Do information asymmetry proxies measure information asymmetry?

Ramzi Abdul-Baki

A Thesis

In

The John Molson
School of Business

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Science (Administration) at
Concordia University
Montreal, Quebec, Canada

July 2013

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Ramzi Abdul-Baki**

Entitled: **Do information asymmetry proxies measure information asymmetry?**

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Administration)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Navneet Vidyarthi	Chair
Gregory Lypny	Examiner
Sandra Betton	Examiner
Rahul Ravi	Supervisor

Approved by ______________________________

Chair of Department or Graduate Program Director

____

Dean of Faculty

Date ______________________________

Abstract

Do information asymmetry proxies measure information asymmetry?

Ramzi Abdul-Baki

Market microstructure based proxies for information asymmetry have been widely used for over two decades. However, their empirical validation is surprisingly scarce. We attempt to address this gap by empirically testing two of the more popular proxies, namely, Glosten and Harris (1988) adverse selection cost component of the bid-ask spread (lambda), and Easley, Hvidkjaer, and O'Hara (2002) probability of informed trading (PIN). We estimate these proxies across three portfolios: broad-based ETFs, sector ETFs, and common stocks. Arguably, information asymmetry about broad-based ETFs should be primarily related to market wide information asymmetry, while information asymmetry about stocks should be mostly related to firm characteristics. We find that while PIN is highest for the broad-based portfolio and lowest for the stock portfolio, lambda is highest for the stock portfolio and lowest for the broad-based portfolio. To the extent that information asymmetry about stocks should be greater than information asymmetry about systematic factors, we provide evidence in support of lambda as a measure of information asymmetry. PIN is more closely related to liquidity.

Contents

1. Introduction

The issue of asymmetric information in the financial markets has received considerable attention in both accounting and finance literature. Since, by definition, information asymmetry is not directly observable, empirical research has relied almost entirely on theoretical proxies for studying it. However, it is rather surprising that very little research exists testing the empirical validity of these measures. Consequently, little consensus exists among empiricists on either the acceptability or the desirability of any of these proxies. This paper attempts to partially address this gap in literature by taking a closer look at two relatively popular proxies of information asymmetry: first, the Glosten and Harris (1988) spread decomposition model based proxy and second, the Easley, Hvidkjaer, and O'Hara (2002) probability of informed trading (PIN) measure.

The Glosten and Harris (1988) spread decomposition model was among the earliest attempts at developing a proxy for measuring information asymmetry among investors. This model decomposes the quoted bid-ask spreads in the market into a permanent component (information asymmetry) and a transitory component. Subsequently, several other proxies were also developed based on spread decomposition. These included Lin, Sanger, and Booth (1995), Huang and Stoll (1997) and Madhavan, Richardson, and Roomans (1997). While differences exist across these models, Huang and Stoll (1997) theoretically highlighted the underlying similarities among them. Clarke and Shastri (2000) found significant positive correlation among these measures, thus lending support to the same conclusion. The spread decomposition based proxies of information asymmetry were the measure of choice among empiricists throughout the 90s and the early years of the new millennium. However, this approach has fallen out of grace in recent times.

The use of spread decomposition based proxies for information asymmetry has declined significantly. A survey of recent literature brought out only a handful of papers using these proxies. These include, Armstrong, Balakrishnan, Cohen (2012), Armstrong, Core, Taylor, and Verrecchia (2011), Verrecchia and Weber (2006), Affleck-Graves, Callahan, and Chipalkatti (2002), and Sadka (2006). Spread decomposition models of information asymmetry were developed with an underlying assumption of a prominent role by the monopolist specialist in determining the quotes. As markets evolved, the diminishing involvement of market makers and the increased role of limit books and other markets in the process of price formation has cast doubts on the validity of these models, and consequently, on the proxies of information asymmetry developed through them. It may be possible to make a naïve extension of these models by arguing that information asymmetry between the pool of buyers and the pool of sellers in the market might still have a role to play in determining the magnitude of the emerging best bid and the best offer (and thereby the prevailing bid-ask spread). However, new empirical evidence in support of these proxies is necessary before they can once again be used as valid proxies for information asymmetry.

Easley, Kiefer, O'Hara, and Paperman (1996) developed the probability of informed trading (PIN) proxy for measuring information. This approach attempts to estimate the market maker's belief about the probability of informed trading, using observed trade imbalances (the difference between the number of buy and the number of sell orders). Since its inception, the popularity of PIN among empirical researchers has been increasing rapidly. Some of the recent papers using this approach for measuring information asymmetry include, Hwang, Lee, Lim, and Park (2013), Armstrong, Balakrishnan, and Cohen (2012), Brown, Hillegeist, and Lo (2009), Jayaraman (2008), Ferreira, Ferreira, and Raposo (2011), Li, Wang, Wu, and He (2009), Chen Goldstein, and Jiang (2007). However, there is a growing debate casting doubts on PIN as a measure of information asymmetry. Duarte and Young (2009) find that only the liquidity component of PIN is priced, while the asymmetric information component is not, thereby suggesting that PIN is a proxy for liquidity. Aktas, de Bodt, Declerck, and Van Oppens (2007) also cast doubts on PIN as a measure of information asymmetry.

Akay, Cyree, Griffiths, and Winters (2012) deconstruct the PIN proxy to highlight the dichotomy between the empirical and the theoretical aspect of PIN as a measure of information asymmetry. According to them, the empirical estimate of PIN is designed to separate abnormal trading from normal trading and the PIN proxy is simply the ratio of the abnormal trading to total trading. The theoretical model, at this point takes a leap by assuming that while uninformed trades will remain un-clustered, the abnormal trading (clustering) will occur due to the trading activities of informed agents in the market. Under this assumption, the PIN becomes the ratio of informed trades to total trades. To the extent that, one-sided order-flow can arise for reasons other than informed trading, this underlying assumption puts the PIN proxy on rather shaky ground. Akay et al (2012) examine the PIN measure for T-bills in an attempt to answer the question, "What does PIN identify?" They argue that the T-bill market should have little or no information asymmetry. Therefore, PIN must approach zero if it is identifying the probability of informed trading. They find that T-bills not only have a non-zero PIN, but their PINs are significantly higher than equity PINs reported in extant literature. They conclude that PIN simply identifies liquidity-based clusters in T-bills markets.

While the above analysis is fairly thorough, there are at least two potential weaknesses, which could benefit from additional research. First, to the extent that there could be asymmetric information about systematic components of asset returns, there could still be non-zero asymmetric information in the T-bill market.[1] Second, microstructure measures are often sensitive to the market structure. Therefore, comparing numbers across different markets could lead to potential problems. Some tests based on assets from the equity market might be warranted before we can extend the conclusions of the above paper to the equity markets.

We use intraday trade and quote data on all broad-based and sector ETFs trading on the NYSE between 2001 and 2007 and a set of matched stocks. The

---

[1] Subrahmanyam (1991) entertains the possibility of asymmetric information about systematic components of asset returns and includes an informed traders factor in his model. Easley, Hvidkjaer, and O'Hara (2002) also allow for a common component in private information. This commonality could potentially be caused by asymmetric information about systematic factors. Finally, most textbooks on investing include sections on "top down" strategies and tactical asset allocation. Such approaches to investing rely on investors being able to avoid (select) asset categories or industries that will do relatively poorly (well). Asymmetry about systematic factors is implicit in these approaches.

general idea is that the broad-based ETFs should reflect primarily market wide uncertainty and therefore have low, but potentially positive levels of asymmetric information. The level of a broad-based ETF's asymmetric information should be related to its level of diversification. Sector ETFs should have market wide asymmetry and also asymmetry related to the industry component of their returns. This suggests that sector ETFs should have higher asymmetric information than broad-based ETFs. Finally, the individual equities that trade on U.S. exchanges should have three components to their information environment: market-wide information, industry information, and idiosyncratic information. Therefore, we expect the asymmetric information for individual equities to be higher than that in both, broad-based and sector ETFs.

We begin our analysis by estimating the spread decomposition measure of information asymmetry using the Glosten and Harris (1988) model and estimating the probability of informed trading (PIN) using the Easley, Hvidkjaer, and O'Hara (2002) model for each of the ETFs and the stocks in our sample, for each year. Our average PIN estimate for broad-based ETFs is approximately 32%, which is substantially greater than the average PIN measure for the set of matched stocks (11%). Similarly the average PIN for sector ETF is approximately 29%. This pattern is consistent across each year in our analysis. We find that the average Glosten and Harris (1988) adverse selection component for broad-based ETF is 14% of the spread, which is significantly less than the average estimate for the set of matched stocks (31%). Similarly the average adverse selection component for the sector ETF is 15%. Since the spread is a function of the stock price, it might be more informative to consider the estimated lambda as a fraction of each dollar traded. We find that the average adverse selection cost of trading, per dollar invested is estimated to be 2.68 basis points for the broad-based ETFs, 3.77 basis points for the sector ETFs and 4.15 basis points for the stocks. Thus, we find that while the spread decomposition measure is consistent with our prediction, for each year in our sample, the PIN measure does not seem to be correctly identifying information asymmetry.

The remainder of the paper continues as follows. In section 2 we present some background discussion and our empirical predictions. Section 3 discusses our data and our empirical specifications. In section 4 we present our results. We conclude in section 5.

2. Background and research design

The degree of information asymmetry is not directly observable and therefore researchers need to rely on proxy variables. Unfortunately the same reason also makes it difficult to empirically test the appropriateness of the various proxies. In our attempt to assess the validity of PIN and Glosten and Harris (1988) λ, we examine the relative levels of information asymmetry in the market for individual stocks and contrast it with the levels of information asymmetry in the diversified ETFs (Exchange traded Funds). We argue that the level of information asymmetry for stocks should be higher than that of diversified ETFs. This section provides some background discussion on information asymmetry and its potential sources and then discusses the arguments for differing levels of information asymmetry across individual stocks and diversified ETF portfolios. Finally, it briefly outlines the theory behind PIN and λ.

2.1. Background

Bagehot (1971) attributes the revenues of the market maker to the transactions by three categories of market participants. The first group consists of traders with special (private) information, the second group comprises of liquidity motivated traders, and finally, the third group of traders act upon publically available information, incorrectly believing that the information has not yet been fully discounted in the market prices. Here, the market maker will invariably lose to the first group. However, he will always gain from transacting with liquidity traders as well as the traders in the third category because they are trading against the market maker's spread. Stoll (1992) identified three components of the market maker's bid-ask spread. The first component is order-processing cost, consisting of all fixed and variable costs such as cost of space, communications, labor, etc., incurred by the market maker. The second component is the inventory carrying cost of the market maker, a consequence of bearing the risk of carrying excess or inadequate inventory. The third component is the adverse information cost which arises from the market maker's disadvantage in transacting with traders who possess special (private) information.

A significant volume of market-microstructure research is devoted to measuring the third component of the transaction cost (portion of the bid-ask spread), which arises from the market maker's informational disadvantage (information asymmetry). Early literature tends to equate informed traders with firm insiders and to that extent the private information was understood as firm characteristics.[2] More recently, the definition of private information has been modified, whereby it now includes sophisticated trading resulting from any informational advantage.[3] In terms of deriving informational advantage, earlier literature tended to view investors as two homogenous groups (Kyle, 1985). One group receives the information and thus becomes informed while the other group remains uninformed. Subsequently, Papers by Harris and Raviv (1993), Kandel and Pearson (1995), Kim and Verrecchia (1994, 1997) and Lundholm (1991) explored the idea of differential interpretations of common signals, thus allowing for heterogeneity across investors in terms of information endowment, resulting as a consequence of their varying levels of processing abilities.

2.2. Diversified ETF portfolios vs. individual stock information asymmetries

Let us start with the view of private information whereby informed traders are privy to firm specific information such as a pending merger or product development, etc. This idiosyncratic information would be diversified away in a large portfolio and knowledge about any one firm in the portfolio would not prove very useful in predicting the return on the portfolio.[4] Allowing for information

---

[2] Kyle (1985) defines informed traders as insiders with unique access to a private observation of the ex post liquidation value of the single risky asset.

[3] Brandt and Kavajecz (2004). Green (2004) use government bonds data to estimate market wide information asymmetry (systematic information asymmetry).

[4] Subrahmanyam (1991) and Gorton and Pennacchi (1993) present models where the bundling of claims on individual assets into composite claims reduces informed traders' informational advantage. We have compared the Glosten and Harris (1988) lambda with the average lambda accross all the S&P 500 constituent stocks with the SPDR lambda for each year from 1993 through 2007. We find

asymmetry about systematic factors, we argue that as portfolios increase in size, the impact of asset specific information becomes arbitrarily small and the adverse selection component of liquidity will have to largely come through asymmetric information about the systematic factors. Therefore, as the level of diversification increases, the set of potentially useful information that could give rise to informational advantage about the asset diminishes.

Furthermore, public information can be about systematic factors (macroeconomic news) and/or idiosyncratic in nature (analysts' forecasts, financial statements, firm news releases). While both sources of information could potentially improve the forecasting ability with regards to individual stocks, only systematic information should lead to the generation of information advantage in fully diversified portfolios. Therefore, once again, as the level of diversification increases, the pool of useful information diminishes. Accordingly, as the level of diversification increases, the adverse selection problem surrounding the asset in the market should decline.

2.3. PIN and λ as two measures of informed trading

This paper attempts to test the validity of two proxies of information asymmetry using the approach outlined above. The first measure is the probability of informed trading (PIN) and the second measure is an estimate of the permanent component of the market maker's bid-ask spread (λ).

---

that SPDR lambda is consistently lower than the average lambda calculated across the constituent stocks.

Easley, Hvidkjaer, and O'Hara (2002) define PIN as the ratio of informed trades to total trades.

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_S + \varepsilon_B} \tag{1}$$

Where α is the frequency of information events (the probability of the arrival of new information), μ is the arrival rate of informed orders, ε_S is the arrival rate of uninformed sell orders, and ε_B is the arrival rate of uninformed buy orders. To the extent that both informed and uninformed trading simultaneously exist in the financial market, identifying one from the other is impossible. As explained in Akay, Cyree, Griffiths, and Winters (2012), the empirical estimation of PIN uses secondary market trading data to estimate normal and abnormal clustering of trades. Thereby, the PIN estimate becomes the ratio of abnormal trading to total trading. The underlying assumption being that only informed trading should give rise to clustering of trades. This assumption is key to PIN being interpreted as a measure of informed trading and thereby a proxy for information asymmetry in the market.[5] Apart from the potential weakness of this assumption, Hwang, Lee, Lim and Park (2013) also point towards potential error in proxy arising from misclassification of buys and sells, thus casting doubt on all proxies of information asymmetry which rely on identification of trade direction.

The second proxy of information asymmetry examined in this study is that

[5] Duarte and Young (2009) point out that trade clustering can also arise from liquidity shocks which may be unrelated to any information based trading.

proposed by Glosten and Harris (1988), which decomposes the quoted spread into a permanent component and a transitory component. The permanent component is interpreted as the innovation in the market maker's beliefs due to trading with informed traders. The model used for the estimation of this proxy relates the change in transaction price to the transaction volume and to a buy/sell indicator in the following specification:

$$\Delta P_t = c_0 \Delta I_t + c_1 \Delta I_t V_t + z_0 I_t + z_1 I_t V_t + \varepsilon_t \tag{2}$$

Here, I_t is a trade indicator that equals 1 if the t^th transaction is buyer-initiated and negative one (-1) if it is seller-initiated; P_t is the transaction price for the t^th trade; V_t is the volume traded; and ε_t captures public news. In this model, the permanent (information asymmetry) component is $2\left(z_0 + z_1 V_t\right)$, and the transitory (inventory-holding and order-processing) component is given by $2\left(c_0 + c_1 V_t\right)$. The adverse selection component as a percentage of the spread is given by:

$$IA = \frac{2\left(z_0 + z_1 \overline{V}\right)}{2\left(c_0 + c_1 \overline{V}\right) + 2\left(z_0 + z_1 \overline{V}\right)} \tag{3}$$

where $\overline{V}$ is the average transaction volume over the estimation period.

The Glosten and Harris (1988) proxy for information asymmetry, as all other proxies based on spread decomposition, was designed for a quote driven market and relied heavily on the dominant role of the market maker in determining the bid-ask spread. Over time, the markets have increasingly moved away from being primarily quote driven to becoming order driven. In an order driven market, traders and their unexecuted limit orders are the primary providers of liquidity. With the decreasing role of the market maker in the market, the spread decomposition proxies have seemingly lost their empirical appeal. We will try to argue here that the spread decomposition proxies may still be valid and to that extent are worth closer examination.

As argued by Bagehot (1971), the market maker can be viewed as a conduit through which money flows from liquidity-motivated traders to traders with special information. He further argues that, "market makers of all kinds make surprisingly little use of fundamental information. Instead they observe the relative pressure of buy and sell orders and attempt to find a price that equilibrates these pressures". We argue here that limit books and other markets, while matching trades and facilitating transactions, are de facto fulfilling the same role as the above described market maker. The question that now demands an answer is what determines the magnitude of the bid-ask spread in this new market?

Traders provide liquidity in an order driven market through their market orders and unfilled limit orders. Arguably, as providers of liquidity, these traders can still be expected to require compensation for order handling costs. However, unlike a market maker, no trader has an obligation to make the market and take the opposite side of a trade. Therefore, inventory risks are less likely to be important. Glosten (1994) presents a theoretical argument suggesting that limit order markets will have a positive bid-ask spread arising from the possibility of at least some market participants trading on their private information. Similarly, Handa, Schwartz and Tiwari (2003) in their theoretical model show that the bid-ask spread in an order-driven market is likely to be a function of the differences in share valuations across various market participants. These theoretical models suggest that the adverse selection cost (information asymmetry) is still a component of the bid ask spread in order-driven markets. In an equilibrium model, stock price must reflect the value of the firm and to the extent that the value of the firm gets into the stock price through the resolution of the information asymmetry across the traders, this cost must still be the permanent price component of the bid-ask spread; all other costs could constitute the transitory costs.[6]

2.4. Idiosyncratic volatility

We use a measure of idiosyncratic volatility as an alternate measure of information asymmetry. Idiosyncratic volatility is a measure of the amount of price variability due to firm specific information. This should be directly related to the level of informed trading in the market and thereby the level of information asymmetry. We estimate this measure using two approaches: first, following Rajgopal and Venkatachalam (2010), we estimate the idiosyncratic volatility as the standard deviation of the residuals $\left(\sigma_{\varepsilon}\right)$ from the Fama French (1993) three factor model:

$$r_{i,t} = \alpha_i + \beta_{1,i} \times RM_t + \beta_{2,i} \times SMB_t + \beta_{3,i} \times HML_t + \varepsilon_{i,t} \tag{4}$$

Following Hutton, Marcus, and Tehranian (2009) we estimate an alternate measure

[6] The misclassification problem outlined by Hwang, Lee, Lim and Park (2013) also applies to all spread decomposition models because these are also dependent on the researcher's ability to reliably classify trades into buyer and seller initiated.

of idiosyncratic volatility, estimated as:[7]

$$IDIOSYNC = \ln\left(\frac{1-R^2}{R^2}\right) \tag{5}$$

where, R^2 is the coefficient of determination from regression (5).

3. Data

We employ all broad-based equity ETFs, sector ETFs, and a set of matched stocks in this study. Our sample covers the period from January 2001 to December 2007. We start in 2001 because although SPDRs began trading on the AMEX in February 1993, sector ETFs did not begin trading until December 1998. With a sample size of less than 20 ETFs trading before 2001, any statistical analysis would be suspect at best.

We begin with the entire population of all ETFs trading in the US market before applying screens to the data. Our first screen requires the ETF to be trading on the NYSE or the AMEX and to be a pure equity ETF. We also require the average trading price of the ETF in a given year to be at least $5.00 and it must have remained listed on NYSE/AMEX for all 12 months of the year. Our starting sample consists of a total of 631 ETFs (336 broad-based and 295 sector). The lowest number is 45 (23 broad and 22 sector) in 2001 and it increases to 631 (336 broad and 295 sector) in 2007. We next require all ETFs in the sample to have data on the

[7] While both methods of estimating idiosyncratic volatility have been widely used in extant research, the results have been often contradictory. Li, Rajgopal and Venkatachalam (2012) argue that the assertion of idiosyncratic volatility measuring firm specific information is valid only if both these measures provide consistent results for a given situation.

NYSE Trade-and-Quote database (TAQ) and the Center for Research in Security Prices (CRSP) database. The final sample consists of 41 ETFs in 2001, 44 in 2002, 47 in 2003, 55 in 2004, 81 in 2005, 141 in 2006 and 214 in 2007.

Table 1, Panels A and B, presents more detailed descriptive statistics for the sample. In 2001 the typical broad-based ETF managed $1.36 billion while a typical sector ETF market cap was $193 million. Contrasting these numbers with 2007, where the average size of broad-based and sector ETFs increase to $3.75 billion and $737 million respectively; we observe tremendous growth in ETFs through time.[8] Table 1 also presents information on the trading volume of the ETFs. Broad-based ETF total trading volume in 2007 is 3.52 times higher than the trading volume in 2001. For sector ETFs the increases are 2.86 times.

We also employ matched samples of common stocks traded on the NYSE and AMEX. In this sample common stocks are matched to ETFs in each year based on the average daily share price, average daily trading volume, and the standard deviation of daily returns. For each ETF we select the common stock with characteristics that minimize the following equation.

$$Score_{j,k} = \left(\frac{P_{stock,i,k} - P_{ETF,j,k}}{P_{stock,i,k} + P_{ETF,j,k}} \right)^2 + \left(\frac{V_{stock,i,k} - V_{ETF,j,k}}{V_{stock,i,k} + V_{ETF,j,k}} \right)^2 + \left(\frac{\sigma_{stock,i,k} - \sigma_{ETF,j,k}}{\sigma_{stock,i,k} + \sigma_{ETF,j,k}} \right)^2 \qquad (6)$$

This matching is designed to reduce disparity in the inventory and order

[8] While we stopped our analysis in 2007 to avoid the financial crash and potential confounding effects in our results, we do have data showing that the typical ETF sizes dropped in 2008 and 2009 by roughly 17% and 16% respectively. ETF sizes surged 23% in 2010, 10% in 2011 and about 14.4% in 2012.

processing components of trading costs across the ETF sample and the matched stock sample. While searching for the best match, we include only common stocks, which have remained listed on either the NYSE or the AMEX for all 12 months of the year. We also require these stocks to have data available on the NYSE Trade-and-Quote database (TAQ) and the Center for Research in Security Prices (CRSP) database. To avoid undue influence from extreme observations, we exclude all stocks with an average monthly price less than $5 and greater than $500. Table 1, Panel C presents information on the matches. The broad-based matching stocks are closer in terms of trading volume than the sector matching equities. Finally, we note that the matching process yields substantial differences in all three dimensions: the standard deviation of returns, trading volume, and stock price. Due to imperfections in our matching, we control for differences in these characteristics in our cross-sectional examination of levels.

Transaction level trade and quote data for all ETFs and stocks are retrieved from TAQ, while the daily closing price, return volatility, and trading volume are obtained from the CRSP daily data. To avoid the influence of any possible recording errors in TAQ, we exclude all quotes with a raw spread greater than $6.5 and with a percentage spread greater than 10%. The TAQ database does not eliminate auto-quotes (passive quotes by secondary market dealers). This can cause quoted spreads to be artificially inflated. Since there is no reliable way to filter out auto-quotes in TAQ, only NBBO (national best bid or offer) eligible primary market (NYSE/AMEX) quotes are used. Quotes established before the opening of the market or after the close are discarded. Negative bid-ask spread quotations, negative transaction prices, and negative quoted depths are discarded. Trades with non-standard settlement conditions are excluded. The first trade of each day is discarded to avoid the effects of the opening procedure.

4. Results

4.1. Univariate comparison of PIN and λ across the sample groups

Table 2 presents a univariate comparison of PIN (probability of informed trading, as calculated using the Easley, Hvidkjaer, and O'Hara, 2002 model) and lambda (adverse selection cost component of the spread, as calculated using the Glosten and Harris, 1988 model) across broad-based and sector ETFs and a set of matched stocks. Panel A presents the average PIN within each category for each year in the sample period. The average PIN for a broad-based ETF is estimated to be 31.57%. The corresponding number for a sector ETF is estimated to be 28.81%. The PIN for a sample of matched stocks is considerably lower at 11.13%. The higher PIN for broad-based and sector ETFs compared to the PIN for the matched set of stocks is inconsistent with broad-based and sector ETFs having lower information asymmetry than the individual stocks.

Panel B presents the lambdas for the sample. Average broad-based ETF lambda is estimated to be 13.9% of the spread. Average sector ETFs lambda is estimated to be marginally higher at 14.55% of the spread and the average lambda for the set of stocks is significantly higher at 31.23% of the spread. Since these numbers are a percentage of the spread, they are multiplied by the spread and divided by the price of the asset to obtain the adverse selection cost of trading per dollar invested. Since these costs are very small, we multiply them by 100 to get the cost per dollar invested in cents. Panel C of table 2 presents these numbers. We find that average costs across the seven years are 2.68 basis points for the broad-based ETFs, 3.77 basis points for the sector ETFs and 4.15 basis points for the stocks. The higher adverse selection cost for stocks vis-à-vis the sector or the broad-Based ETFs is consistent with the expectation that the information asymmetry in individual stocks should be greater than the information asymmetry in a broadly diversified portfolio. We do find that in 2003 the sector ETF cost is higher than the individual stocks. A possible reason for this could be imperfect matching and the fact that Table 2 is a univariate comparison. Following sections will carry out the comparison in a multivariate setup. Another pair contradicting our hypothesis is the broad vs. sector costs in 2007.

Panels D and E of Table 2 present the univariate comparison in terms of idiosyncratic volatility. This is an alternate measure of information asymmetry in the market. We use two related measures of idiosyncratic volatility, as defined in equations (4) and (5). While Panel D presents the coefficient of determination (R^2) based measure, Panel E presents the standard deviation of the residuals from equation (4). The R^2 based measure is a logarithmic transform allowing the estimated number to range from negative infinity to positive infinity. Higher numbers suggest greater levels of information asymmetry. We find that the broad-based ETF group average across the seven years is -2.4827, which is smaller than the sector group average of -0.5181. The highest level of information asymmetry is for the set of matched stocks. This number is estimated to be 1.6219. The alternate measure, which is the standard deviation of the residuals from the asset pricing equation defined in equation (4), provides consistent results with smallest information asymmetry for the broad-based ETFs (0.0268) followed by the sector ETFs (0.0377) and the highest level for stocks (0.0415). These results are consistent with our argument that diversified portfolios should have lower information asymmetry than individual stocks.

Table 3 presents the Pearson's correlations (right triangle) and Spearman's rank correlations (left triangle) across PIN, Lambda (scaled by price), two measures of idiosyncratic volatilities (IV(R^2), and IV(σ)), the three matching criterions (return volatility($\sigma_{\square\square\square}$), trading volume, and price), and a measure of illiquidity (ILLIQ) as defined by Amihud (2012).[9] To the extent that adverse selection is related to liquidity, we introduce this measure at this stage of our paper as a potential explanatory variable in subsequent sections for both PIN and lambda. The Amihud (2002) illiquidity ratio (ILLIQ) is defined as

$$ILLIQ = \frac{1}{D_i}\sum_{t=1}^{D_i}\frac{|r_{i,t}|}{Vol_{i,t}} \times 10^6 \qquad (7)$$

The annual average of the daily ratio between a stock´s absolute return and its dollar volume; D_t is the annual number of valid observation days for stock *i* and

[9] Lambda scaled by the stock price (λ/Price) may be interpreted as the adverse selection cost per dollar traded for a given asset.

$Vol_{i,t}$ is the dollar value of the daily trading volume.[10] According to Amihud (2002), this ratio measures "the daily price response associated with one dollar trading volume". In a horse race run across various proxies of liquidity, Goyenko, Holden and Trzcinka (2009) find the Amihud (2002) illiquidity ratio to be the most stable and reliable liquidity proxy.

The correlation between PIN and scaled lambda is found to be negative and significant. This may seem surprising and rather curious. However, the sign is consistent with the correlations reported in Clarke and Shastri (2000). Scaled lambda is positively correlated with both measures of idiosyncratic volatility (IV), suggesting that greater level of IV is associated with higher information asymmetry in the market. Scaled Lambda is also positively associated with the Amihud illiquidity ratio. Thus suggesting that lower liquidity is associated with higher levels of information asymmetry. However, once again PIN is negatively correlated with both measures of IV and positively related with the level of illiquidity. Return volatility also changes sign in terms of its association with PIN vs. scaled lambda. Remaining correlations are consistent with the extant literature. Both PIN and lambda are negatively correlated with trading volume and stock price. Higher trading volume is associated with lower information asymmetry and lower adverse selection cost of trading. Similarly, larger stocks on average have lower adverse selection (per dollar traded) compared to smaller stocks. Table 3 supports the implications of Table 2, whereby the spread decomposition measure of information

---

[10] We multiply the ratio by 10^6 to change the scale and get rid of too many zeros in the estimated ratio.

asymmetry displays characteristics consistent with it being a proxy for information asymmetry. PIN seems to consistently display diametrically opposite characteristics.

4.2. Multivariate comparison of PIN and λ across the sample groups

Transaction costs are affected by prices, trading volume, and volatility. Our matching process attempted to control for these characteristics, but as shown in Table 1 (Panels A and B), ETFs and their matched equities still differ along these dimensions. We utilize a multivariate regression to test for differences in the levels of PIN and lambda controlling for differences in, trading volume, and volatility. We estimate the following two regressions

$$PIN_i = \alpha + \beta_0 \times \text{Price}_i + \beta_1 \times \sigma_{ret,i} + \beta_2 \times \ln\left(vol\right)_i + \beta_3 \times D_{cat} + \varepsilon \qquad (8)$$

$$\lambda_i \big/ \text{Price}_i = \alpha + \beta_0 \times \text{Price}_i + \beta_1 \times \sigma_{ret,i} + \beta_2 \times \ln\left(vol\right)_i + \beta_3 \times D_{cat} + \varepsilon \qquad (9)$$

The dependent variable is the PIN in equation (8), and the adverse selection cost per dollar traded (lambda scaled by share price) in equation (9). The independent variables are share price (Price), Share volatility (σ_{ret}), natural logarithm of the trading volume (ln(vol)), and dummy variables that take the value 0 for the base case category (stocks) and the value 1 for the comparison category (broad-based and sector ETFs).

Table 4 presents estimates of β_3 (the coefficient on D_{Broad} and D_{Sector}) and their corresponding t-statistics. Panels A and B present the results of running the regression specified in equation (8). Panel A compares the broad-based ETFs and the sector ETFs to the set of matched stocks. Panel B repeats the analysis, in comparing broad-based ETFs to the set of sector ETFs. Panel A reinforces the conclusions of Table 2, robust to controlling for trading volume (ln(vol)), return volatility (σ_{ret}), and stock price. Controlling for these factors, broad-based ETFs' average PIN is about 13.4% greater than the average PIN for the set of stocks. The sector ETF PIN is on average larger than the stock PIN by about 3.6%. The results suggest that the PIN for diversified portfolios is greater than the PIN for individual stocks even in a multivariate setup. These results once again contradict the interpretation of PIN as a proxy for information asymmetry in the market. Extending the diversification argument from section 2 of this paper, we expect the broad-based ETFs to have lower information asymmetry than the sector ETFs. While the results of Panel B are not consistently significant, we do notice that in most years, broad based PIN is greater than the sector PIN.

Panels C and D of Table 4 present the results of the regression specified in equation (9). Panel C compares the broad-based ETFs and the sector ETFs to the set of matched stocks. Panel D repeats the analysis, in comparing broad-based ETFs to the set of sector ETFs. In concurrence with the information hypothesis, we find that controlling for trading volume, return volatility and price, the per dollar cost of trading (due to adverse selection) is lower for the set of broad-based ETFs by 38.5% as compared to the corresponding cost of trading stocks. The per dollar cost of trading (due to adverse selection) is lower for the set of sector ETFs by 34.2% as compared to the corresponding cost of trading stocks. Panel D presents the result of comparing broad-based ETFs with sector ETFs in terms of lambda per dollar. Although the results are not always significant, the adverse selection cost of trading broad-based ETF is found to be lower than the adverse selection cost of trading sector ETFs. These results support the interpretation of lambda as a proxy for information asymmetry.

4.3. Some factors affecting PIN and λ

Our results thus far suggest that λ seems to be a valid proxy for the adverse selection cost of trading (information asymmetry). It also seems to cast doubts over the use of PIN as a valid proxy for information asymmetry. Akay, Cyree, Griffiths and Winters (2012) in their study of PIN have suggested that in the T-bill markets, PIN could be picking up the activities of discretionary liquidity traders. Drawing upon this, we attempt to better understand PIN and λ by exploring their association with a measure of liquidity (Amihud (2002) Illiquidity ratio, *ILLIQ*) and an alternate measure of information asymmetry (idiosyncratic volatility, *IV*). We control for price, trading volume, and return volatility. The models estimated are:

$$PIN_i = \alpha + \beta_1 \times IV + \beta_2 \times ILLIQ + \beta_3 \times \ln\left(vol\right)_i + \beta_4 \times \sigma_{ret,i} + \beta_5 \times \text{Price}_i + \beta_{6\ldots11} \times \text{D}_{1\ldots6} + \varepsilon \quad (10)$$

$$\lambda_i \big/ \text{Price}_i = \alpha + \beta_1 \times IV + \beta_2 \times ILLIQ + \beta_3 \times \ln\left(vol\right)_i + \beta_4 \times \sigma_{ret,i} + \beta_5 \times \text{Price}_i + \beta_{6\ldots11} \times \text{D}_{1\ldots6} + \varepsilon \quad (11)$$

D_1 through D_6 represent six year-dummies representing 2002 through 2007 (2001 is the base case).

Table 5; Panel A reports the results of our analysis of PIN (Equation 10). We find that controlling for stock price, trading volume, and return volatility, PIN is positively associated with Amihud illiquidity ratio. This suggests that higher PIN is more likely to be associated with more illiquid assets (assets with lower liquidity).

To the extent that increased levels of adverse selection risk is expected to be associated with wider spreads and lower liquidity, this finding does not contradict the interpretation of PIN as an information proxy. However, we also find that PIN is negatively associated with idiosyncratic volatility, thereby suggesting that stocks with relatively higher idiosyncratic volatility are likely to be those with relatively lower PIN.[11] If idiosyncratic volatility is a measure of information asymmetry about the firm, the negative coefficient suggests that PIN is at best a measure of a lack of information asymmetry (higher PIN suggesting lower information asymmetry). However, we have to be careful in this generalization of our finding because the negative coefficient in this case could simply be a result of the design of our study. ETFs have lower idiosyncratic volatility and higher PIN than the corresponding matched stocks.

Table 5; Panel B reports the results of our analysis of lambda (Equation 11). We find that scaled lambda, controlling for stock price, trading volume and return volatility, is positively associated with Amihud illiquidity ratio. In this respect lambda behaved identical to PIN, in that, stocks with higher illiquidity (or lower liquidity) are likely to have greater λ per dollar traded. We also find that lambda is positively associated with idiosyncratic volatility. Therefore, stocks with greater idiosyncratic volatility or information asymmetry are likely to have higher λ per dollar traded. Thus, the association of λ per dollar traded with both a measure of illiquidity and a measure of information asymmetry seem to be in concurrence with

---

[11] Results using both measures of idiosyncratic volatility (described in section 2.4) are identical. For sake of brevity, we report the results pertaining to idiosyncratic volatility defined in equation (5).

the interpretation of this measure as a proxy for information asymmetry.

5. Conclusion

Do microstructure based proxies of information asymmetry correctly identify informed trading? We conducted this study as an attempt to answer the question. Since by definition, information asymmetry cannot be directly measured, we could not simply compare these proxies with the measure of true information asymmetry to find our answer. We attempted to get around this problem by comparing the level of information asymmetry for a set of broad-based and sector ETFs with a set of matched stocks. The stocks are matched with the corresponding ETF based on the exchange on which they trade (exchange mechanism is known to affect information asymmetry), the size of the stock, the level of trading volume and the level of return volatility. We argue that more diversified assets must have lower information asymmetry than the corresponding relatively less diversified asset.

We select two measures of information asymmetry: PIN (probability of informed trading), a relatively popular measure, which has come under increased scrutiny in the last few years and the Glosten and Harris (1988) permanent component of the spread. While the Glosten and Harris (1988) measure (or a modified version of it) was rather popular in the 90's, it was discarded in more recent times owing to its dependence on the role of the market maker as the determinant of the bid-ask spread. The fact that recent markets have moved away from being quote driven towards being more order driven has cast doubts over the validity of these models. We believe that to the extent that information asymmetry is still the permanent component of the order-driven market quote, proxies of information asymmetry that are based on the decomposition of these quotes into a permanent and transitory component are still likely to be valid.

Our results reinforce those of Akay, Cyree, Griffiths, and Winters (2012) in suggesting that while PIN is related to liquidity, it does not seem to be a valid proxy for information asymmetry. The permanent component of the spread λ, calculated using the Glosten and Harris (1988) spread decomposition model, on the other hand, performs well in our setup and it does seem to be correctly identifying information asymmetry. Following up the theoretical arguments of Huang and Stoll (1997), our result would suggest that the spread decomposition models for identifying information asymmetry are valid in the current order-driven markets. However, we leave the empirical test of this assertion as potential future research.

References

Affleck-Graves, J., Callahan, C., Chipalkatti, N., 2002. Earnings Predictability, Information Asymmetry, and Market Liquidity. Journal of Accounting Research 40, 561-583.

Akay, O., Cyree K., Griffiths, M., Winters, D., 2012. What does PIN identify? Evidence from the T-bill market. Journal of Financial Markets 15, 29-46.

Aktas, N., de Bodt, E., Declerck, F., Van Oppens, H., 2007. The PIN anomaly around M&A announcements. Journal of Financial Markets 10, 169-191.

Amihud, Y.,2002.Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5,31–56.

Armstrong, C., Balakrishnan, K., Cohen, D., 2012. Corporate governance and the information environment: evidence from state antitakeover laws. Journal of Accounting and Economics 53, 185-204.

Armstrong, C., Core, J., Verrecchia, R., 2011. When does information asymmetry affect cost of capital? Journal of Accounting Research 49, 1-40.

Bagehot, W., 1971. The only game in town. Financial Analysts Journal 22, 12-14.

Brandt, M., Kavajecz, K., 2004. Price Discovery in the U.S. Treasury Market: The Impact of Orderflow and Liquidity on the Yield Curve. The Journal of Finance 59, 2623-2654.

Brown, S., Hillegeist, S., Lo, K., 2009. The effect of earnings surprises on information asymmetry. Journal of Accounting and Economics 47, 208-225.

Chen, Q., Goldstein, I., Jiang, W., 2007. Price informativeness and investment sensitivity to stock price. Review of Financial Studies 20, 619-650.

Clarke, J., Shastri, K., 2000. On Information Asymmetry Metrics. University of Pittsburgh Working Paper.

Duarte, J., Young, L., 2009. Why is PIN priced? Journal of Financial Economics 91, 119-138.

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

Easley, D., Kiefer, N., O'Hara, M., Paperman, J., 1996. Liquidity, information, and infrequently traded stocks. Journal of Finance 51, 1405-1436.

Fama, Eugene F., 1993, Common risk factors in the returns on stocks and bonds, Journal of Financial Economics 33, 3–56.

Ferreira, D., Ferreira, M., Raposo, C., 2011. Board Structure and Price Informativeness. Journal of Financial Economics 99, 523-545.

Glosten, L., 1994, Is the Electronic Open Limit Order Book Inevitable?, Journal of Finance 49, 1127-1161.

Glosten, L., Harris, L., 1988. Estimating the Components of the Bid/Ask Spread. Journal of Financial Economics 21, 123-142.

Gorton, G., Pennacchi, G., 1993. Security Baskets and Index-Linked Securities. The Journal of Business 66, 1-27.

Goyenko,R.Y., Holden,C.W., Trzcinka,C.A., 2009. Do liquidity measures measure liquidity? .Journal of Financial Economics 92, 153–181

Green, T., 2004. Economic News and the Impact of Trading on Bond Prices. The Journal of Finance 59, 1201-1234.

Handa, P.; Schwartz, R. Tiwari, A., 2003, Quote Setting and Price Formation in an Order Driven Market, Journal of Financial Markets 6, 461-489.

Harris, M., Raviv, A., 1993. Differences of opinion make a horse race. Review of Financial Studies 6, 473-506.

Huang, R., Stoll, H., 1997. The Components of the Bid-Ask Spread: A General Approach. Review of Financial Studies 10, 995-1034.

Hutton, A., Marcus, A., Tehranian, H., 2009. Opaque financial reports, R2, and crash risk. Journal of Financial Economics 94, 67-86.

Hwang, L., Lee, W., Lim, S., Park, K., 2013. Does information risk affect the implied cost of equity capital? An analysis of PIN and adjusted PIN. Journal of Accounting and Economics 55, 148-167.

Jayaraman, S., 2008. Earnings Volatility, Cash Flow Volatility, and Informed Trading. Journal of Accounting Research 46, 809-851.

Kandel E., Pearson, N., 1995. Differential Interpretation of Public Signals and Trade in Speculative Markets. Journal of Political Economy 103, 831-872.

Kim, O., Verrecchia, R., 1994. Market liquidity and volume around earnings announcements. Journal of Accounting and Economics 17, 41-67.

Kim, O., Verrecchia, R., 1997. Pre-announcement and event-period private information. Journal of Accounting and Economics 24, 395-419.

Kyle, A., 1985. Continuous auctions and insider trading. Econometrica 6, 1315-1335.

Li, Bin, Rajgopal, Shivaram and Venkatachalam, Mohan, R2 and Idiosyncratic Risk are not Interchangeable (May 8, 2013). Available at SSRN: http://ssrn.com/abstract=2269203 or http://dx.doi.org/10.2139/ssrn.2269203

Li, H., Wang, J., Wu, C., He, Y., 2009. Are Liquidity and Information Risks Priced in the Treasury Bond Market? The Journal of Finance 64, 467-503.

Lin, J., Sanger, G., Booth, G., 1995. Trade Size and Components of the Bid-Ask Spread. Review of Financial Studies 8, 1153-1183.

Lundholm, R., 1991. What Affects the Efficiency of a Market? Some Answers from the Laboratory. The Accounting Review 66, 486-515.

Madhavan, A., Richardson, M., Roomans, M., 1997. Why do Security Prices Change? A Transaction Level Analysis of NYSE Stock. Review of Financial Studies 10, 1035-1064.

Rajgopal, S., and M. Ventatachalam. 2010. Financial reporting quality and idiosyncratic return volatility. Journal of Accounting and Economics 51, 1-20.

Sadka, R., 2006. Liquidity risk and asset pricing. Journal of Financial Economics 80, 309-349.

Stoll, H.R., 1992. The Economics of Market Making, The Nasdaq Handbook: The Stock Market for the Next Hundred Years. Probus Publishing Company.

Subrahmanyam A., 1991, A theory of trading in stock index futures, Review of Financial Studies, 4, 17-51.

Verrecchia, R., Weber, J., 2006. Redacted Disclosure. Journal of Accounting Research 44, 791-814.

Table 1: Descriptive statistics

The following table provides descriptive information with regards to each of the three sample categories used in our study: broad-based ETFs, sector ETFs, and category matched stocks. The information in the table is presented for each year in our sample period (2001-2007). Values are calculated based on the corresponding sample group for each category in each year. Market capitalization is presented in thousands of dollars; trade volume is calculated as an annual average of daily trading volume; price is presented as an average of the daily closing price reported on CRSP; return volatility is calculated as the standard deviation of daily returns.

Panel A: Broad-based ETF sample

	Sample Size	Market Cap (in ,000)	trade Volume (daily mean)	Price	return Volatility
2001	20	$1,245,554.43	922,656	$82.35	0.01482
2002	22	$2,223,957.11	1,905,435	$73.07	0.01743
2003	25	$2,971,308.40	2,199,758	$73.22	0.01270
2004	30	$3,156,245.95	2,067,940	$81.02	0.01014
2005	46	$2,842,344.52	2,258,267	$70.88	0.00882
2006	65	$2,910,972.96	2,382,919	$66.92	0.00933
2007	99	$2,908,672.24	3,248,878	$69.87	0.00950

Panel B: Sector ETF sample

	Sample Size	Market Cap (in ,000)	trade Volume (daily mean)	Price	return Volatility
2001	21	$193,021.92	254,725	$53.29	0.02906
2002	22	$169,441.15	460,744	$47.71	0.02447
2003	22	$290,258.99	617,666	$53.92	0.01717
2004	25	$368,832.09	1,158,438	$59.96	0.01408
2005	35	$373,464.31	1,142,564	$62.89	0.01131
2006	56	$424,864.78	822,116	$52.05	0.01235
2007	104	$478,528.52	727,244	$54.06	0.01052

Panel C: ETF category matched stocks

	Broad ETF merged stock			Sector ETF merged stock		
	Trade Volume (daily mean)	Price	Return Volatility	Trade Volume (daily mean)	Price	Return Volatility
2001	761,159	$61.48	0.02720	1,042,999	$46.19	0.03602
2002	913,839	$65.92	0.02546	1,869,071	$44.87	0.03401
2003	550,255	$64.56	0.02249	1,071,873	$46.85	0.03093
2004	1,064,402	$62.53	0.01866	789,977	$46.30	0.03023
2005	1,952,258	$61.05	0.01504	1,740,310	$55.80	0.02118
2006	1,816,413	$55.16	0.01266	1,127,660	$44.64	0.01774
2007	2,459,526	$50.27	0.01283	2,985,863	$50.16	0.01607

Table 2: Univariate Comparisons
The following table presents annual average values for PIN, lambda, price-scaled lambda, and two measures of idiosyncratic volatility (IV(R^2), and IV(σ)). The annual average values are calculated for each category.

Panel A

PIN			
	Broad ETF	*Sector ETF*	*Individual Stock*
2001	36.30%	30.79%	14.86%
2002	30.46%	33.79%	10.87%
2003	29.83%	24.91%	11.24%
2004	27.53%	24.20%	14.14%
2005	28.45%	25.86%	9.50%
2006	32.38%	33.07%	10.55%
2007	36.06%	29.03%	6.75%

Panel B

Lambda			
	Broad ETF	*Sector ETF*	*Individual Stock*
2001	15.03%	15.70%	40.06%
2002	13.42%	16.89%	34.96%
2003	13.00%	16.72%	39.65%
2004	12.83%	11.58%	37.40%
2005	14.48%	14.58%	28.79%
2006	13.61%	13.40%	22.42%
2007	14.95%	13.01%	15.36%

Panel C

100*(Lambda*Effective Spread)/Price			
	Broad ETF	*Sector ETF*	*Individual Stock*
2001	0.0400	0.0634	0.0653
2002	0.0370	0.0691	0.0712
2003	0.0304	0.0471	0.0351
2004	0.0228	0.0241	0.0408
2005	0.0214	0.0248	0.0287
2006	0.0168	0.0198	0.0279
2007	0.0191	0.0153	0.0214

Panel D

Idiosyncratic Volatility $\ln[(1-R^2)/R^2]$			
	Broad ETF	*Sector ETF*	*Individual Stock*
2001	-2.3648	-0.3958	2.2558
2002	-2.7475	-0.5018	1.7078
2003	-2.5398	-0.4788	0.7918
2004	-2.2068	-0.1668	1.4647
2005	-2.2188	-0.3618	1.9789
2006	-2.4258	-0.5258	2.1895
2007	-2.8768	-1.1988	0.9662

Panel E

Idiosyncratic Volatility (Residual σ)			
	Broad ETF	*Sector ETF*	*Individual Stock*
2001	0.0046	0.0165	0.0310
2002	0.0041	0.0148	0.0218
2003	0.0030	0.0093	0.0139
2004	0.0027	0.0081	0.0145
2005	0.0024	0.0065	0.0138
2006	0.0024	0.0066	0.0126
2007	0.0030	0.0069	0.0084

Table 3: Correlation coefficients

This table presents the Pearson correlations (right triangle) and Spearman rank correlations (left triangle) across PIN, price-scaled lambda, two measures of idiosyncratic volatilities (IV(R^2) and IV(σ)), return volatility($\sigma_{\square\square\square}$), trading volume (vol), price, and a measure of illiquidity (ILLIQ) as defined by Amihud (2012).

Pearson Correlation (columns); Spearman's rank correlation (rows)

	PIN	λ/Price	IV(R^2)	IV (σ)	Illiq	σ_{ret}	vol	Price
PIN	1	-0.1289	-0.4487	-0.4203	0.3601	-0.3320	-0.3301	-0.1199
λ/Price	-0.3230	1	0.4161	0.5753	0.4039	0.5411	-0.0781	-0.5529
IV(R^2)	-0.5160	0.6330	1	0.7284	0.0699	0.3533	-0.0596	-0.3120
IV (σ)	-0.4860	0.6630	0.8790	1	0.1720	0.7841	-0.0143	-0.3953
Illiq	0.6910	0.0850	0.1540	0.0150	1	0.1156	-0.0993	-0.3266
σ_{ret}	-0.4100	0.4440	0.4260	0.6690	0.2010	1	0.1117	-0.3343
vol	-0.7610	-0.1490	-0.3350	-0.3570	-0.8570	0.5450	1	0.1333
Price	-0.1127	-0.5964	-0.3526	-0.4261	-0.4554	-0.2646	0.0924	1

Table 4: Difference in PIN and lambda magnitude across the three portfolios
The following table contains the results corresponding to two multivariate regressions used to test for differences in the levels of PIN and lambda:

$$PIN_i = \alpha + \beta_0 \times \text{Price}_i + \beta_1 \times \sigma_{ret,i} + \beta_2 \times \ln(vol)_i + \beta_3 \times D_{cat} + \varepsilon$$

$$\lambda_i / \text{Price}_i = \alpha + \beta_0 \times \text{Price}_i + \beta_1 \times \sigma_{ret,i} + \beta_2 \times \ln(vol)_i + \beta_3 \times D_{cat} + \varepsilon$$

The regressions control for differences in share price (Price), volatility (σ_{ret}), natural logarithm of the trading volume (ln(vol)), and dummy variables that take the value 0 for the base case category (stocks) and the value 1 for the comparison category (broad-based and sector ETFs). Panels A and C present the results corresponding to the multivariate regression containing two dummy variables, D_{Broad} and D_{Sector}, distinguishing between broad-based ETFs, sector ETFs, and stocks. Panels B and D present the results associated with the multivariate regression containing one dummy variable, D_{Broad}, distinguishing between broad-based and sector ETFs; individual stocks are excluded in Panels B and D. Statistical significance at a 1%, 5%, or 10% level is denoted by ***, **, or *, respectively.

Panel A: PIN as dependent variable

	2001	2002	2003	2004	2005	2006	2007
α	1.008***	0.802***	0.696***	0.558***	0.512***	0.612***	0.346
	10.27	7.953	12.062	8.141	8.052	10.247	1.331
Ln(vol)	-0.053***	-0.051***	-0.045***	-0.035***	-0.027***	-0.03***	-0.012**
	-5.892	-6.295	-9.057	-6.391	-5.4	-6.352	-2.38
σ_{ret}	-2.023*	1.571	1.912	2.549*	0.606	-2.245*	0.606
	-1.981	1.414	1.425	1.907	0.359	-1.859	0.083
Price	-0.001***	-0.001*	0	0	-0.001***	-0.001***	-0.001
	-2.829	-1.781	-1.115	-0.843	-2.787	-4.659	-1.605
D_{Broad}	0.134***	0.154***	0.145***	0.125***	0.135***	0.149***	0.251***
	3.909	5.19	8.144	5.399	5.357	7.609	3.227
D_{Sector}	0.036	0.109***	0.068***	0.079***	0.102***	0.147***	0.183**
	1.043	3.532	3.867	3.62	4.199	7.638	2.263
Adj. R^2	0.632	0.556	0.614	0.448	0.464	0.61	0.515
N	81	88	94	110	155	220	375

Panel B: PIN as dependent variable

	2001	2002	2003	2004	2005	2006	2007
α	1.171***	1.047***	0.885***	0.719***	0.647***	0.776***	0.588**
	12.453	8.927	14.121	8.93	9.674	12.655	2.197
Ln(vol)	-0.058***	-0.063***	-0.044***	-0.035***	-0.025***	-0.028***	-0.012**
	-5.532	-5.225	-6.893	-4.788	-3.605	-4.466	-0.445
σ_{ret}	-2.379*	3.286	-1.245	-0.13	-0.534	-5.586**	2.883
	-1.776	1.365	-0.492	-0.042	-0.147	-2.38	0.224
Price	-0.001*	0	-0.001	0	-0.001**	-0.001***	-0.002
	-1.707	-0.674	-1.315	-0.933	-2.214	-3.857	-1.431
D_{Broad}	0.084**	0.061*	0.064***	0.036	0.031	-0.007	0.068
	2.669	1.733	3.095	1.361	1.178	-0.361	1.106
Adj. R^2	0.604	0.245	0.455	0.208	0.199	0.31	0.253
N	41	44	47	55	76	106	180

Panel C: (λ/Price) as dependent variable

	2001	2002	2003	2004	2005	2006	2007
α	0.986***	0.016***	1.314***	2.202***	1.394***	1.133***	0.614***
	2.775	2.771	2.831	3.898	4.236	5.511	3.18
Ln(vol)	-0.016	-0.002***	-0.07*	-0.145***	-0.081***	-0.041**	-0.028*
	-0.477	-3.345	-1.729	-3.243	-3.115	-2.499	-2.032
σ_{ret}	10.432***	0.61***	41.384***	66.389***	59.035***	17.524***	7.712
	2.822	9.345	3.833	6.03	6.748	4.215	1.43
Price	-0.005**	-0.038***	-0.007***	-0.008***	-0.009***	-0.006***	-0.040***
	-2.629	-6.456	-3.176	-3.801	-5.897	-8.245	-6.59
D_{Broad}	-0.385***	-0.004**	-0.325**	-0.312*	-0.19	-0.14**	-0.163***
	-3.101	-2.271	-2.275	-1.638	-0.996	-2.075	-2.827
D_{Sector}	-0.342***	-0.003***	-0.286**	-0.304***	-0.181	-0.112*	-0.141***
	-3.545	-3.212	-2.009	-2.93	-1.429	-1.684	-3.341
Adj. R^2	0.645	0.71	0.505	0.524	0.475	0.421	0.446
N	81	88	94	110	155	220	375

Panel D: (λ/Price) as dependent variable

	2001	2002	2003	2004	2005	2006	2007
α	0.474**	0.004	0.782*	0.727*	0.888***	0.794***	0.333**
	2.611	0.862	1.696	1.991	3.764	4.528	2.247
Ln(vol)	-0.045**	-0.001***	-0.151***	-0.101***	-0.094***	-0.056***	-0.026*
	-2.243	-2.911	-3.197	-3.063	-3.78	-3.117	-1.732
σ_{ret}	20.236***	0.695***	97.329***	95.302***	99.132***	48.245***	13.606*
	7.825	8.027	5.229	6.839	7.718	7.19	1.904
Price	-0.001	-0.0045	-0.001	-0.004***	-0.007***	-0.005***	-0.0034
	-0.946	-0.172	-0.467	-2.68	-4.724	-6.262	-0.264
D_{Broad}	-0.117*	-0.001	-0.148	-0.265**	-0.146	-0.055	-0.044
	-1.933	-0.96	-0.971	-2.184	-1.586	-0.979	-1.283
Adj. R^2	0.776	0.707	0.425	0.533	0.478	0.509	0.481
N	41	44	47	55	76	106	180

Table 5: Explaining PIN and Lambda (λ)

The following table contains the results corresponding to two multivariate regressions used to explore the association of PIN and lambda with the Amihud (2002) Illiquidity ratio (Illiq) and idiosyncratic volatility, while controlling for price, trading volume, and return volatility.

$$PIN_i = \alpha + \beta_1 \times IV + \beta_2 \times ILLIQ + \beta_3 \times \ln(vol)_i + \beta_4 \times \sigma_{ret,i} + \beta_5 \times \mathrm{Price}_i + \beta_{6\ldots11} \times \mathrm{D}_{1\ldots6} + \varepsilon$$

$$\lambda_i / \mathrm{Price}_i = \alpha + \beta_1 \times IV + \beta_2 \times ILLIQ + \beta_3 \times \ln(vol)_i + \beta_4 \times \sigma_{ret,i} + \beta_5 \times \mathrm{Price}_i + \beta_{6\ldots11} \times \mathrm{D}_{1\ldots6} + \varepsilon$$

D1 through D6 represent six year-dummies (2002 through 2007). For sake of brevity these dummies are not reported in the tables below.

Statistical significance at a 1%, 5%, or 10% level is denoted by ***, **, or *, respectively. Panels A and B present the results corresponding to the the full and reduced models.

Panel A: PIN

	Full Model	Reduced Models				
α	0.4467***	0.2765***	0.2532***	0.2888***	0.42***	0.3221***
	18.1515	16.1613	14.2553	16.3167	17.3533	15.2992
IV(R^2)	-0.0197***	-0.0213***				
	-9.9482	-9.3681				
Illiq	1.069***		1.0934***			
	9.352		8.1579			
Ln(vol)	-0.0408***			-0.0525***		
	-5.9116			-6.6327		
σ_{ret}	-5.0872***				-4.9216***	
	-9.3532				-8.1103	
Price	-0.0009***					-0.0006***
	-5.8051					-3.6164
Adj. R^2	0.566	0.142	0.114	0.509	0.113	0.035
N	1123	1123	1123	1123	1123	1123

Panel B: Lambda

	Full Model	Reduced Models				
Intercept	0.4522***	0.5208***	0.391***	0.5081***	-0.2149**	1.0597***
	4.3901	7.2314	5.3566	6.8407	-2.2712	14.3926
IV(R^2)	0.027***	0.0623***				
	3.2605	6.4992				
Illiq	1.8975***		3.9745***			
	3.9665		7.2151			
Ln(vol)	-0.0086**			-0.0017**		
	-2.1048			-2.1588		
σ_{ret}	14.9759***				25.7921***	
	6.5792				10.8709	
Price	-0.0067***					-0.0092***
	-10.4599					-14.9336
Adj. R^2	0.412	0.142	0.149	0.082	0.229	0.329
N	1123	1123	1123	1123	1123	1123