Short Selling, Informed Trading, and Stock Returns

Tyler R. Henry[†]
University of Georgia

This Draft: May 2006

Abstract

This paper considers the effect of private information on the returns to stocks with high levels of short interest. I use a measure from the market microstructure literature to proxy for levels of asymmetric information. Among highly shorted firms, portfolios with high levels of informed trading generally underperform, while those with low levels of informed trading do not. The results suggest that the underperformance of high short interest stocks is driven by firms that have high levels of informed trading. However, this negative relationship between informed trading and returns is reversed for stocks with low to moderate short interest levels.

This paper is based on a section of my dissertation completed at the University of Washington. I would like to thank my committee members for discussions and guidance, including Jonathan Karpoff, Ed Rice, Lance Young and especially Jennifer Koski (Chair). I also appreciate the valuable comments of Bart Danielsen, Wayne Ferson, Jaehoon Hahn, Andrew Karolyi, Ronnie Sadka, and seminar participants at the Universities of British Columbia, Georgia, Kentucky, Missouri, Southern California, and Washington, the 2004 FMA Annual Meetings, the 2004 FMA Doctoral Consortium, and the 2005 Atlanta Federal Reserve All-Georgia Finance Conference. I thank the NYSE for providing short interest data, and Lew Thorson for considerable programming assistance. I am responsible for any remaining errors.

[†] Assistant Professor, Department of Banking and Finance, Terry College of Business, University of Georgia, Athens, GA 30602-6253. Email: trhenry@terry.uga.edu.

1. Introduction

This paper investigates the relationship between asymmetric information, short selling, and stock returns. I achieve this analysis by examining the intersection of short interest and levels of informed trading. Two key insights are gained from this intersection. First, the underperformance of high short interest stocks could be driven by firms which also have high levels of informed trading. Second, the positive relation between informed trading and average returns revealed in the literature [Easley, Hvidkjaer, and O'Hara (2002)] reverses to a negative relation at high levels of short interest. Along these lines, the paper unites two independent streams of research, with a contribution to each.

The first contribution is to the empirical short selling literature, specifically, the link between short interest levels and subsequent returns. I use a measure of the probability of informed trading (PIN) derived from a microstructure trade model [Easley, Kiefer, and O'Hara (1997b)] to help explain underperformance in stocks with high levels of short interest. The results imply that the consideration of information asymmetries improve the usefulness of short interest as a return predictor. This result is not surprising given both informed and uninformed motivations for short selling.

The second contribution is to the literature that examines the role of microstructure measures of asymmetric information for average returns. Easley, Hvidkjaer, and O'Hara (2002) show a positive relation between the probability of informed trading and the cross section of returns. My finding that the sign of the PIN-return relationship reverses from positive to negative for highly shorted stocks is consistent with the interpretation of high short interest as a noisy proxy for adverse information. Regardless of a positive or negative relation between PIN and returns, the results support the influence of informed trading on stock returns.

To a large extent, the empirical short interest literature is guided by Figlewski's (1981) assertion that short interest is a proxy for negative information, and that the amount of negative information excluded from the market (thus allowing optimists to set prices) is increasing in the level of short interest. Under this pretext, overvaluation should increase with the amount of short interest. There are two concerns with this assumption. First, as pointed out by D'Avolio (2002) and Chen, Hong, and Stein (2002), cross-sectional variation in short interest levels could reflect differences in the transaction costs of short selling rather than differences in information.

Secondly, this assumption disregards the volume of short selling related to uninformed trading strategies.[1] The latter issue is where this paper makes new inroads. Specifically, one contribution of this paper is to more finely target the underperformance of high short interest stocks by controlling for levels of informed trading.

I advocate that, on a standalone basis, the amount of noise in the short interest data hinders its ability as a return predictor. To refine the assumption of Figlewski (1981), short interest is a *noisy* proxy for adverse information. Because some short selling is informed while some is related to uninformed arbitrage or hedging strategies, the link between short interest levels and informed trading is not clear cut. More importantly, any relationship between short interest and returns likely depends on this link.

While short interest is an imperfect proxy for shorting demand, high short interest generally implies high demand for shorting.[2] Still, some of this demand originates from uninformed trading strategies unrelated to valuation. Using a proprietary dataset of stock loans, Cohen, Diether, and Malloy (2005) isolate the demand and supply effects in the shorting market, and show that outward shifts in the demand to short sell drive the low future returns of these stocks. Further, they demonstrate that these demand shifts are unlikely to be driven by public information. Their results suggest two important considerations when using short interest data as an indicator of low subsequent returns. First, any relation between short interest and future returns is likely to exist primarily when shorting demand is high. This finding supports the focus in the literature on predominantly high short interest stocks. Second, the ability to distinguish informed short selling from uninformed short selling among high short interest (e.g., high shorting demand) stocks will likely improve the information content of short interest. These considerations help to motivate the empirical approach taken in this paper: to filter high short interest stocks that have high *informed* demand from those that have high *uninformed* demand.

Indeed, the gains from the ability to partition short sales into two groups based on trade motivation is frequently referenced in the short interest literature.[3] To tackle the empirical

[1] Examples of such trading strategies include index arbitrage, risk (takeover) arbitrage, convertible bond arbitrage, pairs trading, and tax related short selling or "shorting against the box" (which was outlawed by the Taxpayer Relief Act of 1997). Additionally, some short selling may originate from underwriters trying to reduce price volatility in public offerings and specialists trying to offset temporary inventory positions.

[2] However, a firm with low short interest does not necessarily imply low demand, since low supply of lendable shares or binding constraints could prohibit short selling, despite high demand. The 3Com/Palm equity carveout is one notable example where demand was very high but short interest was low due to limited supply.

[3] For examples, see Senchack and Starks (1993), Desai et al. (2002), Asquith, Pathak, and Ritter (2005).

challenge of identifying informed short selling within the short interest data, I borrow a measure from the market microstructure literature that identifies the level of informed trading in a firm's stock. I form portfolios based on this measure, the probability of informed trading (PIN), in an effort to better target future underperformance. Thus, instead of testing the impact of short interest levels on future returns, I test the impact of PIN on future returns for a sample of stocks with high short interest.

If constraints on short selling lead to overpricing by optimists [Miller (1977)], and informed traders can identify the most overpriced stocks, then those stocks with higher amounts of informed short selling should have a stronger relation with future returns. Namely, I hypothesize that stocks with high levels of short interest *and* high levels of informed trading (high PINs) should have lower subsequent returns than high short interest stocks with low PINs.

In support of this hypothesis, I find that size-neutral, high PIN portfolios realize subsequent negative abnormal returns. These returns are both economically meaningful and statistically significant, with high PIN portfolios displaying abnormal returns ranging from -0.471% to -1.106% per month. Except for firms in the lowest size quartile, low PIN portfolios do not statistically underperform. Additionally, the returns to a zero investment portfolio long high PIN stocks and short low PIN stocks confirm that the high PIN portfolio underperforms the low PIN portfolio in most cases. The results are generally robust when portfolio returns are value-weighted.

To address the apparent contradiction between my finding of a negative relation between PIN and returns, and previous literature that shows a positive relation [Easley, Hvidkjaer, and O'Hara (2002)], I expand the sample to include stocks with all levels of short interest. Forming portfolios based on short interest deciles and PIN, I find that PIN-based portfolios in the highest short interest deciles display a negative relationship with returns as predicted by my hypothesis. The portfolios in the lower short interest deciles generally show a positive relationship, consistent with the existing literature. This result corroborates the interpretation of short interest as a negative information signal, but a noisy signal whose precision is increased by controlling for informed trading.

Finally, I perform firm-level cross-sectional regressions of risk-adjusted returns on various characteristics, using the methodology of Brennan, Chordia, and Subrahmanyam (1998). In regressions using a sample of only high short interest stocks, the results confirm a negative relation between PIN and returns. Also, the effect of PIN on returns seems to dominate the effect of the short interest ratio. Using a sample of stocks with all levels of short interest, a significantly positive relation between PIN and returns is revealed. Together, the tests support a relation between PIN and average returns, but the direction of this relation reverses at high short interest levels.

Overall, the results support a few main conclusions. First, some stocks with high levels of short interest do underperform, but this underperformance could be limited to stocks that have high levels of informed trading. Such a finding is consistent with the existence of two types of short selling [Asquith, Pathak, and Ritter (2005)], and suggests the use of microstructure based proxies to distinguish informed from uninformed short selling among stocks with high shorting demand. Finally, if PIN is a valid proxy for informed trading, the results support the impact of asymmetric information on average returns. However, the direction of this relation is negative at high levels of short interest levels.

The remainder of the paper proceeds as follows. Section 2 provides the motivation for an information-based approach. Section 3 describes the data and discusses estimation of PIN. Section 4 examines abnormal returns to portfolios formed on the basis of short interest and PIN. Section 5 performs cross-sectional regressions at the individual firm level. Section 6 concludes.

2. Motivation

The empirical literature examining the relationship between short selling, short sale constraints, and returns is large and growing. Much of this literature is directed at tests of Miller's (1977) overvaluation hypothesis in the presence of both short sale constraints and divergence of opinion. Several papers in this literature use monthly short interest levels to identify overvalued stocks, with short interest levels serving as either a proxy for the demand to short sell or the level of short sale constraints.[4] Despite its popularity, the deficiencies of short interest as a standalone proxy for either demand or constraints are well documented in the literature. Jones and Lamont (2002), provide a nice treatment of these issues, and Lamont (2004) points out that measures of shorting demand and measures of short sale constraints tend to be

[4] Two recent examples are Asquith, Pathak, and Ritter (2005) who use short interest as a proxy for demand to short sell, and Boehme, Danielsen, and Sorescu (2005) who use short interest as a proxy for short sale constraints.

highly correlated in practice, since both are driven by the same underlying mechanism. Recently, several papers have moved away from using short interest in favor of other measures of short sale constraints or demand, such as breadth of ownership [Chen, Hong, and Stein (2002)], residual institutional ownership [Nagel (2005)] or proprietary datasets of stock lending or rebate rates [Reed (2002), D'Avolio (2002), Cohen, Diether, and Malloy (2005)]. While these alternative measures or proprietary datasets provide valuable insights and offer some measurement benefits, the underlying advantages of the short interest data remain that they are widely available, directly related to short selling intensity, and accessible to both practitioners and researchers.

Another reason for the migration away from short interest data, in addition to its limitations as a proxy of shorting indicators, is the conflicting or unconvincing empirical results from some previous studies. Early papers in this literature did not find a strong relationship between short interest and subsequent returns, and claimed that a considerable amount of short selling was arbitrage or hedging related [Brent, Morse, and Stice (1990), Figlewski and Webb (1993), Senchack and Starks (1993), Woolridge and Dickinson (1994)]. Such trades are not motivated by information and should not have a strong relation with future returns. Later papers found that stocks with high levels of short interest generate subsequent negative abnormal returns, and argued that short sellers do have information [Asquith and Meulbroek (1996), Gintschel (2001), Desai et al. (2002)].

Certainly, if most short selling is uninformed, the usefulness of short interest as a return predictor is limited. Empirical evidence consistent with the view that at least some short sellers are informed includes Dechow et al. (2001), who demonstrate that short sellers are able to identify low book-to-market stocks due to overpricing, and Christophe, Ferri, and Angel (2004), who find increased short selling intensity in certain stocks prior to unfavorable earnings announcements. The evidence in these papers is consistent with short sellers using information in anticipation of lower future returns. Thus, the existence of both informed and uninformed motivations for short selling present a challenge to researchers using short interest data. Jones and Larsen (2004) argue that, despite the evidence of overpricing present in stocks that are costly to short, "there definitely is information content in short interest data, although it may be difficult to exploit." (p. 254).

The limitations that result from an inability to distinguish informed from uninformed short selling is frequently referenced in the short interest literature. For example, although Senchack and Starks (1993) find a weak relation between short interest and returns, they point out that an inability to purge uninformed short sales from the short interest numbers may confound any empirical tests. They remark that "the observed market reaction to short interest announcements may be underestimated due to noninformational short sales…" (p.186). Desai et al. (2002) note that short interest has become a noisier proxy for adverse information over time, due to the increasing use of short selling in arbitrage and hedging transactions. This point is consistent with Asquith, Pathak, and Ritter (2005), who find a weaker relation between short interest and returns in the later portion of their sample period.[5] They use firms with convertible bonds as a proxy for arbitrage based shorting, and suggest the negative abnormal returns they find are driven by "valuation" shorting. Their general approach, while useful, illustrates the empirical difficulty of identifying short selling related to uninformed strategies. As they acknowledge, there could be informed short selling occurring in stocks with convertible bonds, but they are unable to identify these situations.

Lately, the notion that most high short interest stocks underperform has come under renewed scrutiny. Asquith, Pathak, and Ritter (2005) claim to "find more ambiguous patterns than the previous literature suggests" (p.245). Boehme, Danielsen, and Sorescu (2005) argue that the varied results of this literature stem from the failure of most studies to account for both of the theoretical requirements of overvaluation; constrained short selling *and* heterogeneous beliefs (e.g., differences in opinion). By accounting for both, they offer a more direct test of Miller (1977). They report that only stocks with *both* high constraints and high dispersion of opinion are overvalued. Stocks lacking either condition are not overvalued.

The results of the short interest literature offer two conclusions. First, short selling can be either informed or uninformed, and any relation between short interest and returns likely depends on this distinction. Second, asset pricing tests that do not directly control for these disparate motives should produce weak empirical results. An ability to identify which stocks contain greater amounts of informed short selling could increase the information content of short interest.

[5] Arnold et al. (2005), however, find a stronger negative relation between short interest and subsequent returns post 1997 after the Taxpayer Relief Act of 1997 increased the costs of short selling by prohibiting uninformed, tax-based short-selling .

If the constraints on short sales restrict the private negative information of some investors from the trading process, these constraints reduce levels of informed trading. This restriction on informed traders should impact prices. One method to estimate informed trading has been devised in a series of papers by Easley et al. (1996, 1997a, 1997b). The measure derives from a sequential trade model with both informed and uninformed traders. Estimating the structural parameters of such a model provides an empirical measure of the probability of information based trade. Easley, Hvidkjaer, and O'Hara (2002, 2004) use this measure to argue that information risk is priced in the cross-section of asset returns, and that size-neutral portfolios formed on the basis of PIN earn different returns. They advocate the empirical importance of asymmetric information for asset prices.

Existing tests of the performance of high short interest stocks are undertaken without definitive regard for asymmetric information. Given the different motivations for short selling, it is not immediately clear whether high short interest stocks will necessarily have higher levels of informed trading. In other words, short interest alone is not a successful indicator of future underperformance. If PIN is an accurate measure of asymmetric information, it could help to identify which highly shorted stocks have higher amounts of informed trading. Combining these two measures should increase the precision of the information signal. Specifically, high short interest indicates negative information conditional on the existence of information, and high PIN indicates the existence of information. Thus, among a sample of high short interest firms, high PIN stocks should have larger negative abnormal returns than low PIN stocks.

3. Empirical Methodology

3.1. Data and Short Interest Subsamples

There are two subsamples of short interest data that I use in the study. The first subsample includes only stocks with high levels of short interest, and the second subsample includes stocks with all levels of short interest. High short interest stocks are defined as those that have a short interest ratio greater than 2.5% for *at least one* month during the sample period. The short interest ratio is defined as the number of currently outstanding short positions divided by the number of shares outstanding.[6] The sample period includes the 131 months from January 1992 to December 2002.[7] The source of the data is the NYSE. I use only NYSE stocks because the structural model used to estimate PIN more closely describes the price discovery process of a single market maker, similar to that of the NYSE specialist. Many of the hypotheses in this paper are directed specifically at firms with high shorting demand, proxied by high levels of short interest. The 2.5% cutoff as a selection criteria for high short interest is consistent with the recent literature [Asquith and Meulbroek (1996), Desai et al. (2001), Asquith, Pathak, and Ritter (2005)]. I exclude from the short interest sample preferred shares, warrants, ADRs, convertible securities, and unit investment trusts. Shares outstanding and monthly return data are from CRSP.

To estimate PIN (discussed in the next section) requires daily trade and quote data from the NYSE TAQ database starting in 1993 and from the ISSM database prior to 1993. For the subsample of high short interest stocks, I estimate a monthly PIN for each month, regardless of whether or not the short interest exceeds 2.5% in that specific month. Initially, I have 128,917 firm-month observations of firms that have short interest greater than 2.5% during at least one month. For the asset pricing tests in later sections, I include only those firm-months where the short interest is greater than 2.5% in the *previous* month. After combining this restricted sample with the monthly return data from CRSP and the trade and quote data from ISSM and TAQ, I am left with 39,617 firm-month observations where the short interest ratio is actually greater than 2.5% in the previous month.

The second short interest subsample includes stocks with all levels of short interest, so it is not limited to high short interest stocks, but does include the high short interest stocks. This sample is maintained for robustness checks of my hypotheses, and comparisons between high SI stocks and low SI stocks. For this sample, I use the annual PIN estimates from Easley, Hvidkjaer, and O'Hara (2002), rather than a monthly PIN estimate.

---

[6] Another short interest indicator frequently used by the practitioner community is the days to cover ratio. This is the level of short interest divided by the average daily share volume, and represents, on average, how many days of trading would be required to cover all outstanding short positions. Asquith, Pathak, and Ritter (2005) point out that the days to cover ratio is the more appropriate measure if short interest is viewed as a signal of future buying pressure (as many practitioners believe it to be), rather than as a signal of potentially negative information. Regardless, the two measures are correlated. Moreover, some practitioner analysts are beginning to use the short interest ratio as a predictive indicator.

[7] The month of September 2001 is excluded from my sample. Because the exchange was closed for several days after September 11th, there are not enough trading days in this month for an accurate estimate of PIN.

Table 1 includes descriptive statistics for the high short interest firms. Panel A contains mean monthly statistics for the firm-month observations where the short interest is above 2.5% (N=39,617), by year. These observations are those that will be used to form portfolios in later sections. This high short interest sample contains 1,812 unique firms. Naturally, some firms jump in and out of the high short interest sample as short interest changes from month to month. Panel B contains statistics for the same high short interest firms, but during months in which their short interest is less than 2.5%. On average, a firm is in the total sample (e.g., Panel A plus Panel B) for an average of 81.7 months but has short interest actually greater than 2.5% (Panel A) for an average of 20.9 months, or 28.5% of the time. Some obvious patterns emerge from these two panels. First, the average short interest ratio is much larger for these firms when they meet the 2.5% inclusion criteria. Over the sample period, the average short interest for these firms when they are included in the high short interest sample is 6.34%, but drops to 0.99% when they fall out of the high sample. Average price and market value of equity are smaller when these firms are included in the high short interest sample, and their subsequent returns are lower. For example, the average subsequent month's raw return is 0.50% when firms are in the high short interest sample and 1.15% when they are not. The short interest ratio for the high short interest sample exhibits high autocorrelation over one lag of 0.76. Panel C combines the observations from Panel A and Panel B, providing a summary for the high short interest firms during all months.

For comparison to Panel A, Panel D shows statistics for *all* firm-months (N=176,964), where a firm's short interest is between zero and 2.5% in a given month. Some of the firms in this panel never enter the high short interest sample (e.g., are not included in Panel A, B, or C). From this panel it is apparent that firms in the high short interest sample are smaller and trade at lower prices. Asquith and Meulbroek (1996) find that their sample of highly shorted NYSE and AMEX firms are smaller than a randomly matched sample, but Desai et al. (2002) find that their sample of highly shorted Nasdaq firms is larger than average. R_{t+1}, the subsequent monthly raw return, averages 1.01% for the low short interest stocks over the entire 131 month period.

3.2. Estimating the Probability of Informed Trading

To measure the probability of informed trading, I apply the discrete time sequential trade model of Easley, Kiefer, and O'Hara (1997b).[8] The model depicts the learning problem of a market maker as he observes the sequencing of trades throughout the day. The model has two trader types: informed traders and uninformed traders. There are four structural parameters to the model: α is the probability that an information event occurs at the beginning of trading, δ is the probability that an information event is negative, μ is the probability that a trader is informed, and ε is the probability that an uninformed traders transacts after arriving to trade. For each trader that arrives to trade, one of three possible outcomes will occur; a buy, a sell, or a no-trade. Estimating the structural parameters of the model requires knowledge only of the daily number of buys, sells, and no-trade outcomes, $\left(B_d, S_d, N_d\right)$. Since researchers cannot observe short selling in the order flow, regular sales and short sales are pooled together in the estimation procedure. The decision tree that describes the trade model is displayed in Figure 1.

The likelihood function of observing the number of buys, sells, and no-trade outcomes on a given day is $\Pr\left\{\left(B,S,N\right)|\theta\right\}$ where $\theta=\{\alpha,\delta,\mu,\varepsilon\}$. Assuming independence across trading days, the likelihood of observing the history of buys, sells and no-trades across the estimation period is found by taking the product across days such that

$$\Pr\left\{\left(B_d, S_d, N_d\right)_{d=1}^{D} | \theta\right\} = \prod_{d=1}^{D} \Pr\left\{\left(B_d, S_d, N_d\right) | \theta\right\} \qquad (1)$$

where D is the number of trading days in the estimation period.[9] Easley, Kiefer, and O'Hara (1997b) show that this is computed by maximizing the following likelihood function:

[8] There is a corresponding continuous time model with similar design. (See Easley, Kiefer, O'Hara and Paperman (1996)). My choice to use the discrete time model is motivated by the allowance for no-trade outcomes in the discrete model. Short selling constraints may lead to an increased incidence of no-trade outcomes, which is one of the main points of Diamond and Verrecchia (1987). Since these are modeled in the discrete form model, it is the more suitable choice.

[9] Easley, Kiefer, and O'Hara (1997b) test the assumption of independence across trading days and are unable to reject the assumed independence.

$$\sum_{d=1}^{D}\log\left\{\alpha(1-\delta)\left(1+\frac{\mu}{x}\right)^{B}+\alpha\delta\left(1+\frac{\mu}{x}\right)^{S}+(1-\alpha)\left(\frac{1}{1-\mu}\right)^{B+S+N}\right\}+\sum_{d=1}^{D}\log\left\{x^{B+S}\left[(1-\mu)(1-\varepsilon)\right]^{N}\right\} \tag{2}$$

where $x=\frac{1}{2}(\varepsilon-\mu)$. Once the parameters have been estimated, the probability of informed trading, or PIN, is then calculated as the probability that a trade comes from an informed trader divided by the probability that a trade occurs:

$$PIN=\frac{\alpha\mu}{\alpha\mu+(1-\alpha\mu)\varepsilon} \tag{3}$$

When counting buys and sells to estimate the likelihood function in (2), trade direction is determined using the Lee and Ready (1991) algorithm where a trade that occurs above (below) the prevailing quote midpoint is classified as buyer-initiated (seller-initiated).[10] A trade that occurs at the midpoint is buyer-initiated (seller-initiated) if the last trade was an uptick (downtick). I allow for a five-second lag in matching quotes to trades as is common in the literature to account for potential delays in trade reporting [see Lee and Ready (1991)].[11] I include only trades and quotes that occur during normal trading hours on the NYSE and exclude quotes and trades from the regional exchanges.

To count the number of no-trade outcomes in a day, I first calculate the average number of daily trades for a given stock and divide by the total number of seconds in a trading day. This time interval is then used as the no-trade period, and if this amount of time passes without a trade, a no-trade outcome is counted (if twice this time interval passes, I count two no-trade outcomes, and so on until a trade takes place). I compute this time interval for each stock individually and for each month in the sample. This approach accounts for differences in trade

[10] Several studies test the effectiveness of the Lee and Ready algorithm. Odders-White (2000) finds the algorithm to be 85 percent accurate for NYSE TORQ data, while Lee and Radhakrishna (2000) find 93 percent accuracy for trades that can be unambiguously classified. Ellis, Michaely, and O'Hara (2000) report 81 percent accuracy for Nasdaq data. Despite its deficiencies, the algorithm is reasonably accurate and remains the standard in the literature.
[11] Bessembinder (2003) and Peterson and Sirri (2003) find that the accuracy of the trade direction algorithm may be improved when no timing lag is implemented. I keep the lag to be consistent with other papers that estimate PIN using TAQ data. My results are not sensitive to this convention.

frequency across stocks and time varying trade volume, as opposed to assigning an arbitrary length of time for all stocks and months.[12]

I estimate the model parameters for each stock on a monthly basis, by maximizing the likelihood function in equation (2) with the daily number of buys, sells and no-trades over the month. The parameter estimates then give a monthly measure of the probability of informed trading from equation (3). Additionally, I calculate the monthly average of the daily trade-weighted quoted half-spread, relative half-spread, and effective half-spread. The relative half-spread is one half of the quoted spread divided by the spread midpoint. The effective half-spread is one half of the absolute value of the trade price minus the quoted midpoint prior to the trade. In the structural trade model, there are no inventory or order processing costs. Thus, the spread measures are alternative proxies for information asymmetry. Easley et al. (1996), Easley, O'Hara, and Paperman (1998), and Chung and Li (2003) all find a significant positive relationship between PIN and the adverse selection component from various spread decompositions. Heidle and Huang (2002) report an association between changes in spreads and changes in PIN for firms that undergo an exchange listing relocation.

Table 2 shows the time series average of the cross-sectional correlations of the variables of interest in the study for the high short interest subsample. The interrelations between PIN and the other variables are of primary interest. PIN is weakly positively correlated to the quoted half-spread, but displays stronger positive correlation to the relative spread (0.262). Additionally, correlation between PIN and short interest is positive but weak (0.062), implying no clear cut relationship between the two. This finding is not surprising given the previous discussion about various motives for short selling. That these two variables are not strongly correlated further motivates the approach in this paper to use PIN instead of the short interest level alone as a predictor of returns to highly shorted stocks. PIN is negatively correlated with both firm size (-0.162) and the daily number of trades (-0.216), consistent with the findings of Easley et al. (1996) that smaller firms and low volume stocks have higher levels of private information. Lastly, PIN is negatively correlated with subsequent excess returns. This relation is in contrast to the results of Easley, Hvidkjaer, and O'Hara (2004) who find positive correlation, but is expected given my sample of only high short interest stocks. The negative correlation

[12] While this seems like a logical approach to tabulating no-trade outcomes, there are other approaches. Chung, Li, and McInish (2005) experiment with different methods of calculating this interval, and find their estimation results to be robust to different specifications.

highlights one of many key differences in the two studies; my high short interest sample generally includes firms where an information event is likely to contain a negative signal. Explanations for this negative relationship among high short interest stocks and comparisons to stocks with low levels of short interest are explored in later sections.

One concern with estimating PIN at a monthly frequency is the precision of the parameter estimates. Most existing studies that use PIN estimate over longer time intervals to address this concern. Easley et al. use a minimum of 60 trading days in most of their studies, but Easley, Hvidkjaer and O'Hara (2002) discuss the possible value of estimating over monthly horizons. Vega (2005) estimates PIN over 40 trading days, and Fu (2002) over 45 days. I choose the shorter monthly interval to capture changes in monthly short interest positions.[13]

Given the timing of the short interest data, some thought must be given to the appropriate time interval over which to estimate the monthly PIN. Short interest levels are collected by the exchanges on the fifteenth day of every month (if it is a business day) and represent short positions outstanding as of the settlement date. The public announcement date occurs a few days later, between the 19th and 25th day of the month during my sample. So, if the 15th is a trading day, the short interest level reported that month represents the amount of short interest outstanding three trading days prior.[14] I then begin my estimation period for month *t* two trading days prior to the 15th, and end it three trading days before the next month's collection date. Therefore, my monthly PIN estimate corresponds with the period over which the short interest amount changed from month *t-1* to month *t*. Figure 2 provides a graphical description of the intervals over which the variables are measured.

4. PIN-Based Portfolio Returns

Empirical studies of the underperformance of high short interest stocks usually test the relation between short interest levels and subsequent monthly returns, and are motivated by an

[13] I perform unreported diagnostic tests on estimating PIN for both monthly and quarterly horizons, and find the estimates to be very similar across the two horizons. Some precision is sacrificed with the monthly horizon, but this is necessary to capture monthly fluctuations in PIN that may correspond to changes in short interest. Results available upon request.

[14] Settlement was reduced to *t+3* from *t+5* in June, 1995. This is during my sample, and I account for this change accordingly. See Asquith, Pathak, and Ritter (2005) for more details concerning the timing of short interest collection dates and public announcement dates.

overpricing story in the spirit of Miller (1977).[15] The most recent of these empirical tests of underperformance is Asquith, Pathak, and Ritter (2005). They show that underperformance of high short interest stocks is not robust when portfolio returns are value-weighted (VW). They also show that underperformance in portfolios of stocks with short interest greater than 2.5% is driven by a small subset of stocks. When portfolios are "truncated" to create independent samples, this underperformance weakens. In general, they argue that the evidence on the underperformance of high short interest stocks is more ambiguous than previously thought. This sentiment exactly describes my motivation for suggesting a new approach to detecting any existing underperformance.

In unreported experiments, I find similar results as those in Asquith, Pathak, and Ritter (2005) for my sample when portfolios are formed on the basis of short interest only.[16] Namely, when returns are value-weighted and portfolios are independent, some of the underperformance of high short interest stocks disappears. Yet, some portfolios still have negative abnormal returns. Notably, this underperformance is not necessarily present in the most heavily shorted stocks. Perhaps then, the precision of any information signal contained in the short interest numbers can be improved by conditioning on levels of informed trading. Such a finding would suggest a finer indicator of overvaluation in highly shorted stocks.

4.1. PIN-Based Portfolio Construction

Since the constraints on short selling restrict the trading of negatively informed agents, a measure of asymmetric information may help to explain returns to highly shorted stocks. The probability of informed trading, or PIN, is such a measure of asymmetric information. Easley, Hvidkjaer, and O'Hara (2004) show that size-neutral zero investment portfolios formed on the basis of PIN produce positive abnormal returns. In this section, I examine whether or not highly shorted stocks with greater levels of private information have larger abnormal returns. Since this subsample contains *only* high short interest stocks, any information should be negative and I

[15] Miller's (1977) theory produces hypotheses related to the level of short sale *constraints* and overvaluation. He makes no predictions regarding short interest levels and returns. The goal of this paper is not to measure the level of short sale constraints. In other words, my hypotheses apply specifically to stocks with high levels of short interest (high shorting demand), which is not necessarily synonymous with high short sale constraints. However, to the degree that shorting demand and short sale constraints are correlated, my measure may indirectly be picking up the effect of constraints, as they relate to high shorting demand.
[16] Results available upon request.

expect these realized abnormal returns to be negative. My hypothesis is that high PIN portfolios will underperform low PIN portfolios.

To test this hypothesis, I construct portfolios that are sorted on firm size and PIN. Sorting by size is necessary to isolate the effect of PIN, since PIN and size are strongly negatively correlated (Table 2). Each month, I sort all high short interest stocks into quartiles based on the previous month's market capitalization. Within each size quartile, I sort into terciles based on the prior month's PIN. This process gives me 12 portfolios every month. Sorting by PIN within quartile allows me to have an approximately equal number of stocks in each portfolio. Given my small sample size (only high short interest stocks), ensuring there are a reasonable number of firms in each portfolio is especially important.

Table 3 gives the results of these portfolio sorts, including the average number of stocks in each monthly portfolio, the average size, PIN, short interest, quoted and relative half-spread, and the estimated parameters from the structural trading model. These figures illustrate the effectiveness of the sorts and the characteristics of the different portfolios. Within each size quartile, I check for differences in means across the low and high PIN portfolios with a nonparametric Wilcoxon rank sum test (Z-statistic). The sorts appear to be fairly successful in controlling for firm size, and there is significant variation in PIN within each size quartile. Admittedly, sorting into a larger number of portfolios would better control for size, though at the expense of reduced power. On average, I am left with less than 30 firms per portfolio each month. Both the average portfolio PIN and the difference between the low and high PIN decrease with higher size quartiles, confirming the negative relation between size and PIN. For example, the high PIN portfolio for small stocks (portfolio 13) has an average monthly PIN of 0.279 while the high PIN portfolio for large stocks (portfolio 43) has an average PIN of 0.192. Within the low size quartile, the difference between the low and high PIN portfolios is 0.168. This difference falls to 0.102 in the high size quartile. Thus, there is less variation in PIN for larger firms. Easley, Hvidkjaer, and O'Hara (2004) report a similar finding.

Notably, the variation of average short interest across PIN portfolios is quite low. The Wilcoxon test fails to reject a difference in mean short interest between the low PIN and high PIN portfolio for all but the highest size group. Furthermore, the average short interest for the high PIN portfolios is lower than that of the corresponding low PIN portfolio in the first two size quartiles (though not statistically different). Initially at least, this indicates that the level of informed trading is not necessarily increasing in the level of short interest, and further motivates a PIN-based test. As expected, both the quoted spread and the relative spread are decreasing in size and increasing in PIN. Wider spreads are expected in stocks with higher levels of asymmetric information. The structural parameters from the trading model display most of the expected relationships as well. The probability than an information event occurred (α), is increasing in PIN. This parameter drives the difference in PIN across terciles. Interestingly, the probability that new information contains a low signal (δ) is increasing in PIN in the lower three size quartiles. The difference in means for δ in the lower two size quartiles is significant, with differences of 0.058 and 0.041, respectively. Normally, in a random sample of stocks we would not have any theoretical prior for a difference in the probability than new information was bad news [Easley et al.(1996)]. Since I am looking only at firms with large amounts of short selling, this finding here is not surprising. Overall, the portfolio sorts give wide variation in PIN within size quartiles with little variation in short interest, allowing for a clean test of the effect of PIN on abnormal returns. Next, I examine whether or not PIN can predict differences in returns.

4.2. Regressions of PIN-Based Portfolios

If my hypothesis is correct, high PIN portfolios of high short interest stocks should produce negative abnormal returns. Furthermore, high PIN portfolios should underperform low PIN portfolios. Consistent with the literature, I apply the calendar-time portfolio approach advocated by Fama (1998) and Mitchell and Stafford (2000). I regress the time series of the monthly portfolio excess returns on the three Fama and French (1993) factors and a momentum factor [Carhart (1997)]. The excess return for a portfolio is regressed on the four factors in the following equation:

$$R_{pt} - R_{ft} = a + bRMRF_t + sSMB_t + hHML_t + mMOM_t + \varepsilon_t , \qquad (4)$$

where $R_p - R_f$ is the monthly portfolio return in excess of the one-month Treasury bill, $RMRF$ is the market factor, SMB is the size factor, HML is the book-to-market factor, and

MOM is the momentum factor.[17] These return factors are described in detail in Fama and French (1993) and Carhart (1997). The intercept, *a*, is the measure of abnormal performance. The various portfolios contain different subsets of high short interest stocks. If a certain high short interest portfolio underperforms, *a* should be negative and significantly different from zero.

Table 4 shows estimated coefficients and *t*-statistics for regressions of monthly portfolio excess returns on the four return factors from equation (4). The intercept is the measure of monthly abnormal performance. Panel A shows results for equal-weighted portfolios. The high PIN portfolio in three of the four size quartiles produces significantly negative intercepts (two at the 5% level, one at the 10% level), ranging from -0.561 to -1.122% per month. In these cases, PIN is able to predict economically significant underperformance for even large-size equal-weighted portfolios in my sample. Also, within these three size quartiles, the low PIN portfolios do not statistically underperform. Two of the four low PIN portfolios even have positive intercepts, though not significant. Within the lowest size quartile, the low and middle PIN portfolios produce abnormal returns, but the high PIN portfolio does not. The low and middle PIN portfolios have large significant intercepts of -1.242 and -0.988 per month, respectively.

With the exception of the low size portfolios, the results from these regressions support my hypothesis. Specifically, increasingly negative returns as we move from low PIN to high PIN portfolios, and the absence of significantly abnormal returns for the low PIN portfolios in the top three size quartiles are consistent with my hypothesis. The absence of underperformance in the low size, high PIN portfolio, however, is inconsistent with the remaining results. The GRS F-statistic strongly rejects the null that all 12 intercepts are jointly equal to zero, with a value of 309039 (*p-value* = 0.0001).

A more direct test of my hypothesis, in addition to looking at the intercepts of the low and high PIN portfolios individually, is to form a zero investment portfolio that is long the high PIN portfolio and short the low PIN portfolio. If the high PIN portfolios statistically underperforms the low PIN portfolio, the intercept from a regression of the hedge portfolio returns on the four factors will be negative and significant. Panel C of Table 4 contains intercepts and *t*-statistics from these regressions. Consistent with the results from Panel A, the intercept for the three largest size quartiles are negative, indicating underperformance of the high

[17] I am grateful to Kenneth French for making the return factors publicly available on his website. Construction of the return factors is also discussed in great detail on the website.

PIN portfolio. Two of the three are significant at the 5% level, with monthly underperformance of -0.9%. The third portfolio is significant at the 10% level with underperformance of -0.6% per month. Also consistent with the Panel A results, for the lowest size quartile the hedge portfolio delivers positive abnormal returns of 1.1% per month. The evidence from the three largest size quartiles supports my hypothesis, while the lowest size quartile contradicts it.

The underperformance of high PIN portfolios is generally robust to value-weighted returns as well. Panel B of Table 4 shows the coefficient estimates from these regressions. Most of the results from Panel A hold. One notable change is that the high size , high PIN portfolio is now significant at the lower 10% level (*t*-statistic = 1.89). Due to the smaller variation in PIN across larger firms, detecting significant performance is less likely. Easley, Hvidkjaer, and O'Hara (2004) do not find abnormal returns in large size, PIN-based portfolios. The magnitude of underperformance is similar to the equal-weighted portfolios, with monthly abnormal returns ranging from -0.471 to -1.106% per month. The increasingly negative performance as we move from low to high PIN portfolios still holds for the three largest size quartiles. The high size, low PIN portfolios has a positive intercepts, though not significant. Again, the exception to this pattern of increasingly negative performance is the low size, high PIN portfolio. The GRS F-test again rejects the null hypothesis that all 12 intercepts are jointly equal to zero (*p-value* = 0.0007).

The results from Table 4 indicate that PIN is reasonably successful at predicting negative abnormal returns in high short interest stocks, even for value-weighted portfolios. Consistent with my hypothesis, high PIN portfolios generally have larger negative abnormal returns than low PIN portfolios when controlling for size, book-to-market, and momentum. The abnormal returns are economically meaningful. This result represents a new finding for the short interest literature, and is a key result of this paper. Specifically, the relationship between the level of short interest and subsequent returns is unclear, and may be responsible for the ambiguous findings cited by Asquith, Pathak, and Ritter (2005). On the other hand, a measure of the probability of informed trade from the microstructure literature seems to give predictions about underperformance in high short interest stocks. The results of this section also complement the literature that examines the link between information and asset returns [Easley, Hvidkjaer, and O'Hara (2002, 2004)]. If PIN is a successful measure for asymmetric information, then applying such a measure seems to be a useful framework for investigating the returns to high short interest stocks.

4.3. Contemporaneous Returns and Overpricing

The results from Table 4 lend support to an overpricing hypothesis similar to Miller (1977). This is the story usually offered to explain the existence of abnormal returns in the month subsequent to the short interest announcement.[18] Miller shows that under constrained short selling, optimists are price setters and prices are overvalued when informed pessimists are prevented from trading. If the overpricing explanation is accurate, there should be evidence of price increases prior to the short interest announcement. Looking at contemporaneous returns that occur during the month of the short interest announcement should shed light on this issue (recall that the announcement comes near the end of the month). Further, if high PIN portfolios identify stocks in which negatively informed traders are exploiting this overpricing, the high PIN portfolios should have the largest price run-ups.

I repeat the PIN-based portfolio regressions, but with contemporaneous, rather than subsequent, monthly returns. Table 5 shows the results from these regressions. Focusing on the value-weighted returns from Panel B, there is a clear pattern of increasing portfolio returns with higher PIN portfolios, across all four size quartiles. For the two lowest size quartiles, the low PIN portfolios have significantly negative abnormal returns, and the high PIN portfolios have positive abnormal returns (*t*-statistics of 1.72). Size quartile three has significantly positive returns for the middle and high PIN portfolios. For the largest size quartile, all three portfolios have positive and significant intercepts, with the size of the returns increasing with PIN. The magnitudes of the returns in the largest firms is quite large, sometimes in excess of 1.5% per month. This evidence suggests large price run-ups in the month for which the short interest announcement occurs.

A possible scenario that describes this pattern is that negatively informed traders are constrained from shorting, and sit on the sidelines as optimistic traders drive up prices. As the prices increase, the constraint on these negatively informed traders is relaxed when the profit from a short position after prices revert to fundamentals will exceed the costs. As these traders observe the price run-up, they take short positions. When the short interest is announced near the

[18] In the rational expectations model of Diamond and Verrecchia (1987), price adjustments to unexpected changes in short interest would occur immediately after the short interest announcement. Senchack and Starks (1993) offer a more direct test of their predictions, which requires a model of expected short interest.

end of the month, their information is revealed, assuming traders can identify that the short positions come from informed agents (a key assumption). As the market processes this information, subsequent realized returns are low as the overpricing is corrected. In fact, the inability of traders to separate informed from uninformed short selling is another explanation for the delayed reaction of prices to this public information. Investors implementing a trading strategy based on this scenario could infer that high PIN, high short interest stocks are those that are the most overvalued.

Panel C shows the returns to a hedge portfolio long the high PIN portfolio and short the low PIN portfolio. For the smallest three size groups, these intercepts are quite large, between 1.4% and 2.4% per month (t-statistics in excess of four). However, this trading strategy cannot be implemented since the information used to form the portfolios is revealed at the end of the return horizon. Regardless, the results are instructive in that they point to large positive returns in the month before underperformance is detected, suggestive of an overpricing story.

4.4. PIN and Average Returns

One unique finding of this paper is the negative relationship between PIN and returns. That is, I find that abnormal returns for high short interest stocks become more negative with increasing levels of PIN. The existing literature in this area documents a positive relationship between PIN and returns [Easley, Hvidkjaer, and O'Hara (2002, 2004)]. They suggest that PIN could be a priced risk factor, inducing traders to require a positive risk premium for assuming the risk of high information asymmetries. Despite this usual interpretation, it is intuitive to think that, among a sample of high short interest stocks, those with higher levels of informed trading would have lower realized returns. While the intuition is appealing, a theoretical underpinning for this relationship would strengthen the claim.

Irvine (2004) models the returns to highly shorted stocks as a function of short squeeze risk (e.g., the risk of a rapid increase in price due to heavy short covering). He argues that short squeeze risk imposes a liquidity cost on traders with a short position. Therefore, short sellers demand a liquidity premium related to the increased short squeeze risk, which takes the form of *lower* future returns. He models this risk as a peso problem, such that it will affect subsequent returns even if short squeezes rarely occur. Further, he argues that since this risk is negligible at most short interest levels, it is only with high short interest levels that this risk affects returns. Thus, his model provides a theoretical basis for a risk premium that takes the form of lower, not higher, future returns for stocks with high short interest ratios.

The suggestion that PIN will affect highly shorted stocks differently than those with low levels of short interest can be verified empirically. To do this, I look at a sample of all NYSE stocks with reported short interest, rather than just those with a short interest ratio greater than 2.5% in a given month. For comparison to their paper, I use the annual PIN computed by Easley, Hvidkjaer, and O'Hara (2002), rather than a monthly PIN.[19] This data ends in 2001, so the sample period for these tests is January 1992 to December 2001.[20] Each month, I form 30 portfolios from sorts on the short interest ratio and PIN. First, I create short interest deciles, and within each decile, sort into PIN terciles. On average, this gives 1,424 firms per month. Since the monthly cutoff for the high short interest sample (2.5% short interest ratio) varies between the 80th and 90th percentile, the high short interest sample used in the previous tests of this paper fall into the top two short interest deciles.

Each month, I repeat the regression from equation (4) for the 30 portfolios. Given the finding of a positive relationship between PIN and returns from previous papers, and the negative relationship I discover for high short interest stocks, I expect the difference in risk-adjusted abnormal returns for the high PIN minus the low PIN portfolio to be positive for the bottom eight short interest deciles and to be negative for the top two deciles.

Table 6 shows the results of these tests; several interesting results are worth highlighting. From the equal-weighted portfolio returns of Panel A, only three of the 30 intercepts are significant, and two of them are in the highest short interest decile. The high SI, middle PIN portfolio and the high SI, high PIN portfolio have monthly underperformance of -0.9524% and -0.8145% monthly, with *t*-statistics above three. In addition to being two of the only three significant intercepts, they are also the two largest in absolute magnitude of all 30 intercepts. The other significant intercept is the high PIN portfolio for SI decile three (0.5080%

---

[19] I thank Soeren Hvidkjaer for making this data available on his website. Note that this version of PIN is calculated from the continuous time version of their sequential trade model, where $PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_S + \varepsilon_B}$. See Easley, Hvidkjaer, and O'Hara (2002) for details.

[20] The correlation between the annual average of the monthly PIN estimates and the annual PIN from Easley, Hvidkjaer, and O'Hara (2002) among the sample of stocks for which both are available (the high short interest stocks) is 0.64. The mean PIN among this sample is 0.16 for both PIN measures.

monthly). When looking at the intercepts from a hedge portfolio that is long high PIN and short low PIN stocks within each short interest decile, the intercept for both deciles nine and ten (the high short interest stocks) are negative. Furthermore, the top decile is significant, with monthly underperformance of -0.5747%, again the largest in magnitude of all the hedge portfolio intercepts. Only one other hedge portfolio intercept is significant; short interest decile three has a positive abnormal return of 0.4721% per month, which supports the findings of Easley, Hvidkjaer, and O'Hara (2002). Finally, when looking at the intercept on a portfolio that is long high SI stocks and short low SI stocks within PIN group, the results are equally illustrative. For example, the high SI minus low SI portfolio for high PIN stocks has a significant negative return of -1.13% per month (*t*-stat = -4.05). Within the low PIN group, the high SI minus low SI intercept is not significant. This non-result is equally supportive of my hypothesis. If the underperformance of high short interest stocks is driven by stocks with high levels of informed trading, then high SI stocks should not underperform low SI stocks when informed trading is low. So, in addition to PIN being important for high SI stocks, the reverse is also true. Short interest levels are important for high PIN stocks, but not for low PIN stocks. This finding strengthens the argument that PIN is important for stock returns.

The value-weighted results from Panel B are similar, but not quite as strong. Again, there are only two out of 30 significant intercepts, one of which is in the high SI decile. None of the hedge portfolios deliver abnormal returns significant at the 5% level, but the portfolio from decile nine is the largest in magnitude (-0.659%), and the strongest statistically (*t*-statistic of -1.80). Both of the two highest SI decile hedge portfolios deliver negative returns.

In general, these results endorse the notion of an information-based risk for high short interest stocks that takes the form of lower, not higher, future returns. Further, the hypothesis that high PIN portfolios with high levels of short interest should underperform low PIN portfolios with high short interest is supported. Additionally, high SI portfolios underperform low SI portfolios when PIN is high, but not when PIN is low. I believe the results from this section scratch the surface of an area with considerable promise for future research.

5. Cross-Sectional Analysis

In this section, I consider the role of PIN as a potential determinant of the cross-section of returns to high short interest stocks. The objective is to establish whether or not PIN makes a marginal contribution to the determinants of returns, above and beyond other factors and characteristics known to influence average returns (e.g., size, book-to-market, momentum, liquidity). Rather than implement a portfolio approach as in previous sections, I examine returns at the individual firm level. Since portfolio formation can conceal variation in firm characteristics that are relevant to returns [Roll (1977)], this approach should offer increased power to identify a relation between PIN and the returns to high short interest stocks. Looking at returns to individual firms also allows the investigation of several firm characteristics simultaneously. Finally, this approach alleviates some of the concerns of using attribute sorted portfolios in asset pricing tests, as described by Ferson et al. (1999).

5.1. A Model of Returns

Given a model with *K* risk factors and *M* firm characteristics, the standard Fama and MacBeth (1973) regression in month *t* is expressed as:

$$R_{it} - R_{ft} = a_{0t} + \sum_{k=1}^{K} \beta_{kit} F_{kt} + \sum_{m=1}^{M} \gamma_{mt} Z_{mit} + e_{it} \,, \qquad (5)$$

where $R_i - R_f$ is the excess return for firm *i*, β_{ki} is firm *i*'s loading on risk factor F_k, and γ_m is the slope coefficient (or characteristic reward) for firm characteristic Z_{mi}. Under the null hypothesis, returns are independent of the firm characteristics and $\gamma_m = 0$, $m = 1,2,\ldots M$. Since the true factor loadings are not known, they must be estimated in a first pass regression. The measurement error from these estimates, which are independent variables in the second pass regression, leads to the familiar errors-in-variables problem of potentially biased standard errors and inflated *t*-statistics. To avoid this problem, I follow the procedure of Brennan et al. (1998) in which *risk-adjusted* returns are the dependent variables. In this context, risk-adjusted returns are defined as:

$$R_{it}^{*} \equiv R_{it} - R_{ft} - \sum_{k=1}^{K} \widehat{\beta}_{kit} F_{kt} \qquad (6)$$

where the K risk factors are given by the Fama and French three-factor model, and $\widehat{\beta}_{ki}$ are the estimated factor loadings. Under this process, measurement error from the factor loading estimates is now contained in the dependent variable. The coefficients on the firm characteristics are now estimated with the following monthly regression:

$$R_{it}^{*} = a_{0t} + \sum_{m=1}^{M} \gamma_{mt} Z_{mit} + e_{it}^{*}, \qquad (7)$$

and the error term e_{it}^{*} contains measurement error arising from the factor loading estimates. To the extent that this measurement error is correlated with the firm characteristics Z_{mi}, the estimates of γ_{mt} are correlated with the factor realizations, introducing bias to the time-series average of these monthly estimates. To account for this possibility, two separate values are reported for each γ_m. The first is the time-series average of the monthly coefficient estimates, and is simply the standard Fama-MacBeth estimator: $\hat{\gamma}_m = \sum_{t=1}^{T} \hat{\gamma}_{mt}$. The second reported value is the constant term from the following OLS regression of the monthly characteristic slope estimates on the three Fama-French factor realizations:

$$\hat{\gamma}_{mt} = \gamma_{mp} + \sum_{k=1}^{K} \beta_k F_{kt} + u_t \qquad (8)$$

where $\hat{\gamma}_{mt}$ is the estimate of γ_m in month t. The intercept from this regression, γ_{mp}, is purged of any effects from the factor realizations and is unbiased even when the beta measurement error is correlated with the firm characteristics [see Brennan et al. (1998)]. This "purged" estimate was first used by Black and Scholes (1974).

5.2. Data and Estimation Procedure

In month *t*, factor loadings are estimated for all firms which had at least 24 monthly returns over the previous 60 months, by regressing the firm's excess return on the three Fama-French factors, *RMRF*, *SMB*, and *HML*. To account for potential problems from thin trading, the betas are estimated with one lag, following Dimson (1979). These factor loading estimates are used to calculate the monthly risk-adjusted return from equation (6). Then, these risk-adjusted returns are regressed on other firm characteristics, to check their marginal contribution to the determinants of average returns. My objective is to test the explanatory power of PIN and the short interest ratio for the returns to high short interest stocks. Because these variables could be correlated with other firm characteristics shown to influence expected returns, I control for those characteristics in the regressions. Specifically, the empirical asset pricing literature has identified the effects of size [Banz (1981)], book-to-market [Stattman (1980), Rosenberg et al. (1985), Chan and Chen (1991), Fama and French (1992)], return momentum [Jegadeesh and Titman (1993)], and liquidity [Amihud and Mendelson (1986), Brennan and Subrahmanyam (1996)] on the cross-section of stock returns.

Following Fama and French (1992), size (*SIZE*) is defined as the market value of equity in June of year *t*, and the book-to-market ratio (*BM)* is defined as the book value of equity from COMPUSTAT in year *t-1* divided by the market value of equity in December of year *t-1*. These variables are matched to the monthly stock returns in July of year *t* to June of year *t+1*. Book-to-market values are trimmed at the 0.5% and 99.5% levels, as in Fama and French (1992). Return momentum (*MOM*) is defined as the sum of the six monthly returns ending two months prior to the dependent variable return. I use two different measures to control for liquidity: a measure of share turnover (*TURN*), defined as the number of shares traded in the previous month divided by the shares outstanding two months prior; and the dollar volume of shares (*DVOL*) traded in the month two months prior to the dependent variable return.[21] Finally, I include the PIN estimate and the short interest ratio (*SI*) from the previous month, and an interaction term that is the product of PIN and *SI* (*PINxSI)*. These three variables are the focus of the tests. With the

[21] Lee and Swaminathan (2000) find turnover to be a predictor of future returns. The use of trading volume is motivated by Stoll (1978) who finds volume to be a major determinant of the bid-ask spread, and Brennan and Subrahmanyam (1995), who document its role as a measure of liquidity.

exception of return momentum, all independent variables enter the regression as the natural logarithm of the calculated value.

5.3. Regressions of High Short Interest Portfolio Firms

Table 7 contains the results of these cross-sectional regressions for the high short interest stocks that make up the portfolio sample in previous sections (e.g., stocks with a short interest ratio greater than 2.5% the previous month). The "Raw" estimates are the standard Fama-MacBeth estimator, and the "Purged" estimates are the intercept from equation (8). The baseline case from the top two rows of the table regresses the risk-adjusted returns on *SIZE* and *BM*. These two variables are unable to completely explain returns, as indicated by the significantly negative intercept. Once other firm characteristics are added to the regressions, the intercepts are insignificant. When *PIN* and *SI* are included as independent variables, *PIN* offers statistically significant explanatory power (*t*-statistic of -2.02) while *SI* does not. Notably, the coefficient on *PIN* is negative, consistent with the portfolio results from previous sections. The negative sign supports the hypothesis that PIN and returns are negatively related for high short interest stocks. Some of the statistical significance is lost in the purged estimate. When the regressions include *MOM* and *TURN* , the coefficient on *PIN* is still negative, but only marginally significant at the 10% level for the raw estimates. The purged estimates lack significance. When *DVOL* is used instead of *TURN* to control for liquidity effects, *PIN* is significant (*t*-statistic of -2.00) for the raw estimate. The sign on *DVOL* is negative, but insignificant. The next two sets of tests find a positive sign on *DVOL*. Karpoff (1988) argues that the positive volume-return relation in equity markets is due to costly short selling. Finally, when *SI* and *PIN* are replaced with the interaction term *PINxSI*, the purged coefficient estimate is negative and significant, even when *MOM* and *TURN* are included as explanatory variables.

Table 7 confirms the negative relationship between PIN and returns for high short interest stocks, since all coefficient estimates for PIN are negative, and several are significant. Given the small sample size, I suspect that the results would be even stronger with a longer time series of data. Regardless, *PIN* seems to offer more explanatory power for the average returns to these high short interest stocks than the short interest ratio, supporting the hypotheses put forth earlier in the paper. However, the results are strongest when short interest and PIN are considered jointly as an interaction term. Given the significant contribution of these characteristic variables, the null hypothesis that expected returns are completely described by the factor model is rejected. One interpretation of this finding is that expected returns for these high short interest stocks are determined by non-risk characteristics in addition to the priced risk factors. An alternate explanation is that the risk model is not fully specified, and the firm characteristics proxy for some omitted risk factor.

Since PIN is an estimated variable, there is potential for an errors-in-variables problem due to estimation error. Easley, Hvidkjaer, and O'Hara (2002) explore this issue, and find that correcting for any potential bias is immaterial. Also, they state that any bias would be against finding significance for the PIN coefficient. Nonetheless, this potential bias implies caution when interpreting the results of these coefficient estimates.

5.4. Regressions of High Short Interest Firms: All Months

Table 8 reports the coefficient estimates when the cross-sectional regressions are repeated, but for the high short interest stocks in all months that they fall into the sample. For example, the stocks included in these regressions need not have had a short interest ratio greater than 2.5% in the preceding month, but must have crossed that threshold at some point during the sample period. The increased number of firms per month in these regressions should lead to increased statistical power. As in Table 7, all coefficient estimates for *PIN* are negative, and significant in most cases. However, a key difference in this table is that *SI* achieves strong statistical significance (*t*-statistics of 4.6 and greater) in every regression, implying an important role in explaining the cross-section of returns to these stocks. Further, the magnitude of the *SI* coefficients are, on average, slightly larger than the coefficients for *PIN*. The increased explanatory power of *SI* for this sample could be due to the high first-order autocorrelation of short interest ratios, whereas the monthly PIN estimates are not as highly autocorrelated. The interaction term *PINxSI* is also negatively significant in all cases, and economically large. The results from this larger sample size substantiate the inferences from the previous section. Namely, I interpret these results as confirming a role for both PIN and the short interest ratio in explaining average returns to high short interest stocks. The larger sample size used in these regressions delivers increased statistical power.

5.5. Regressions of All Short Interest Firms

Finally, I perform cross-sectional regressions on a sample of stocks that is not limited to high levels of short interest. In these tests, I again use the annual PIN measure from Easley, Hvidkjaer, and O'Hara (2002), which means the sample period ends in 2001 rather than 2002. These tests will illustrate whether the results from the previous two sections are unique to a sample of high short interest stocks. Specifically, the tests will reveal if the short interest ratio continues to offer explanatory power for average returns, and whether there remains a relation between PIN and returns. My prior is that short interest will be insignificant, and PIN will have a positive relation with returns. Examination of Table 9 shows that in fact, *SI* continues to have significant explanatory power for the returns to stocks will all levels of short interest. The coefficient estimates for *SI* are all negative, as expected, but with high *t*-statistics. The negative sign is quite surprising given the low short interest ratios of most stocks. Also, previous studies that included a random sample of short interest stocks (rather than a sample of only high short interest stocks) failed to detect a relation between short interest and returns [e.g., Woolridge and Dickinson (1994)]. Estimates for the coefficient on *PIN* are positive in every case, and statistically significant in most cases. The positive sign supports the results of Table 6 that most stocks display a positive PIN-return relationship, and complements the findings of Easley, Hvidkjaer, and O'Hara (2002). The interaction term between short interest and PIN is also negative and significant. Again, the significance of this result is surprising giving the low levels of short interest on average for this sample. The finding that short interest seems to have an important role in explaining the cross-section of returns presents opportunities for further research.

Overall, the cross-sectional regressions offer compelling evidence on the role of PIN for the expected returns to both high short interest stocks, and all stocks in general. As hypothesized, PIN has a negative relation with the returns to high short interest stocks, but a positive relation to stocks with all levels of short interest. The importance of considering the effects of asymmetric information in explaining the returns to high short interest stocks is bolstered by these findings.

6. Concluding Remarks

This paper advocates an alternate approach to testing the information content of short interest. I estimate the probability of informed trading for a sample of high short interest stocks to see if differences in PIN help to explain different returns for size-neutral portfolios. In general, the results support the use of PIN as a measure of informed trading in highly shorted stocks. I find that high PIN stocks have significant negative abnormal returns. Low PIN stocks, on the other hand, generally do not underperform. The differences in PIN can be motivated by different reasons for short selling, such as arbitrage versus information-based short selling. Cross-sectional tests confirm a negative relation between PIN and average returns for high short interest stocks. When stocks with low to moderate short interest are included, this relationship reverts to a positive one.

There are several opportunities for further study. Easley, Hvidkjaer, and O'Hara (2004) create a traded factor based on PIN and incorporate this measure into a factor asset pricing model. They find some evidence that such a factor helps explain returns to portfolios formed on the basis of size and PIN. A natural extension of this study would be to apply such a factor to high short interest stocks, but should require a longer time series of data.

Given the renewed interest in tests of Miller (1977), it would be interesting to further investigate the degree to which PIN may be picking up effects related to either divergence of opinion or short sale constraints. This presents an opportunity to disentangle, if possible, the effects of private information from heterogeneous expectations based on public information. This task would require some understanding about the different effects of asymmetric information and divergence of opinion. Also, this would require a more direct measure of short sale constraints. While no exact measure is publicly available, suitable proxies, such as institutional ownership or breadth of ownership, exist.

Lastly, this study does not incorporate trading costs into the analysis, so it remains to be seen whether or not such a trading strategy is profitable after costs. The answer to this question depends partly on how long-lived the information in the high PIN portfolios are. For example, by altering the selection criteria, and allowing a stock to remain in the portfolio even after its PIN falls below a certain cutoff value, we could test whether or not the abnormal returns are persistent. Such an approach is applied in Asquith, Pathak, and Ritter (2005), and Desai et al.(2000). Not surprisingly, they find different results. This contradiction presents an occasion to examine if high PIN stocks also have long-lived information.

References

Amihud, Y., and H. Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics* 17, 223-249.

Arnold, T., A. W. Butler, T. F. Crack, and Y. Zhang, 2005, The information content of short interest: A natural experiment, *Journal of Business* 78, 1307-1335.

Asquith, P., and L. Meulbroek, 1996, An empirical investigation of short interest, Working Paper 96-012, Harvard Business School.

Asquith, P., P. A. Pathak, and J. R. Ritter, 2005, Short interest, institutional ownership, and stock returns, *Journal of Financial Economics* 78, 243-276.

Banz, Rolf W., 1981, The relationship between return and market value of common-stocks, *Journal of Financial Economics* 9, 3-18.

Black, F., and M. S. Scholes, 1974, The effects of dividend yield and dividend policy on common stock prices and returns, *Journal of Financial Economics* 1, 1-22.

Boehme, R. D., B. R. Danielsen, and S. M. Sorescu, 2005, Short-sale constraints, dispersion of opinion and overvaluation, *Journal of Financial and Quantitative Analysis* Forthcoming.

Brennan, M. J., T. Chordia, and A. Subrahmanyam, 1998, Alternative factor specifications, security characteristics, and the cross-section of expected stock returns., *Journal of Financial Economics* 49, 345-373.

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

Brennan, M. J., and A. Subrahmanyam, 1996, Market microstructure and asset pricing: On the compensation for illiquidity in stock returns, *Journal of Financial Economics* 41, 441-464.

Brent, A., D. Morse, and E. K. Stice, 1990, Short Interest - Explanations and Tests, *Journal of Financial and Quantitative Analysis* 25, 273-289.

Carhart, M., 1997, On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 57-82.

Chan, K. C., and N. Chen, 1991, Structural and Return Characteristics of Small and Large Firms, *Journal of Finance* 46, 1467-1484.

Chen, J., H. Hong, and J. C. Stein, 2002, Breadth of ownership and stock returns, *Journal of Financial Economics* 66, 171-205.

Christophe, S. E., M. G. Ferri, and J. J. Angel, 2004, Short-selling prior to earnings announcements, *Journal of Finance* 59, 1845-1875.

Chung, K. H., and M. Li, 2003, Adverse selection costs and the probability of information-based trading, *The Financial Review* 38, 257-272.

Chung, K. H., M. Li, and T. H. McInish, 2005, Information-based trading, price impact of trades, and trade autocorrelation, *Journal of Banking and Finance* 29, 1645-1669.

Cohen, K., K. B. Diether, and C. J. Malloy, 2005, Supply and demand shifts in the shorting market, Working Paper, Yale University, Ohio State University, and London Business School.

D'Avolio, G., 2002, The market for borrowing stock, *Journal of Financial Economics* 66, 271-306.

Dechow, P. M., A. P. Hutton, L. Meulbroek, and R. G. Sloan, 2001, Short-sellers, fundamental analysis and stock returns, *Journal of Financial Economics* 61, 77-106.

Desai, H., K. Ramesh, S. R. Thiagarajan, and B. V. Balachandran, 2002, An investigation of the informational role of short interest in the Nasdaq market, *Journal of Finance* 57, 2263-2287.

Diamond, D. W., and R. E. Verrecchia, 1987, Constraints on Short-Selling and Asset Price Adjustment to Private Information, *Journal of Financial Economics* 18, 277-311.

Dimson, E., 1979, Risk measurement when shares are subject to infrequent trading, *Journal of Financial Economics* 7, 197-226.

Easley, D., M. O'Hara, J. Paperman, 1998, Financial analysts and information-based trade, *Journal of Financial Markets* 1, 175-201.

Easley, D., S. Hvidkjaer, and M. O'Hara, 2002, Is information risk a determinant of asset returns?, *Journal of Finance* 57, 2185-2221.

Easley, D., S. Hvidkjaer, and M. O'Hara, 2004, Factoring information into returns, Working Paper, Cornell University and University of Maryland.

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

Easley, D., N. M. Kiefer, M. O'Hara, and J. B. Paperman, 1996, Liquidity, Information, and Infrequently Traded Stocks, *Journal of Finance* 51, 1405-1436.

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

Ellis, K., R. Michaely, and M. O'Hara, 2000, The accuracy of trade classification rules: Evidence from Nasdaq, *Journal of Financial and Quantitative Analysis* 35, 529-551.

Fama, E. F., 1998, Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics* 49, 283-306.

Fama, E. F., and K. R. French, 1992, The Cross-Section of Expected Stock Returns, *Journal of Finance* 47, 427-465.

Fama, E. F., and K. R. French, 1993, Common Risk-Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics* 33, 3-56.

Ferson, W., S. Sarkissian, and T. Simin, 1999, The alpha factor asset pricing model: A parable, *Journal of Financial Markets* 2, 49-68.

Figlewski, S., 1981, The Informational Effects of Restrictions on Short Sales - Some Empirical-Evidence, *Journal of Financial and Quantitative Analysis* 16, 463-476.

Figlewski, S., and G. P. Webb, 1993, Options, Short Sales, and Market Completeness, *Journal of Finance* 48, 761-777.

Fu, H., 2002, Information asymmetry and the market reaction to equity carve-outs, Working Paper, University of Minnesota.

Gintschel, A., 2001, Short interest on Nasdaq, Working Paper, Emory University.

Heidle, H. G., and R. D. Huang, 2002, Information-based trading in dealer and auction markets: An analysis of exchange listings, *Journal of Financial and Quantitative Analysis* 37, 391-424.

Irvine, P., 2004, Liquidity and asset prices: The case of the short squeeze and the returns to the short position, Working Paper, University of Georgia.

Jegadeesh, N., and S. Titman, 1993, Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance* 48, 65-91.

Jones, C. M., and O. A. Lamont, 2002, Short sale constraints and stock returns, *Journal of Financial Economics* 66, 207-239.

Jones, S. L., and G. Larsen, 2004, The information content of short sales, in Frank J. Fabozzi, ed.: *Short selling: Strategies, risks, and rewards* (John Wiley & Sons, Inc., Hoboken, NJ).

Karpoff, J. M., 1988, Costly short sales and the correlation of returns with volume, *journal of Financial Research* 11, 173-188.

Lamont, O., 2004, Short sale constraints and overpricing, in Frank J. Fabozzi, ed.: *Short selling: Strategies, risks, and rewards* (John Wiley & Sons, Inc., Hoboken, NJ).

Lee, C. M. C., and B. Radhakrishna, 2000, Inferring investor behavior: Evidence from TORQ Data, *Journal of Financial Markets* 3, 83-111.

Lee, C. M. C., and M. J. Ready, 1991, Inferring Trade Direction from Intraday Data, *Journal of Finance* 46, 733-746.

Lee, C. M. C., and B. Swaminathan, 2000, Price momentum and trading volume, *Journal of Finance* 55, 2017-2069.

Miller, E. M., 1977, Risk, Uncertainty, and Divergence of Opinion, *Journal of Finance* 32, 1151-1168.

Mitchell, M. L., and E. Stafford, 2000, Managerial decisions and long-term stock price performance, *Journal of Business* 73, 287-329.

Nagel, S., 2005, Short sales, institutional investors and the cross-section of stock returns, *Journal of Financial Economics* 78, 277-309.

Odders-White, E. R., 2000, On the occurrence and consequences of inaccurate trade classification, *Journal of Financial Markets* 3, 259-286.

Peterson, M. A., and E. R. Sirri, 2003, Order preferencing and market quality on US equity exchanges, *Review of Financial Studies* 16, 385-415.

Reed, A. V., 2002, Costly short-selling and stock price adjustment to earnings announcements, Working Paper, University of North Carolina.

Roll, R., 1977, Critique of Asset Pricing Theory Tests .1. Past and Potential Testability of Theory, *Journal of Financial Economics* 4, 129-176.

Rosenberg, B., K. Reid, and R. Lanstein, 1985, Persuasive evidence of market inefficiency, *Journal of Portfolio Management* 11, 9-16.

Senchack, A. J., and L. T. Starks, 1993, Short-Sale Restrictions and Market Reaction to Short-Interest Announcements, *Journal of Financial and Quantitative Analysis* 28, 177-194.

Stattman, D., 1980, Book values and stock returns, *The Chicago MBA: A Journal of Selected Papers* 4, 25-45.

Stoll, H. R., 1978, The Supply of Dealer Services in Securities Markets, *Journal of Finance* 33, 1133-1151.

Vega, C, 2005, Stock price reaction to public and private information, *Journal of Financial*

Economics Forthcoming.

Woolridge, J. R., A. Dickinson, 1994, Short selling and common stock prices, *Financial Analysts Journal* 50, 30-38.

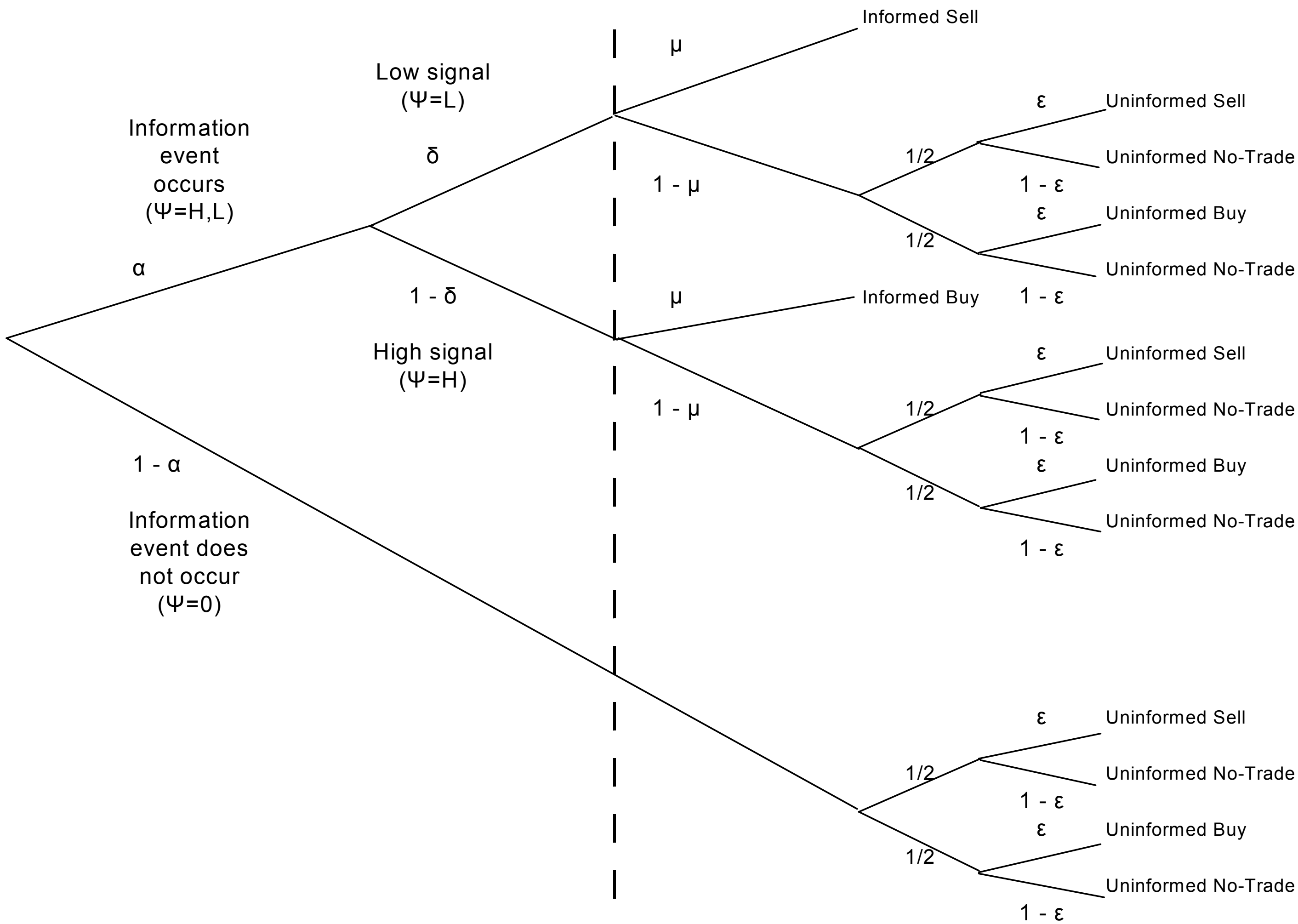

Figure 1: Decision Tree Diagram of the Trading Process
From Easley, Kiefer, and O’Hara (1997b)

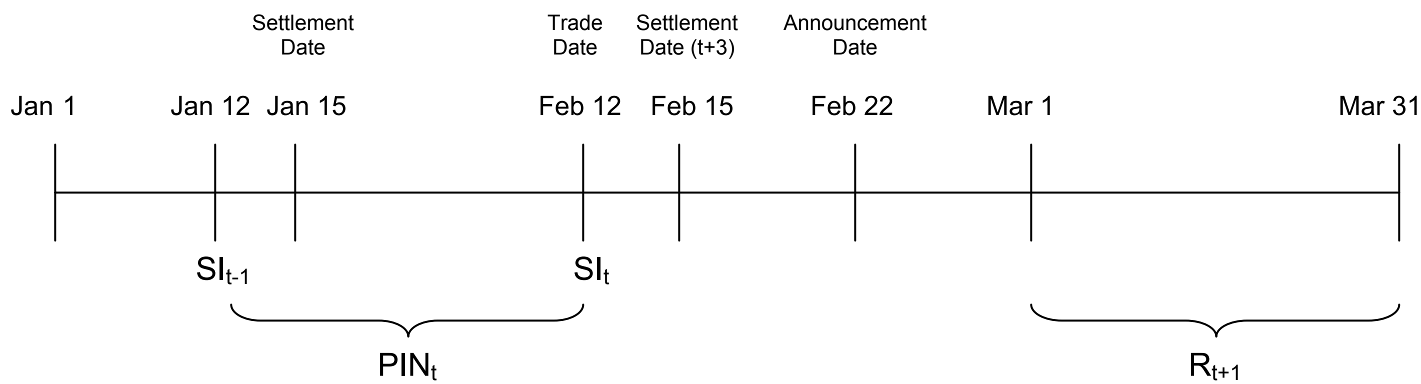

Figure 2: Timeline of Variable Measurement

The above timeline gives an example of the measurement intervals for the variables used in the study. The above example shows the timing of February short interest. February's short interest is collected for trades that settle as of February 15th. This represents all outstanding short positions as of February 12th, since settlement is t+3. Therefore, the PIN that accompanies February short interest is measured over the interval between January's short interest trade date (January 12th) and February's short interest trade date (February 12th). The public announcement of February short interest occurs on February 22nd. Portfolios are formed at the end of February and subsequent performance is calculated with the March monthly return.

Table 1
Descriptive Statistics for High Short Interest Firms

Mean monthly statistics for all firms that fall into the high short interest sample for at least one month during the sample period, January 1992 – December 2002, by year. Panel A includes only firm-months where short interest is greater than or equal to 2.5%. Panel B contains only firm months where short interest is less than 2.5%, and Panel C includes all months. For comparison, Panel D contains statistics for all firm-months (not just high short interest firms) where short interest is less than 2.5%. No. Firms per Month is the monthly average number of firms in the sample. Short interest ratio is the monthly short interest divided by the number of shares outstanding. Size is the market value of equity, in millions of dollars. R_{t+1} is the subsequent month's return and $HPR_{(t-12,t-1)}$ is the prior 12 months' holding period return.

Panel A: High SI Portfolio Months: SI >= 2.5% (N=39,617)

Year	No. Firms per Month	Short Interest Ratio	Size	Price	R_{t+1}	$HPR_{(t-12,t-1)}$
1992	133.9	0.0571	1151.81	25.63	0.0107	0.3244
1993	178.3	0.0573	1478.89	28.36	0.0154	0.2690
1994	229.0	0.0606	1592.45	26.51	-0.0092	0.1705
1995	262.6	0.0634	1941.68	27.55	0.0207	0.1727
1996	306.4	0.0631	2220.69	29.18	0.0149	0.3290
1997	400.2	0.0645	2710.62	30.73	0.0122	0.2563
1998	406.9	0.0666	2995.43	29.23	-0.0044	0.1187
1999	320.7	0.0649	3597.82	28.53	-0.0014	0.0301
2000	296.1	0.0672	3419.81	30.79	0.0169	0.1554
2001	369.0	0.0648	3381.66	31.66	-0.0032	0.1880
2002	429.1	0.0675	2823.90	28.96	-0.0182	0.0510
Total	302.4	0.0634	2476.30	28.81	0.0050	0.1877

Panel B: High SI: SI < 2.5% (N=89,300)

Year	No. Firms per Month	Short Interest Ratio	Size	Price	R_{t+1}	$HPR_{(t-12,t-1)}$
1992	596.9	0.0070	2442.85	29.72	0.0131	0.2301
1993	747.1	0.0073	2506.65	30.55	0.0154	0.2519
1994	787.4	0.0080	2428.61	28.47	-0.0009	0.1069
1995	677.2	0.0092	2790.60	65.03	0.0249	0.1443
1996	695.5	0.0102	3270.49	49.15	0.0202	0.2805
1997	654.7	0.0112	3983.75	36.70	0.0197	0.2903
1998	608.5	0.0112	5175.06	35.07	0.0031	0.2008
1999	632.0	0.0111	6016.43	30.73	0.0083	0.0684
2000	847.9	0.0099	5962.00	29.85	0.0232	0.1253
2001	702.1	0.0112	5862.63	28.82	0.0070	0.2281
2002	550.9	0.0123	5959.93	27.65	-0.0078	0.1223
Total	681.7	0.0099	4205.54	35.67	0.0115	0.1859

Table 1 (continued)

Panel C: High SI All Months (N=128,917)

Year	No. Firms per Month	Short Interest Ratio	Size	Price	R_{t+1}	$HPR_{(t-12,t-1)}$
1992	730.8	0.0161	2206.43	28.97	0.0129	0.2466
1993	925.4	0.0169	2309.93	30.14	0.0154	0.2557
1994	1016.4	0.0199	2240.65	28.03	-0.0028	0.1213
1995	939.8	0.0243	2552.94	54.51	0.0238	0.1525
1996	1001.9	0.0264	2949.79	43.26	0.0186	0.2952
1997	1054.8	0.0314	3496.80	34.43	0.0169	0.2767
1998	1015.4	0.0334	4299.39	32.72	0.0000	0.1674
1999	952.7	0.0292	5201.01	30.01	0.0051	0.0557
2000	1144.0	0.0248	5290.35	30.09	0.0218	0.1325
2001	1071.1	0.0298	5008.78	29.80	0.0037	0.2136
2002	980.0	0.0365	4589.48	28.23	-0.0120	0.0917
Total	984.1	0.0262	3639.22	33.68	0.0094	0.1824

Panel D : All Firms: SI < 2.5% (N=176,964)

Year	No. Firms per Month	Short Interest Ratio	Size	Price	R_{t+1}	$HPR_{(t-12,t-1)}$
1992	1070.3	0.0056	2919.34	30.23	0.0118	0.2138
1993	1251.8	0.0056	2811.68	37.16	0.0144	0.2291
1994	1379.8	0.0060	2602.33	40.30	-0.0017	0.0964
1995	1415.4	0.0065	2930.34	46.53	0.0216	0.1223
1996	1477.1	0.0070	3499.79	38.48	0.0186	0.2431
1997	1499.4	0.0077	4269.70	33.84	0.0191	0.2758
1998	1509.1	0.0078	5234.74	32.74	0.0015	0.1895
1999	1528.4	0.0076	6180.36	29.68	0.0037	0.0376
2000	1373.1	0.0082	7411.61	29.27	0.0211	0.0979
2001	1183.1	0.0096	7817.27	28.33	0.0072	0.2128
2002	1059.5	0.0104	7471.94	27.35	-0.0063	0.1196
Total	1340.6	0.0074	4831.74	33.99	0.0101	0.1671

Table 2
Cross-Sectional Correlations

Time-series average of the monthly cross-sectional correlation between variables. SI is the monthly short interest ratio, Spr ($) is the monthly average of the trade-weighted daily quoted half-spread, Spr (%) is the monthly average of the trade-weighted daily relative half-spread, and Spread (Eff.) is the effective spread. PIN is the probability of informed trading, calculated as $\frac{\alpha\mu}{\alpha\mu + (1-\alpha\mu)\varepsilon}$. α is the probability than an information event occurred, δ is the probability that an information signal is a low signal, ε is the probability that an uninformed traders transacts if selected to trade, and μ is the probability that a trader is informed. Size is the market value of equity, No. Trades is the monthly average of the daily number of trades, and R_{t+1} is the subsequent month's return.

	SI	Spr ($)	Spr (%)	Spread (Eff.)	PIN	α	δ	ε	μ	Size	No. Trades	R_{t+1}
SI	1											
Spread ($)	0.081	1										
Spread (%)	0.100	0.031	1									
Spread (Eff.)	0.071	0.839	0.046	1								
PIN	0.062	0.086	0.262	0.104	1							
α	-0.024	-0.030	-0.024	-0.011	0.694	1						
δ	-0.027	-0.025	0.198	0.008	0.059	0.038	1					
ε	-0.074	-0.142	-0.243	-0.118	-0.566	-0.262	-0.109	1				
μ	0.064	0.099	0.259	0.104	-0.083	-0.599	0.070	-0.013	1			
Size	-0.127	-0.114	-0.346	-0.101	-0.162	0.095	-0.112	0.203	-0.244	1		
No. Trades	-0.027	-0.185	-0.370	-0.143	-0.216	0.086	-0.190	0.330	-0.297	0.690	1	
R_{t+1}	-0.021	-0.004	-0.015	-0.012	-0.001	0.005	-0.004	-0.010	-0.010	0.010	-0.004	1

Table 3
Summary Statistics for Portfolios Sorted by Firm Size and PIN

This table describes the composition of the 12 portfolios created from monthly sorts of firm size and PIN. Firms are grouped first by size quartile, and within quartile, by PIN. For example, portfolio 23 contains the firms in the highest PIN group within the second size quartile. Nfirms is the monthly average number of firms in a given portfolio, Size is the market value of equity, PIN is the probability of informed trading, Spr ($), Spr (%), and Spr (Eff.) is the monthly average of the trade-weighted daily quoted half-spread, relative spread, and effective spread, respectively. α is the probability than an information event occurred, δ is the probability that an information signal is a low signal, ε is the probability that an uninformed traders transacts if selected to trade, μ is the probability that a trader is informed. The numbers are monthly averages, and for each size quartile a Wilcoxon Rank Sum test (Zstat) is performed to test for difference in means across the low and high PIN portfolios.

Firm Size	**PIN**	Portfolio	nfirms	Size	PIN	SI	Spr ($)	Spr (%)	Spr (Eff.)	α	δ	ε	μ
	Low	11	23.7	232.1	0.1109	0.0778	0.0842	0.0120	0.0599	0.2681	0.3426	0.5575	0.2864
Low		12	23.6	228.0	0.1788	0.0763	0.0830	0.0120	0.0593	0.4315	0.3657	0.5363	0.2623
	High	13	22.6	198.7	0.2791	0.0754	0.0878	0.0129	0.0634	0.5861	0.4005	0.4950	0.2880
		13-11		-33.45	0.1682	-0.002	0.0036	0.0009	0.0035	0.3180	0.0578	-0.0625	0.0016
		Zstat		(-3.88)	(13.99)	(-1.11)	(0.52)	(1.61)	(1.01)	(13.99)	(3.73)	(-12.36)	(0.73)
		21	23.4	663.1	0.1063	0.0647	0.0791	0.0078	0.0555	0.2949	0.3018	0.5651	0.2446
		22	23.5	660.3	0.1640	0.0658	0.0797	0.0078	0.0562	0.4593	0.3215	0.5455	0.2245
		23	22.4	635.4	0.2388	0.0642	0.0813	0.0080	0.0586	0.5907	0.3424	0.5115	0.2429
		23-21		-27.74	0.1325	-0.001	0.0022	0.0002	0.0031	0.2959	0.0407	-0.0536	-0.0017
		Zstat		(-1.32)	(13.99)	(-0.35)	(0.65)	(0.82)	(0.89)	(13.99)	(2.56)	(-12.09)	(-0.99)
		31	23.5	1611.7	0.0979	0.0585	0.0755	0.0056	0.0526	0.3160	0.2788	0.5736	0.2143
		32	23.5	1571.5	0.1480	0.0595	0.0763	0.0057	0.0531	0.4731	0.2771	0.5570	0.1981
		33	22.5	1509.8	0.2162	0.0585	0.0764	0.0058	0.0550	0.6082	0.3046	0.5317	0.2171
		33-31		-101.95	0.1183	0.000	0.0009	0.0002	0.0024	0.2922	0.0259	-0.0418	0.0028
		Zstat		(-2.05)	(13.99)	(0.14)	(0.41)	(1.29)	(0.92)	(13.99)	(1.25)	(-12.23)	(0.38)
	Low	41	23.4	6552.9	0.0894	0.0487	0.0749	0.0039	0.0530	0.3506	0.2525	0.5820	0.1765
High		42	23.4	5785.1	0.1318	0.0499	0.0756	0.0042	0.0525	0.4965	0.2376	0.5685	0.1693
	High	43	22.4	5751.7	0.1916	0.0516	0.0737	0.0044	0.0551	0.6273	0.2357	0.5503	0.1895
		43-41		-801.17	0.1022	0.003	-0.0012	0.0005	0.0021	0.2767	-0.0168	-0.0317	0.0130
		Zstat		(-2.88)	(13.98)	(3.47)	(-0.23)	(2.12)	(1.37)	(13.98)	(-1.90)	(-10.61)	(2.57)
	All Firms		277.9	2110.5	0.1616	0.0626	0.0790	0.0075	0.0562	0.4560	0.3051	0.5482	0.2262

Table 4
Four Factor Estimates for Portfolios Sorted by Size and PIN

Each month, 12 portfolios are formed based on a sort of the previous month's market value of equity and PIN. Firms are grouped first by size quartile, and then by PIN within each size quartile. For example, portfolio 23 contains firms in the second size quartile, and within that quartile, the highest PINs. Regression coefficients from a time-series regression of monthly portfolio returns (in excess of one-month T-bill rate) on the three Fama and French (1993) factors and a fourth momentum factor from Carhart (1997) are below. The intercept from the following equation is the measure of the portfolio monthly abnormal return:

$$R_{p,t} - R_{f,t} = a + bRMRF_t + sSMB_t + hHML_t + mMOM_t + \varepsilon_t\ ,$$

where $R_{p,t} - R_{f,t}$ is the monthly excess portfolio return, $RMRF_t$ is the market factor, SMB_t is the size factor, HML_t is the book-to-market factor and MOM_t is the momentum factor. Each regression contains 131 monthly observations over the sample period January 1992 – December 2002. Portfolio returns are equal weighted in Panel A and value weighted in Panel B. The *t*-statistics are in parentheses. GRS is the F-test of Gibbons, Ross and Shanken (1989) and the corresponding p-value. Panel C contains intercepts and *t*-statistics for the regressions of zero investment portfolio returns that are long the high PIN portfolio and short the low PIN portfolio on the four factors.

			Panel A: Equal Weighted						Panel B: Value Weighted					
Firm Size	PIN	Port	Int	RMRF	SMB	HML	MOM	Adj. R^2	Int	RMRF	SMB	HML	MOM	Adj. R^2
	Low	11	-1.242	1.375	1.096	0.833	-0.345	0.74	-1.353	1.405	1.007	0.873	-0.313	0.72
			(-3.36)	(14.09)	(11.00)	(6.76)	(-4.99)		(-3.54)	(13.95)	(9.80)	(6.87)	(-4.39)	
Low		12	-0.988	1.158	1.143	0.824	-0.238	0.78	-1.086	1.216	1.149	0.911	-0.214	0.76
			(-3.33)	(14.80)	(14.32)	(8.33)	(-4.30)		(-3.50)	(14.83)	(13.72)	(8.80)	(-3.69)	
	High	13	-0.129	1.024	0.831	0.728	-0.313	0.65	-0.162	1.124	0.823	0.719	-0.308	0.64
			(-0.38)	(11.40)	(9.06)	(6.41)	(-4.91)		(-0.44)	(11.53)	(8.27)	(5.84)	(-4.47)	
		21	-0.171	1.143	0.523	0.426	-0.162	0.60	-0.179	1.135	0.497	0.429	-0.161	0.58
			(-0.44)	(11.28)	(5.06)	(3.33)	(-2.26)		(-0.46)	(10.96)	(4.71)	(3.28)	(-2.20)	
		22	-0.374	1.105	0.677	0.717	-0.250	0.65	-0.431	1.121	0.658	0.724	-0.211	0.63
			(-1.12)	(12.50)	(7.50)	(6.42)	(-3.99)		(-1.25)	(12.28)	(7.06)	(6.28)	(-3.26)	
		23	-1.122	1.135	0.700	0.573	-0.127	0.71	-1.106	1.119	0.674	0.555	-0.120	0.70
			(-3.71)	(14.24)	(8.60)	(5.69)	(-2.26)		(-3.70)	(14.18)	(8.36)	(5.57)	(-2.15)	
		31	0.006	1.379	0.359	0.651	-0.323	0.77	-0.046	1.377	0.347	0.639	-0.337	0.76
			(0.02)	(17.36)	(4.43)	(6.48)	(-5.73)		(-0.15)	(16.92)	(4.17)	(6.22)	(-5.84)	
		32	-0.349	1.227	0.461	0.658	-0.189	0.67	-0.444	1.227	0.458	0.680	-0.190	0.68
			(-1.07)	(14.31)	(5.26)	(6.07)	(-3.11)		(-1.38)	(14.48)	(5.29)	(6.36)	(-3.16)	
		33	-0.561	1.182	0.189	0.640	0.028	0.64	-0.512	1.178	0.225	0.660	0.027	0.64
			(-1.85)	(14.73)	(2.31)	(6.31)	(0.49)		(-1.69)	(14.73)	(2.76)	(6.53)	(0.48)	
	Low	41	0.321	1.148	0.091	0.336	-0.265	0.72	0.200	1.116	-0.042	0.210	-0.239	0.71
			(1.09)	(14.82)	(1.15)	(3.43)	(-4.83)		(0.68)	(14.29)	(-0.52)	(2.13)	(-4.32)	
High		42	-0.391	1.211	0.264	0.378	-0.082	0.84	-0.140	1.150	0.249	0.218	-0.048	0.81
			(-1.95)	(22.91)	(4.89)	(5.65)	(-2.20)		(-0.63)	(19.65)	(4.18)	(2.95)	(-1.15)	
	High	43	-0.620	1.110	0.209	0.429	0.019	0.74	-0.471	1.120	0.079	0.333	0.082	0.73
			(-2.60)	(17.66)	(3.25)	(5.40)	(0.42)		(-1.89)	(16.99)	(1.18)	(4.00)	(1.77)	

GRS: 3.9039 (p-value = 0.0001)

GRS: 3.1143 (p-value = 0.0007)

Table 4 (continued)

Panel C: Intercept for Return to a Zero Investment Portfolio: High PIN - Low PIN

Size	EW	VW
Low	1.113 (2.36)	1.1904 (2.41)
2	-0.951 (-2.32)	-0.928 (-2.19)
3	-0.568 (-1.71)	-0.466 (-1.36)
High	-0.941 (-2.94)	-0.671 (-1.95)

Table 5
Portfolios Sorted by Size and PIN: Contemporaneous Returns

Each month, 12 portfolios are formed based on a sort of market value of equity and PIN. Firms are grouped first by size quartile, and then by PIN within each size quartile. For example, portfolio 23 contains firms in the second size quartile, and within that quartile, the highest PINs. Regression coefficients from a time-series regression of *contemporaneous* monthly portfolio returns (in excess of one-month T-bill rate) on the three Fama and French (1993) factors and a fourth momentum factor from Carhart (1997) are below. The intercept from the following equation is the measure of the portfolio monthly abnormal return:

$$R_{p,t} - R_{f,t} = a + bRMRF_t + sSMB_t + hHML_t + mMOM_t + \varepsilon_t ,$$

where $R_{p,t} - R_{f,t}$ is the monthly excess portfolio return, $RMRF_t$ is the market factor, SMB_t is the size factor, HML_t is the book-to-market factor and MOM_t is the momentum factor. Each regression contains 131 monthly observations over the sample period January 1992 – December 2002. Portfolio returns are equal weighted in Panel A and value weighted in Panel B. The *t*-statistics are in parentheses. Panel C contains intercepts and *t*-statistics for the regressions of zero investment portfolio returns that are long the high PIN portfolio and short the low PIN portfolio on the four factors.

			Panel A: Equal Weighted						Panel B: Value Weighted					
Firm Size	PIN	Port	Int	RMRF	SMB	HML	MOM	Adj. R^2	Int	RMRF	SMB	HML	MOM	Adj. R^2
	Low	11	-1.766 (-5.39)	1.262 (14.28)	0.943 (10.88)	0.708 (6.57)	-0.356 (-5.88)	0.74	-1.693 (-4.51)	1.332 (13.14)	0.840 (8.45)	0.750 (6.07)	-0.342 (-4.92)	0.68
Low		12	-0.991 (-2.79)	1.147 (11.98)	1.065 (11.34)	0.840 (7.19)	-0.307 (-4.67)	0.68	-0.939 (-2.60)	1.213 (12.44)	1.103 (11.41)	0.940 (7.90)	-0.296 (-4.40)	0.69
	High	13	0.308 (0.82)	1.166 (11.58)	0.946 (9.58)	0.795 (6.47)	-0.416 (-6.02)	0.66	0.700 (1.72)	1.221 (11.01)	0.982 (9.13)	0.891 (6.64)	-0.487 (-6.43)	0.64
		21	-1.117 (-3.56)	1.209 (14.30)	0.648 (7.82)	0.540 (5.24)	-0.202 (-3.48)	0.70	-1.004 (-3.03)	1.206 (13.51)	0.585 (6.69)	0.530 (4.87)	-0.197 (-3.22)	0.67
		22	0.004 (0.01)	1.119 (11.93)	0.751 (8.17)	0.616 (5.38)	-0.150 (-2.33)	0.63	0.197 (0.55)	1.122 (11.65)	0.741 (7.84)	0.625 (5.32)	-0.161 (-2.43)	0.62
		23	0.469 (1.28)	1.141 (11.56)	0.519 (5.36)	0.588 (4.88)	-0.145 (-2.14)	0.57	0.626 (1.72)	1.137 (11.61)	0.483 (5.03)	0.557 (4.66)	-0.156 (-2.33)	0.58
		31	0.004 (0.01)	1.181 (15.18)	0.275 (3.60)	0.519 (5.47)	-0.280 (-5.25)	0.71	0.172 (0.57)	1.168 (14.35)	0.260 (3.25)	0.461 (4.65)	-0.317 (-5.69)	0.70
		32	0.626 (2.02)	1.289 (15.43)	0.442 (5.39)	0.654 (6.42)	-0.218 (-3.81)	0.70	0.689 (2.25)	1.267 (15.49)	0.399 (4.92)	0.606 (5.98)	-0.207 (-3.67)	0.70
		33	1.627 (5.38)	1.079 (13.24)	0.296 (3.70)	0.459 (4.62)	-0.163 (-2.92)	0.63	1.604 (5.32)	1.079 (13.27)	0.278 (3.48)	0.467 (4.71)	-0.150 (-2.69)	0.63
	Low	41	1.068 (3.54)	1.121 (13.76)	0.185 (2.32)	0.262 (2.63)	-0.253 (-4.53)	0.69	1.328 (4.15)	1.127 (13.04)	0.093 (1.10)	0.181 (1.72)	-0.206 (-3.47)	0.67
High		42	1.120 (4.55)	1.166 (17.57)	0.159 (2.44)	0.324 (4.00)	-0.165 (-3.62)	0.76	1.529 (5.65)	1.090 (14.95)	0.154 (2.03)	0.195 (2.19)	-0.110 (-2.15)	0.71
	High	43	1.372 (5.63)	1.045 (15.89)	0.334 (5.18)	0.436 (5.44)	0.032 (0.72)	0.70	1.676 (5.96)	1.015 (13.38)	0.265 (3.52)	0.345 (3.72)	0.107 (2.01)	0.62

GRS: 12.138 (p-value = <0.0001)

GRS: 12.490 (p-value = <0.0001)

Table 5 (continued)

Panel C: Intercept for Return to a Zero Investment Portfolio: High PIN - Low PIN

Size	EW	VW
Low	2.0744 (4.60)	2.4004 (4.79)
2	1.5857 (4.16)	1.6294 (4.29)
3	1.6226 (4.76)	1.4319 (4.14)
High	0.3038 (0.92)	0.3342 (0.87)

Table 6
All Short Interest Stocks: Portfolios sorted by Short Interest Ratio and PIN

Each month, 30 portfolios are formed based on a sort of the previous month's short interest ratio and PIN. Firms are grouped first by SI decile, and then within each decile by PIN. The sample contains all short interest stocks (not just high short interest stocks). The high short interest stocks from earlier tables are contained in SI deciles 9 and 10. The tests use annual PIN estimates from Easley, Hvidkjaer, and O'Hara (2002). Each regression contains 120 monthly observations over the sampler period January 1992 – December 2001. Corresponding intercepts and *t*-statistics for monthly time-series regressions of the 30 portfolios from the following regression is reported:

$$R_{p,t} - R_{f,t} = a + bRMRF_t + sSMB_t + hHML_t + mMOM_t + \varepsilon_t \ ,$$

where $R_{p,t} - R_{f,t}$ is the monthly excess portfolio return, $RMRF_t$ is the market factor, SMB_t is the size factor, HML_t is the book-to-market factor and MOM_t is the momentum factor. Panel A reports results from equal-weighted portfolio returns, and Panel B contains the results from value-weighted portfolios. Results include the intercept and *t*-statistic from a zero investment portfolio long the high PIN portfolio and short the low PIN portfolio, and a zero investment portfolio long the high short interest portfolio and short the low short interest portfolio.

Panel A: Equal Weighted Returns

	Short Interest Ratio										
PIN	Low	2	3	4	5	6	7	8	9	High	High - Low
Low	0.1327	0.0618	0.0359	0.0594	0.0341	-0.0499	-0.1520	-0.0597	-0.1748	-0.2398	-0.3725
	(0.72)	(0.32)	(0.20)	(0.31)	(0.21)	(-0.28)	(-0.82)	(-0.33)	(-0.83)	(-1.03)	(-1.49)
	0.1556	0.2872	-0.2401	-0.1085	-0.1282	0.1563	-0.1223	-0.1053	-0.3219	-0.9524	-1.1079
	(0.85)	(1.44)	(-1.12)	(-0.54)	(-0.57)	(0.78)	(-0.63)	(-0.51)	(-1.35)	(-3.57)	(-3.83)
High	0.3188	-0.0847	0.5080	0.1914	0.0912	0.2500	-0.2576	0.0512	-0.3086	-0.8145	-1.1332
	(1.55)	(-0.37)	(2.23)	(0.87)	(0.41)	(0.94)	(-1.04)	(0.19)	(-1.09)	(-3.00)	(-4.05)
High - Low	0.1861	-0.1465	0.4721	0.1320	0.0571	0.2999	-0.1056	0.1109	-0.1338	-0.5747	
	(0.87)	(-0.57)	(2.12)	(0.58)	(0.27)	(1.22)	(-0.45)	(0.45)	(-0.46)	(-2.06)	

Panel B: Value Weighted Returns

	Short Interest Ratio										
PIN	Low	2	3	4	5	6	7	8	9	High	High - Low
Low	-0.0220	0.0119	0.2915	0.1139	-0.0199	0.0609	0.0500	0.0652	0.3653	-0.1418	-0.1198
	(-0.07)	(0.03)	(1.19)	(0.55)	(-0.10)	(0.34)	(0.25)	(0.36)	(1.61)	(-0.57)	(-0.29)
	0.0027	-0.1685	-0.3638	0.2400	-0.0359	0.0549	-0.1094	-0.1043	-0.4140	-0.9251	-0.9277
	(0.01)	(-0.66)	(-1.19)	(0.90)	(-0.15)	(0.23)	(-0.57)	(-0.47)	(-1.81)	(-2.91)	(-2.33)
High	0.4541	-0.1638	0.8137	0.4912	0.3786	0.4361	-0.2436	0.2892	-0.2937	-0.3799	-0.8341
	(1.89)	(-0.52)	(2.42)	(1.31)	(1.29)	(1.02)	(-0.76)	(0.98)	(-0.87)	(-0.98)	(-2.01)
High - Low	0.4762	-0.1757	0.5222	0.3773	0.3986	0.3752	-0.2936	0.2240	-0.6590	-0.2381	
	(1.30)	(-0.41)	(1.33)	(0.92)	(1.07)	(0.83)	(-0.79)	(0.75)	(-1.84)	(-0.56)	

Table 7
Cross-Sectional Regressions of High Short Interest Stocks: Portfolio Months

Coefficients are time-series averages of cross-sectional regressions of individual firm risk-adjusted returns on various firm characteristics. The sample includes all firms with a short interest ratio greater than 2.5 percent the previous month. *SIZE* is the market value of equity, *BM* is the book-to-market ratio, *MOM* is the summed return over the 6 months ending 2 months prior to the dependent variable return, *TURN* is the previous month's share volume divided by shares outstanding from two months prior, *DVOL* is the dollar volume of trading from the month two months prior, *SI* is the short interest ratio from the previous month, *PIN* is the probability of informed trading, and *PINxSI* is an interaction term. All independent variables except for MOM are natural logarithms. The cross-sectional regressions are run over 131 months from January 1992 – December 2002. Raw estimate is the time-series average of the monthly estimates, and purged estimate is the intercept from the regression in equation (13).

	Intercept	SIZE	BM	MOM	TURN	DVOL	SI	PIN	PINxSI
Raw	-0.46	0.1413	-0.0049						
	(-2.21)	(1.49)	(-0.03)						
Purged	-0.55	0.2129	0.0160						
	(-2.57)	(2.24)	(0.11)						
Raw	1.53	0.0629	0.0427				-0.2519	-0.5775	
	(1.62)	(0.63)	(0.29)				(-1.08)	(-2.02)	
Purged	1.43	0.1277	0.0644				-0.3274	-0.5294	
	(1.46)	(1.24)	(0.42)				(-1.34)	(-1.79)	
Raw	1.15	0.0822	0.0251	0.0036			-0.2358	-0.5204	
	(1.26)	(0.83)	(0.17)	(0.55)			(-1.00)	(-1.86)	
Purged	0.99	0.1503	0.0274	0.0076			-0.3036	-0.4816	
	(1.05)	(1.48)	(0.18)	(1.13)			(-1.23)	(-1.65)	
Raw	1.97	0.0465	0.0135	0.0040	0.1905		-0.3045	-0.4717	
	(1.32)	(0.47)	(0.09)	(0.61)	(0.95)		(-1.15)	(-1.71)	
Purged	2.62	0.1052	0.0253	0.0078	0.3676		-0.4443	-0.3908	
	(1.74)	(1.03)	(0.16)	(1.15)	(1.89)		(-1.64)	(-1.36)	
Raw	1.24	0.0972	0.0031	0.0046	0.1278			-0.4900	
	(1.04)	(1.01)	(0.02)	(0.70)	(0.72)			(-1.81)	
Purged	1.49	0.1791	0.0152	0.0086	0.2645			-0.4161	
	(1.23)	(1.86)	(0.10)	(1.27)	(1.51)			(-1.47)	
Raw	1.54	0.1955	0.0251	0.0039		-0.0994	-0.1450	-0.5458	
	(1.45)	(0.93)	(0.17)	(0.57)		(-0.52)	(-0.56)	(-2.00)	
Purged	0.79	0.1187	0.0345	0.0067		0.0337	-0.2699	-0.4817	
	(0.72)	(0.55)	(0.22)	(0.96)		(0.17)	(-1.02)	(-1.68)	
Raw	1.09	0.0590	-0.0178						-0.3670
	(1.29)	(0.58)	(-0.12)						(-1.94)
Purged	1.17	0.1185	0.0014						-0.4114
	(1.33)	(1.13)	(0.01)						(-2.09)
Raw	1.96	0.0454	-0.0322	0.0040	0.2280				-0.3763
	(1.34)	(0.46)	(-0.21)	(0.60)	(1.20)				(-1.94)
Purged	2.70	0.1031	-0.0227	0.0078	0.3714				-0.4387
	(1.83)	(1.01)	(-0.14)	(1.16)	(2.00)				(-2.20)

Table 8
Cross-Sectional Regressions of High Short Interest Stocks: All Months

Coefficients are time-series averages of cross-sectional regressions of individual firm risk-adjusted returns on various firm characteristics. The sample includes all firms with a short interest ratio greater than 2.5 percent any month during the sample period. *SIZE* is the market value of equity, *BM* is the book-to-market ratio, *MOM* is the summed return over the 6 months ending 2 months prior to the dependent variable return, *TURN* is the previous month's share volume divided by shares outstanding from two months prior, *DVOL* is the dollar volume of trading from the month two months prior, *SI* is the short interest ratio from the previous month, *PIN* is the probability of informed trading, and *PINxSI* is an interaction term. All independent variables except for MOM are natural logarithms. The cross-sectional regressions are run over 131 months from January 1992 – December 2002. Raw estimate is the time-series average of the monthly estimates, and purged estimate is the intercept from the regression in equation (13).

	Intercept	SIZE	BM	MOM	TURN	DVOL	SI	PIN	PINxSI
Raw	-0.16	-0.0291	-0.0174						
	(-1.01)	(-0.60)	(-0.19)						
Purged	-0.23	-0.0198	-0.0129						
	(-1.45)	(-0.40)	(-0.14)						
Raw	0.63	-0.0342	-0.0409				-0.2503	-0.2594	
	(1.82)	(-0.69)	(-0.46)				(-4.88)	(-2.36)	
Purged	0.57	-0.0217	-0.0454				-0.3060	-0.2560	
	(1.62)	(-0.42)	(-0.49)				(-5.94)	(-2.31)	
Raw	0.40	-0.0245	-0.0437	0.0078			-0.2600	-0.2358	
	(1.17)	(-0.51)	(-0.49)	(1.41)			(-5.27)	(-2.19)	
Purged	0.28	-0.0080	-0.0618	0.0111			-0.3078	-0.2329	
	(0.80)	(-0.16)	(-0.67)	(2.02)			(-6.15)	(-2.13)	
Raw	1.83	-0.0295	-0.0304	0.0073	0.3119		-0.3322	-0.1838	
	(2.27)	(-0.63)	(-0.34)	(1.33)	(2.32)		(-5.72)	(-1.79)	
Purged	2.03	-0.0149	-0.0441	0.0105	0.3809		-0.3967	-0.1730	
	(2.58)	(-0.30)	(-0.48)	(1.91)	(2.91)		(-7.09)	(-1.64)	
Raw	0.53	-0.0438	-0.0104	0.0081	0.0729			-0.1756	
	(0.74)	(-0.93)	(-0.12)	(1.46)	(0.62)			(-1.74)	
Purged	0.47	-0.0311	-0.0172	0.0114	0.0930			-0.1656	
	(0.66)	(-0.63)	(-0.19)	(2.09)	(0.79)			(-1.59)	
Raw	0.11	-0.0753	-0.0408	0.0064		0.0545	-0.2592	-0.2163	
	(0.19)	(-0.58)	(-0.46)	(1.15)		(0.44)	(-4.66)	(-2.09)	
Purged	-0.31	-0.1289	-0.0553	0.0090		0.1226	-0.3222	-0.2062	
	(-0.52)	(-1.01)	(-0.60)	(1.62)		(1.02)	(-6.03)	(-1.93)	
Raw	0.59	-0.0392	-0.0470						-0.2514
	(3.13)	(-0.81)	(-0.52)						(-5.21)
Purged	0.67	-0.0320	-0.0489						-0.3008
	(3.49)	(-0.64)	(-0.52)						(-6.20)
Raw	1.99	-0.0501	-0.0294	0.0074	0.2812				-0.3104
	(2.55)	(-1.11)	(-0.33)	(1.34)	(2.10)				(-5.66)
Purged	2.29	-0.0418	-0.0413	0.0106	0.3360				-0.3623
	(3.00)	(-0.89)	(-0.45)	(1.93)	(2.58)				(-6.81)

Table 9
Cross-Sectional Regressions of All Short Interest Stocks

Coefficients are time-series averages of cross-sectional regressions of individual firm risk-adjusted returns on various firm characteristics. The sample includes all short interest firms. *SIZE* is the market value of equity, *BM* is the book-to-market ratio, *MOM* is the summed return over the 6 months ending 2 months prior to the dependent variable return, *TURN* is the previous month's share volume divided by shares outstanding from two months prior, *DVOL* is the dollar volume of trading from the month two months prior, *SI* is the short interest ratio from the previous month, *PIN* is an annual estimate of the probability of informed trading, and *PINxSI* is an interaction term. All independent variables except for MOM are natural logarithms. The cross-sectional regressions are run over 120 months from January 1992 – December 2001. Raw estimate is the time-series average of the monthly estimates, and purged estimate is the intercept from the regression in equation (13).

	Intercept	SIZE	BM	MOM	TURN	DVOL	SI	PIN	PINxSI
Raw	-0.11	0.0127	-0.0081						
	(-0.62)	(0.24)	(-0.08)						
Purged	-0.16	0.0147	-0.0024						
	(-0.86)	(0.26)	(-0.02)						
Raw	-1.05	0.0970	-0.0171				-0.1293	0.3234	
	(-1.59)	(1.40)	(-0.16)				(-3.52)	(1.43)	
Purged	-1.78	0.1491	-0.0091				-0.1567	0.5710	
	(-2.66)	(2.06)	(-0.08)				(-4.22)	(2.45)	
Raw	-1.10	0.1219	0.0277	0.0007			-0.1377	0.3317	
	(-1.92)	(2.15)	(0.27)	(0.11)			(-3.76)	(1.64)	
Purged	-1.66	0.1546	0.0193	0.0013			-0.1661	0.5046	
	(-2.87)	(2.61)	(0.18)	(0.21)			(-4.48)	(2.41)	
Raw	0.81	0.1224	0.0538	0.0009	0.4164		-0.2330	0.4163	
	(1.00)	(2.19)	(0.52)	(0.15)	(3.53)		(-6.47)	(2.01)	
Purged	0.35	0.1555	0.0452	0.0015	0.4346		-0.2669	0.5881	
	(0.42)	(2.65)	(0.42)	(0.25)	(3.71)		(-7.10)	(2.74)	
Raw	-0.55	0.0863	0.0624	0.0012	0.1921			0.5259	
	(-0.69)	(1.57)	(0.61)	(0.20)	(1.74)			(2.54)	
Purged	-1.24	0.1103	0.0543	0.0019	0.1738			0.7144	
	(-1.54)	(1.89)	(0.51)	(0.31)	(1.59)			(3.34)	
Raw	-1.82	-0.0230	0.0359	-0.0011		0.1441	-0.1676	0.3813	
	(-2.38)	(-0.19)	(0.35)	(-0.18)		(1.38)	(-4.54)	(1.80)	
Purged	-2.46	-0.0113	0.0277	-0.0009		0.1646	-0.2001	0.5525	
	(-3.19)	(-0.09)	(0.26)	(-0.14)		(1.58)	(-5.17)	(2.52)	
Raw	0.17	0.0330	-0.0264						-0.1254
	(1.05)	(0.61)	(-0.25)						(-3.41)
Purged	0.19	0.0424	-0.0239						-0.1530
	(1.09)	(0.74)	(-0.22)						(-4.10)
Raw	2.39	0.0278	0.0365	0.0011	0.3888				-0.2221
	(3.48)	(0.57)	(0.36)	(0.18)	(3.38)				(-6.27)
Purged	2.44	0.0306	0.0216	0.0020	0.3985				-0.2534
	(3.50)	(0.58)	(0.20)	(0.33)	(3.47)				(-6.82)