}e^{-\epsilon_b}\frac{\epsilon_s^{S_j}}{S_j!}, \end{aligned} \tag{1}$$

where $\theta = (\alpha, \delta, \mu, \epsilon_b, \epsilon_s)$ is a vector of the parameters defined above. Assuming that trading days are independent, we can express the joint likelihood of observing a series of daily buys and sells over trading days $j = 1, 2, \ldots, J$ as the product of the daily likelihoods:

$$L(M \mid \theta) = \prod_{j=1}^{J} L(\theta \mid B_j, S_j), \tag{2}$$

where $M = ((B_1, S_1), \ldots, (B_J, S_J))$ is the data set. The parameter vector θ can be estimated by maximizing the joint likelihood defined in Equation (2). *PIN*, the probability that a trade will come from an informed trader, is defined by

$$PIN = \frac{\alpha\mu}{\alpha\mu + \epsilon_b + \epsilon_s}. \tag{3}$$

PIN does not distinguish between informed buying on good news and informed selling on bad news. Therefore, to analyze the behavior of investors trading on the two types of private information and its consequences for required returns and costs of capital, we decompose *PIN* into the probabilities of informed trading based on good news and informed trading based on bad news as follows:

$$PIN = \frac{\alpha\mu}{\alpha\mu + \epsilon_b + \epsilon_s} \equiv PIN\_G + PIN\_B, \tag{4}$$

where

$$PIN\_G \equiv \frac{\alpha\mu(1-\delta)}{\alpha\mu + \epsilon_b + \epsilon_s} \tag{5}$$

and

$$PIN\_B \equiv \frac{\alpha\mu\delta}{\alpha\mu + \epsilon_b + \epsilon_s}. \tag{6}$$

In the above equations, $\alpha\delta$ is the probability of a bad-news event and $\alpha(1-\delta)$ is the probability of a good-news event, so that *PIN_G* is the probability of informed trading (buying) based on good news, and *PIN_B* is the probability of informed trading (selling) based on bad news. For expositional purposes we write δ, the probability that an information event contains bad news, as *DELTA*.

2.2. Estimation of *PIN*-Related Parameters

Estimation of *PIN_B*, *PIN_G*, and *PIN* requires that each trade be classified as a buy or sell order. To match trades and quotes and then classify each trade as buyer- or seller-initiated using the Lee and Ready (1991) algorithm, we ignored any quote less than five seconds before the trade and retained the first one at least five seconds before the trade for the years from 1983 to 1998. Since it is believed that timing differences in recording trades and quotes have declined dramatically in recent years, we do not impose this five-second-delay rule from 1999 to 2010. Instead, for this period the quote that is closest in time to the transaction, with a time stamp of two seconds or more before the transaction, is retained. The transactions data are then signed as follows. If a trade occurs above (below) the prevailing quote midpoint, it is regarded as buyer-initiated (seller-initiated). To minimize possible signing errors in processing order flows, we discard approximately 5% of the trades that occur exactly at the quote midpoint, following Sadka (2006).

Although the Lee and Ready (1991) algorithm is imperfect, Lee and Radhakrishna (2000) and Odders-White (2000) show that its accuracy in the 1990s is approximately 85%. O'Hara et al. (2014) find that imbalance measures constructed from the NYSE Trades and Automated Quotations (TAQ) database are subject to error, because TAQ excludes odd-lot trades that have increased substantially in the recent high-frequency-trading era. However, they show that *PIN* estimation is not much affected by the classification errors of the algorithm.[1] Chakrabarty et al. (2012) report that errors in the proportions of daily buys and sells due to misclassification by the Lee–Ready algorithm are close to 0, because buy-side errors and sell-side errors offset each other in the daily aggregation. This suggests that the effect of classification errors on the *PIN* parameters is likely to be small since the estimates rely on the daily totals of intraday buy and sell orders.[2] As a robustness check, however, we also report results from data that exclude the years 2007–2010, during which period high-frequency-trading volume is known to be substantial.[3] To be included in our sample, stocks must have at least 40 positive volume days within each quarter.

We estimate the *PIN* parameters (θ) using transaction-level data from the Institute for the Study of Security Markets (ISSM) and the NYSE Trades and Automated Quotations (TAQ) over the 336 months from January 1983 to December 2010. We restrict our attention to NYSE/AMEX-listed stocks because of limitations on the availability of transaction-level data for NASDAQ-listed stocks and because the NASDAQ market has different trading protocols (Atkins and Dyl 1997). Trades and quotes in the ISSM/TAQ database that are out of sequence, recorded before the open or after the close, or involved in errors or corrections are ignored.

Brown et al. (2004) and Yan and Zhang (2012) express concerns about estimating the parameter vector θ by numerical maximization of the likelihood function in Equation (2): in some cases the estimates may depend on the initial values for the search algorithm and not yield the global maximum; in other cases, the estimates may be corner solutions or *PIN* cannot be computed because of numerical "overflow" or "underflow." To reduce these problems, the parameters are estimated using the Yan and Zhang (2012) algorithm as follows: we (i) construct 125 sets of prespecified initial parameter values; (ii) run the maximization procedure for all acceptable sets of initial values and record the solutions; and, (iii) if all solutions are on the boundaries, choose a boundary solution θ that maximizes the objective function and otherwise exclude all boundary solutions and then choose one of the nonboundary solutions that maximizes the objective function. Yan and Zhang (2012) show that the algorithm alleviates the boundary-solution bias by expanding the parameter space effectively. We estimate the five *PIN* parameters in θ on a quarterly basis to allow for the time variation in the parameters and to improve the power of our empirical tests.

Summary statistics (unreported) on the data used to estimate the *PIN* parameters and the components of trading costs reported in Tables 1 and 2 show that the total number of trades classified as buys or sells (by matching them to bid–ask quotes) is 6.77 billion over the 28-year period, excluding trades executed at the quote midpoints. On average there were 20.14 million trades each month, and the highest number of trades in a given month was 127.13 million in August 2007. For each firm there was an average of 6,870.6 trades in a month, excluding trades executed at the quote midpoints, but some firms such as JPMorgan Chase & Co. (JPM) and ExxonMobil Corporation (XOM) have experienced much more intense trading

Table 1 **Descriptive Statistics of the *PIN*-Related Measures for NYSE/AMEX-Listed Stocks**

Variables	Mean	Median	STD	Skewness	Kurtosis	*N*
PIN	0.195	0.121	0.121	1.97	5.97	2,595.52
DELTA	0.386	0.271	0.271	0.51	−0.64	2,595.52
PIN_G	0.117	0.099	0.099	2.45	9.44	2,595.07
PIN_B	0.078	0.089	0.089	2.64	9.80	2,595.16

Notes. This table reports descriptive statistics (mean, median, standard deviation (STD), skewness, and kurtosis) for the *PIN* components. The cross-sectional value for each statistic is calculated each month, and the time-series average of those values is reported. The *PIN* measures are estimated quarterly for each stock based on Easley et al. (2002) using the numbers of buyer-initiated trades and seller-initiated trades each day (after intradaily order flows are processed via the Lee and Ready (1991) algorithm), lagged by one quarter, and then converted into monthly series. To survive in the sample for the *PIN* measures, stocks must have at least 40 positive-volume days within each quarter. The variables are defined as follows. *PIN* is the probability of informed trading defined by $\alpha\mu/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, where α is the probability with which a private information event occurs on a given day, μ is the rate at which orders from informed traders arrive if the information event does occur, ε_b is the rate at which orders from uninformed buyers arrive, and ε_s is the rate at which orders from uninformed sellers arrive. *DELTA* is the probability (δ) with which a private information event, if it occurs on a given day, contains bad news. *PIN_G* is good-news or buy-side *PIN* defined by $\alpha\mu(1-\delta)/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, and *PIN_B* is bad-news or sell-side *PIN* defined by $\alpha\mu\delta/(\alpha\mu+\varepsilon_b+\varepsilon_s)$. *N* is the average number of component stocks used each month. The sample period is the 336 months 1983:01–2010:12 (28 years) for NYSE/AMEX stocks. The average number of component stocks used each month (*N*) is approximately 2,596.

[1] See the paper by O'Hara et al. (2014, §4 and Table 8).

[2] Hwang et al. (2013) show that the Lee–Ready algorithm causes a downward bias in the *PIN* estimates and that the bias is greatest among infrequently traded stocks.

[3] There is evidence that high-frequency trading has substantially decreased since 2011. See Philips (2013).

Table 2 Correlations of *PIN*, *AdjPIN*, and *PSOS* and Measures of Transaction Costs

Measures	*PIN*	$\bar{\varphi}^{GH}$	$\bar{\varphi}^{FV}$	λ^{GH}	λ^{FV}	A^0	*A*	*PSOS*	*AdjPIN*
PIN	1								
$\bar{\varphi}^{GH}$	0.048	1							
$\bar{\varphi}^{FV}$	0.047	0.997	1						
λ^{GH}	0.265	0.003	−0.006	1					
λ^{FV}	0.262	0.003	−0.009	0.993	1				
A^0	0.110	−0.099	−0.098	0.243	0.240	1			
A	0.142	−0.059	−0.059	0.210	0.208	0.492	1		
PSOS	0.260	0.036	0.034	0.228	0.225	0.067	0.096	1	
AdjPIN	0.479	0.051	0.050	0.181	0.179	0.072	0.106	−0.031	1

Notes. This table reports correlations between *PIN* and other illiquidity measures for NYSE/AMEX stocks. The cross-sectional correlation coefficients are first calculated each month, and the time-series averages of these correlations are reported here. The definitions of the variables are as follows. *PIN* is the probability of informed trading defined by $\alpha\mu/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, where α is the probability with which a private information event occurs on a given day, μ is the rate at which orders from informed traders arrive if the information event does occur, ε_b is the rate at which orders from uninformed buyers arrive, and ε_s is the rate at which orders from uninformed sellers arrive. $\bar{\varphi}^{GH}$ is illiquidity caused by the noninformation component (order processing and inventory maintenance), which is estimated based on Glosten and Harris (1988) using intradaily dollar order flows available within each month. $\bar{\varphi}^{FV}$ is illiquidity caused by the noninformation component (order processing and inventory maintenance), which is estimated based on Foster and Viswanathan (1993) using intradaily (unexpected) dollar order flows available within each month. λ^{GH} is illiquidity caused by adverse selection, which is estimated based on Glosten and Harris (1988) using intradaily dollar order flows available within each month. λ^{FV} is illiquidity caused by adverse selection, which is estimated based on Foster and Viswanathan (1993) using intradaily (unexpected) dollar order flows available within each month. A^0 is the original Amihud (2002) measure defined as the monthly average of daily ratios of the absolute return to dollar volume in \$1,000. *A* is the turnover version of the Amihud measure, which is the monthly average of daily ratios of the absolute return to share turnover computed by using the Center for Research in Security Prices (CRSP) daily file. *PSOS* is the probability of symmetric order-flow shock (illiquidity "unrelated to information asymmetry") estimated based on Duarte and Young (2009). *AdjPIN* is the probability of informed trading "purely" related to adverse selection, estimated based on Duarte and Young (2009). $\bar{\varphi}$'s and λ's are multiplied by 10^2 and 10^6, respectively, before computing the correlations. The sample period is the 336 months from 1983 to 2010. The average number of component stocks used each month is 2,723.1.

in recent months: the number of trades in JPMorgan Chase & Co. was 948,233 in March 2009. The total number of firm-month observations is 985,007.

Table 1 reports descriptive statistics on the estimates of *PIN*, *PIN_G*, *PIN_B*, and *DELTA*, which are defined as follows:

PIN is the probability of informed trading defined in Equation (3).

DELTA is the probability (δ in Equation (1)) with which a private information event contains bad news, when the information event occurs. So $(1-\delta)$ is the probability with which a private information event, if it occurs, contains good news.

PIN_G is the probability of informed trading on good news defined in Equation (5).

PIN_B is the probability of informed trading on bad news defined in Equation (6).

For our subsequent analyses, we assume that the *PIN* statistics are the same for each month of the quarter and construct a monthly series from the quarterly estimates.

In Table 1, the cross-sectional value for each of the six statistics is calculated each month, and the time-series average of these values is reported. The average number of stocks included each month is approximately 2,596. The average probability of informed trading on a given day, *PIN*, is 19.5%. The average value of *DELTA* is 38.6%, so that, given an information event, the probability that the event contains good news is 61.4%, which is much higher than the 38.6% probability of the event containing bad news. This difference may reflect costs of short selling as well as the greater difficulty of arguing that short sales are undertaken for other than speculative reasons; this is a concern for investors who trade on inside information and wish to disguise their motives. Taking account of the probability that an information event occurs, the average unconditional probability of informed trading on good news, *PIN_G*, is 11.7%, whereas the average unconditional probability of informed trading on bad news, *PIN_B*, is only 7.8%. The average cross-sectional standard deviations of *PIN_G* and *PIN_B* are 9%–10%.

3. Hypothesis and Test Procedure

3.1. Asset-Pricing Hypothesis

As discussed in the introduction, our basic hypothesis is that it is *PIN_B* that causes the pricing of *PIN*. The hypothesis is based on the conjecture that investors are more concerned about informed trading on bad news than about informed trading on good news. The asymmetry arises because investors holding or taking a long position will be concerned about their ability to liquidate their position at a fair price at short notice. Consequently, they will be more concerned about informed traders with negative information whose trading would depress the sale price than about traders with positive information whose trading would raise the sale price. Thus the price that they will be willing to pay for a security will depend negatively on the probability of informed trading on bad news and will be less strongly related to the probability of informed trading on good news. This leads to our hypothesis, which we state in alternative form as follows.

Hypothesis 1 (H1). *The pricing of PIN and its effect on the cost of equity capital are entirely due to the pricing of PIN_B, the probability of informed trading on bad news; PIN_G, the probability of trading on good news, is not priced.*

3.2. Asset-Pricing Test Procedure

To test this hypothesis, we follow the Brennan et al. (1998) approach, which uses individual stock data. This avoids the data-snooping biases inherent in portfolio-based approaches (Lo and MacKinlay 1990) and eliminates the errors-in-variables bias caused by errors in estimating the loadings on risk factors.[4]

The dependent variable in the estimation is the return adjusted for four factors (F4): the three (Fama and French (1993)) factors (MKT_t, SMB_t, and HML_t) and the Carhart (1997) momentum factor (UMD_t).[5] We estimate the loadings on the four factors in two ways, obtaining two different estimates of the F4-adjusted returns. First, we use the entire sample period, denoting the resulting F4-adjusted excess return by R^{e1}_{jt}. Second, for each month we use rolling estimates of the four factor loadings (β_j) that are estimated from the time-series data over the past 60 (at least 24) months; this method allows for the time variation in the factor loadings. We denote the corresponding estimate of the F4-adjusted return by R^{e2}_{jt}. Then, letting R^{eh}_{jt}, $h=\{1,2\}$, denote the two measures of risk-adjusted returns, we can write the F4-adjusted returns as

$$R^{eh}_{jt} = R_{jt} - R_{Ft} - (\hat{\beta}_{j1} MKT_t + \hat{\beta}_{j2} SMB_t + \hat{\beta}_{j3} HML_t + \hat{\beta}_{j4} UMD_t), \qquad (7)$$

where R_{Ft} denotes the risk-free rate. Each of the two sets of F4-adjusted returns is then used as the dependent variable in Fama and MacBeth (1973) cross-sectional regressions,

$$R^{eh}_{jt} = c_0 + \sum_{i=1}^{M} \psi_i \Lambda_{ijt-l} + \sum_{n=1}^{N} c_n Z_{njt-l} + e_{jt}, \qquad (8)$$

where Λ_{ijt} $(i=1,2,\ldots,M)$ are *PIN* and its components for stock j in month t; Z_{njt} $(n=1,\ldots,N)$ are control variables for firm j in month t; and l is the number of lags used for the explanatory variables. For comparison purposes, we also use the risk-unadjusted excess return (the monthly stock return less the one-month T-bill rate, denoted by R^{e0}_{jt}) as the dependent variable. We control for firm characteristics such as firm size (*SIZE*), book-to-market equity (*BTM*), and past returns (*RET*01-*RET*0712) in the regression since, according to Avramov and Chordia (2006), a constant-beta version of the Fama and French (1993) three-factor model is not able to capture adequately the predictive ability of firm characteristics in stock returns.

To avoid a possible look-ahead bias, the *PIN* measures (Λ_{ijt-l}) are lagged by one quarter before we convert them to monthly series by filling the intervening months each quarter with the most recent quarterly estimates. Thus, for example, the return regressions for April, May, and June all use the *PIN* estimates over the quarter spanning January to March, and so on.[6] The control variables and their lags are discussed in the next section.

Following Fama and MacBeth (1973), we estimate the vector of coefficients $c = [c_0, \psi_1, \psi_2, \ldots, \psi_M, c_1, c_2, \ldots, c_N]'$ in Equation (8) each month by ordinary least-squares (OLS) regressions, and the final estimator is the time-series average of the monthly coefficients. The standard error of this estimator is taken from the time series of the monthly coefficient estimates.

3.3. Data and Control Variables

The factor-adjusted excess returns (R^{e1} and R^{e2}) defined above are estimated using stock returns from the CRSP monthly file and the four factors (*MKT*, *SMB*, *HML*, and *UMD*) from Kenneth French's website.[7] For R^{e1}, stocks must have returns for at least 24 months during the entire sample period. Similarly, for R^{e2}, stocks must have at least 24 monthly returns in the past 60 months, which excludes some illiquid firms from the sample. For example, Table 5 in §5.1 shows that use of R^{e2} reduces the average number of stocks in the sample each month from approximately 2,600 to approximately 2,260. The 336-month sample period is from January 1983 to December 2010.

The control variables used in the asset-pricing regressions are defined as follows.

SIZE is the natural logarithm of the market value of equity (*MV*), which is the stock price times the number of shares outstanding in million dollars (as of month $t-2$).

BTM is the natural logarithm of the book-to-market ratio if it exists and is positive, and 0 otherwise.

[4] Ang et al. (2008) also argue that using individual stocks provides more efficient tests of whether factors are priced.

[5] For details about factor-adjusted returns, see Brennan et al. (1998) and Chordia et al. (2009).

[6] This method is often used when a variable is available only at a lower frequency. See, for example, Chordia et al. (2007) and Asparouhova et al. (2010).

[7] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#HistBenchmarks provides links to these factors.

BMDUM is a dummy variable that is unity if the book-to-market ratio is positive, and 0 if the ratio is missing or negative. These variables follow Pontiff and Woodgate (2008). As in Fama and French (1992), the quarterly book-to-market ratio is lagged by two quarters (assuming a lag of six months before the data are known to investors). We then fill the corresponding three months with the quarterly value to convert the quarterly data into a monthly series.

*RET*01, *RET*0203, *RET*0406, and *RET*0712 are the compounded holding period returns of a stock over the most recent month (month $t-1$), from month $t-2$ to month $t-3$, from month $t-4$ to month $t-6$, and from month $t-7$ to month $t-12$, respectively. These variables together capture the monthly reversal (Jegadeesh 1990), and longer-term momentum (Jegadeesh and Titman 1993) effects.

The book-to-market ratio is constructed using the Compustat quarterly files. R^{e0} and other firm characteristics or related variables (*MV*, *SIZE*, and *RET*01–*RET*0712) are calculated using the CRSP monthly file.

4. *PIN* Statistics as Measures of Informed Trading

Since there is controversy about whether *PIN* really captures informed trading, in this section we assess the validity of *PIN* statistics as measures of informed trading, before addressing the principal issue of how the two components of *PIN* affect the cost of equity capital. First, we examine whether the quarterly estimate of *PIN* is correlated with the adverse-selection component of trading costs or with the illiquidity component of trading costs that is unrelated to information asymmetry. We also estimate *AdjPIN* and *PSOS* of Duarte and Young (2009) at a quarterly frequency and include them in the correlation analyses. We then examine whether the *PIN* statistics reflect informed trading before earnings announcements.

4.1. Correlations of *PIN*, *AdjPIN*, and *PSOS* with Measures of Illiquidity

We start by examining the relation between estimates of *PIN* and estimates of the adverse-selection and noninformation components of transaction costs that are based on Glosten and Harris (1988) and Foster and Viswanathan (1993).

The Glosten and Harris (1988) model may be written as

$$\Delta P_{i,t,m}=\bar{\varphi}^{GH}_{i,m}(S_{i,t,m}-S_{i,t-1,m})+\lambda^{GH}_{i,m}S_{i,t,m}V_{i,t,m}+\xi_{i,t,m}, \tag{9}$$

where $\Delta P_{i,t,m}$ is the change in the price of stock i at intraday trade t in month m; $S_{i,t,m}$ is the sign of the trade ($S=+1$ if the trade is buyer-initiated and $S=-1$ if it is seller-initiated); $V_{i,t,m}$ is the volume of the trade; $S_{i,t,m}V_{i,t,m}$ is signed volume (order flow); and $\xi_{i,t,m}$ is the unobservable error term. $\bar{\varphi}^{GH}_{i,m}$ is the noninformation component of trading costs (i.e., order-processing and inventory-holding cost), and $\lambda^{GH}_{i,m}$ is the adverse-selection component of trading costs.

The Foster and Viswanathan (1993) model is similar except that the (raw) order flow is replaced by the unexpected order flow (τ_t), which is estimated by fitting an auto-regressive process with order 5, AR(5), to the order flow:

$$\Delta P_{i,t,m}=\bar{\varphi}^{FV}_{i,m}(S_{i,t,m}-S_{i,t-1,m})+\lambda^{FV}_{i,m}\tau_{i,t,m}+\xi'_{i,t,m}, \tag{10}$$

where $\bar{\varphi}^{FV}_{i,m}$ is again the noninformation cost and $\lambda^{FV}_{i,m}$ is the adverse-selection cost.

The parameters $\bar{\varphi}^{GH}$, λ^{GH}, $\bar{\varphi}^{FV}$, and λ^{FV} are estimated each month for each stock by time-series regressions (9) and (10) using all the intradaily order flows available within the month. The order flow is calculated by combining the signed transaction data described above with the size of each transaction taken from ISSM and TAQ. For details on estimating the two components of trading costs (λ and $\bar{\varphi}$), see Huh (2014) and Chung and Huh (2015), who document that the adverse-selection component (λ^{GH} and λ^{FV}) is strongly positively priced in the cross section of stock returns.

Duarte and Young (2009) decompose *PIN* into a component that is intended to capture pure information asymmetry, which they denote by *AdjPIN*, and a component that is therefore unrelated to information asymmetry, which they denote by *PSOS*. They argue that the pricing of *PIN* is solely attributable to the noninformation-related component, *PSOS*, suggesting that information asymmetry is irrelevant in asset pricing and that *PIN* fails to measure informed trading. To further understand the issues surrounding the *PIN* measure, we estimate *AdjPIN* and *PSOS* on a quarterly basis, following the method used in Hwang et al. (2013),[8] who use model 1 described in panel A of Table 3 in Duarte and Young (2009). The two measures are defined as

$$AdjPIN\equiv\frac{\alpha\mu}{\alpha\mu+2(1-\alpha)\omega\Delta+\epsilon_b+\epsilon_s} \tag{11}$$

and

$$PSOS\equiv\frac{2(1-\alpha)\omega\Delta}{\alpha\mu+2(1-\alpha)\omega\Delta+\epsilon_b+\epsilon_s}, \tag{12}$$

where ω is the probability with which there is a symmetric order-flow shock when no private information event has occurred, and Δ is the additional arrival rate of buy or sell orders in the event of a symmetric order-flow shock; other parameters are already defined in §2.

[8] We gratefully appreciate the invaluable assistance of Lee-Seok Hwang and Woo-Jong Lee in estimating *AdjPIN* and *PSOS*.

Table 2 reports time-series averages of monthly cross-sectional correlations of *PIN*, *AdjPIN*, and *PSOS* with the cost components (λ and $\bar{\varphi}$) from the two different models, as well as with the original Amihud (2002) measure (denoted by A^0) and its turnover version (denoted by A).[9] The estimate of *PIN* is significantly correlated (0.26–0.27) with the two measures of illiquidity that are associated with adverse selection, λ^{GH} and λ^{FV}, and it is only weakly correlated (<0.05) with the noninformation component estimates, $\bar{\varphi}^{GH}$ and $\bar{\varphi}^{FV}$. In contrast, the estimate of *AdjPIN* has significantly lower correlations (approximately 0.18) with the estimated adverse-selection components, λ^{GH} and λ^{FV}, than does *PIN* (0.26–0.27), and its correlations with the noninformation component estimates, $\bar{\varphi}^{GH}$ and $\bar{\varphi}^{FV}$, are slightly higher than those of *PIN*. This suggests that the estimate of *PIN* is a better measure of informed trading than the estimate of *AdjPIN*.

Moreover, although *PSOS* is intended to capture the noninformation-related component of illiquidity, it is significantly correlated (approximately 0.23) with the adverse-selection components, λ^{GH} and λ^{FV}, and is scarcely correlated (0.03–0.04) with the noninformation components, $\bar{\varphi}^{GH}$ and $\bar{\varphi}^{FV}$). The estimate of *PSOS* thus fails to isolate the component of illiquidity that is unrelated to adverse selection from *PIN*, which is not consistent with Duarte and Young (2009). In addition, *PIN* has higher correlations with the Amihud measure than does *AdjPIN*, whereas *PSOS* and *AdjPIN* have very similar correlations with the Amihud measure.[10]

Thus, at least when the *PIN*-related parameters are estimated at a quarterly frequency, there is no reason to believe that *AdjPIN* does a better job in capturing the intensity of informed trading than does *PIN*. As noted above, Duarte and Young (2009) find no evidence that *AdjPIN* is priced but do find that *PSOS* is priced. Therefore, in examining the effect of informed trading on the cost of equity capital in §5, we shall concentrate on *PIN* and its good- and bad-news components.

4.2. Informed Trading Around Earnings Announcements

Further evidence on the effectiveness of *PIN* as a measure of informed trading is provided by its behavior around earnings announcements. Rendleman et al. (1982) show that positive earnings announcements are preceded by positive returns and vice versa, suggesting that informed trading takes place before the announcement. To the extent that the intensity of informed trading depends on the magnitude of the earnings surprise and the *PIN* variables reflect this intensity, we should expect that *PIN* estimated over a given quarter would be higher for those firms that have the highest earnings surprises (positive or negative), and that *PIN_G* (*PIN_B*) would be increasing (decreasing) in the earnings surprise.

We investigate this using two earnings-surprise proxies. The first is the cumulative abnormal stock return (*CAR*) around the earnings announcement. The daily abnormal return is computed as the difference between the stock return and the CRSP value-weighted average return. $CAR(-1, +1)$ is the cumulative abnormal return over three trading days (days -1, 0, and $+1$) around the quarterly earnings announcement date (EAD).[11] Quarterly EAD data are obtained from the CRSP/Compustat Merged (CCM) file and I/B/E/S, and daily individual stock returns and CRSP value-weighted returns are from the CRSP daily stock file.

The second surprise proxy is the quarterly standardized unexpected earnings (*SUE*), which is based on the difference between actual and forecast earnings.[12] Following Livnat and Mendenhall (2006), we compute *SUE* by $SUE_{it} = (EPS_{it} - EPS_{it-4})/P_{it}$, where EPS_{it} is the "street" earnings per share for firm i in quarter t that excludes special items from the Compustat-reported EPS (as in Abarbanell and Lehavy (2007)), P_{it} is the stock price at the end of quarter t, and EPS_{it-4} is the EPS at the end of quarter $t-4$ (adjusted for stock splits and stock dividends). *SUE* is thus computed by assuming, for forecasting purposes, that EPS follows a seasonal random walk. The advantage of this *SUE* definition is that quarterly earning surprises can be estimated for almost all firms, unlike other *SUE* definitions that require analysts' forecasts.[13]

Each quarter, firms are assigned to 1 of 10 deciles by sorting on each of the two earnings-surprise proxies (*CAR* or *SUE*), and the average characteristics of each decile are calculated. These are then averaged over the sample period from the first quarter of 1983 to the last quarter of 2010 and the results are reported

[9] The turnover-version measure, A, is defined as the monthly average of daily ratios of the absolute return to share turnover. Brennan et al. (2013) and Lou and Shu (2014) report that the turnover-version Amihud measure is more significant for asset pricing.

[10] Duarte and Young (2009) conclude that *PIN* is not a useful measure of information asymmetry because *PSOS* is positively priced but *AdjPIN* is not, and *PSOS* is more highly correlated (0.18) with the Amihud (2002) measure than is *AdjPIN* (0.11). (See Table 6 of Duarte and Young 2009.) We address the pricing of *AdjPIN* and *PSOS* in §6.1.

[11] Our results are robust to using different estimation windows in computing the indirect surprise measure, i.e., $CAR(-2, +2)$ and $CAR(-5, +5)$.

[12] We thank Denys Glushkov at Wharton Research Data Services for assistance in programming our estimation of the *SUE* measure.

[13] When the *SUE* definition is based on analysts' forecasts, the average number of component stocks in each decile in each quarter is too small. Therefore, we use the *SUE* measure as defined in Livnat and Mendenhall (2006).

Table 3 Average Values of *PIN* Measures in the 10 Portfolios Formed by Sorting on *CAR* and *SUE*

	Panel A: Formed on $CAR(-1,+1)$											
	Low									High	(High − Low)	
Variable	CAR1	CAR2	CAR3	CAR4	CAR5	CAR6	CAR7	CAR8	CAR9	CAR10	Value	(*t*-stat.)
$CAR(-1,+1)$	−0.1236	−0.0516	−0.0296	−0.0158	−0.0049	0.0054	0.0166	0.0310	0.0538	0.1336	0.2572	(*39.61*)
PIN	0.1970	0.1861	0.1833	0.1814	0.1827	0.1800	0.1800	0.1787	0.1800	0.1962	−0.0009	(*−1.01*)
PIN_G	0.1230	0.1143	0.1120	0.1097	0.1096	0.1093	0.1095	0.1100	0.1154	0.1313	0.0083	(*6.71*)
PIN_B	0.0741	0.0719	0.0712	0.0717	0.0730	0.0707	0.0706	0.0687	0.0647	0.0649	−0.0092	(*−9.50*)
	Panel B: Formed on *SUE*											
	Low									High	(High − Low)	
Variable	SUE1	SUE2	SUE3	SUE4	SUE5	SUE6	SUE7	SUE8	SUE9	SUE10	Value	(*t*-stat.)
SUE	−0.3333	−0.0176	−0.0065	−0.0020	0.0004	0.0020	0.0038	0.0069	0.0148	0.2135	0.5469	(*9.04*)
PIN	0.2155	0.1929	0.1801	0.1739	0.1628	0.1614	0.1684	0.1783	0.1901	0.2204	0.0049	(*1.94*)
PIN_G	0.1005	0.1014	0.1090	0.1056	0.1057	0.1162	0.1130	0.1216	0.1296	0.1412	0.0407	(*27.54*)
PIN_B	0.0861	0.0764	0.0788	0.0717	0.0683	0.0684	0.0629	0.0656	0.0621	0.0601	−0.0260	(*−18.03*)

Notes. This table reports the average values of the cumulative abnormal return around the earnings announcement date (*CAR*), the standardized unexpected earnings (*SUE*), *PIN*, *PIN_G*, and *PIN_B* for decile portfolios formed by sorting on *CAR* or *SUE*. In panel A, the component stocks are split into 10 groups (with equal number of stocks) each quarter after being sorted in ascending order by $CAR(-1,+1)$, which is the cumulative abnormal return over three trading days (days −1, 0, +1) around the quarterly earnings announcement date (EAD). Then for each of the portfolios we report the time-series averages of the quarterly cross-sectional means for the four variables. Panel B does the same, but the 10 portfolios are formed by sorting on *SUE*, which is the standardized unexpected earnings, defined by the quarterly earnings surprise computed by assuming, for forecasting purposes, that the earnings per share (EPS) follows a seasonal random walk process (following Abarbanell and Lehavy (2007), special items are excluded from the Compustat-reported EPS). *PIN* is the probability of informed trading defined by $\alpha\mu/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, where α is the probability with which a private information event occurs on a given day, μ is the rate at which orders from informed traders arrive if the information event does occur, ε_b is the rate at which orders from uninformed buyers arrive, and ε_s is the rate at which orders from uninformed sellers arrive. *PIN_G* is the probability of informed trading on good news defined by $\alpha\mu(1-\delta)/(\alpha\mu+\varepsilon_b+\varepsilon_s)$. *PIN_B* is the probability of informed trading on bad news defined by $\alpha\mu\delta/(\alpha\mu+\varepsilon_b+\varepsilon_s)$. Each panel also shows the time-series average values for the differential (*High − Low*) between values of *CAR*, *SUE*, *PIN*, *PIN_G*, and *PIN_B* across the highest *CAR* (in panel A) or *SUE* (in panel B) portfolio (CAR10 or SUE10) and the lowest *CAR* or *SUE* portfolio (CAR1 or SUE1), together with the t-statistics to test the null hypothesis that the time-series average of the differences equals 0. The t-values for the three *PIN*-related measures (*PIN*, *PIN_G*, and *PIN_B*) are computed using logistically transformed values. The average number of component stocks used in each quarter is 2,090.1 (hence 209.01 stocks in each portfolio in each quarter) in ppanel A, and it is 1,795.1 (hence 179.51 stocks in each portfolio in each quarter) in panel B. The sample period is from the first quarter of 1983 to the last quarter of 2010 (1983:Q1–2010:Q4) for NYSE/AMEX stocks.

in Table 3, with the two panels corresponding to the earnings-surprise proxies (*CAR* and *SUE*) used to form the deciles. We report t-tests of the hypothesis that there is no difference in the characteristics of the extreme deciles. The average number of component stocks used in each quarter is 2,090.1 in panel A and 1,795.1 in panel B.

Panels A and B of Table 3 reveal a U-shaped relation between *PIN* and the earnings-surprise proxies, $CAR(-1,+1)$ and *SUE*: the biggest surprises in absolute value are associated with the most intense informed trading as measured by *PIN* estimated in the quarter before the earnings are announced. This is evidence that *PIN* captures the intensity of informed trading before earnings announcements, although *PIN* does not account for the direction of trading.

The bottom two rows in each panel in Table 3 show the average values of *PIN_G* and *PIN_B* across the deciles. In both panels, the decile with the most positive earnings surprise (CAR10 and SUE10) has the highest value of *PIN_G*, and the difference between the average values of *PIN_G* for the most positive and most negative surprise deciles is positive and highly significant:[14] this is consistent with more informed trading on good news in the firms with larger earnings surprises. Similarly, in both panels, the decile with the most negative earnings surprise (CAR1 and SUE1) has the highest value of *PIN_B*, and the (*High − Low*) value of *PIN_B* is negative and again highly significant in each case. Thus, the U-shaped relation between *PIN* and the earnings surprise results from the sum of a positive relation for *PIN_G* and a negative relation for *PIN_B*.

Overall, the behavior of the *PIN* statistics before earnings announcements provides further evidence that these statistics do indeed capture the intensity of informed trading. We turn next to their role in asset pricing.

[14] We transform *PIN* and its components logistically before computing the t-statistic, to accord with the unbounded support of the t-distribution, although results are similar without the transformation.

5. The Effect of *PIN* and Its Components on the Cost of Equity Capital

Bhattacharya et al. (2012) have shown that information asymmetry as measured by *PIN* and the price-impact measure of Huang and Stoll (1997) affects an ex ante measure of the cost of equity capital, and Levi and Zhang (2015) have shown that increases in information asymmetry before earnings announcements as measured by the price impact of trades lead to a higher cost of equity capital. In this section, we investigate the effects of *PIN* and its two components on the risk-adjusted expected return or cost of equity capital.

5.1. Cross-Sectional Regressions

We start by providing confirmatory evidence that *PIN* itself is a priced variable that affects the cost of equity capital, and then we examine whether or not the effect is solely due to *PIN_B*, as our hypothesis predicts. Tables 4 and 5 report monthly Fama and MacBeth (1973) estimates of Equation (8) with and without the control variables, *SIZE*, *BTM*, and *RET*01–*RET*0712. The regression coefficients are multiplied by 100 throughout the asset-pricing regressions. The average number of stocks used each month in the cross-sectional regressions ranges from 2,255.8 to 2,628.6.

Panels A–C in Table 4 show that the coefficient of *PIN* is positive and highly significant (at the 1% level) in all of the regressions, and the point estimate is little affected by how the return is risk adjusted, although using the simple excess return (R^{e0}) instead of the risk-adjusted excess return (R^{e1} or R^{e2}) as the dependent variable increases the standard error

Table 4 The Pricing of *PIN* for NYSE/AMEX Stocks

	Panel A: Dep. var. = R^{e0}		Panel B: Dep. var. = R^{e1}		Panel C: Dep. var. = R^{e2}	
Explanatory variables	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Intercept	0.459	0.496	−0.142*	0.198	−0.130*	0.199
	1.57	*1.33*	*−2.18*	*0.88*	*−1.99*	*0.77*
PIN	1.258**	0.910**	1.342**	0.913**	1.586**	1.107**
	2.86	*2.69*	*4.13*	*3.94*	*4.05*	*3.63*
SIZE		−0.073		−0.077*		−0.080*
		−1.66		*−2.56*		*−2.48*
BTM		0.154**		0.067		0.038
		2.64		*1.59*		*0.84*
BMDUM		0.308**		0.113		0.171
		2.57		*1.21*		*1.64*
*RET*01		−0.508		−0.651		−1.015
		−1.07		*−1.75*		*−1.91*
*RET*0203		0.543		0.619*		0.633
		1.47		*2.12*		*1.61*
*RET*0406		0.805**		0.803**		0.724**
		2.64		*3.56*		*2.67*
*RET*0712		0.726**		0.704**		0.595**
		3.18		*4.07*		*2.65*
Avg R^2	0.003	0.048	0.002	0.028	0.003	0.035
Avg Obs	2,624.8	2,479.3	2,598.4	2,469.5	2,261.2	2,255.8

Notes. This table reports the results of monthly Fama and MacBeth (1973) cross-sectional regressions using *PIN* for NYSE/AMEX stocks over the 336 months 1983:01–2010:12. The dependent variable is R^{e0}, R^{e1}, or R^{e2}. R^{e0} (used in panel A) is the unadjusted excess return (the monthly return less the one-month T-bill rate). R^{e1} and R^{e2} used in panels B and C, respectively, are the four-factor (F4)-adjusted excess returns using the three Fama–French (FF) factors and Carhart's (1997) momentum factor. For R^{e1} the factor loadings are estimated from the entire time series of data, and stocks must have at least 24 returns to be included. For R^{e2} the factor loadings are estimated each month using the past 60 months of returns, and stocks must have at least 24 returns in the 60-month period to be included. The variables are defined as follows. *PIN* is the probability of informed trading defined by $\alpha\mu/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, where α is the probability with which a private information event occurs on a given day, μ is the rate at which orders from informed traders arrive if the information event does occur, ε_b is the rate at which orders from uninformed buyers arrive, and ε_s is the rate at which orders from uninformed sellers arrive. *SIZE* is the natural logarithm of the market value of equity in million dollars. *BTM* is the natural logarithm of the book-to-market ratio (BM) if BM is positive and 0 if BM is negative or missing. *BMDUM* is 1 if BM is positive and 0 if BM is negative or missing. *RET*01 is the holding period return of a stock over the most recent month (month $t-1$). *RET*0203 is the compounded holding period return over the two months (from month $t-2$ to month $t-3$). *RET*0406 is the compounded holding period return over the three months from month $t-4$ to month $t-6$. *RET*0712 is the compounded holding period return over the six months from month $t-7$ to month $t-12$. The values in the first row for each explanatory variable are time-series averages of coefficients obtained from the monthly cross-sectional regressions, and the values italicized in the second row of each variable are t-statistics computed based on Fama and MacBeth (1973). The sample loses the first four observations in the cross-sectional regressions. To avoid a look-ahead bias, *PIN* is lagged by a quarter (equivalent to three months). All coefficients are multiplied by 100. *Avg* R^2 is the average of adjusted R^2. *Avg Obs* is the average number of companies used each month in the cross-sectional regressions. To survive in the sample for *PIN*, stocks should have at least 40 positive-volume days within each quarter.

* and ** indicate statistical significance at the 5% and 1% levels, respectively.

Table 5 The Pricing of *PIN_G* and *PIN_B* for NYSE/AMEX Stocks

Explanatory variables	(i)	(ii)	(iii)	(iv)	(v)	(vi)
			Panel A: Dep. var. $= R^{e0}$			
Intercept	0.678*	0.808*	0.581*	0.626	0.536	0.580
	2.52	*2.50*	*2.01*	*1.81*	*1.83*	*1.55*
PIN_G	0.295	−0.045			0.552	0.233
	0.59	*−0.11*			*1.15*	*0.53*
PIN_B			1.477**	1.104**	1.291*	0.901*
			2.57	*2.74*	*2.33*	*2.02*
SIZE		−0.093*		−0.079		−0.075
		−2.24		*−1.88*		*−1.73*
BTM		0.152**		0.152**		0.154**
		2.58		*2.58*		*2.63*
BMDUM		0.277*		0.283*		0.299*
		2.22		*2.37*		*2.53*
*RET*01		−0.438		−0.473		−0.430
		−0.94		*−1.00*		*−0.92*
*RET*0203		0.598		0.571		0.615
		1.63		*1.54*		*1.70*
*RET*0406		0.870**		0.827**		0.869**
		2.82		*2.69*		*2.85*
*RET*0712		0.747**		0.731**		0.744**
		3.24		*3.21*		*3.29*
Avg R^2	0.002	0.048	0.003	0.048	0.005	0.049
Avg Obs	2,627.6	2,482.1	2,628.6	2,483.0	2,627.1	2,481.5
			Panel B: Dep. var. $= R^{e1}$			
Intercept	0.089	0.503*	−0.021	0.295	−0.069	0.271
	1.49	*2.35*	*−0.36*	*1.36*	*−1.10*	*1.20*
PIN_G	0.228	−0.119			0.505	0.182
	0.64	*−0.45*			*1.48*	*0.65*
PIN_B			1.697**	1.190**	1.588**	1.032**
			4.40	*4.37*	*4.40*	*3.47*
SIZE		−0.096**		−0.079**		−0.078**
		−3.35		*−2.73*		*−2.62*
BTM		0.068		0.068		0.067
		1.59		*1.58*		*1.57*
BMDUM		0.089		0.091		0.105
		0.94		*0.98*		*1.12*
*RET*01		−0.571		−0.618		−0.577
		−1.55		*−1.65*		*−1.56*
*RET*0203		0.667*		0.644*		0.683*
		2.32		*2.21*		*2.39*
*RET*0406		0.857**		0.821**		0.856**
		3.79		*3.62*		*3.79*
*RET*0712		0.713**		0.701**		0.714**
		4.11		*4.05*		*4.16*
Avg R^2	0.001	0.027	0.002	0.027	0.002	0.028
Avg Obs	2,601.2	2,472.3	2,602.2	2,473.2	2,600.7	2,471.7

of the estimated coefficient by approximately 45%. When the seven control variables are included in the regressions, the point estimate of the coefficient on *PIN* falls by approximately 30%, but the coefficient remains highly significant. When the standard deviation of *PIN* reported in Table 1 is used, the estimated coefficient in specification (vi), for example, implies that a one-standard-deviation increase in *PIN* will increase the expected return or cost of equity capital by ($1.107 \times 0.121 =$) 0.13% per month or 1.61% per year. This is economically significant, considering that the average monthly excess return in our sample (unreported) is 0.7%.

Consistent with the prior literature, the size effect is negative and statistically significant when the return is F4-adjusted (R^{e1} and R^{e2}). The coefficients on *BTM*

Table 5 (Continued)

Explanatory variables	(i)	(ii)	(iii)	(iv)	(v)	(vi)
			Panel C: Dep. var. $= R^{e2}$			
Intercept	0.141*	0.635**	−0.007	0.308	−0.045	0.279
	2.33	*2.68*	*−0.12*	*1.27*	*−0.71*	*1.09*
PIN_G	0.261	−0.197			0.499	0.207
	0.61	*−0.69*			*1.25*	*0.65*
PIN_B			2.180**	1.539**	1.997**	1.350**
			4.48	*4.31*	*4.41*	*3.42*
SIZE		−0.113**		−0.084**		−0.082*
		−3.65		*−2.70*		*−2.56*
BTM		0.036		0.038		0.037
		0.77		*0.81*		*0.81*
BMDUM		0.135		0.151		0.169
		1.26		*1.45*		*1.63*
*RET*01		−0.918		−0.967		−0.924
		−1.74		*−1.82*		*−1.75*
*RET*0203		0.701		0.662		0.711
		1.82		*1.69*		*1.85*
*RET*0406		0.808**		0.747**		0.797**
		2.99		*2.74*		*2.95*
*RET*0712		0.613**		0.590**		0.610**
		2.74		*2.62*		*2.74*
Avg R^2	0.001	0.034	0.002	0.034	0.003	0.035
Avg Obs	2,263.9	2,258.4	2,264.7	2,259.2	2,263.3	2,257.9

Notes. This table reports the results of the monthly Fama and MacBeth (1973) cross-sectional regressions using *PIN_G* and *PIN_B* for NYSE/AMEX stocks over the 336 months 1983:01–2010:12. The dependent variable is R^{e0}, R^{e1}, or R^{e2}. R^{e0}, used in panel A, is the unadjusted excess return (the monthly return less the one-month T-bill rate). R^{e1} and R^{e2} used in panels B and C, respectively, are the four-factor (F4)-adjusted excess returns using the three Fama–French (FF) factors and Carhart's (1997) momentum factor. For R^{e1} the factor loadings are estimated from the entire time series of data, and stocks must have at least 24 returns to be included. For R^{e2} the factor loadings are estimated each month using the past 60 months of returns, and stocks must have at least 24 returns in the past 60-month period. The variables are defined as follows. *PIN_G* is the probability of informed trading on good news defined by $\alpha\mu(1-\delta)/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, where α is the probability with which a private information event occurs on a given day, μ is the rate at which orders from informed traders arrive if the information event does occur, ε_b is the rate at which orders from uninformed buyers arrive, and ε_s is the rate at which orders from uninformed sellers arrive. *PIN_B* is the probability of informed trading on bad news defined by $\alpha\mu\delta/(\alpha\mu+\varepsilon_b+\varepsilon_s)$. *SIZE* is the natural logarithm of the market value of equity in million dollars. *BTM* is the natural logarithm of the book-to-market ratio (BM) if BM is positive, and 0 if BM is negative or missing. *BMDUM* is 1 if BM is positive, and 0 if BM is negative or missing. *RET*01 is the holding period return of a stock over the most recent month (month $t-1$). *RET*0203 is the compounded holding period return over the most recent two months (from month $t-2$ to month $t-3$). *RET*0406 is the compounded holding period return over the three months from month $t-4$ to month $t-6$. *RET*0712: is the compounded holding period return over the six months from month $t-7$ to month $t-12$. The values in the first row for each explanatory variable are time-series averages of coefficients obtained from the monthly cross-sectional regressions, and the values italicized in the second row of each variable are t-statistics computed based on Fama and MacBeth (1973). The sample loses the first four observations in the cross-sectional regressions. *PIN_G* and *PIN_B* are lagged by a quarter (equivalent to three months). All coefficients are multiplied by 100. *Avg* R^2 is the average of adjusted R^2. *Avg Obs* is the average number of companies used each month in the cross-sectional regressions. To survive in the sample for the two *PIN* components, stocks should have at least 40 positive-volume days within each quarter.

* and ** indicate statistical significance at the 5% and 1% levels, respectively.

and *BMDUM* are positive but are significant only if the return is not F4 adjusted (R^{e0}). The last two momentum variables (*RET*0406 and *RET*0712) are strongly positively related to the current month return, consistent with Jegadeesh and Titman (1993). We find that the return tends to reverse in the short run (at a one-month horizon), although the coefficient on *RET*01 is not significant.

To determine whether the effect of *PIN* on the cost of equity capital is due solely to the effect of *PIN_B*, Equation (8) is now reestimated with *PIN_G* and *PIN_B* as independent variables. The results reported in Table 5 (panel A for R^{e0}, panel B for R^{e1}, and panel C for R^{e2}) are striking. Regressions (i) and (ii) of panels A–C show no evidence that *PIN_G* is priced; in fact, the point estimates of the coefficient on *PIN_G* are negative when the control variables are included in the regression. In contrast, regressions (iii) and (iv) in all three panels provide strong evidence that *PIN_B* commands a positive return premium and increases the cost of equity capital. The point estimates of the coefficient on *PIN_B* in regressions (iii) and (iv) of panel C, for example, are at least 36% larger than that on *PIN* itself reported in the corresponding regressions, (v) and (vi), of Table 4; additionally, the t-statistic for the coefficient on *PIN_B* is in excess of 3.4 across all the specifications that use the F4-adjusted returns in panels B and C.[15]

When both components of *PIN* are included in specifications (v) and (vi) of panels A–C of Table 5, the point estimates of the coefficient on *PIN_B* decrease slightly but remain significant in all cases, whereas the estimated coefficient of *PIN_G* remains small and insignificant: in specification (v), the coefficient of *PIN_G* is only approximately 19%–35% as large as the coefficient of *PIN_B* in the three panels.

To summarize, the results reveal a strong asymmetry between the effects of *PIN_B* and *PIN_G* on returns, with only *PIN_B* significantly affecting the cost of equity capital. Thus, we cannot reject H1, which posits that the pricing of *PIN* is entirely due to the pricing of *PIN_B*, the probability of informed trading on bad news, and that *PIN_G*, the probability of trading on good news, is not priced. The point estimate of the coefficient on *PIN_B* in specification (iv) of panel C of Table 5 implies that a one-standard-deviation increase in *PIN_B* will increase the expected return or cost of equity capital by $(1.539 \times 0.089 =)$ 0.14% per month or 1.64% per year, which is consistent with the entire effect of *PIN* being attributable to *PIN_B*.

Although *PIN* and its components capture the probability of informed trading, they do not measure the adverse-selection costs incurred by uninformed investors trading with informed investors, and it is these costs that we expect to determine the cost of equity capital. Arbel and Strebel (1983) and Arbel et al. (1983) argue that small firms are "neglected" by many investors; this may be because of the high adverse-selection costs that investors face in trading such firms. Brennan and Wang (2010) show that larger firms tend to be more efficiently priced. Further, the analysis of Hong et al. (2000) suggests that adverse news diffuses more slowly across small firms, on which public information is not readily available. For these reasons, we expect the adverse-selection costs to be greater for smaller firms for a given level of *PIN*, and therefore the probability of bad-news-informed trading to have a larger effect on equilibrium returns for small firms.

To investigate this issue, we interact *PIN_B* with *SIZE* in the regressions and report the results of regressions that use the F4-adjusted returns in Table 6. Including the interaction term increases the estimated coefficient of *PIN_B* itself by a factor of three and makes the coefficient on *SIZE* insignificant, suggesting that the residual size effect after allowing for the four-factor risks may reflect the costs of trading with informed investors. More strikingly, the coefficient of the interaction term (*PIN_B* × *SIZE*) is negative and highly significant, which implies that the return premium for a given level of *PIN_B* is decreasing in firm size, as the above arguments imply.

[15] Harvey et al. (2013) caution against overemphasizing variables that are only marginally significant in asset-pricing tests. When the return is F4-adjusted, the *t*-statistics on *PIN* and *PIN_B* in Tables 4 and 5 are well above their recommended critical threshold of 3.

Table 6 The Pricing of *PIN_B* and Its Interaction with *SIZE*

Explanatory variables	Panel A: Dep. var. $= R^{e1}$ (vii)	(viii)	Panel B: Dep. var. $= R^{e2}$ (ix)	(x)
Intercept	0.137	0.126	0.231	0.113
	0.66	*0.61*	*0.99*	*0.48*
PIN_B	4.269**	4.351**	5.406**	5.253**
	4.33	*4.49*	*4.42*	*4.62*
SIZE	−0.013	−0.039	−0.022	−0.037
	−0.45	*−1.43*	*−0.70*	*−1.25*
PIN_B × *SIZE*	−0.751**	−0.754**	−0.948**	−0.882**
	−3.57	*−3.81*	*−3.71*	*−3.93*
BTM		0.073		0.046
		1.72		*0.99*
BMDUM		0.077		0.132
		0.83		*1.26*
RET01		−0.617		−0.966
		−1.65		*−1.82*
RET0203		0.671*		0.693
		2.32		*1.78*
RET0406		0.860**		0.790**
		3.81		*2.90*
RET0712		0.714**		0.609**
		4.15		*2.72*
Avg R^2	0.008	0.028	0.009	0.036
Avg Obs	2,599.1	2,473.2	2,263.8	2,259.2

Notes. This table reports the results of the monthly Fama and MacBeth (1973) cross-sectional regressions that include a term interacting between *PIN_B* and *SIZE* for NYSE/AMEX stocks over the 336 months 1983:01–2010:12. The dependent variable is R^{e1} in panel A and R^{e2} in panel B. For R^{e1} the factor loadings are estimated from the entire time series of data, and stocks must have at least 24 returns to be included. For R^{e2} the factor loadings are estimated each month using the past 60 months of returns, and stocks must have at least 24 returns in the 60-month period to be included. The variables are defined as follows. *PIN_B* is the the probability of informed trading on bad news defined by $\alpha\mu\delta/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, where α is the probability with which a private information event occurs on a given day, μ is the rate at which orders from informed traders arrive if the information event does occur, ε_b is the rate at which orders from uninformed buyers arrive, and ε_s is the rate at which orders from uninformed sellers arrive. *SIZE* is the natural logarithm of the market value of equity in million dollars. *PIN_B* × *SIZE* is *PIN_B* times *SIZE*. Other variables are defined in Table 4. The values in the first row for each explanatory variable are time-series averages of coefficients obtained from the monthly cross-sectional regressions, and the values italicized in the second row of each variable are *t*-statistics computed based on Fama and MacBeth (1973). The sample loses the first four observations in the cross-sectional regressions. All coefficients are multiplied by 100. *Avg* R^2 is the average of adjusted R^2. *Avg Obs* is the average number of companies used each month in the cross-sectional regressions. To survive in the sample for *PIN_B*, stocks should have at least 40 positive-volume days within each quarter.

* and ** indicate statistical significance at the 5% and 1% levels, respectively.

Table 7 Abnormal Returns (*F4-Alphas*) for Portfolios Formed by Sorting on Firm Size and *PIN_G* or *PIN_B*

	Firm-size (*MV*) group					
	Small	Medium	Big	Small	Medium	Big
PIN_G or *PIN_B* group	Panel A: Sorted on *PIN_G*			Panel B: Sorted on *PIN_B*		
1 Low	0.0054	0.0001	−0.0006	0.0004	−0.0036	−0.0002
	4.00	*0.10*	*−0.75*	*0.28*	*−3.45*	*−0.28*
2	0.0037	0.0006	0.0011	0.0025	−0.0018	−0.0007
	2.21	*0.63*	*1.50*	*1.59*	*−1.83*	*−0.86*
3	0.0029	−0.0004	0.0008	0.0028	0.0002	0.0011
	1.72	*−0.41*	*1.05*	*1.78*	*0.22*	*1.23*
4	0.0015	0.0008	0.0007	0.0046	0.0022	0.0007
	0.96	*0.75*	*0.83*	*2.55*	*2.04*	*0.92*
5 High	0.0031	−0.0009	−0.0001	0.0082	0.0032	0.0011
	1.97	*−0.79*	*−0.12*	*4.54*	*2.75*	*1.24*
High − Low	−0.0022	−0.0010	0.0005	0.0078	0.0068	0.0013
	−1.90	*−0.80*	*0.51*	*5.40*	*5.47*	*1.53*

GRS test for H0: *F4-Alphas* for the three (High − Low) portfolios are jointly 0

Null hypothesis	*F*-statistic	*p*-value	*F*-statistic	*p*-value
H0	*1.04*	*0.38*	*76.21*	*0.00*

Notes. This table reports the abnormal returns for the 15 portfolios sorted on firm size and *PIN_G* or *PIN_B*. The portfolios are formed by sorting the component stocks each month, first on firm size, i.e., market value of equity (*MV*), to split them into three groups (*Small*, *Medium*, and *Big*), and then sorting the stocks in each of the three groups again on *PIN_G* (in panel A) or *PIN_B* (in panel B) to split into five groups, which results in the 15 portfolios. For each portfolio, the abnormal return (denoted by *F4-Alpha*) is obtained as the intercept from a time-series regression of the portfolio excess returns (in excess of the one-month T-bill rate) on the three Fama–French (FF) factors and Carhart's (1997) momentum factor. *PIN_G* is the probability of informed trading on good information defined by $\alpha\mu(1-\delta)/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, where α is the probability with which a private information event occurs on a given day, μ is the rate at which orders from informed traders arrive if the information event does occur, ε_b is the rate at which orders from uninformed buyers arrive, and ε_s is the rate at which orders from uninformed sellers arrive. *PIN_B* is the probability of informed trading on bad information defined by $\alpha\mu\delta/(\alpha\mu+\varepsilon_b+\varepsilon_s)$. In the upper part of each panel, *t*-statistics are italicized. The *F*-statistic (and *p*-value) in the lower part of each panel, which tests the hypothesis that the *F4-Alpha*s for the three (*High* − *Low*) portfolios are jointly 0, is based on Gibbons et al. (1989) (H0 is the null hypothesis). The sample period is the 336 months 1983:01–2010:12 for NYSE/AMEX stocks. The average number of component stocks used in each portfolio in each month is 172.3.

5.2. Portfolio Analyses

To provide further detail on the effect of firm size on the relation between equilibrium expected returns and the *PIN* components, we construct 15 portfolios by sorting first on firm size and then on the *PIN* variable. Each month firms are assigned to three equal-sized portfolios (*Small*, *Medium*, and *Big*) based on the firm size (i.e., *MV*) at the end of the previous month. Firms in each size portfolio are then allocated to five equal-sized portfolios according to their value, first of *PIN_G* and then of *PIN_B*. This results in 15 portfolios based on size and *PIN_G* and 15 portfolios based on size and *PIN_B*. Table 7 reports the abnormal returns on the portfolios (denoted by *F4-Alpha*), which is the intercept from a time-series regression of the portfolio excess return on the three Fama–French (FF) factors and Carhart's (1997) momentum factor (i.e., *MKT*, *SMB*, *HML*, and *UMD*). Panel A reports the results for the portfolios formed on firm size (*MV*) and *PIN_G*, and panel B for the portfolios formed on *MV* and *PIN_B*.

Panel A of Table 7 shows that the *t*-statistic for the *F4-Alpha* is less than 2 except for the two lowest *PIN_G* quintiles within the *Small* size group, where the abnormal return is positive and significant. For both the *Small* and *Medium* size groups, the highest *PIN_G* portfolio actually has a lower abnormal return than does the lowest *PIN_G* portfolio, although the difference is not significant, and the Gibbons et al. (1989) (GRS) test shown in the lower part of panel A does not reject the null hypothesis that the *F4-Alphas* for the three (*High* − *Low*) portfolios are jointly 0.[16] In summary, there is no evidence that *PIN_G* is priced in any size group of stocks, which confirms our regression analyses.

The results in panel B of Table 7, where the portfolios are formed by sorting on *PIN_B*, present a striking contrast. For the *Small* and *Medium* size groups, the abnormal return (*F4-Alpha*) is monotonically increasing in *PIN_B*, and the difference in the abnormal

[16] Let there be M time-series observations, G portfolios, and $L-1$ factors (excluding the intercept). Further, let X denote the matrix of regressors. Then the GRS test statistic is given by $A'^{-1}A((M-L-G+1)/(G(M-L)\omega_{1,1}))$, where A is the column vector of *F4-Alphas*, Σ is the covariance matrix of the residuals from the time-series regressions, and $\omega_{1,1}$ is the diagonal element of $(X'X)^{-1}$ corresponding to the intercept. Under the null hypothesis, this statistic follows an F-distribution with G and $M-L-G+1$ degrees of freedom.

return between the highest and lowest *PIN_B* portfolios is of the order of 0.7%–0.8% per month with a *t*-statistic in excess of 5. For the *Big* firm size group, the abnormal returns on the two lowest *PIN_B* quintiles are negative whereas those on the three highest *PIN_B* quintiles are positive, although the differences are not statistically significant and the point estimate of the difference in the abnormal return between the highest and lowest *PIN_B* quintiles is only 0.1% per month. These results confirm the important role of *PIN_B* in asset pricing and show that the effect is concentrated in the two terciles of smaller firm size.

6. Robustness Tests

In this section, we report the results of a series of robustness tests. For brevity we present only the results from the regressions that use R^{e2} as the dependent variable (except for panel A in Table 8), so that they are directly comparable to those in panel C of Tables 4 and 5.[17]

6.1. With *AdjPIN* and *PSOS*

As mentioned in §4.1, Duarte and Young (2009) cast doubt on the role of *PIN* as a measure of the intensity of informed trading. Although we have shown that *PIN* is related to the adverse-selection component of transaction costs and behaves as we would expect a measure of informed trading to behave around earnings announcements, the claim of Duarte and Young (2009) raises important issues for the interpretation of our results. This claim is based on their finding no evidence that their estimate of *AdjPIN*, the component of *PIN* that is intended to capture pure information asymmetry, has a significant effect on expected returns. They find, on the other hand, that *PSOS*, which is intended to capture the noninformation-related component of *PIN*, does affect returns.

There are two principal differences between the empirical approach of Duarte and Young (2009) and ours: Duarte and Young (2009) use (lagged) annual estimates of *PIN* and *AdjPIN*, and they use raw returns rather than risk-adjusted or excess returns in their regressions. To explore the effect of these differences, we conduct Fama–MacBeth regressions using the raw return (R^0) as in Duarte and Young (2009), as well as the risk-adjusted excess return (R^{e2}), as the dependent variable, together with their (lagged) annual estimates of *PIN*, *AdjPIN*, and *PSOS*, obtained from Jefferson Duarte's website.[18] The monthly cross-sectional regressions using the estimates of Duarte and Young (2009) are run over the same sample period: January 1984 to December 2004. The seven control variables used in the previous tables are included in the regressions, but their coefficients and *t*-values are not reported for expositional convenience.

The results are contained in the first two panels of Table 8: the ones with R^0 in panel A and those with R^{e2} in panel B, where *PIN_DY*, *AdjPIN_DY*, and *PSOS_DY* denote the annual estimates of Duarte and Young (2009). Specifications (i)–(v) in panel A show that when the raw return (R^0) is used as in Duarte and Young (2009), none of the coefficients of *PIN_DY*, *AdjPIN_DY*, and *PSOS_DY* are statistically significant. Specifications (vi) and (vii) in panel B in which the dependent variable is the risk-adjusted return (R^{e2}) are consistent with the finding of Duarte and Young (2009) that the coefficient on *PSOS_DY* is significant whereas the coefficient of *AdjPIN_DY* is not significant. However, in specification (x) in which *PIN_DY* is included in the regression along with *PSOS_DY*, the coefficient of *PIN_DY* is significant but the coefficient of *PSOS_DY* is no longer significant. Moreover, in specification (viii) in which *PIN_DY* is the only informed-trading variable in the regression, the *t*-statistic for its coefficient is 3.45, whereas the *t*-statistic for the coefficient on *PSOS_DY* in the corresponding specification (i.e., regression (vi)) is only 2.98. Overall, we conclude from this analysis that *PSOS_DY* has no marginal explanatory power for the risk-adjusted return once the effect of *PIN_DY* is taken into account.

To determine whether *AdjPIN* and *PSOS* affect our finding that it is *PIN_B* that accounts for the pricing of *PIN*, regressions (xi)–(xiii) in panel C of Table 8 include our quarterly estimates (used in Table 2) of the Duarte and Young (2009) variables in the regressions, together with *PIN_B* and *PIN_G*. Specification (xi) shows that when *PIN_G* is included along with the quarterly estimates of the Duarte and Young (2009) variables (*AdjPIN* and *PSOS*), the coefficient on *PIN_G* continues to be insignificant as in Table 5, whereas the coefficient of *AdjPIN* is significant and the coefficient on *PSOS* is not. In specification (xii), *PIN_G* is replaced by *PIN_B*. The coefficient of *PIN_B* is highly significant as in Table 5, but neither of the variables in Duarte and Young (2009)(*AdjPIN* and *PSOS*) plays a significant role. Similar results are found in regression (xiii), which includes both *PIN_G* and *PIN_B*, together with the two variables of Duarte and Young (2009). The quarterly estimate of *PSOS* is never priced in panel C,[19] and our results are robust to the inclusion of *AdjPIN* and *PSOS*.

[17] The results using R^{e1} are qualitatively similar. Full results are available upon request.

[18] The annual estimates used in Duarte and Young (2009) are available for the 1983–2004 period at http://www.owlnet.rice.edu/~jd10/publications.htm.

[19] Lai et al. (2014) report that *PSOS* is negatively priced.

Table 8 *AdjPIN*, *PSOS*, **and the Pricing of** *PIN_G* **and** *PIN_B*

	With the annual measures of Duarte and Young (2009) from Duarte's website (1984–2004)										With quarterly measures (1983–2010)		
	Panel A: Dep. var. $= R^0$					Panel B: Dep. var. $= R^{e2}$					Panel C: Dep. var. $= R^{e2}$		
Exp. var.	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)	(xi)	(xii)	(xiii)
Intercept	0.823	0.646	0.721	0.889	0.872	−0.124	−0.429	−0.267	−0.029	−0.199	0.449	0.295	0.282
	1.36	*0.91*	*1.05*	*1.35*	*1.29*	*−0.30*	*−0.89*	*−0.57*	*−0.06*	*−0.47*	*1.73*	*1.07*	*1.01*
PIN_DY			0.765		0.622			1.196**		0.942*			
			1.77		*1.20*			*3.45*		*2.18*			
AdjPIN_DY		0.563		0.558			0.922		0.922				
		1.10		*1.09*			*1.89*		*1.91*				
PSOS_DY	0.405	0.413			0.282	0.680**	0.683**			0.365			
	1.63	*1.65*			*0.95*	*2.98*	*2.96*			*1.25*			
PIN_G											−0.413		0.062
											−1.46		*0.21*
PIN_B												1.447**	1.205**
												4.47	*3.35*
AdjPIN											0.836*	0.329	0.400
											2.32	*1.01*	*1.24*
PSOS											−0.090	−0.257	−0.212
											−0.47	*−1.39*	*−1.18*
Avg R^2	0.048	0.049	0.048	0.048	0.043	0.042	0.043	0.043	0.043	0.036	0.035	0.035	0.035
Avg Obs	1,931.5	1,931.5	1,931.5	1,931.5	1,924.7	1,773.9	1,773.9	1,773.9	1,773.9	1,773.6	2,257.8	2,258.6	2,257.2

Notes. This table reports the results of the monthly Fama and MacBeth (1973) cross-sectional regressions using the *PIN* and its components estimated at both annual and quarterly frequencies for NYSE/AMEX stocks. Panels A and B contains the results from regressions (over the 252 months from January 1984 to December 2004) that use *PIN_DY*, *AdjPIN_DY*, and *PSOS_DY* estimated on an annual basis and used in Duarte and Young (2009) (obtained from Jefferson Duarte's website). Panel C contains the results from regressions (over the period from May 1983 to December 2010) that use *AdjPIN* and *PSOS*, together with *PIN_G* and *PIN_B*, which are all estimated on a quarterly basis. The dependent variable is R^0, the (raw) monthly returns, in panel A, and R^{e2}, the four-factor (F4)-adjusted excess returns using the three Fama–French (FF) factors and Carhart's (1997) momentum factor, in panels B and C. For R^{e2} the factor loadings are estimated each month using the past 60 months of returns, and stocks must have at least 24 returns in the past 60-month period. The variables are defined as follows. *PIN_DY*, *AdjPIN_DY*, and *PSOS_DY* are the annual estimates of *PIN*, *AdjPIN*, and *PSOS* that are used in Duarte and Young (2009) (taken from the website of Jefferson Duarte and then lagged by one year before converting them to monthly series, following Duarte and Young (2009)). *PIN_G* is the probability of informed trading on good news defined by $\alpha\mu(1-\delta)/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, where α is the probability with which a private information event occurs on a given day, μ is the rate at which orders from informed traders arrive if the information event does occur, ε_b is the rate at which orders from uninformed buyers arrive, and ε_s is the rate at which orders from uninformed sellers arrive. *PIN_B* is the probability of informed trading on bad news defined by $\alpha\mu\delta/(\alpha\mu+\varepsilon_b+\varepsilon_s)$. *AdjPIN* is the probability of informed trading "purely" related to adverse selection, estimated based on Duarte and Young (2009). *PSOS* is the probability of symmetric order-flow shock (illiquidity "unrelated to information asymmetry") estimated based on Duarte and Young (2009). *SIZE* is the natural logarithm of the market value of equity in million dollars. *BTM* is the natural logarithm of the book-to-market ratio (BM) if BM is positive, and 0 if BM is negative or missing. *BMDUM* is 1 if BM is positive, and 0 if BM is negative or missing. *RET*01 is the holding period return of a stock over the most recent month (month $t-1$). *RET*0203 is the compounded holding period return over the next recent two months (from month $t-2$ to month $t-3$). *RET*0406 is the compounded holding period return over the three months from month $t-4$ to month $t-6$, and *RET*0712 is the compounded holding period return over the six months from month $t-7$ to month $t-12$. The values in the first row for each explanatory variable are time-series averages of coefficients obtained from the monthly cross-sectional regressions, and the values italicized in the second row of each variable are t-statistics computed based on Fama and MacBeth (1973). *PIN_G* and *PIN_B* are lagged by a quarter (equivalent to three months). All coefficients are multiplied by 100. For expositional convenience, the coefficients on the control variables and their t-values are not reported in the table. *Avg* R^2 is the average of adjusted R^2. *Avg Obs* is the average number of companies used each month in the cross-sectional regressions. To survive in the sample for the two *PIN* components (*PIN_B* and *PIN_B*), stocks should have at least 40 positive-volume days within each quarter.

* and ** indicate statistical significance at the 5% and 1% levels, respectively.

6.2. Weighted Least-Squares Regressions

There is a concern that CRSP returns can induce microstructure biases in asset-pricing tests because of the bid–ask bounce. Asparouhova et al. (2010) suggest that weighted least-squares (WLS) regressions that use the prior-month gross return as the weighting variable can reduce this problem. Therefore, we repeat the Fama–MacBeth tests with WLS regressions using the prior-month gross return as a weighting variable.

The results in Table 9 confirm our earlier findings. In specification (i), the coefficient of *PIN* is similar to that of specification (vi) reported in Table 4 and is highly significant. In regression (ii), the coefficient of *PIN_G* is negative and insignificant as it was in the corresponding regression in panel C of Table 5. In regressions (iii) and (iv), the estimate of the coefficient on *PIN_B* is close to that reported in specifications (iv) and (vi) in panel C of Table 5 and continues to be highly significant. Comparing Table 9 with Tables 4 and 5, we also find that the WLS regressions weaken the size effect but increase the momentum effect.

Table 9 Weighted Least-Squares Regressions

	Dep. var. $= R^{e2}$			
Explanatory variables	(i)	(ii)	(iii)	(iv)
Intercept	0.000	0.428	0.093	0.077
	0.00	*1.85*	*0.39*	*0.30*
PIN	1.102**			
	3.49			
PIN_G		−0.191		0.201
		−0.64		*0.62*
PIN_B			1.544**	1.363**
			4.23	*3.39*
SIZE	−0.052	−0.084**	−0.054	−0.054
	−1.60	*−2.79*	*−1.75*	*−1.69*
BTM	0.069	0.070	0.070	0.070
	1.51	*1.50*	*1.50*	*1.52*
BMDUM	0.177	0.142	0.160	0.176
	1.68	*1.31*	*1.53*	*1.69*
*RET*01	−0.449	−0.364	−0.394	−0.368
	−0.85	*−0.69*	*−0.74*	*−0.70*
*RET*0203	1.163**	1.219**	1.192**	1.228**
	2.98	*3.20*	*3.08*	*3.22*
*RET*0406	0.973**	1.041**	0.984**	1.029**
	3.53	*3.80*	*3.56*	*3.76*
*RET*0712	0.785**	0.803**	0.784**	0.798**
	3.58	*3.64*	*3.55*	*3.64*
Avg R^2	0.036	0.035	0.035	0.036
Avg Obs	2,255.8	2,258.4	2,259.2	2,257.9

Notes. This table reports the results of monthly Fama and MacBeth (1973) cross-sectional regressions using weighted least-squares (WLS) estimation for NYSE/AMEX stocks over 336 months 1983:01–2010:12. Following Asparouhova et al. (2010), the prior-month gross return (one plus the return of month $t-1$) is used as the weighting variable. The dependent variable is R^{e2}, the F4-adjusted excess return using the four factors with factor loadings estimated from the past 60-month rolling regressions (monthly returns should be available for at least 24 months). The variables are defined as follows. *PIN* is the probability of informed trading defined by $\alpha\mu/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, where α is the probability with which a private information event occurs on a given day, μ is the rate at which orders from informed traders arrive if the information event does occur, ε_b is the rate at which orders from uninformed buyers arrive, and ε_s is the rate at which orders from uninformed sellers arrive. *PIN_G* is the probability of informed trading on good news defined by $\alpha\mu(1-\delta)/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, and *PIN_B* is the probability of informed trading on bad news defined by $\alpha\mu\delta/(\alpha\mu+\varepsilon_b+\varepsilon_s)$. Other variables are defined in Table 4. The values in the first row for each explanatory variable are the time-series averages of coefficients obtained from the monthly cross-sectional regressions, and the values italicized in the second row of each variable are t-statistics computed based on Fama and MacBeth (1973). The sample loses the first four observations in the cross-sectional regressions. All coefficients are multiplied by 100. *Avg* R^2 is the average of adjusted R^2 values. *Avg Obs* is the average number of companies used each month in the cross-sectional regressions. To survive in the sample for the *PIN*-related measures, stocks should have at least 40 positive-volume days within each quarter.

* and ** indicate statistical significance at the 5% and 1% levels, respectively.

6.3. With the Sample from the TAQ Period Only

The accuracy of the ISSM database, which is used to calculate *PIN*, is also a potential source of concern. In the early years of ISSM, the data were entered manually, which could have caused errors. It has also been noted that some of the condition codes for identifying different kinds of trades used by TAQ are not exactly the same as those used by ISSM. To address the issue of potential errors in this database, we now restrict the sample to the 18-year TAQ period, 1993–2010, and examine how the empirical patterns observed in earlier sections are affected. Excluding the 10-year non-TAQ period also allows us to verify that the results are robust to the changes that have taken place in financial markets since the beginning of the 1990s.

The results are reported in panel A of Table 10. Regression (i) shows that when the 10 years of the ISSM period are excluded, the estimated coefficient of *PIN* increases, compared to that in specification (vi) of Table 4, and remains highly significant. In regressions (ii) and (iv), the estimated coefficient of *PIN_G* is insignificant as in panel C of Table 5. In regressions (iii) and (iv), the estimated coefficient of *PIN_B* is positive and highly significant, and is much larger than in Table 5. Thus, our results are robust to the exclusion of the ISSM years.

6.4. Excluding the High-Frequency-Trading Years

To alleviate possible concerns about higher classification error rates from the Lee and Ready (1991) algorithm in the high-frequency-trading era, we report results based on the sample up to 2006 only, excluding the years 2007–2010, during which high-frequency trading became more important. Panel B of Table 10 reports the results, again using R^{e2} as the dependent

Table 10 Different Sample Periods

Panel A: TAQ period (1993:01–2010:12) only, dep. var. = R^{e2}

Explanatory variables	(i)	(ii)	(iii)	(iv)
Intercept	−0.018	0.519	0.098	0.045
	−0.05	*1.71*	*0.32*	*0.13*
PIN	1.395**			
	3.18			
PIN_G		−0.107		0.402
		−0.27		*0.89*
PIN_B			1.899**	1.815**
			3.68	*3.19*
SIZE	−0.048	−0.089*	−0.051	−0.049
	−1.12	*−2.23*	*−1.25*	*−1.13*
BTM	0.068	0.067	0.064	0.067
	1.21	*1.17*	*1.12*	*1.19*
BMDUM	0.175	0.123	0.157	0.172
	1.24	*0.84*	*1.13*	*1.24*
RET01	−0.378	−0.362	−0.339	−0.348
	−0.54	*−0.52*	*−0.48*	*−0.50*
RET0203	0.931	0.951	0.977*	0.966
	1.88	*1.92*	*1.98*	*1.95*
RET0406	0.670	0.672	0.658	0.668
	1.92	*1.93*	*1.88*	*1.91*
RET0712	0.657*	0.656*	0.647*	0.649*
	2.44	*2.42*	*2.39*	*2.41*
Avg R^2	0.038	0.038	0.038	0.038
Avg Obs	2,554.3	2,558.0	2,558.3	2,557.1

Table 10 (Continued)

Panel B: Excluding the high-frequency-trading years (2007:01–2010:12), dep. var. = R^{e2}

Explanatory variables	(i)	(ii)	(iii)	(iv)
Intercept	0.279	0.541*	0.269	0.326
	1.12	*2.42*	*1.15*	*1.35*
PIN	0.551*			
	2.26			
PIN_G		−0.514*		−0.223
		−1.98		*−0.82*
PIN_B			1.173**	0.970**
			4.18	*3.16*
SIZE	−0.079*	−0.098**	−0.076*	−0.081**
	−2.51	*−3.27*	*−2.50*	*−2.60*
BTM	0.089	0.092	0.090	0.088
	1.85	*1.88*	*1.85*	*1.82*
BMDUM	0.206*	0.207	0.210*	0.218*
	1.96	*1.95*	*1.99*	*2.07*
RET01	−0.701	−0.601	−0.633	−0.591
	−1.28	*−1.09*	*−1.15*	*−1.08*
RET0203	1.577**	1.655**	1.612**	1.662**
	4.14	*4.45*	*4.24*	*4.48*
RET0406	1.227**	1.316**	1.252**	1.301**
	4.18	*4.51*	*4.25*	*4.47*
RET0712	0.793**	0.820**	0.801**	0.816**
	3.43	*3.55*	*3.46*	*3.54*
Avg R^2	0.034	0.033	0.034	0.034
Avg Obs	2,272.2	2,273.4	2,274.1	2,273.4

Notes. This table reports the results of the monthly Fama and MacBeth (1973) cross-sectional regressions for NYSE/AMEX stocks over two different sample periods. Panel A uses the sample from the TAQ period only (216 months: 1993:01–2010:12), excluding the ISSM period (1983–1992), whereas panel B uses the sample from the period (1983–2006) that excludes the last four years (2007–2010), during which high-frequency-trading volume was substantial. In each panel the dependent variable is R^{e2}, the F4-adjusted excess return using the four factors with factor loadings estimated from the past 60-month rolling regressions (monthly returns should be available for at least 24 months). The variables are defined as follows. *PIN* is the probability of informed trading defined by $\alpha\mu/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, where α is the probability with which a private information event occurs on a given day, μ is the rate at which orders from informed traders arrive if the information event does occur, ε_b is the rate at which orders from uninformed buyers arrive, and ε_s is the rate at which orders from uninformed sellers arrive. *PIN_G* is the probability of informed trading on good news defined by $\alpha\mu(1-\delta)/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, and *PIN_B* is the probability of informed trading on bad news defined by $\alpha\mu\delta/(\alpha\mu+\varepsilon_b+\varepsilon_s)$. Other variables are defined in Table 4. The values in the first row for each explanatory variable are the time-series averages of coefficients obtained from the month-by-month cross-sectional regressions, and the values italicized in the second row of each variable are t-statistics computed based on Fama and MacBeth (1973). All coefficients are multiplied by 100. *Avg* R^2 is the average of adjusted R^2 values. *Avg Obs* is the average number of companies used each month in the cross-sectional regressions. To survive in the sample for the *PIN*-related measures, stocks should have at least 40 positive-volume days within each quarter.

* and ** indicate statistical significance at the 5% and 1% levels, respectively.

variable. The coefficient of *PIN* continues to be significant, albeit smaller than in Table 4. The coefficient of *PIN_G* is negative and statistically significant in regression (ii), but it becomes insignificant when combined with *PIN_B* in regression (iv). Most importantly, the estimated coefficient of *PIN_B* in regressions (iii) and (iv) continues to be positive and highly significant, albeit again somewhat smaller than in Table 5. We also find that by excluding the four most recent years from the sample, the size, book-to-market, and momentum effects all become stronger, consistent with Chordia et al. (2014).

6.5. Different Trading Regimes

The NYSE and AMEX reduced the minimum tick size from \$1/8 to \$1/16 on June 24, 1997 and then decimalized trading prices in January 2001.[20] Since this change is likely to have affected trading behavior, our concern is whether it also affected the way that our *PIN* measures are priced. The two subperiods that we consider are the \$1/8th era from January 1983 to May 1997, during which the minimum tick size is \$1/8, and the decimal era from February 2001 to December 2010. The intervening period from July 1997 to December 2000 (the \$1/16 regime) is not considered

[20] The AMEX started to apply the \$1/16 rule from 1992 (to stocks priced between \$0.25 and \$5) and extended the rule in 1997 to all stocks trading at or above \$0.25. (For details, see U.S. Securities and Exchange Commission 2012.)

Table 11 Different Trading Regimes

Panel A: \$1/8th era (1983:01–1997:05), dep. var. = R^{e2}				
Explanatory variables	(i)	(ii)	(iii)	(iv)
Intercept	0.071	0.313	0.140	0.137
	0.23	*1.16*	*0.48*	*0.47*
PIN	0.650*			
	2.38			
PIN_G		−0.208		0.023
		−0.71		*0.08*
PIN_B			1.031**	0.829*
			3.45	*2.55*
SIZE	−0.065	−0.081*	−0.068	−0.068
	−1.69	*−2.31*	*−1.84*	*−1.85*
BTM	0.113	0.117	0.117	0.114
	1.83	*1.86*	*1.88*	*1.82*
BMDUM	0.262*	0.264*	0.256*	0.268*
	2.24	*2.24*	*2.18*	*2.28*
RET01	−0.035	0.116	0.044	0.102
	−0.06	*0.19*	*0.07*	*0.17*
RET0203	1.446**	1.535**	1.470**	1.543**
	3.03	*3.38*	*3.08*	*3.40*
RET0406	1.482**	1.623**	1.528**	1.600**
	4.09	*4.55*	*4.20*	*4.49*
RET0712	0.879**	0.919**	0.891**	0.914**
	2.81	*2.94*	*2.83*	*2.93*
Avg R^2	0.029	0.029	0.029	0.029
Avg Obs	1,937.3	1,937.7	1,938.9	1,937.7

in the analysis because it is too short to compute meaningful Fama–MacBeth statistics.

The results are shown in Table 11. In the results for the \$1/8th era (1983:01–1997:05) reported in panel A, we continue to see that *PIN* and *PIN_B* are significantly priced, while *PIN_G* commands a negative and insignificant premium. The results for the decimal era (2001:02–2010:12) reported in panel B show that both *PIN* and *PIN_B* are significantly priced in specifications (i) and (iii), whereas *PIN_G* does not command a significant return premium in specifications (ii) and (iv).[21]

7. Conclusion

In this study, we have provided new evidence of the effects of information-based trading on the cost of equity capital. Our analysis is motivated by the observation that the price that an uninformed investor who is subject to random liquidation shocks is willing to pay depends on the price at which the investor will be able to sell stocks, which in turn depends

[21] We also conduct robustness tests, in which we control for the Amihud measure and its two components. We find that *PIN* and *PIN_B* remain significant at the 1% level across all specifications, whereas *PIN_G* is not significant. These results are available from the authors upon request.

Table 11 (Continued)

Panel B: Decimal era (2001:02–2010:12), dep. var. = R^{e2}				
Explanatory variables	(i)	(ii)	(iii)	(iv)
Intercept	−0.078	0.725	0.162	0.052
	−0.15	*1.60*	*0.35*	*0.10*
PIN	2.202**			
	2.94			
PIN_G		0.118		0.818
		0.18		*1.08*
PIN_B			2.458**	2.341*
			2.87	*2.44*
SIZE	−0.039	−0.100	−0.045	−0.040
	−0.61	*−1.77*	*−0.79*	*−0.64*
BTM	0.042	0.040	0.040	0.046
	0.53	*0.48*	*0.49*	*0.57*
BMDUM	0.154	0.055	0.107	0.137
	0.78	*0.26*	*0.55*	*0.71*
RET01	−0.436	−0.425	−0.421	−0.437
	−0.46	*−0.45*	*−0.45*	*−0.47*
RET0203	0.129	0.150	0.178	0.169
	0.18	*0.21*	*0.25*	*0.23*
RET0406	−0.142	−0.149	−0.164	−0.145
	−0.31	*−0.33*	*−0.36*	*−0.32*
RET0712	0.319	0.316	0.300	0.307
	0.93	*0.90*	*0.86*	*0.89*
Avg R^2	0.041	0.040	0.040	0.041
Avg Obs	2,510.9	2,517.6	2,518.2	2,516.1

Notes. This table reports the results of the monthly Fama and MacBeth (1973) cross-sectional regressions for NYSE/AMEX stocks under two different trading regimes. The first subperiod shown in panel A is the \$1/8th era (1983:01–1997:05), during which the minimum tick size is \$1/8, and the second subperiod shown in panel B is the decimal era (2001:02–2010:12), during which the minimum tick size is \$0.01. In each panel the dependent variable is R^{e2}, the F4-adjusted excess return using the four factors with factor loadings estimated from the past 60-month rolling regressions (monthly returns should be available for at least 24 months). The variables are defined as follows. *PIN* is the probability of informed trading defined by $\alpha\mu/(\alpha\mu + \varepsilon_b + \varepsilon_s)$, where α is the probability with which a private information event occurs on a given day, μ is the rate at which orders from informed traders arrive if the information event does occur, ε_b is the rate at which orders from uninformed buyers arrive, and ε_s is the rate at which orders from uninformed sellers arrive. *PIN_G* is the probability of informed trading on good news defined by $\alpha\mu(1-\delta)/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, and *PIN_B* is the probability of informed trading on bad news defined by $\alpha\mu\delta/(\alpha\mu + \varepsilon_b + \varepsilon_s)$. Other variables are defined in Table 4. The values in the first row for each explanatory variable are the time-series averages of coefficients obtained from the month-by-month cross-sectional regressions, and the values italicized in the second row of each variable are *t*-statistics computed based on Fama and MacBeth (1973). The sample loses the first four observations in the cross-sectional regressions in panel A. All coefficients are multiplied by 100. *Avg* R^2 is the average of adjusted R^2 values. *Avg Obs* is the average number of companies used each month in the cross-sectional regressions. To survive in the sample for the *PIN*-related measures, stocks should have at least 40 positive-volume days within each quarter.

* and ** indicate statistical significance at the 5% and 1% levels, respectively.

on the degree of adverse selection or intensity of informed trading in the sales market for the security. This leads us to decompose the *PIN* measure of Easley et al. (2002) into the probability of trading on good private information (informed purchases), *PIN_G*, and the probability of trading on bad private information (informed sales), *PIN_B*. To capture time variation in the measures, *PIN* and its components are estimated quarterly, in contrast to the annual estimates that have been used in much of the earlier literature.

We validate the *PIN* measures of informed trading by relating them to estimates of the adverse-selection and noninformation components of transaction costs that are based on Glosten and Harris (1988) and Foster and Viswanathan (1993) and contrast them with the measures of informed trading proposed by Duarte and Young (2009). We show that *PIN* is more highly correlated with the adverse-selection component of trading costs than *AdjPIN* estimated quarterly based on Duarte and Young (2009), and that *PSOS* based on Duarte and Young (2009) is only moderately correlated with the noninformation components, whereas it is highly correlated with the adverse-selection component. We then examine the behavior of the *PIN* measures relative to the surprise element of quarterly earnings announcements. *PIN* is increasing in the absolute value of the quarterly earnings-surprise proxies, which is consistent with more informed trading before larger earnings surprises. We also find that the estimate of *PIN_G* is increasing in the earnings surprise, and the estimate of *PIN_B* is decreasing in the surprise, which implies more informed buying before good news and more informed selling before bad news.

More importantly, after verifying that *PIN* is priced in the cross section of stock returns during the 28-year period, we find that the cost of equity capital is asymmetrically influenced by the components of *PIN*. In particular, although the probability of informed trading on bad news (*PIN_B*) strongly influences the cost of capital, there is no evidence that the probability of informed trading on good news (*PIN_G*) commands a positive return premium. The pricing of *PIN_B* is strongly dependent on firm size, with the strongest effect among smaller firms. This is consistent with less efficient pricing of small firms, implying a higher cost of adverse selection for a given level of informed trading. Our findings are robust to different subsamples that account for changes in the trading regime, and to the inclusion of the measures of informed trading proposed by Duarte and Young (2009).

The finding that informed trading on bad news commands a greater return premium than does trading on good news raises several issues for future research. First, it would be useful to explore firm-specific time variation in the probability of trading on negative information. For example, as a firm becomes distressed, the probability may rise, leading to an increase in the cost of capital. Second, it would be interesting to explore whether phenomena like earnings management are more common in firms with greater probability of bad-news informed trading, since earnings management will increase information asymmetry between management and retail investors, creating greater opportunities for informed trading. Third, common variation across firms in the probability of informed trading on bad news may be important in explaining common return effects. These and other issues are left for future research.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mnsc.2015.2250.

Acknowledgments

The authors are grateful for the insightful comments of three anonymous referees, an associate editor, and department editor Neng Wang. The authors also thank Lee-Seok Hwang and Woo-Jong Lee for advice on estimating the *AdjPIN* and *PSOS* measures of Duarte and Young (2009); Stephen Brown, Søren Hvidkjaer, Yuxing Yan, and Mao Ye for suggestions on estimating the *PIN* measure; Jefferson Duarte and Lance Young for discussions on the analysis methods used in Duarte and Young (2009); Denys Glushkov at Wharton Research Data Services for assistance in calculating the earnings surprise (*SUE*) measure; and Joon Chae, Miriam Marra, Keebong Park, Masahiro Watanabe, and seminar/conference participants at the 2014 Financial Management Association (FMA) Asian Conference and the 2014 Korea-America Finance Association (KAFA) and Korea finance associations joint conference for valuable feedback. Research assistance was ably provided by Seung-Oh Han and Yue Wang. S.-W. Huh acknowledges financial support from the Social Sciences and Humanities Research Council of Canada and the Ministry of Research and Innovation of Ontario. All errors are solely the authors' responsibility.

References

Abarbanell J, Lehavy R (2007) Letting the "tail wag the dog": The debate over GAAP versus street earnings revisited. *Contemporary Accounting Res.* 24:675–723.

Acharya V, Pedersen L (2005) Asset pricing with liquidity risk. *J. Financial Econom.* 77:385–410.

Aktas N, de Bodt E, Declerck F, Van Oppens H (2007) The *PIN* anomaly around M&A announcements. *J. Financial Markets* 10:169–191.

Amihud Y (2002) Illiquidity and stock returns: Cross-section and time-series effects. *J. Financial Markets* 5:31–56.

Amihud Y, Mendelson H (1986) Asset pricing and the bid-ask spread. *J. Financial Econom.* 17:223–249.

Ang A, Liu J, Schwarz K (2008) Using stocks or portfolios in tests of factor models. Working paper, Columbia University, New York.

Arbel A, Strebel P (1983) Pay attention to neglected firms! *J. Portfolio Management* 9:37–42.

Arbel A, Carvel S, Strebel P (1983) Giraffes, institutions, and neglected firms. *Financial Anal. J.* 39:57–63.

Asparouhova E, Bessembinder H, Kalcheva I (2010) Liquidity biases in asset pricing tests. *J. Financial Econom.* 96:215–237.

Atkins A, Dyl E (1997) Market structure and reported trading volume: NASDAQ versus the NYSE. *J. Financial Res.* 20:291–304.
Avramov D, Chordia T (2006) Asset pricing models and financial market anomalies. *Rev. Financial Stud.* 19:1002–1040.
Benos E, Jochec M (2007) Testing the PIN variable. Working paper, University of Illinois at Urbana–Champaign, Champaign.
Bhattacharya N, Ecker F, Olsson P, Schipper K (2012) Direct and mediated associations among earnings quality, information asymmetry, and the cost of equity. *Accounting Rev.* 87:449–482.
Brennan M, Subrahmanyam A (1996) Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *J. Financial Econom.* 41:441–464.
Brennan M, Wang A (2010) The mispricing return premium. *Rev. Financial Stud.* 23:3437–3468.
Brennan M, Chordia T, Subrahmanyam A (1998) Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *J. Financial Econom.* 49:345–373.
Brennan M, Huh S, Subrahmanyam A (2013) An analysis of the Amihud illiquidity premium. *Rev. Asset Pricing Stud.* 3: 133–176.
Brennan M, Chordia T, Subrahmanyam A, Tong Q (2012) Sell-order liquidity and the cross-section of expected stock returns. *J. Financial Econom.* 105:523–541.
Brown S, Hillegeist S (2007) How disclosure quality affects the level of information asymmetry. *Rev. Accounting Stud.* 12:443–477.
Brown S, Hillegeist S, Lo K (2004) Conference calls and information asymmetry. *J. Accounting Econom.* 37:343–366.
Carhart M (1997) On persistence in mutual fund performance. *J. Finance* 52:57–82.
Chakrabarty B, Moulton P, Shkilko A (2012) Short sales, long sales, and the Lee-Ready trade classification algorithm revisited. *J. Financial Markets* 15:467–491.
Chen Q, Goldstein I, Jiang W (2007) Price informativeness and investment sensitivity to stock price. *Rev. Financial Stud.* 20:619–650.
Chordia T, Huh S, Subrahmanyam A (2007) The cross-section of expected trading activity. *Rev. Financial Stud.* 20:709–740.
Chordia T, Huh S, Subrahmanyam A (2009) Theory-based illiquidity and asset pricing. *Rev. Financial Stud.* 22:3629–3668.
Chordia T, Subrahmanyam A, Tong Q (2014) Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *J. Accounting Econom.* 58:41–58.
Chung K, Huh S (2015) The non-information cost of trading and its relative importance in asset pricing. Working paper, University at Buffalo, Buffalo, NY.
Duarte J, Young L (2009) Why is *PIN* priced? *J. Financial Econom.* 91:119–138.
Easley D, Hvidkjaer S, O'Hara M (2002) Is information risk a determinant of asset returns? *J. Finance* 57:2185–2221.
Easley D, Kiefer N, O'Hara M, Paperman J (1996) Liquidity, information, and infrequently traded stocks. *J. Finance* 51: 1405–1436.
Ellul A, Pagano M (2006) IPO underpricing and after-market liquidity. *Rev. Financial Stud.* 19:381–421.
Ellul A, Panayides M (2013) Do financial analysts restrain insiders' informational advantage? Working paper, Indiana University, Bloomington.
Fama E, French K (1992) The cross-section of expected stock returns. *J. Finance* 47:427–466.
Fama E, French K (1993) Common risk factors in the returns on stocks and bonds. *J. Financial Econom.* 33:3–56.
Fama E, MacBeth J (1973) Risk, return, and equilibrium: Empirical tests. *J. Political Econom.* 81:607–636.
Foster D, Viswanathan S (1993) Variations in trading volume, return volatility, and trading costs: Evidence on recent price formation models. *J. Finance* 48:187–211.
Gibbons M, Ross S, Shanken J (1989) A test of the efficiency of a given portfolio. *Econometrica* 57:1121–1152.
Glosten L, Harris L (1988) Estimating the components of the bid-ask spread. *J. Financial Econom.* 21:123–142.
Glosten L, Milgrom P (1985) Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *J. Financial Econom.* 14:71–100.
Harvey C, Liu Y, Zhu H (2013) . . .and the cross-section of expected stock returns. Working paper, Duke University, Durham, NC.
Hong H, Lim T, Stein J (2000) Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *J. Finance* 55:265–295.
Huang R, Stoll H (1997) The components of the bid-ask spread: A general approach. *Rev. Financial Stud.* 10:995–1034.
Huh SW (2014) Price impact and asset pricing. *J. Financial Markets* 19:1–38.
Hwang L, Lee W, Lim S, Park K (2013) Does information risk affect the implied cost of equity capital? An analysis of *PIN* and adjusted *PIN*. *J. Accounting Econom.* 55:148–167.
Jegadeesh N (1990) Evidence of predictable behavior in security returns. *J. Finance* 45:881–898.
Jegadeesh N, Titman S (1993) Returns to buying winners and selling losers: Implications for stock market efficiency. *J. Finance* 48:65–92.
Kothari S, Shu S, Wysocki P (2008) Do managers withhold bad news? *J. Accounting Res.* 41:241–276.
Kyle A (1985) Continuous auctions and insider trading. *Econometrica* 53:1315–1335.
Lai S, Ng L, Zhang B (2014) Does PIN affect equity prices around the world? *J. Financial Econom.* 114:178–195.
Lee C, Ready M (1991) Inferring trade direction from intraday data. *J. Finance* 46:733–747.
Lee C, Radhakrishna B (2000) Inferring investor behavior: Evidence from TORQ data. *J. Financial Markets* 3:83–111.
Levi S, Zhang X-J (2015) Do temporary increases in information asymmetry affect the cost of equity? *Management Sci.* 61: 354–371.
Livnat J, Mendenhall R (2006) Comparing the post-earnings announcement drift for surprises calculated from analysts and time series forecasts. *J. Accounting Res.* 44:177–205.
Lo A, MacKinlay C (1990) Data-snooping biases in tests of financial asset pricing models. *Rev. Financial Stud.* 3:431–468.
Lou X, Shu T (2014) Why is the Amihud (2002) illiquidity measure priced? Working paper, University of Delaware, Newark.
Mohanram P, Rajgopal S (2009) Is *PIN* priced risk? *J. Accounting Econom.* 47:226–243.
Odders-White E (2000) On the occurrence and consequences of inaccurate trade classification. *J. Financial Markets* 3:205–332.
O'Hara M, Yao C, Ye M (2014) What's not there: The odd-lot bias in TAQ data. *J. Finance* 69:2199–2236.
Pástor L, Stambaugh R (2003) Liquidity risk and expected stock returns. *J. Political Econom.* 113:642–685.
Philips M (2013) How the robots lost: High-frequency trading's rise and fall. *Bloomberg Businessweek* (June 6), http://www.bloomberg.com/bw/articles/2013-06-06/how-the-robots-lost-high-frequency-tradings-rise-and-fall.
Pontiff J, Woodgate A (2008) Shares issuance and cross-sectional returns. *J. Finance* 43:921–945.
Rendleman R Jr, Jones C, Latané H (1982) Empirical anomalies based on unexpected earnings and the importance of risk adjustments. *J. Financial Econom.* 10:269–287.
Sadka R (2006) Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk. *J. Financial Econom.* 80: 309–349.
U.S. Securities and Exchange Commission (2012) Report to Congress on decimalization. Report, U.S. Securities and Exchange Commission, Washington, DC.
Vega C (2006) Stock price reaction to public and private information. *J. Financial Econom.* 82:103–133.
Yan Y, Zhang S (2012) An improved estimation method and empirical properties of the probability of informed trading. *J. Banking Finance* 36:454–467.