Does PIN affect equity prices around the world?

Sandy Lai, Lilian Ng, and Bohui Zhang*

Current Version: August 29, 2013

*Lai is from Faculty of Business and Economics, University of Hong Kong, K.K. Leung Building, Pokfulam Road, Hong Kong, Ng is from the Sheldon B. Lubar School of Business, University of Wisconsin, Milwaukee, P.O. Box 742, WI 53201-0742, U.S.A., and Zhang is from Australian School of Business, University of New South Wales, Sydney, NSW 2052, Australia. Authors' Information: Lai: sandy_lai@hku.hk, +852-39174180; Ng: lng@uwm.edu, +1-414-2295925; Zhang: Bohui.zhang@unsw.edu.au, +61-2-93855834.
We thank Ekkehart Boehmer, Andrew Ellul, Audrey Hackett, Campbell Harvey, Chuan-Yang Hwang, Eric Hughson, Andrew Karolyi, Patrick Kelley, Michael Lemmon, Ernst Maug, Randall Morck, Maureen O'Hara, the anonymous referee, Mark Seasholes, Duane Seppi, Editor Bill Schwert, Kumar Venkataraman, Akiko Watanabe, Masahiro Watanabe, and participants of the Central University of Finance and Economics (Beijing), 2012 Frontiers in Finance (University of Alberta), Hong Kong University of Science and Technology, Massey University, Singapore Management University, University of Technology, Sydney, University of New South Wales, University of Sydney, Temple University, York University, 2009 Nippon Finance Association Conference in Tokyo, 2009 NTU International Conference on Economics, Finance and Accounting in Taiwan, 2009 European Finance Association meetings in Norway, 2010 China International Conference in Finance, and 2013 Asian Finance Association Annual Meeting in China for many helpful comments and suggestions. We also thank the 2013 Asian Finance Association's program committee for awarding us the SIRCA Best Paper Award. Bohui Zhang acknowledges the 2009 research grant from the Australian School of Business at the University of New South Wales.

Electronic copy available at: http://ssrn.com/abstract=2207579

Does PIN affect equity prices around the world?

Abstract

This study examines the empirical controversy over the pricing effect of Easley, Hvidkjaer, and O'Hara's (2002) probability of information-based trading, PIN, on a sample of 30,095 firms from 47 countries worldwide. Contrary to the empirical evidence of Easley, Hvidkjaer, and O'Hara, but consistent with that of Duarte and Young (2009), we find no evidence that PIN exhibits a positive effect on a cross-section of expected stock returns in international markets. Alternative information-based trading measures also display no effect on expected stock returns, corroborating our finding that information risk proxied by PIN, in general, has no pricing effect in world markets.

Keywords: International Markets, Information Risk, PIN, Asset Pricing

JEL Classification Number: G11, G12, G23

Electronic copy available at: http://ssrn.com/abstract=2207579

1. Introduction

Easley and O'Hara (2004) suggest that information risk arising from information asymmetry between informed and uninformed investors is systematic and non-diversifiable. Using a rational expectations asset pricing model, they show that more information asymmetry increases the risk faced by uninformed investors since informed investors can shift their portfolio weights to adjust for new information. All else equal, uninformed investors demand a premium to hold shares in firms with higher information asymmetry, since the uninformed expect to lose to the informed and therefore demand to be compensated for this expected loss. Based on a structural microstructure model, Easley, Hvidkjaer, and O'Hara (2002) derive a measure of private information-based trading, the PIN measure, and find a strong positive cross-sectional relationship between expected stock returns and PIN, suggesting that information asymmetry, as measured by PIN, is priced.

Recent theoretical and empirical studies, however, provide results that challenge the evidence that asymmetric information risk embodied in PIN has a pricing effect. Theoretically, Hughes, Liu, and Liu (2007) and Lambert, Leuz, and Verrecchia (2007) yield empirical implications that are at variance with those in Easley and O'Hara (2004). Specifically, their models imply that information risk is potentially idiosyncratic in nature and hence, fully diversifiable. Empirically, Duarte and Young (2009) find no evidence that supports Easley, Hvidkjaer, and O'Hara's (2002) finding that PIN is associated with priced information risk.[1] They decompose PIN into two components, one related to asymmetric information and one related to illiquidity, and find that only the PIN component related to illiquidity is priced. They therefore argue that liquidity effects unrelated to information asymmetry explain the cross-sectional relation between PIN and expected returns.

Given the extensive applications of PIN, implicitly and explicitly, as a proxy for priced information risk in both finance and accounting literatures,[2] it is imperative that we investigate this contentious issue by subjecting PIN to robust out-of-sample analyses. Thus far, existing empirical studies focus only on the US market, and it is therefore important that we examine the asset

[1]Mohanram and Rajgopal (2009) replicate Easley, Hvidkjaer, and O'Hara's study and report that the evidence in the latter is not robust to alternative specifications and time periods. The effect of PIN on expected returns becomes negative and insignificant in an extended period from 1999 to 2002.

[2]See Appendix A of Mohanram and Rajgopal (2009) for a detailed list of references.

Electronic copy available at: http://ssrn.com/abstract=2207579

pricing implications of PIN in non-US markets. Specifically, to resolve the debatable issue of whether information risk measured by PIN is priced, we need to test whether the Easley, Hvidkjaer, and O'Hara (2002) PIN (hereafter PIN_{EHO}), the asymmetric information component of PIN (hereafter PIN_{DY}), as derived by Duarte and Young (2009), or both systematically explain cross-sectional variation in expected stock returns across international markets.

Our study begins by estimating PIN_{EHO} and PIN_{DY} using the methodologies developed by Easley, Hvidkjaer, and O'Hara (2002) and Duarte and Young (2009) on a sample of 30,095 international stocks across 47 countries worldwide. Our estimates of the probability of informed trading for each stock are based on the information in the newly available global intradaily stock transactions data provided by Thomson Reuters Tick History database (TRTH) for the period from 1996 to 2010. While our study represents the first to estimate PINs for this large cross-section of international firms, one concern is that stocks of these firms are mostly traded on electronic order-driven markets, which might be inconsistent with the market microstructure model of market making in which PIN is derived. As a result, it is possible that our PIN estimates may not actually capture the probability of informed trading for our sample of stocks that we have expected. To address this issue, we conduct two different tests to assess the quality of our PIN estimates.[3]

First, following Easley et al. (1996), we show how well our PIN estimates predict different measures of spreads. Theoretical studies have shown that spreads widen as adverse selection costs caused by informed trading become larger. Thus, we use spreads as a means to verify the quality of our PIN estimates, while controlling for trading volume. Next, we examine the association between PIN estimates and several other proxies of information asymmetry at firm and country levels. If the PIN estimates capture the level of private information, then they should be strongly correlated with other proxies of information asymmetry commonly adopted in the existing literature. Our firm-level proxies for information asymmetry include analysts following, analyst forecast dispersion, press coverage, firm age, index membership, and closely-held ownership, while country-level proxies are a country's accounting standard index, disclosure requirement index, newspapers circulation, capital market governance, and financial transparency factor. We find that our PIN estimates are

[3]We thank the referee for this excellent suggestion.

strongly correlated with spreads and with firm- and country-level asymmetric-information proxies in predictable ways, indicating the reasonableness of our estimates of the probability of informed trading using order flows from automated trading systems. Even though these analyses suggest that our findings are quite robust, some concerns about the adequacy of PIN estimates still remain. Our evidence should therefore be interpreted cautiously, keeping these concerns in mind.

We next turn to examining whether the information risk captured by PIN can systematically explain cross-sectional variation in expected stock returns. We conduct two different asset pricing tests. First, we form portfolios of stocks single-sorted on PIN and also double-sorted on a firm's market capitalization and PIN and then compute excess returns and risk-adjusted returns on each of these portfolios. Results indicate no significant differences in excess returns or in risk-adjusted returns between high and low PIN-formed portfolios, even after controlling for the market capitalization of the portfolios. Second, using Fama-MacBeth's (1973) approach, we find that PIN_{EHO} exhibits no significant positive relationship with future realized stock returns. These results are robust to orders submitted by algorithm trading implemented in a multiplicity of markets. Furthermore, consistent with Duarte and Young (2009), we also find that the asymmetric information component of PIN_{DY} exhibits no significant impact on the cross-section of expected stock returns. All this evidence therefore provides no support that PIN reflects information risk systematically priced by investors.

Finally, if information risk related to PIN is diversifiable, it is possible that we can find similar evidence when we use alternative information-based trading measures in place of PIN in our asset pricing tests. We exploit the richness of our database to estimate four alternative information-based trading measures drawn from the existing literature, namely Hasbrouck's (1991) measure of relative trade informativeness, Huang and Stoll's (1996, 1997) percentage price impact measure and adverse selection component, and Madhavan, Richardson, and Roomans's (1997) asymmetric information parameter. We repeat our asset pricing tests using these four measures, separately, as well as using the first principal components extracted from these four measures with different combinations of PIN_{EHO} and PIN_{DY}. Our evidence remains robust that information risk proxied by trading-based measures has no effect on the cross-section of expected stock returns in international markets.

Our research contributes to several strands of finance and accounting literatures. First, our study represents the first to examine the pricing of PIN in an international setting, and such an analysis should provide sufficiently robust evidence to help resolve the debate on whether PIN is a priced information risk. We show that the pricing effect of PIN in Easley, Hvidkjaer, and O'Hara (2002) is neither robust to the time period of our study, nor is it robust across our sample of equity markets. Our results further corroborate the findings of Duarte and Young (2009) who focus on US equity markets and also provide no evidence that PIN reflects information risk priced by investors. In addition, our exploratory analysis using four other popular measures of information-based trading reinforces our overall evidence that information risk proxied by PIN, in general, has no pricing effect in international equity markets.

Second, our work adds to a growing empirical literature that successfully applies PIN to explaining various information-based regularities. This measure is used to study informed trading across different markets (Easley, O'Hara, and Srinivas, 1998) and types of securities (Easley et al., 1996), stock price reactions to public and private news surprises (Vega, 2006), the information effect of IPO underpricing (Ellul and Pagano, 2006), the corporate investment sensitivity to stock prices (Chen, Goldstein, and Jiang, 2007), the impact of Regulation FD on information asymmetry (Duarte et al., 2008), among others. Our study contributes to this literature by showing that PIN, while not priced, is strongly associated with various proxies of information asymmetry at both firm and country levels.

The remainder of the paper is organized as follows. Section 2 briefly discusses the methodologies and estimation of PIN_{EHO} and PIN_{DY} for our sample of 30,095 firms from 47 countries worldwide and then assesses the quality of the two PIN estimates. Section 3 investigates the asset pricing implications of PIN, and Section 4 examines the relation between other trading-based information asymmetry measures and equity prices. The final section summarizes the paper.

2. The Estimation of PIN_{EHO} and PIN_{DY} Models

This section first describes PIN, which is derived from the market microstructure model of Easley et al. (1996) and Easley, Hvidkjaer, and O'Hara (2002), and its extension by Duarte and Young (2009). It then discusses the methodologies and global intradaily transactions data employed in estimating the two measures of PIN, followed by cross-country summary statistics of their estimates. In this section, we also perform several tests to assess the quality of these estimates.

2.1. The PIN Model and its Extension

PIN is derived from the structural microstructure model of Easley et al. (1996) and Easley, Hvidkjaer, and O'Hara (2002) and is based on the imbalance between buy and sell orders among investors.[4] The premise of their model is that order imbalances reflect active trading of informed investors, resulting from the arrival of private information. Otherwise, a more stable and balanced order flow is observed if trading is not driven by private information. Therefore, PIN is a firm-level estimate of the probability that an observed trade originates from a privately informed investor, who may have advance knowledge of analysts' reports, proprietary industry or macro forecasts, insider information, superior ability to process public information, among others.

Easley, Hvidkjaer, and O'Hara (2002) compute PIN_{EHO} as a fraction of orders that arises from informed investors relative to the overall order flow,[5] as follows.

$$PIN_{EHO} = \frac{\alpha \cdot \mu}{\alpha \cdot \mu + \varepsilon_S + \varepsilon_B}, \tag{1}$$

where α is the probability that a private information event occurs at the beginning of the trading day, μ is the daily arrival rate of orders from informed investors, and ε_B and ε_S are the daily arrival rates of buy and sell orders from uninformed investors.

Duarte and Young (2009), however, show that the PIN_{EHO} model does not capture the prevalent positive correlation between buyer- and seller-initiated order flows or the large variances of these

[4] PIN takes into account patterns in the number of trades, but not trade size. Easley, Hvidkjaer, and O'Hara (2002) show that trade volume reveals little information beyond the number of trades, suggesting that PIN is an adequate proxy for the degree of informed trading.

[5] A more detailed discussion of PIN is contained in Easley, Hvidkjaer, and O'Hara (2002).

order flows. The two authors extend the PIN_{EHO} model to account for the observed volatility and positive correlation between buyer- and seller-initiated order flows by allowing for simultaneous positive shocks to both order flows. This extended model allows them to compute an adjusted measure of asymmetric information (hereafter PIN_{DY}),

$$PIN_{DY} = \frac{\alpha \cdot (d \cdot \mu_B + (1-d) \cdot \mu_S)}{\alpha \cdot (d \cdot \mu_B + (1-d) \cdot \mu_S) + (\Delta_B + \Delta_S) \cdot (\alpha \cdot \theta' + (1-\alpha) \cdot \theta) + \varepsilon_S + \varepsilon_B}, \tag{2}$$

where d is the probability that informed traders receive a positive signal if a private information event occurs on a specific day, μ_B is the arrival rate of informed buyers, μ_S is the arrival rate of informed sellers, and θ is the probability that a symmetric order shock occurs in the absence of private information, whereas θ' is the probability that a symmetric order shock occurs when private information arrives. In the event of symmetric order flow shocks, the additional arrival rate of buys is Δ_B and of sells is Δ_S.

Duarte and Young's (2009) extended model also gives rise to an associated probability, $PSOS$, the unconditional probability that a given trade will come from a shock to both buy and sell order flows,

$$PSOS = \frac{(\Delta_B + \Delta_S) \cdot (\alpha \cdot \theta' + (1-\alpha) \cdot \theta)}{\alpha \cdot (d \cdot \mu_b + (1-d) \cdot \mu_s) + (\Delta_B + \Delta_S) \cdot (\alpha \cdot \theta' + (1-\alpha) \cdot \theta) + \varepsilon_S + \varepsilon_B}. \tag{3}$$

They find that firms with high $PSOS$ tend to have high Amihud (2002) illiquidity measures on most days, but experience large increases in both buy and sell orders on days with the release of public information. Shocks to both buy and sell orders may occur when traders disagree about the interpretation of a public news event, or when traders coordinate their trades on certain days to reduce transaction costs. Duarte and Young therefore argue that $PSOS$ is effectively a proxy for illiquidity unrelated to asymmetric information.

As the PIN_{DY} model contains twice as many parameters as the PIN_{EHO} model, we follow Duarte and Young (2009) by estimating a parsimonious specification of PIN_{DY} with θ equals θ'. Throughout this study, our analysis employs this model specification as it facilitates the estimation of PIN_{DY} in that its maximum likelihood estimation tends to converge more easily.

2.2. PIN Methodology and Global Intraday Data

Based on the maximum likelihood estimation procedure, we estimate both PIN_{EHO} and PIN_{DY} for every available stock using global intradaily stock transactions data from 47 countries worldwide over a 15-year period from January 2, 1996 to December 31, 2010. For a majority of the countries, the global transactions data are available from 1996 onwards. Appendix A lists the starting date of the data for each country.

The global intradaily transactions data are from TRTH,[6] managed by the Securities Industry Research Center of Asia-Pacific (SIRCA). TRTH provides millisecond-time-stamped tick data of over 5 million equity and equity derivatives instruments worldwide since January 2, 1996, and such data are sourced from the Reuters Integrated Data Network, which obtains feeds directly from the exchanges. TRTH has an equity coverage of 250 regular stock exchanges in more than 100 countries. As constrained by the availability of price data from Datastream and financial information from the Worldscope, our study only focuses on all securities listed in the main exchanges of 47 countries, and these stock exchanges are listed in Appendix A. For China, Japan, and the United States, we include stocks listed in their two main exchanges given their equal importance in the countries.It is necessary to emphasize that while the NASDAQ market is the second largest in the United States in terms of market capitalization, our sample excludes stocks traded in this market for two reasons. One, it allows us to compare our results with those of existing US studies that focus on only NYSE and AMEX stocks. Two, the NASDAQ market is a multiple-dealer market and its multiple trades based on the same order might affect the recorded number of buys and sells and hence, PIN estimates.

The initial sample covers 57,892 securities. We merge these securities with the Datastream database to obtain their basic firm-level information by using codes provided by Thomson Reuters terminals. For those securities that cannot be matched by Thomson Reuters codes, we manually match them by firm names. In total, we are able to match 44,760 securities. Next, we apply filters provided by Datastream to eliminate American Depositary Receipts, Global Depositary Receipts,

[6] The database was formerly known as the global TaqTic.

warrants, trusts, funds, and non-equity securities from our sample. After filtering, our sample is reduced to 30,095 domestic stocks that belong to their respective major share class of firms and whose primary listings are in the main stock exchange(s) of the country.

When estimating PIN, we require trades and quotes submitted during the regular trading hours of each stock exchange. TRTH provides information on trade qualifiers. Thus, trades identified as irregular trades or with negative trading prices are excluded. For quotes, we eliminate those with bid-ask spreads that are greater than half their mid-point quote prices. We employ the Lee and Ready (1991) algorithm to identify buyer- or seller-initiated trades. If quotes are missing during a trading day, we use tick tests to classify trades and then estimate the yearly PIN parameters using the maximum likelihood approach. It is noted that consistent with Duarte and Young (2009), our untabulated results also show that buyer- and seller-initiated orders are positively and significantly correlated, with mean (median) correlation coefficients of 0.543 (0.581) for stocks from developed markets, 0.645 (0.692) for those from emerging markets, and 0.597 (0.640) for the full sample. The magnitude of the correlation coefficients is comparable with the 0.50 median correlation coefficient reported in Table 1 of Duarte and Young (p. 121) for their US sample. The observed correlation coefficients in our sample of buyer- and seller-initiated orders suggest that our international analysis ought to employ Duarte and Young's approach to estimating PIN.

To avoid corner and local optimal solutions in our maximum likelihood estimations, we try a set of 7,776 (i.e., 6 different initial values for each of the 5 parameters) initial values for each maximization algorithm of PIN_{EHO} and a set of 19,683 (i.e., 3 different initial values for each of the 9 parameters) initial values for each maximization algorithm of PIN_{DY} and pick the parameters associated with the largest maximum likelihood value. Finally, we exclude observations with PIN estimates of zero or one, and these observations constitute on average about 5.2% of our total sample size for PIN_{EHO} estimates and 6.7% for PIN_{DY} estimates. As a result, our final sample covers 30,095 firms across 47 countries.

It is important to stress that we have made several checks on the accuracy of the newly, untested TRTH. First, we compare the trades from TRTH and TAQ databases for NYSE stocks reported in our sample period. After screening out duplicate trades reported in TAQ data, the trades from these two databases are identical. Note that Thomson Reuters has already filtered their data in TRTH by eliminating duplicate trades from the raw exchange data before making their data available to SIRCA. Second, we also compare trades and quotes information between TRTH and other transactions data collected from local stock exchanges that are available to us, namely the Australian stock exchange and Shanghai and Shenzhen stock exchanges. We find the information from TRTH and the two exchanges to be substantially the same. Third, we also cross-check our mean and median PIN estimates with those reported in existing studies. For example, the mean and median PIN_{EHO} estimates for NYSE stocks for the period of 1983-1998 are 0.191 and 0.185 in Easley, Hvidkjaer, and O'Hara (2002), NYSE and AMEX stocks for the period of 1983-1999 are 0.211 and 0.191 in Aslan et al. (2011), and our sample of NYSE and AMEX stocks for the period of 1996-2010 are 0.190 (mean) and 0.161 (median). Similarly, the median PIN_{DY} estimate for NYSE and AMEX stocks is 0.17 in Duarte and Young (2009), and the mean (median) PIN_{DY} estimate in our sample is 0.170 (0.151). While the PIN estimates are in the same order of magnitude, the decreasing trend in the PIN estimate probably reflects the increasing financial transparency of US markets and the implementation of an automated trading system in 2000. All these various checks reinforce our level of confidence in the accuracy of SIRCA's TRTH.

Table 1 presents the distributions of PIN_{EHO} and PIN_{DY} estimates, together with the number of sample firms, by country. Specifically, it reports their respective mean, standard deviation, quartiles 1 and 3, and median value. We estimate the two PINs for each firm-year across a sample of 16,840 firms from 22 developed countries and 13,255 firms from 25 emerging countries. The number of firms from each country is generally proportional to the size of its economy. Among the developed markets, Japan, the United States, and the United Kingdom have the largest number of firms, with each having at least 2,000 firms included in our sample, whereas Luxembourg has the smallest with only 10 firms. With the exception of India with 2,739 firms in our sample, the largest number of sample firms from emerging economies such as China, Taiwan and Malaysia is fewer than 2,000.

The table shows striking contrasts between the two PINs and across developed and emerging markets. Overall, the mean, median, and both quartiles of PIN_{EHO} are consistently larger than their PIN_{DY} counterparts. Consistent with Duarte and Young's (2009) expectation, the larger PIN_{EHO} estimate reflects not only the probability of informed trading, but also illiquidity effects unrelated to information asymmetry. For the full sample of countries, average differences between PIN_{EHO} and PIN_{DY} estimates are 0.054 (0.061) for the mean (median). Despite the difference in their sizes, PIN_{EHO} and PIN_{DY} estimates are highly correlated. The untabulated cross-country correlation coefficient of the mean (median) estimate between PIN_{EHO} and PIN_{DY} is 79.2% (80.3%).

The means of PIN_{EHO} and PIN_{DY}, with few exceptions of the latter, are at least twice the size of their respective standard deviations. The statistics indicate that emerging markets have a larger PIN than do developed markets. Based on the mean and median values, PIN_{EHO} is about 13.4%-14.6% larger in emerging than in developed markets, and PIN_{DY} is about 12.3%-13.5% larger. Unreported $p-$values from the $t-$test for the mean differentials in PIN_{EHO} and PIN_{DY} are 0.019 and 0.029, respectively, and from the Kruskal-Wallis test for their median differentials are 0.008 and 0.012, indicating that stocks from developed and emerging markets have statistically different PINs. Among the developed markets, the United States has the smallest PIN estimates of 0.190 for PIN_{EHO} and 0.170 for PIN_{DY}, while among the emerging markets, China has the smallest PINs of 0.175 for PIN_{EHO} and 0.146 for PIN_{DY}. Unlike US equity markets, Chinese equity markets are mainly dominated by individual investors, who make up of 99.5% of the total number of investor accounts in the markets (Ng and Wu, 2006). It is plausible that the low PIN estimates for China predominantly arise from individual investor trading.

2.3. The Quality of PIN as a Measure of Information Asymmetry

In our study, a majority of stock exchanges have implemented automated electronic trading systems during our sample period from January 1996 to December 2010. Only the stock exchanges of Egypt, Ireland, Israel, Jordan, Pakistan, Sri Lanka, the U.K., and the United States (i.e., NYSE) started automated trading after 1996. Many of these electronic markets are organized as electronic limit order books. This form of market structure typically has no designated liquidity provider such as a specialist or a dealer. We recognize that such electronic order-driven markets are inconsistent with the market structure type assumed in a PIN model with a central market maker.[7]

In this subsection, we examine whether PIN estimated using order flows from electronic limit order books actually perform as a measure of information asymmetry. We perform two different sets of tests to evaluate the quality of PIN estimates. One test follows Easley et al. (1996) by investigating whether PIN estimates have predictive power for spreads, and the other test examines whether PIN estimates are associated with other measures of information asymmetry at the firm and country levels.

2.3.1. PIN and Spreads

Easley et al. (1996) contend that if the quality of PIN estimates is adequate, then PIN should have a positive effect on bid-ask spreads. They investigate this issue by regressing spreads on PIN. If their model accurately estimates the probability of informed trading, they would expect the coefficient on PIN to be positive, implying that the larger the probability of informed trading, the wider are spreads. In addition, their regression analysis also includes trading volume to account for any inventory effect on spreads, and if such effects matter, then trading volume would have a negative impact on spreads.

Following Easley et al. (1996), we conduct pooled cross-country regressions of spreads, *Spread*, on both PIN and stock turnover, $Turnover$, as follows.

$$Spread = \delta_0 + \delta_1 PIN + \delta_2 Turnover + Controls + \epsilon. \tag{4}$$

We compute two different measures of spreads, the effective spread (*ESpread*) and quoted spread (*QSpread*), and for each measure, we calculate an equal-weighted and a volume-weighted average of daily percentage spreads. We also compute the correlations between these spreads and PIN. Untabulated results indicate that the average correlations between PIN_{EHO} and *ESpread* are 31.7% (full sample), 25.2% (developed markets), and 41.1% (emerging markets), and those between PIN_{DY} and *ESpread* are 19.1%, 15.8%, and 24.1%. Correspondingly, the average correlations between PIN_{EHO} and *QSpread* are 31.7% (full sample), 24.4% (developed markets), and 44.3%

[7] While there is no market making in electronic automated trading systems, the experimental study of Bloomfield, O'Hara, and Saar (2005) shows that a market-making role still arises endogenously in the electronic markets.

(emerging markets), and those between PIN_{DY} and $QSpread$ are 19.8%, 15.6%, and 26.9%. All these statistics suggest that both PIN measures are positively and strongly correlated with spreads, consistent with those of Easley et al. (1996). Results in Table 2 further reinforce these findings, thereby validating the quality of PIN estimates. Both PIN_{EHO} and PIN_{DY} have strong positive effects on the two different measures of spreads, while $Turnover$ displays a strong negative effect.

2.3.2. PIN and Proxies for Information Asymmetry at Firm and Country Levels

We now turn to testing the quality of PIN by verifying whether PIN is strongly associated with other measures of information asymmetry that are extensively employed in extant empirical studies. If PIN actually provides an estimate of the probability of information-based trading for each stock, then it should be highly correlated with other measures of information asymmetry. To address this issue, we regress PIN on several firm- and country-level information proxies, separately, while controlling for variables that can potentially affect the relationship between PIN and the information proxy in question.

Drawn from the existing literature, the firm-level measures of information asymmetry are the number of analysts following a firm ($Analysts$), analyst forecast dispersion ($FDisp$), press coverage of the firm ($Press$), firm age (Age), MSCI membership ($MSCI$), and closely-held ownership ($CHeld$), with control variables including log of total assets ($TAssets$), log of book-to-market (BM), leverage ($Leverage$), return on total assets (ROA), American Depositary Receipts (ADR), research and development scaled by total assets ($R\&D$), and stock return volatility (σ_{Ret}). The country-level proxies for information asymmetry are a country's accounting standard index ($AcStd$), disclosure requirement index ($DReq$), newspapers circulation ($Newspapers$), capital market governance (CMG), and financial transparency factor ($FTran$), as well as control variables, namely GDP per capita ($GDPC$), stock market capitalization deflated by GDP ($MCap$), ratio of private credit to GDP ($Credit$), annual GDP growth (GDP_g), standard deviation of GDP over the past five years (σ_{GDP}), market segmentation measure (SEG), and law and order index (Law & $Order$). All these variables are defined in Appendix B. Panel A of Table 3 shows pooled cross-country regressions of firm-level PIN on each information proxy as well as control variables at the firm level, whereas Panel B reports regression results of the country-median PIN against each country-level information proxy while controlling for country characteristics and year fixed effects.[8]

Several notable observations emerge from Table 3. Panel A shows that PIN is strongly associated with the level of a firm's information asymmetry measured using the extent of its analyst coverage and the earnings forecasts dispersion ($Analysts$ and $FDisp$). The estimated coefficients on analyst coverage and forecast dispersion are all statistically significant at conventional levels. Similarly, firms with wide press coverage, older firms, firms with $MSCI$ membership, and those that are less closely held ought to be associated with a low level of information asymmetry and hence, have low PINs. For instance, the coefficient estimates of $Press$, Age, $MSCI$, and $CHeld$ in M3-M6, where PIN_{EHO} is the dependent variable, are -0.008 ($t = -18.02$), -0.000 ($t = -8.29$), -0.028 ($t = -33.65$), and 0.020 ($t = 15.94$), respectively. Similar qualitative results are obtained in M9-M12, where PIN_{DY} is the dependent variable. These findings suggest that more serious adverse selection problems are evident in firms with low quality of analyst coverage or press coverage, small firms, firms whose stocks are not index members, and concentrated ownership firms. More importantly, our estimates of PIN_{EHO} and PIN_{DY} are able to reflect these adverse information costs, indicating that the quality of both PIN estimates is reasonable.

Country-level results presented in Panel B further reinforce our earlier findings about the quality of PIN estimates. The panel shows that PIN is strongly and negatively associated with all the different information proxies at the country level, indicating that PIN decreases as the country's level of information asymmetry falls. All the coefficients of these information variables, except for $FTran$ in M10, are statistically significant at the 5% level.

The overall results suggest that the two different PIN measures provide adequate estimates of the probability of information-based trading in stocks from markets with electronic limit order books. Even though these findings are robust, we acknowledge that some concerns about measurement error of PIN still remain. Hence, our evidence should be interpreted with caution.

[8] The results remain qualitatively the same if we use value-weighted PIN as the dependent variable.

3. PIN and Equity Prices Around the World

In this section, we employ two different approaches to testing the pricing of PIN in an international framework: (i) We first look at the distribution of stock returns across portfolios of stocks single-sorted on PIN and double-sorted on $Size$ and then PIN; (ii) We test whether PIN affects cross-sectional expected stock returns using Fama and French's (1992) asset pricing framework.

3.1. Excess Returns and Risk-Adjusted Returns of Portfolios formed on PIN and on Size and PIN

In Table 4, we examine the time-series association between PIN and portfolio excess returns for the period from 1997 to 2011. We compute time-series average monthly excess returns and risk-adjusted returns, *Alphas*, of global portfolios of stocks single-sorted on PIN and of stocks double-sorted on $Size$ and then PIN.

We form single-sorted PIN quintile portfolios as follows. For each year and for each country, we first rank stocks based on their prior-year PIN estimates from the lowest to the highest and then group these stocks into quintiles based on their ranked PINs. We then combine stocks of the same PIN quintile-ranking across all countries into a global PIN-ranked quintile. For example, the Low PIN portfolio consists of stocks in the lowest PIN quintile portfolio from their respective countries, and the High PIN portfolio contains those from the highest PIN quintile portfolio. We repeat this procedure annually. For double-sorted portfolios, we do the same, except that we first form three groups of stocks from each country based on their prior-year market capitalization ($Size$), and within each $Size$ portfolio, we form five groups of stocks based on their prior-year PIN estimates. Similar to single-sorted portfolios, we aggregate all stocks of the same $Size$-PIN rankings across countries into global $Size$-PIN portfolios. For each global portfolio of stocks, we compute its time-series value-weighted average of raw returns in excess of a 30-day US Treasury bill rate.

To obtain the *Alpha* of a portfolio, we regress each monthly global portfolio excess returns against Fama-French global factors for the global market portfolio (MKT^G), market capitalization

(SMB^G), and book-to-market (HML^G),

$$r_{p,t}^G = Alpha + \beta MKT_t^G + hHML_t^G + sSMB_t^G + \varepsilon_t, \tag{5}$$

where $r_{p,t}^G$ is the monthly global portfolio return in excess of a 30-day US Treasury bill rate, the intercept $Alpha$ is the risk-adjusted return, and MKT^G is the global market index excess return. SMB^G and HML^G are constructed as follows. For each country and each year, country-level SMB^C and HML^C factors for July of year t to June of year $t+1$ are constructed using six value-weighted portfolios formed at June-end of year t on the intersection of two $Size$ portfolios and three BM portfolios. The size breakpoint is determined by the median market capitalization of the country at June-end of year t, with firms below the median classified as small firms and those above as big firms. The BM breakpoints are 30th and 70th percentiles of firm BMs of the country at the fiscal year ending $t-1$, with the top 30% of firms grouped as the value portfolio, the middle 40% as the middle portfolio, and the bottom 30% as the growth portfolio. The SMB^C factor is the difference in the monthly average return between the three small portfolios and the three big portfolios, and the HML^C factor is the difference in the monthly average return between the two value portfolios and two growth portfolios. We group country-level HML^C factors together to form the global HML^G factor and country-level SMB^C factors together to construct the global SMB^G factor.

Panel A of Table 4 provides average excess returns and *Alpha*s for PIN-sorted global portfolios of stocks, while Panel B presents those of *Size*-PIN sorted global portfolios. Results of Panel A show no systematic pattern of a positive relationship between PIN and portfolio excess returns. Instead, we find that the average excess returns and *Alpha*s tend to be larger for Low than for High PIN portfolios, and that this pattern persists across portfolios formed on either PIN_{EHO} or PIN_{DY}. Differential excess returns and *Alpha*s are smaller for High$-$Low PIN_{DY} portfolios, compared with those for High$-$Low PIN_{EHO} portfolios.[9] But none of these differential excess returns and *Alpha*s are statistically different from zero, consistent with Duarte and Young's (2009) findings that PIN_{DY} has no effect on expected stock returns.

[9]The differential results are not surprising, because PIN_{EHO} contains both asymmetric information and liquidity components, whereas PIN_{DY} is only associated with the asymmetric information component.

Similar to Panel A, Panel B depicts larger excess returns and *Alpha*s mostly in Low than in High PIN portfolios, holding size constant, but again, the differences are statistically insignificant at conventional levels. Our overall findings suggest no apparent evidence of any correlation between excess returns or *Alpha*s and PIN. While our results differ from those of Easley, Hvidkjaer, and O'Hara (2002), they are consistent with the findings reported in Mohanram and Rajgopal (2009). Easley, Hvidkjaer, and O'Hara show that PIN is positively associated with excess returns for the sample period between 1983 and 1998 and that the difference between high and low PIN excess returns is smaller in small than in large stocks. They argue that private information tends to have a greater impact on price for small stocks than for large stocks. On the other hand, Mohanram and Rajgopal employ a longer sample period of 1984-2002, but find that the spread in returns between the highest and lowest PIN deciles is no longer statistically significant at conventional levels.

Overall, the time-series regression results provide no evidence that asymmetric information proxied by PIN has any effect on equity prices. In subsequent subsections, we provide further analyses to examine whether PIN is priced in a cross-sectional asset pricing framework.

3.2. PIN and the Cross-Section of Expected Equity Returns

In this subsection, we conduct asset pricing tests similar to those employed by Easley, Hvidkjaer, and O'Hara (2002) and Duarte and Young (2009) to examine whether the asymmetric information or illiquidity component of PIN is priced in an international setting. Table 5 reports time-series averages of the estimated coefficients from cross-sectional regressions of excess stock returns against PIN and with combinations of $PSOS$ and $Illiquidity$, while controlling for previously found return predictors, namely the log of book-to-market equity ratio (BM), log of market capitalization ($Size$), country market beta (β_C), and global market beta (β_G). The definitions of all the variables together with their data sources are contained in Appendix B. The table also shows time-series averages of the regression slope coefficients, together with their robust $t-$statistics in parentheses, for the full sample of firms from 47 countries and sub-samples of firms from developed and emerging markets.

The table reveals several distinctive results. First, our findings show a negative and mainly statistically insignificant PIN_{EHO} coefficient. The negative PIN_{EHO} coefficient, however, seems counter-intuitive, because it suggests that the expected return is decreasing in information risk. Although this finding contradicts Easley, Hvidkjaer, and O'Hara's (2002) result that PIN reflects information risk systematically priced by investors, it is consistent with the US evidence documented in Mohanram and Rajgopal (2009). The latter employ the implied cost of capital as a proxy for the expected return and show the PIN_{EHO} coefficient to be negative and not robustly significant at conventional levels. They therefore argue that the pricing of PIN_{EHO} is sensitive to alternative specifications and time periods. Our findings further reinforce their results by also showing that the effect of PIN_{EHO} on the cross-section of expected returns is not robust across international markets.

Second, the results based on PIN_{DY} are broadly consistent with Duarte and Young's (2009) findings that PIN_{DY} exhibits no effect on expected returns, suggesting that asymmetric information associated with PIN_{EHO} is not priced. *Illiquidity* continues to maintain its level of significance in all model specifications. The role of illiquidity in asset prices is not only shown in Duarte and Young's study of US equity markets, but also consistent with the recent evidence documented in Lee (2011) that liquidity risk is priced in international financial markets. We also show that only the book-to-market effect is strongly significant and positive, and that other conventional proxies for firm risk, such as country and global market betas as well as $Size$, are insignificantly related to the cross-section of expected stock returns. These findings are also reported in both Duarte and Young and Lee.

Third, when $PSOS$ and PIN_{EHO} are estimated jointly in model M3, the coefficient of the illiquidity component $PSOS$, not associated with information asymmetry, is statistically significant and negative. A similar result is obtained when PIN_{DY} is used in place of PIN_{EHO} in model M11. These results seem to contradict Duarte and Young's (2009) finding of a positive $PSOS$ impact on expected US stock returns for the period from 1983 to 2004. They interpret that high $PSOS$ stocks tend to be very illiquid and hence, have a positive illiquidity premium. While it is plausible that the difference in results may be due to the different sample periods employed in both studies (our sample period is 1996-2010 and theirs is 1983-2004), we concede that the negative $PSOS$ coefficient is puzzling. We, however, leave this puzzle for future research.

One may argue that our results are likely driven by orders submitted by algorithm trading implemented in a multiplicity of markets around the globe. The increase in high-frequency trading accounts for the majority of trading volume in today's markets (see Easley, López de Prado, and O'Hara, 2012). Such trading algorithms are designed to delay or accelerate trading in reaction to market events within milliseconds. For example, traders may split large orders into multiple small orders, and such orders occurring in short intervals are not truly independent observations. To rule out this alternative interpretation, we calculate the numbers of buyer- and seller-initiated orders by aggregating orders on the same side of the market over short intervals into a single observation in the following ways: (i) aggregating sequential trading at the same price if there is no update in quotes (PIN^1), (ii) aggregating sequential trading within 15 seconds if there is no update in quotes (PIN^2), and (iii) aggregating sequential trading if there is no update in quotes (PIN^3). We replicate key regression models of Table 5 (i.e., M5 and M12) using these revised PIN estimates; the results presented in Table 6 remain materially unaltered, suggesting that our main findings are robust to high-frequency trading.

4. Additional Tests

Consistent with theoretical arguments,[10] our earlier evidence of a generally insignificant PIN effect on expected returns possibly suggests that information-risk measured by PIN is diversifiable. Thus, it is likely that we can find similar evidence using alternative information-based trading measures. This motivates us to exploit the richness of our database to test whether information risk proxied by alternative trading-based information measures can explain the cross-section of expected stock returns in international markets. If the alternative information-based trading measures, while not PIN, have a significant positive effect on expected stock returns, then we argue that PIN may not be a good proxy for information asymmetry. On the other hand, if the alternative trading-based information measures also exhibit no significant impact on expected stock returns, then we interpret that information risk related to trading-based measures, in general, is not priced.

Given that we cannot exhaust the many different measures of informed trading in the existing

[10]See Fama (1991), Hughes, Liu, and Liu (2007), and Lambert, Leuz, and Verrecchia (2007).

literature, we select the following four measures that we consider to be more popularly employed in extant empirical studies. The first measure is Hasbrouck's (1991) measure of relative trade informativeness, R_w^2 (equation (6), p. 577), and

$$R_w^2 \equiv \frac{\sigma_{wx}^2}{\sigma_w^2}. \tag{6}$$

R_w^2 is the coefficient of determination in a regression of price innovation w on trade innovation x. w reflects the market's updates to the available information set, whereas x reflects the market's signal of private information through trading. The second measure is Huang and Stoll's (1996) percentage price impact measure, $\%PImpact$,

$$\%PImpact = \frac{2 \times Q_{it} \times (M_{i,t+30} - M_{it})}{M_{it}}, \tag{7}$$

where Q_{it} is a binary variable that equals $+1$ for buyer-initiated orders and -1 for seller-initiated orders; $M_{i,t+30}$ is the mid-point of the first quote reported at least 30 minutes after the transaction. $\%PImpact$ incorporates liquidity providers' quote revisions following a series of buyer- or seller-initiated orders. We employ Huang and Stoll's (1997) adverse selection component as the third measure (equation (23), p. 1014).

$$\Delta M_t = (\alpha + \beta)\frac{S}{2}Q_{t-1} - \alpha\frac{S}{2}(1 - 2\pi)Q_{t-2} + \epsilon_t, \tag{8}$$

where M_t is the quote midpoint calculated from bid-ask quotes that occur just before a transaction, S is a constant spread, π is the probability of trade reversals, and Q_t is a buy-sell trade indicator that equals $+1$ for a buyer-initiated trade and -1 for a seller-initiated. α is the adverse selection component of the half-spread, and β is the inventory holding component. The conditional expectation of the trade indicator at time $t-1$, given Q_{t-2}, is shown in equation (21) of Huang and Stoll,

$$E(Q_{t-1}|Q_{t-2}) = (1 - 2\pi)Q_{t-2}. \tag{9}$$

Esimating the preceding two equations simultaneously, we obtain an estimate of the adverse selection component, α, and label it α_{HS}. The last measure is the asymmetric information parameter derived from Madhavan, Richardson, and Roomans's (1997) model for transaction price changes (equation (4), p. 1042),[11]

$$p_t - p_{t-1} = (\phi + \theta)x_t - (\phi + \rho\theta)x_{t-1} + \epsilon_t + \xi_t - \xi_{t-1}, \tag{10}$$

where $p_t - p_{t-1}$ is the change in transaction prices, ϕ is the cost of supplying liquidity, θ is the asymmetric information parameter, ρ is the autocorrelation of the order flow, and x_t is the trade initiation variable. To distinguish the different notations used in this study, we use θ_{MRR} to denote the asymmetric information parameter θ.

We proceed to replicate Fama-MacBeth regressions of M2 and M5 from Table 5 using the above four different information-based trading measures, as well as three different first principal components denoted by $PComp^1$, $PComp^2$, and $PComp^3$. $PComp^1$ ($PComp^2$) is the first principal component extracted from performing a principal component analysis on PIN_{EHO} (PIN_{DY}), α_{HS}, θ_{MRR}, $\%PImpact$, and R_W^2, while $PComp^3$ is extracted using all six measures altogether. Results are shown in Table 7.

Consistent with those of Table 5, information-based trading measures, in general, exhibit no strongly significant effect on expected stock returns; only Huang and Stoll's (1996, 1997) private-information measures, $\%PImpact$ and α_{HS}, have a marginally significant effect. The coefficient estimates of $\%PImpact$ and α_{HS} are 18.738 ($t = 1.76$) in M3 and 10.641 ($t = 1.88$) in M5, respectively. But when jointly estimated with *Illiquidity*, they become statistically insignificant. *Illiquidity*, however, continues to have a consistently, positive effect on expected stock returns. Overall, these results suggest that information risk proxied not only by PIN, but also by four alternative trading-based measures, in general, is not robustly priced.

5. Summary

The pricing of information asymmetry has become a recent subject of debate in both theoretical and empirical asset pricing and microstructure literatures. On the one hand, Easley et al. (1996), Easley, Hvidkjaer, and O'Hara (2002), and Easley and O'Hara (2004) provide theoretical arguments, with supporting empirical evidence, that information risk associated with PIN is priced. On the

[11] See Madhavan, Richardson, and Roomans (1997) for the assumptions underlying this model specification.

other hand, theoretical models of Hughes, Liu, and Liu (2007) and Lambert, Leuz, and Verrecchia (2007) yield empirical implications that are at variance with those in Easley and O'Hara (2004). Specifically, their models imply that information risk is potentially idiosyncratic in nature and hence, fully diversifiable. Empirically, Duarte and Young (2009) also find no evidence that PIN is a priced information risk. Our study contributes to this controversy over the pricing effect of PIN by subjecting PIN to more rigorous tests but in an international setting. To the best of our knowledge, our investigation represents the first to examine the asset pricing implications of PIN for a large cross-section of international firms from a wide spectrum of countries around the world.

We estimate both Easley, Hvidkjaer, and O'Hara's (2002) and Duarte and Young's (2009) PINs (PIN_{EHO} and PIN_{DY}) for a sample of 30,095 firms from 47 countries worldwide for which we have intradaily transactions data to estimate their stock-level PINs over a 15-year period from 1996 to 2010. Our international sample expands the US samples employed by Easley, Hvidkjaer, and O'Hara and Duarte and Young, whose sample periods span from 1983 to 1998 and from 1983 to 2004, respectively. During our sample period, all stock exchanges, including those of the United States, have moved to adopt an automated electronic limit order book system. Such a system, however, differs from the market structure with specialists that the PIN model assumes. Therefore, we perform two tests to ensure that PIN indeed captures the probability of informed trading for our sample of stocks as we have expected. Results validate the reasonableness of PIN quality by showing that both PIN_{EHO} and PIN_{DY} predict spreads in accordance with theoretical arguments, and that they are strongly correlated with other measures of firm- and country-level private information widely employed in existing studies.

Our analysis shows robust evidence that PIN exhibits no positive relationship with expected stock returns. This finding not only reinforces Duarte and Young's (2009) result that the information asymmetry associated with PIN is not priced, but also suggests that the pricing of PIN is not robust across international markets. We further explore whether other proxies for information asymmetry, specifically alternative information-based trading measures, have any effect on expected stock returns. Drawn from the existing literature, we employ four widely adopted trading-based measures of asymmetric information and find evidence consistent with our main findings that information risk related to trading-based measures is not systematically priced by investors in international markets. This finding suggests that one needs to be cautious when interpreting results of earlier studies that rely on PIN as a priced information risk.

References

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

Aslan, H., Easley, D., Hvidkjaer, S., O'Hara, M., 2011. Firm characteristics and informed trading: implications for asset pricing. Journal of Empirical Finance 18, 782-801.

Bekaert, G., Harvey, C., Lundblad, C., Siegel, S., 2011. What segments equity markets? Review of Financial Studies 24, 3841-3890.

Bhattacharya, U., Daouk, H., 2002. The world price of insider trading. Journal of Finance 57, 75-107.

Bloomfield, R., O'Hara, M., Saar, G., 2005. The make or take decision in an electronic market: evidence on the evolution of liquidity. Journal of Financial Economics 75, 165-199.

Bushman, R.M., Piotroski, J.D., Smith, A.J., 2004. What determines corporate transparency? Journal of Accounting Research 42, 207-252.

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

Charoenrook, A., Daouk, H., 2005. Market-wide short-selling restrictions. Unpublished working paper. Cornell University.

Duarte, J., Han, X., Harford, J., Young, L., 2008. Information asymmetry, information dissemination and the effect of regulation FD on the cost of capital. Journal of Financial Economics 87, 24-44.

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.M., O'Hara, M., 1997. One day in the life of a very common stock. Review of Financial Studies 10, 805-835.

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

Easley, D., de Prado, M.M.L., O'Hara, M., 2012. Flow toxicity and liquidity in a high-frequency world. Review of Financial Studies 25, 1457-1493.

Easley, D., O'Hara, M., 2004. Information and the cost of capital. Journal of Finance 59, 1553-1583.

Easley, D., O'Hara, M., Srinivas, P.S., 1998. Option volume and stock prices: evidence on where informed traders trade. Journal of Finance 53, 431-465.

Ellul, A., Pagano, M., 2006. IPO underpricing and after-market liquidity. Review of Financial Studies 19, 381-421.

Fama, E., 1991. Efficient capital markets: II. Journal of Finance 46, 1575-1617.

Fama, E., French, K., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427-465.

Fama, E., MacBeth, J., 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy 81, 607-636.

Hasbrouck, J., 1991. The summary informativeness of stock trades: an econometric analysis. Review of Financial Studies 4, 571-595.

Huang, R., Stoll, H., 1996. Dealer versus auction markets: a paired comparison of execution costs on NASDAQ and the NYSE. Journal of Financial Economics 41, 313-357.

Huang, R., Stoll, H., 1997. The components of the bid-ask spread: a general approach. Review of Financial Studies 10, 995-1034.

Hughes, J.S., Liu, J., Liu, J., 2007. Information asymmetry, diversification, and cost of capital. The Accounting Review 82, 705-729.

Jain, P.K., 2005. Financial market design and the equity premium: electronic versus floor trading. Journal of Finance 60, 2955-2985.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R.W., 1998. Law and finance. Journal of Political Economy 106, 1113-1155.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., 2006. What works in securities laws? Journal of Finance 61, 1-32.

Lambert, R., Leuz, C., Verrecchia, R., 2007. Accounting information, disclosure, and the cost of capital. Journal of Accounting Research 45, 385-420.

Lee, C.M.C., Ready, M.J., 1991. Inferring trade direction from intraday data. Journal of Finance 46, 733-746.

Lee, K.H., 2011. The world price of liquidity risk. Journal of Financial Economics 99, 136-161.

Madhavan, A., Richardson, M., Roomans, M., 1997. Why do security prices change? A transaction-level analysis of NYSE stocks. Review of Financial Studies 10, 1035-1064.

Mohanram, P., Rajgopal, S., 2009. Is PIN priced risk? Journal of Accounting and Economics 47, 226-243.

Ng, L., Wu, F., 2006. Revealed stock preferences of individual investors: evidence from Chinese equity markets. Pacific-Basin Finance Journal 14, 175-192.

Vega, C., 2006. Stock price reaction to public and private information. Journal of Financial Economics 82, 103-133.

Table 1
Summary Statistics of PIN_{EHO} and PIN_{DY} by Country

This table reports the mean, median, standard deviation (Std), and quartiles 1 and 3 ($Q1$ and $Q3$) of the probability of informed trading (PIN) constructed using the Easley, Hvidkjaer, and O'Hara (2002) approach (PIN_{EHO}) and of the adjusted PIN suggested by Duarte and Young (2009) (PIN_{DY}). It also presents the type of market (emerging, EMG; or developed, DEV) and the number of firms in our sample ($NFirms$). See Appendix A for the starting year of availability for intraday data used to compute PIN for firms in each country. The sample period is from 1996 to 2010.

	Type of		*Distribution of* PIN_{EHO}					*Distribution of* PIN_{DY}				
Country	*Market*	*NFirms*	*Mean*	*Std*	*Q1*	*Median*	*Q3*	*Mean*	*Std*	*Q1*	*Median*	*Q3*
Argentina	EMG	81	0.354	0.120	0.266	0.338	0.427	0.277	0.129	0.183	0.251	0.335
Australia	DEV	1,946	0.284	0.091	0.228	0.278	0.333	0.184	0.089	0.122	0.168	0.223
Austria	DEV	81	0.276	0.119	0.199	0.263	0.325	0.226	0.101	0.157	0.206	0.266
Belgium	DEV	148	0.265	0.110	0.196	0.249	0.308	0.199	0.095	0.129	0.180	0.244
Brazil	EMG	152	0.288	0.100	0.217	0.270	0.352	0.240	0.096	0.180	0.217	0.275
Canada	DEV	1,210	0.272	0.095	0.208	0.260	0.323	0.239	0.098	0.173	0.221	0.286
Chile	EMG	113	0.318	0.101	0.244	0.316	0.378	0.291	0.114	0.208	0.275	0.366
China	EMG	1,791	0.175	0.072	0.131	0.163	0.201	0.146	0.069	0.100	0.125	0.183
Denmark	DEV	217	0.268	0.097	0.206	0.258	0.310	0.175	0.086	0.115	0.159	0.205
Egypt	EMG	200	0.339	0.120	0.245	0.328	0.423	0.283	0.119	0.194	0.262	0.351
Finland	DEV	148	0.236	0.077	0.189	0.236	0.278	0.204	0.093	0.142	0.186	0.243
France	DEV	829	0.241	0.093	0.183	0.234	0.284	0.206	0.084	0.149	0.193	0.250
Germany	DEV	936	0.206	0.088	0.155	0.191	0.234	0.195	0.091	0.136	0.178	0.229
Greece	EMG	337	0.250	0.091	0.186	0.228	0.291	0.221	0.099	0.152	0.197	0.260
Hong Kong	DEV	1,086	0.278	0.080	0.228	0.273	0.320	0.186	0.079	0.134	0.174	0.219
India	EMG	2,739	0.263	0.088	0.200	0.254	0.316	0.195	0.088	0.138	0.178	0.224
Indonesia	EMG	399	0.385	0.112	0.309	0.375	0.448	0.275	0.096	0.210	0.259	0.320
Ireland	DEV	59	0.262	0.084	0.207	0.252	0.316	0.234	0.103	0.158	0.213	0.288
Israel	EMG	644	0.268	0.093	0.208	0.257	0.309	0.238	0.101	0.169	0.220	0.281
Italy	DEV	344	0.220	0.078	0.168	0.211	0.263	0.173	0.081	0.121	0.154	0.199
Japan	DEV	2,902	0.233	0.087	0.168	0.221	0.285	0.207	0.100	0.134	0.187	0.251
Jordan	EMG	226	0.349	0.101	0.279	0.341	0.411	0.286	0.113	0.202	0.262	0.359
Luxembourg	DEV	10	0.379	0.130	0.287	0.346	0.460	0.222	0.116	0.128	0.207	0.276
Malaysia	EMG	1,145	0.315	0.079	0.260	0.309	0.361	0.234	0.084	0.183	0.220	0.266
Mexico	EMG	115	0.317	0.115	0.226	0.303	0.399	0.269	0.114	0.188	0.251	0.331
Netherlands	DEV	158	0.214	0.083	0.156	0.204	0.258	0.210	0.100	0.140	0.187	0.248
New Zealand	DEV	129	0.364	0.179	0.238	0.312	0.461	0.336	0.160	0.215	0.309	0.442
Norway	DEV	278	0.276	0.092	0.215	0.272	0.326	0.237	0.102	0.164	0.217	0.282
Pakistan	EMG	418	0.314	0.094	0.247	0.305	0.368	0.238	0.098	0.171	0.217	0.290
Peru	EMG	63	0.391	0.099	0.327	0.381	0.436	0.327	0.121	0.240	0.310	0.402
Philippines	EMG	223	0.330	0.090	0.265	0.321	0.388	0.251	0.090	0.196	0.237	0.292
Poland	EMG	392	0.296	0.090	0.241	0.284	0.339	0.241	0.095	0.179	0.224	0.275
Portugal	EMG	57	0.294	0.126	0.200	0.285	0.356	0.251	0.112	0.166	0.231	0.307
Russia	EMG	264	0.288	0.106	0.210	0.279	0.347	0.221	0.090	0.157	0.201	0.266
Saudi Arabia	EMG	137	0.257	0.116	0.183	0.240	0.303	0.200	0.081	0.149	0.181	0.228
Singapore	DEV	801	0.297	0.080	0.245	0.290	0.341	0.180	0.083	0.123	0.165	0.217
South Africa	EMG	449	0.305	0.105	0.231	0.297	0.371	0.260	0.113	0.180	0.241	0.322
South Korea	EMG	802	0.237	0.075	0.185	0.225	0.275	0.202	0.076	0.155	0.185	0.225
Spain	DEV	153	0.208	0.071	0.159	0.199	0.245	0.175	0.088	0.115	0.151	0.201
Sri Lanka	EMG	194	0.320	0.110	0.248	0.306	0.376	0.207	0.095	0.140	0.193	0.255
Sweden	DEV	504	0.241	0.080	0.190	0.230	0.278	0.219	0.096	0.155	0.196	0.255
Switzerland	DEV	275	0.282	0.098	0.216	0.269	0.328	0.246	0.101	0.177	0.225	0.297
Taiwan	EMG	1,413	0.219	0.083	0.161	0.201	0.256	0.178	0.094	0.118	0.147	0.202
Thailand	EMG	572	0.306	0.105	0.237	0.281	0.346	0.217	0.075	0.174	0.205	0.245
Turkey	EMG	329	0.220	0.062	0.181	0.207	0.244	0.180	0.073	0.140	0.162	0.195
United Kingdom	DEV	2,269	0.247	0.098	0.180	0.238	0.297	0.221	0.107	0.148	0.199	0.261
United States	DEV	2,357	0.190	0.098	0.121	0.161	0.235	0.170	0.082	0.112	0.151	0.201
	DEV	16,840	0.261	0.096	0.197	0.248	0.309	0.211	0.097	0.143	0.192	0.254
	EMG	13,255	0.296	0.098	0.228	0.284	0.349	0.237	0.098	0.171	0.218	0.282
	ALL	30,095	0.279	0.097	0.213	0.267	0.330	0.225	0.097	0.158	0.206	0.269

Table 2
PIN Estimates and Spreads

This table presents pooled cross-country regressions of spreads on the probability of information-based trading (PIN), constructed using the Easley, Hvidkjaer, and O'Hara (2002) methodology (PIN_{EHO}) or Duarte and Young's (2009) approach (PIN_{DY}), and stock turnover ($Turnover$), as well as several controls which include year-, industry-, and country-fixed effects (FE).

$$Spread = \delta_0 + \delta_1 PIN + \delta_2 Turnover + Controls + \epsilon$$

We employ two different measures of spreads: effective spreads ($ESpread$) and quoted spreads ($QSpread$). We employ equal-weighted and volume-weighted effective spreads ($ESpread_{EW}$ and $ESpread_{VW}$) and equal-weighted and volume-weighted quoted spreads ($QSpread_{EW}$ and $QSpread_{VW}$). All variables are defined in Appendix B. NObs is the number of observations; $\bar{R}^2$ is the adjusted R^2. Robust $t-$statistics are in parentheses. The sample period is from 1996 to 2010.

Panel A: PIN and $ESpread$

	$PIN = PIN_{EHO}$						$PIN = PIN_{DY}$					
	$ESpread_{EW}$			$ESpread_{VW}$			$ESpread_{EW}$			$ESpread_{VW}$		
	All	DEV	EMG	All	DEV	EMG	All	DEV	EMG	All	DEV	EMG
Variable	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
PIN	0.055	0.044	0.069	0.061	0.052	0.073	0.032	0.028	0.036	0.037	0.035	0.040
	(64.54)	(40.76)	(52.31)	(57.30)	(35.93)	(48.29)	(45.00)	(31.71)	(31.74)	(39.73)	(27.27)	(30.31)
$Turnover$	-0.001	-0.002	-0.001	-0.001	-0.002	-0.001	-0.001	-0.002	-0.001	-0.001	-0.002	-0.001
	(-29.63)	(-20.16)	(-20.41)	(-28.81)	(-19.47)	(-20.58)	(-31.34)	(-21.10)	(-22.18)	(-30.41)	(-20.46)	(-22.02)
NObs	179,663	106,955	72,708	179,854	107,076	72,778	179,663	106,955	72,708	179,854	107,076	72,778
$\bar{R}^2$	28.6%	27.9%	31.7%	24.9%	24.1%	27.6%	25.7%	26.4%	26.1%	23.0%	23.6%	23.3%
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: PIN and $QSpread$

	$PIN = PIN_{EHO}$						$PIN = PIN_{DY}$					
	$QSpread_{EW}$			$QSpread_{VW}$			$QSpread_{EW}$			$QSpread_{VW}$		
Variable	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24
PIN	0.058	0.049	0.071	0.060	0.051	0.072	0.034	0.031	0.037	0.038	0.034	0.042
	(60.26)	(37.67)	(50.98)	(59.40)	(37.33)	(49.68)	(41.81)	(28.67)	(31.50)	(43.75)	(29.78)	(32.79)
$Turnover$	-0.001	-0.002	-0.001	-0.001	-0.003	-0.001	-0.001	-0.002	-0.001	-0.001	-0.003	-0.001
	(-28.61)	(-20.83)	(-17.82)	(-34.24)	(-24.87)	(-21.22)	(-30.21)	(-21.61)	(-19.69)	(-35.23)	(-25.29)	(-22.88)
NObs	179,671	106,939	72,732	179,854	107,076	72,778	179,671	106,939	72,732	179,854	107,076	72,778
$\bar{R}^2$	29.9%	28.8%	33.8%	28.5%	27.3%	32.2%	27.4%	27.5%	28.6%	26.3%	26.1%	27.7%
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 3

PIN and Information Asymmetry Proxies at Firm and Country Levels

Panel A of this table shows the relationship between PIN estimates and different firm-level measures of information asymmetry. We employ PIN constructed using the Easley, Hvidkjaer, and O'Hara (2002) methodology (PIN_{EHO}) or Duarte and Young's (2009) approach (PIN_{DY}). The firm-level measures of information asymmetry include number of analysts following a firm (*Analysts*), analyst forecast dispersion ($FDisp$), press coverage of the firm ($Press$), firm age (Age), MSCI membership ($MSCI$), and closely-held ownership ($CHeld$). The control variables are log of total assets ($TAssets$), log of book-to-market (BM), leverage ($Leverage$), return on total assets (ROA), American Depositary Receipts (ADR), research and development scaled by total assets ($R\&D$), and stock return volatility (σ_{Ret}), as well as untabulated year-, industry-, and country-fixed effects (FE). Panel B presents the relationship between country-median PIN estimates and different country-level measures of information asymmetry, which include a country's accounting standard index ($AcStd$), disclosure requirement index ($DReq$), newspapers circulation ($Newspapers$), capital market governance (CMG), and financial transparency factor ($FTran$). The country-level control variables are log of GDP per capita ($GDPC$), stock market capitalization deflated by GDP ($MCap$), ratio of private credit to GDP ($Credit$), annual GDP growth (GDP_g), standard deviation of GDP over the last 5 years (σ_{GDP}), market segmentation measure (SEG), and law and order index (Law & $Order$), as well as untabulated year-fixed effects (FE). All variables are defined in Appendix B. NObs is the number of observations; $\bar{R}^2$ is the adjusted R^2. Robust $t-$statistics are in parentheses. The sample period is from 1996 to 2010.

Panel A: PIN and Firm-Level Proxies for Information Asymmetry, $Info_{Firm}$

	$PIN = PIN_{EHO}$						$PIN = PIN_{DY}$					
	Definition of $Info_{Firm}$											
	$Analysts$	$FDisp$	$Press$	Age	$MSCI$	$CHeld$	$Analysts$	$FDisp$	$Press$	Age	$MSCI$	$CHeld$
Variable	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
$Info_{Firm}$	-0.002	0.001	-0.008	-0.000	-0.028	0.020	-0.001	0.001	-0.005	-0.000	-0.020	0.012
	(-24.82)	(2.57)	(-18.02)	(-8.29)	(-33.65)	(15.94)	(-15.79)	(2.69)	(-12.33)	(-6.64)	(-26.59)	(10.55)
$TAssets$	-0.013	-0.016	-0.016	-0.017	-0.013	-0.017	-0.008	-0.010	-0.010	-0.010	-0.008	-0.011
	(-53.54)	(-55.10)	(-50.80)	(-76.68)	(-55.53)	(-84.70)	(-38.18)	(-39.12)	(-36.40)	(-55.35)	(-37.44)	(-61.08)
BM	0.007	0.008	0.011	0.009	0.007	0.009	0.006	0.007	0.009	0.007	0.006	0.007
	(19.80)	(15.02)	(22.42)	(26.55)	(20.26)	(26.98)	(18.63)	(12.79)	(19.48)	(23.15)	(18.12)	(23.34)
$Leverage$	0.007	0.011	0.010	0.012	0.008	0.013	0.007	0.009	0.007	0.010	0.007	0.011
	(3.90)	(5.12)	(4.54)	(7.01)	(4.65)	(7.78)	(4.76)	(4.29)	(3.35)	(6.70)	(4.83)	(7.23)
ROA	0.010	-0.004	0.005	0.009	0.011	0.008	0.002	-0.013	-0.001	0.002	0.003	0.001
	(4.63)	(-1.24)	(1.61)	(4.45)	(5.33)	(3.72)	(1.12)	(-3.45)	(-0.43)	(1.01)	(1.61)	(0.58)
ADR	-0.005	-0.013	0.001	-0.011	-0.013	-0.010	-0.006	-0.009	0.003	-0.010	-0.011	-0.009
	(-2.10)	(-5.35)	(0.52)	(-4.96)	(-6.07)	(-4.33)	(-2.99)	(-4.43)	(1.49)	(-4.87)	(-5.77)	(-4.46)
$R\&D$	-0.023	-0.028	-0.038	-0.035	-0.026	-0.038	-0.005	-0.007	-0.001	-0.012	-0.006	-0.013
	(-3.05)	(-2.80)	(-3.67)	(-4.57)	(-3.54)	(-4.96)	(-0.71)	(-0.74)	(-0.08)	(-1.64)	(-0.79)	(-1.88)
σ_{Ret}	0.012	0.004	0.012	0.011	0.013	0.012	-0.007	-0.013	-0.008	-0.008	-0.007	-0.007
	(14.76)	(2.99)	(9.53)	(14.06)	(15.67)	(14.48)	(-9.29)	(-8.79)	(-6.40)	(-9.68)	(-8.58)	(-9.40)
NObs	154,597	70,760	77,050	154,591	154,597	154,597	154,597	70,760	77,050	154,591	154,597	154,597
$\bar{R}^2$	33.6%	33.0%	30.3%	33.0%	34.1%	33.2%	15.6%	15.4%	14.6%	15.4%	16.0%	15.5%
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 3 - Continued

PIN and Information Asymmetry Proxies at Firm and Country Levels

Panel B: PIN and Country-Level Proxies for Information Asymmetry, $Info_{Country}$

	Country Median $PIN = PIN_{EHO}$					Country Median $PIN = PIN_{DY}$				
	Definition of $Info_{Country}$									
	$AcStd$	$DReq$	$Newspapers$	CMG	$FTran$	$AcStd$	$DReq$	$Newspapers$	CMG	$FTran$
Variable	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
$Info_{Country}$	-0.048	-0.029	-0.068	-0.075	-0.020	-0.086	-0.069	-0.044	-0.082	-0.007
	(-1.97)	(-2.52)	(-3.39)	(-4.65)	(-3.82)	(-3.61)	(-6.49)	(-2.23)	(-5.56)	(-1.74)
$GDPC$	-0.018	-0.025	-0.007	-0.004	-0.012	-0.010	-0.018	-0.003	0.004	-0.007
	(-4.93)	(-7.98)	(-2.11)	(-1.24)	(-3.04)	(-2.71)	(-6.05)	(-0.79)	(1.38)	(-1.99)
$MCap$	0.016	0.013	0.015	0.020	0.008	0.001	0.001	-0.003	0.002	-0.011
	(3.49)	(3.30)	(4.00)	(6.25)	(1.97)	(0.16)	(0.18)	(-0.71)	(0.75)	(-2.72)
$Credit$	-0.024	-0.018	-0.043	-0.039	-0.012	-0.006	0.007	-0.020	-0.020	0.004
	(-4.14)	(-3.11)	(-6.83)	(-6.60)	(-2.12)	(-1.19)	(1.36)	(-3.86)	(-4.34)	(0.85)
GDP_g	-0.156	-0.154	-0.657	-0.424	-0.326	-0.198	-0.098	-0.517	-0.360	-0.294
	(-1.13)	(-1.23)	(-5.10)	(-3.59)	(-2.39)	(-1.65)	(-0.94)	(-4.13)	(-3.72)	(-2.38)
σ_{GDP}	0.062	0.398	0.421	-0.409	1.836	-0.204	0.310	0.416	-0.351	1.493
	(0.29)	(1.85)	(1.69)	(-2.56)	(6.57)	(-1.19)	(1.54)	(1.77)	(-2.35)	(5.32)
SEG	0.658	0.358	0.336	0.192	0.050	0.332	0.061	-0.070	-0.006	-0.170
	(4.43)	(2.67)	(1.73)	(1.50)	(0.32)	(2.34)	(0.47)	(-0.37)	(-0.04)	(-1.11)
Law & $Order$	0.033	0.042	0.013	-0.017	0.044	-0.012	0.006	-0.024	-0.051	-0.005
	(1.67)	(2.26)	(0.68)	(-1.00)	(2.13)	(-0.60)	(0.35)	(-1.28)	(-3.35)	(-0.27)
NObs	529	584	522	636	541	529	584	522	636	541
$\bar{R}^2$	25.1%	28.8%	26.3%	24.1%	28.6%	15.7%	17.5%	13.3%	17.4%	14.5%
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 4

Monthly Excess Returns and Risk-Adjusted Returns (*Alphas*) of PIN-Sorted and Size-PIN Sorted Portfolios

This table presents value-weighted average monthly excess returns and risk-adjusted returns (*Alphas*) of stock portfolios sorted on PIN and on *Size* and then PIN. PIN is constructed using the Easley, Hvidkjaer, and O'Hara (2002) methodology (PIN_{EHO}) or Duarte and Young's (2009) approach (PIN_{DY}) The *Alpha* is the intercept obtained from regressing monthly portfolio excess returns (r_p^G) against global Fama-French factors for the global market portfolio (MKT^G), size factor (SMB^G), and book-to-market factor (HML^G).

$$r_{p,t}^G = Alpha + \beta MKT_t^G + hHML_t^G + sSMB_t^G + \varepsilon_t \quad (11)$$

We construct single-sorted PIN quintile portfolios as follows. For each year and for each country, we first rank stocks based on their prior-year PIN estimates from the lowest to the highest and then group these stocks into quintiles based on their ranked PINs. We then combine stocks of the same PIN quintile-ranking from each country into a global PIN-ranked quintile. For example, the Low PIN portfolio consists of stocks in the lowest PIN quintile portfolio from their respective countries, and the High PIN portfolio contains those from the highest PIN quintile portfolio. The remaining portfolios are formed in a similar manner. We repeat this procedure annually. For the double-sorted portfolios, we do the same, except that we first form three groups of stocks from each country based on their prior-year market capitalization (*Size*), and within each size portfolio, we form five groups of stocks based on their prior-year PIN values from the lowest to the highest. Similar to single-sorted portfolios, we combine all stocks of the same *Size*-PIN rankings from their respective countries into global *Size*-PIN portfolios. All $t-$statistics reported in parentheses are based on Newey-West standard errors. Monthly excess returns are from January 1997 to December 2011, whereas PIN estimates are from 1996 to 2010.

Panel A: Single-Sorted Portfolios

	$PIN = PIN_{EHO}$						$PIN = PIN_{DY}$					
	Low	2	3	4	High	High−Low	Low	2	3	4	High	High−Low
Excess Return	0.592	0.387	0.481	0.262	0.479	-0.113	0.586	0.808	0.240	0.290	0.230	-0.356
t-value	(1.28)	(0.84)	(0.97)	(0.55)	(1.06)	(-0.56)	(1.24)	(1.40)	(0.54)	(0.69)	(0.54)	(-1.51)
Alpha	0.049	-0.111	-0.083	-0.318	-0.095	-0.144	0.030	0.298	-0.356	-0.260	-0.266	-0.296
t-value	(0.50)	(-0.65)	(-0.44)	(-2.71)	(-0.55)	(-0.78)	(0.22)	(1.43)	(-2.44)	(-1.93)	(-1.97)	(-1.44)

Panel B: Double-Sorted Portfolios

		$PIN = PIN_{EHO}$						$PIN = PIN_{DY}$					
Size		Low	2	3	4	High	High−Low	Low	2	3	4	High	High−Low
Small	Excess return	0.492	0.597	0.528	0.451	0.332	-0.159	0.479	0.507	0.499	0.546	0.399	-0.081
	t-value	(0.90)	(1.17)	(1.07)	(0.90)	(0.71)	(-0.83)	(0.94)	(0.97)	(0.97)	(1.11)	(0.82)	(-0.64)
	Alpha	-0.109	0.070	0.046	0.013	-0.045	0.064	-0.076	-0.026	0.013	0.105	-0.038	0.038
	t-value	(-1.27)	(0.87)	(0.49)	(0.12)	(-0.35)	(0.40)	(-1.01)	(-0.31)	(0.13)	(0.99)	(-0.36)	(0.32)
2	Excess return	0.472	0.443	0.425	0.448	0.394	-0.079	0.435	0.515	0.382	0.423	0.424	-0.010
	t-value	(0.93)	(0.89)	(0.89)	(0.98)	(0.87)	(-0.62)	(0.89)	(1.04)	(0.81)	(0.90)	(0.90)	(-0.10)
	Alpha	-0.200	-0.203	-0.153	-0.075	-0.114	0.086	-0.190	-0.097	-0.187	-0.135	-0.150	0.040
	t-value	(-2.41)	(-2.79)	(-1.61)	(-1.02)	(-1.32)	(0.82)	(-2.46)	(-1.22)	(-2.38)	(-1.66)	(-1.74)	(0.38)
Big	Excess return	0.373	0.715	0.508	0.470	0.410	0.038	0.747	0.400	0.656	0.420	0.207	-0.540
	t-value	(0.88)	(1.25)	(1.07)	(0.97)	(0.96)	(0.28)	(1.61)	(0.77)	(1.20)	(0.96)	(0.48)	(-2.62)
	Alpha	-0.186	0.209	-0.077	-0.056	-0.170	0.016	0.184	-0.129	0.050	-0.128	-0.314	-0.499
	t-value	(-1.57)	(0.88)	(-0.44)	(-0.29)	(-1.20)	(0.13)	(1.22)	(-0.58)	(0.23)	(-0.71)	(-2.00)	(-2.48)

Table 5

The Effect of PIN on Cross-Sectional Expected Stock Returns

This table shows Fama-MacBeth (1973) cross-sectional regression results for the following model.

$$r_{it+1} = \gamma_0 + \gamma_1 PIN_{it} + \gamma_2 PSOS_{it} + \gamma_3 Illiquidity_{it} + \gamma_4 \beta_{i,Gt} + \gamma_5 \beta_{i,Ct} + \gamma_6 BM_{it} + \gamma_7 Size_{it} + \epsilon_{it+1}. \quad (12)$$

r_{it+1} is the monthly stock return of firm i in excess of a 30-day US Treasury bill rate at time $t+1$; PIN is the probability of information-based trading estimated using the EHO framework (PIN_{EHO}) or Duarte and Young's (2009) approach (PIN_{DY}); $PSOS$ is the probability of symmetric order-flow shocks constructed using Duarte and Young's method; $Illiquidity$ is the Amihud illiquidity measure; β_G (β_C) is the covariance of the stock return with the global (country) market index returns over the past five years divided by the global (country) market index return variance, where the country market index return is orthogonalized to the global market return; BM is the log book-to-market ratio of the firm; $Size$ is the log firm market capitalization; ϵ is a forecast error. All variables are defined in Appendix B. M1-M9 use PIN_{EHO} as PIN, while M10-M14 employ PIN_{DY} as PIN. Results are reported for firms from all countries (All) and from developed (DEV) vs. emerging (EMG) markets. The table presents time-series averages of the estimated slope coefficients from the above regression. Monthly excess returns are from January 1997 to December 2011 and the firm-specific variables are from 1996 to 2010. All $t-$statistics reported in parentheses are based on Newey-West standard errors.

	$PIN = PIN_{EHO}$									$PIN = PIN_{DY}$				
	All Countries							DEV	EMG	All Countries			DEV	EMG
Variable	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14
PIN		-0.405	-0.294		-0.839			-0.929	-0.600	0.110	0.123	-0.059	-0.123	0.182
		(-1.24)	(-0.84)		(-2.89)			(-3.02)	(-1.38)	(0.54)	(0.58)	(-0.34)	(-0.62)	(0.68)
$PSOS$			-0.395			-0.413	-0.438				-0.426			
			(-2.46)			(-2.72)	(-2.86)				(-2.83)			
$Illiquidity$				0.081	0.096		0.087	0.109	0.089			0.081	0.093	0.074
				(2.53)	(3.10)		(2.73)	(3.55)	(2.29)			(2.57)	(3.02)	(1.82)
β_G	0.002	-0.004	0.004	0.031	0.026	0.009	0.039	0.026	0.076	0.002	0.009	0.030	0.029	0.079
	(0.01)	(-0.02)	(0.02)	(0.18)	(0.15)	(0.05)	(0.23)	(0.14)	(0.54)	(0.01)	(0.05)	(0.18)	(0.16)	(0.56)
β_C	0.005	0.003	0.009	0.016	0.014	0.010	0.023	0.023	-0.090	0.005	0.010	0.016	0.025	-0.082
	(0.07)	(0.05)	(0.12)	(0.23)	(0.19)	(0.13)	(0.30)	(0.36)	(-0.44)	(0.07)	(0.13)	(0.22)	(0.39)	(-0.39)
BM	0.433	0.431	0.425	0.439	0.435	0.426	0.433	0.365	0.515	0.433	0.425	0.438	0.367	0.521
	(5.62)	(5.65)	(5.53)	(5.77)	(5.78)	(5.50)	(5.66)	(4.35)	(5.73)	(5.62)	(5.50)	(5.76)	(4.34)	(5.72)
$Size$	0.018	0.010	0.008	0.118	0.122	0.014	0.122	0.153	0.071	0.019	0.015	0.118	0.148	0.068
	(0.41)	(0.23)	(0.18)	(1.86)	(1.93)	(0.32)	(1.94)	(2.18)	(1.08)	(0.44)	(0.36)	(1.85)	(2.09)	(1.02)
Intercept	0.858	1.035	1.164	-0.163	0.012	1.040	-0.046	-0.414	0.246	0.821	1.003	-0.144	-0.565	0.059
	(1.26)	(1.39)	(1.63)	(-0.18)	(0.01)	(1.62)	(-0.05)	(-0.41)	(0.21)	(1.15)	(1.49)	(-0.16)	(-0.56)	(0.05)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 6
Robustness Tests Using PIN Estimates Based on Different Order-Aggregation Methods

This table repeats Fama-MacBeth regressions of M5 and M12 of Table 5 using PIN estimated from the number of buy or sell orders by (i) aggregating sequential trading at the same price if there is no update in quotes (PIN^1), (ii) aggregating sequential trading within 15 seconds if there is no update in quotes (PIN^2), and (iii) aggregating sequential trading if there is no update in quotes (PIN^3). It shows time-series averages of the slope coefficients from the following cross-sectional regression.

$$r_{it+1} = \gamma_0 + \gamma_1 PIN_{it}^{\#} + \gamma_2 Illiquidity_{it} + \gamma_3 \beta_{i,Gt} + \gamma_4 \beta_{i,Ct} + \gamma_5 BM_{it} + \gamma_6 Size_{it} + \epsilon_{it+1}. \quad (13)$$

r_{it+1} is the monthly stock return of firm i in excess of a 30-day US Treasury bill rate at time $t+1$; PIN is the probability of information-based trading estimated using the EHO framework (PIN_{EHO}) or Duarte and Young's (2009) approach (PIN_{DY}); *Illiquidity* is the Amihud illiquidity measure; β_G (β_C) is the covariance of the stock return with the global (country) market returns over the past five years divided by the global market (country) return variance, where country index return is orthogonalized to the global market return; BM is the log book-to-market ratio of the firm; $Size$ is the log firm market capitalization; ϵ is a forecast error. All variables are defined in Appendix B. Monthly excess returns are from January 1997 to December 2011 and firm-specific variables are from 1996 to 2010. All $t-$statistics reported in parentheses are based on Newey-West standard errors.

	PIN_{EHO}^1	PIN_{EHO}^2	PIN_{EHO}^3	PIN_{DY}^1	PIN_{DY}^2	PIN_{DY}^3
Variable	M1	M2	M3	M4	M5	M6
$PIN^{\#}$	-0.638	-1.062	-0.818	-0.265	-0.146	-0.493
	(-3.20)	(-3.90)	(-3.32)	(-1.58)	(-0.65)	(-2.81)
Illiquidity	0.090	0.093	0.090	0.081	0.088	0.090
	(2.79)	(2.84)	(2.74)	(2.53)	(2.68)	(2.82)
β_G	0.036	0.039	0.031	0.027	0.027	0.038
	(0.22)	(0.23)	(0.18)	(0.16)	(0.16)	(0.22)
β_C	0.012	0.006	0.028	0.027	0.027	0.042
	(0.17)	(0.08)	(0.39)	(0.40)	(0.38)	(0.52)
BM	0.438	0.443	0.443	0.435	0.452	0.440
	(5.79)	(5.87)	(5.90)	(5.63)	(5.91)	(5.69)
$Size$	0.130	0.129	0.132	0.109	0.132	0.139
	(2.02)	(1.98)	(2.05)	(1.69)	(2.04)	(2.16)
Intercept	-0.137	-0.040	-0.160	0.004	-0.292	-0.360
	(-0.15)	(-0.04)	(-0.18)	(0.00)	(-0.33)	(-0.40)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 7
Effects of PIN and Other Information-Based Trading Measures on Cross-Sectional Expected Stock Returns

This table repeats Fama-MacBeth regressions of M2 and M5 of Table 5 using different information-based trading measures, $InfAsy_{trade}$, as follows.

$$r_{it+1} = \gamma_0 + \gamma_1 InfAsy_{trade,it} + \gamma_2 Illiquidity_{it} + \gamma_3 \beta_{i,Wt} + \gamma_4 \beta_{i,Ct} + \gamma_5 BM_{it} + \gamma_6 Size_{it} + \text{Controls} + \epsilon_{it+1}. \quad (14)$$

r_{it+1} is the monthly stock excess return of firm i in excess of a 30-day US Treasury bill rate at time $t+1$. $InfAsy_{trade}$ includes a list of information-based trading measures such as the EHO PIN_{EHO}, Duarte and Young's (2009) PIN_{DY}, Hasbrouck's (1991) relative trade informativeness measure R^2_W, Huang and Stoll's (1996) percentage price impact measure, $\%PImpact$, Huang and Stoll's (1997) adverse selection component α_{HS}, and Madhavan, Richardson, and Roomans's (1997) adverse information parameter θ_{MRR}, as well as three different first principal components denoted by $PComp^1$, $PComp^2$, and $PComp^3$. $PComp^1$ ($PComp^2$) is the first principal component extracted from performing a principal component analysis on PIN_{EHO} (PIN_{DY}) and the other four information-based trading measures, namely R^2_W, $\%PImpact$, α_{HS}, and θ_{MRR}, and $PComp^3$ is extracted using all six measures altogether. *Illiquidity* is the Amihud illiquidity measure; β_G (β_C) is the covariance of the stock return with the global (country) market index return over the past five years divided by the global (country) market index return variance; BM is the log book-to-market ratio of the firm; $Size$ is the log firm market capitalization; ϵ is a forecast error. All variables are defined in Appendix B. All $t-$statistics reported in parentheses are based on Newey-West standard errors. Monthly excess returns are from January 1997 to December 2011.

	Definition of Information-Based Trading Measures, $InfAsy_{trade}$													
	R^2_W		$\%PImpact$		α_{HS}		θ_{MRR}		$PComp^1$		$PComp^2$		$PComp^3$	
Variable	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14
$InfAsy_{Trade}$	0.189	0.492	18.738	11.245	10.641	4.633	13.220	-2.102	0.043	-0.081	0.069	-0.027	0.037	-0.095
	(0.45)	(1.22)	(1.76)	(0.94)	(1.88)	(0.85)	(1.17)	(-0.20)	(0.64)	(-1.15)	(1.02)	(-0.38)	(0.59)	(-1.50)
$Illiquidity$		0.091		0.076		0.090		0.092		0.110		0.099		0.113
		(3.02)		(2.23)		(2.96)		(2.95)		(3.19)		(2.90)		(3.35)
β_G	0.007	0.035	-0.002	0.025	0.001	0.031	0.001	0.033	0.003	0.040	0.001	0.036	0.003	0.040
	(0.04)	(0.21)	(-0.01)	(0.15)	(0.01)	(0.17)	(0.00)	(0.19)	(0.01)	(0.22)	(0.00)	(0.20)	(0.02)	(0.22)
β_C	0.007	0.016	0.003	0.013	-0.004	0.006	0.006	0.015	0.003	0.013	0.002	0.011	0.004	0.013
	(0.09)	(0.22)	(0.04)	(0.19)	(-0.06)	(0.08)	(0.08)	(0.21)	(0.04)	(0.18)	(0.03)	(0.15)	(0.05)	(0.18)
BM	0.429	0.436	0.434	0.438	0.436	0.444	0.433	0.437	0.431	0.431	0.432	0.434	0.431	0.431
	(5.58)	(5.75)	(5.68)	(5.81)	(5.18)	(5.33)	(5.32)	(5.43)	(5.20)	(5.27)	(5.19)	(5.27)	(5.21)	(5.28)
$Size$	0.013	0.126	0.035	0.122	0.053	0.157	0.061	0.152	0.057	0.156	0.064	0.158	0.056	0.157
	(0.30)	(2.11)	(0.92)	(1.95)	(1.19)	(2.53)	(1.37)	(2.44)	(1.34)	(2.51)	(1.54)	(2.55)	(1.30)	(2.52)
Intercept	0.846	-0.386	0.588	-0.260	0.383	-0.640	0.240	-0.561	0.343	-0.585	0.256	-0.630	0.359	-0.583
	(1.34)	(-0.49)	(0.93)	(-0.30)	(0.53)	(-0.72)	(0.33)	(-0.64)	(0.49)	(-0.67)	(0.37)	(-0.73)	(0.51)	(-0.67)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Appendix A
Stock Exchange and Starting Years of Automated Trading and Transactions Data by Country

This table lists the exchange(s) whose stocks are included in this study and the starting year of its (their) electronic limit order book system, which is obtained from Jain (2005), and of its global transactions data from TRTH database by country.

		Starting Year	
Country	Stock Exchange(s)	Automated Trading	TRTH Data
Argentina	Buenos Aires Stock Exchange	1995	1998
Australia	Australian Stock Exchange	1987	1996
Austria	Vienna Stock Exchange	1996	1996
Belgium	Euronext Brussels	1996	1996
Brazil	Sao Paolo Stock Exchange	1990	1998
Canada	Toronto Stock Exchange	1977	1996
Chile	Santiago Stock Exchange	1989	2002
China	Shanghai and Shenzhen Stock Exchanges	1990	1996
Denmark	Copenhagen Stock Exchange	1988	1996
Egypt	Cairo Stock Exchange	1997	1996
Finland	Helsinki Stock Exchange	1988	1996
France	Euronext Paris	1986	1996
Germany	Frankfurt Stock Exchange	1991	1996
Greece	Athens Stock Exchange	1992	1996
Hong Kong	Hong Kong Stock Exchange	1986	1996
India	Mumbai Stock Exchange	1995	1996
Indonesia	Jakarta Stock Exchange	1995	1996
Ireland	Irish Stock Exchange	2000	2000
Israel	Tel Aviv Stock Exchange	1997	1996
Italy	Milan Stock Exchange	1994	1996
Japan	Tokyo Stock Exchange and Osaka Securities Exchange	1982	1996
Jordan	Amman Stock Exchange	2000	2000
Korea	Korea Stock Exchange	1988	1997
Luxembourg	Luxembourg Stock Exchange	1991	1999
Malaysia	Kuala Lumpur Stock Exchange	1992	1996
Mexico	Bolsa Mexicana de Volores	1996	1996
Netherlands	Euronext Amsterdam	1994	1996
New Zealand	New Zealand Stock Exchange	1991	1996
Norway	Oslo Stock Exchange	1988	1996
Pakistan	Karachi Stock Exchange	1997	2001
Peru	Lima Stock Exchange	1995	1998
Philippines	Philippine Stock Exchange	1993	1996
Poland	Warsaw Stock Exchange	1996	2001
Portugal	Euronext Lisbon	1991	1996
Russia	Russian Trading System	1994	1996
Saudi Arabia	Saudi Stock Exchange	1990	2002
Singapore	Singapore Stock Exchange	1989	1996
South Africa	Johannesburg Stock Exchange	1996	1996
Spain	SIBE-Mercado Continuo Espanol	1989	1996
Sri Lanka	Colombo Stock Exchange	1997	1998
Sweden	Stockholm Stock Exchange	1989	1996
Switzerland	Swiss Exchange	1996	1996
Taiwan	Taiwan Stock Exchange	1985	1996
Thailand	Thailand Stock Exchange	1991	1996
Turkey	Istanbul Stock Exchange	1993	1996
U.K.	London Stock Exchange	1997	1996
U.S.	AMEX and NYSE	2000 (NYSE)	1996

Appendix B

Variable Definition and Data Source

Variable	Definition	Data Source
PIN **Variables**		
PIN_{EHO}	Probability of informed trading constructed using Easley, Hvidkjaer, and OHara's (2002) approach	TRTH
PIN_{DY}	Probability of informed trading constructed using Duarte and Young's (2009) approach	TRTH
Liquidity Variables		
Illiquidity	Log of average of daily Amihud's (2002) measure calculated as the absolute value of stock return divided by dollar volume on a given day	Datastream
$ESpread_{EW}$	Equal-weighted average of daily percentage effective spreads, $\frac{2*\lvert P-M\rvert}{M}$, where $M = \frac{(Ask+Bid)}{2}$ and P is the transaction price	TRTH
$ESpread_{VW}$	Volume-weighted average of daily percentage effective spreads, $\frac{2*\lvert P-M\rvert}{M}$, where $M = \frac{(Ask+Bid)}{2}$ and P is the transaction price	TRTH
$QSpread_{EW}$	Equal-weighted average of daily percentage quoted spreads, $\frac{(Ask-Bid)}{M}$, where $M = \frac{(Ask+Bid)}{2}$ and P is the transaction price	TRTH
$QSpread_{VW}$	Volume-weighted average of daily percentage quoted spreads, $\frac{(Ask-Bid)}{M}$, where $M = \frac{(Ask+Bid)}{2}$ and P is the transaction price	TRTH
$Turnover$	Log of average of daily total number of shares traded scaled by the number of shares outstanding	Datastream
Asset Pricing Test Variables		
$PSOS$	Probability of symmetric order-flow shocks estimated using Duarte and Young's (2009) approach	TRTH
β_G	Covariance of the stock return with the global market index return over past five years divided by the global market index return variance	Datastream
β_C	Covariance of the stock return with the country index return over past five years divided by country index return variance, where country index return is orthogonalized to the global market index return	Datastream
BM	Log of book-to-market equity ratio	Worldscope
$Size$	Log of market capitalization denominated in US$	Datastream

Appendix B - Continued

Variable Definition and Data Source

Variable	Definition	Data Source
Firm-Level Characteristics		
$TAssets$	Log of total assets denominated in US$	Worldscope
$Leverage$	Ratio of total debt to total assets	Worldscope
ROA	Operating income divided by total assets	Worldscope
ADR	Dummy variable equals one if the firm is cross-listed on a US stock exchange	Multiple sources
$R\&D$	Research and development expenses scaled by total assets	Worldscope
σ_{Ret}	Annualized standard deviation of monthly stock returns	Datastream
Firm-Level Information Proxies		
$Analysts$	Number of financial analysts covering a firm	IBES
$FDisp$	Log of standard deviation of analyst forecasts scaled by stock price	IBES
$Press$	Log of one plus number of press releases on a firm in a given year	RavenPack
Age	Number of years from the listed date to current date	Datastream
$MSCI$	MSCI member dummy, which equals one if the firm is included in an MSCI country index	Datastream
$CHeld$	Fraction of shares closely held by insiders and controlling shareholders	Worldscope
Country-Level Characteristics		
$GDPC$	Log of GDP per capita measured in US$	World Development Indicators
$MCap$	Stock market capitalization deflated by GDP	World Development Indicators
$Credit$	Private credit deflated by GDP. Private credit refers to financial resources available to the private sector, through loans, purchases of non-equity securities, and trade credits and other accounts receivable.	World Development Indicators
GDP_g	Annual GDP growth	World Development Indicators
σ_{GDP}	Standard deviation of annual GDP growth over the last 5 years	World Development Indicators
SEG	Bekaert et al.'s (2011) measure of stock market segmentation	Datastream
Law & $Order$	Law and order index which measures the strength and impartiality of the legal system and popular observance of the law.	International Country Risk Guide

Appendix B - Continued

Variable Definition and Data Source

Variable	Definition	Data Source
Country-Level Information Proxies		
$DReq$	Average score of six disclosure sub-indexes: prospectus delivering, insider compensations, large shareholder ownership, insider ownership, contracts outside the normal course of business, and related parties transactions. All these sub-indexes are dummy variables, and for each sub-index, the value of 1 is assigned to the index if it signifies high quality disclosure and 0 if otherwise	La Porta et al. (2006)
$AcStd$	Accounting standard index that examines and rates companies' 1990 annual reports on 90 items for 36 countries, covering general information, income statements, balance sheets, fund flow statements, accounting standards, stock data, and other special items.	La Porta et al. (1998)
CMG	A composite index that captures the degree of earnings opacity, the enforcement of insider laws, and the effect of removing short-selling restrictions	Bhattacharya and Daouk (2002); Charoenrook and Daouk (2005); Worldscope
$FTran$	Measure of the intensity and timeliness of financial disclosure by firms and interpretation and dissemination of a firm's news by financial analysts and media	Bushman, Piotroski, and Smith (2004)
$Newspapers$	Daily newspapers refer to those published at least four times a week and calculated as the average circulation (or copies printed) per 1,000 people	World Development Indicators
Other Information-Based Trading Measures		
α_{HS}	Huang and Stoll's (1997) measure of adverse selection component	TRTH
θ_{MRR}	Madhavan, Richardson, and Roomans's (1997) measure of adverse information parameter	TRTH
$\%PImpact$	Huang and Stoll's (1996) percentage price impact measure	TRTH
R^2_W	Hasbrouck's (1991) measure of relative trade informativeness, proportion of efficient price variation driven by trades	TRTH