PIN, Adjusted *PIN*, and *PSOS*: Difference of Opinion in the Korean Stock Market†

Kyong Shik Eom‡, Jangkoo Kang§ and Kyung Yoon Kwon*

Abstract

Duarte and Young (2009) decompose *PIN* into adjusted *PIN* (*AdjPIN*) and probability of trading caused by symmetric order flow shocks (*PSOS*). We explore sources of *PSOS* in the Korean stock market and examine the relation between *PSOS* and stock returns. Using transaction data with trader types and initiator information, we find that *AdjPIN* is not priced, while *PSOS* is negatively priced, a finding that Lai, Ng, and Zhang (2014) labeled "puzzling." We find that the negative price of *PSOS* comes from differences of opinion among domestic individual investors on the significance of public news.

Keywords: *PIN*, information risk, asset pricing, difference of opinion, individual investor

JEL classification: G12, G14

---

† The authors are very grateful for helpful discussions with Robert Anderson, Jong-Ho Park, and many seminar participants at Korea Advanced Institute of Science and Technology (KAIST) and the University of Seoul. All errors are the responsibility of the authors.

‡ Center for Risk Management Research, University of California at Berkeley; 530 Evans Hall # 3880, Berkeley, CA 94720-3880, U.S.A.; tel. : +82-70-8761-0562; Email: kseom2@gmail.com

§ College of Business, Korea Advanced Institute of Science and Technology; 85 Hoegiro, Dongdaemoon-gu, Seoul, 130-722, South Korea; tel.: +82-2-958-3521; Email: jkkang@business.kaist.ac.kr

* Corresponding author, College of Business, Korea Advanced Institute of Science and Technology; 85 Hoegiro, Dongdaemoon-gu, Seoul, 130-722, South Korea; tel.: +82-2-958-3693; Email: noldya@business.kaist.ac.kr

1. Introduction

Easley and O'Hara (1992) and their subsequent papers[1] propose a measurement of the probability of information-based trading (*PIN*) and empirically validate it. Duarte and Young (2009) (hereafter DY), however, document that the original *PIN* model fails to explain two well-known stylized facts in the US stock markets: the positive correlation between the numbers of buyer-initiated and seller-initiated trades, and the high volatilities of buyer-initiated and seller-initiated trades. DY suggest an adjusted *PIN* model that replaces the original *PIN* measure with two measures, adjusted probability of informed trading (*AdjPIN*) and probability of trading caused by symmetric order flow shocks (*PSOS*) arising from public information.

[Insert Table 1]

As for the Korean stock market, the existing empirical studies also raise a question of whether *PIN* is an appropriate proxy for the information risk, or whether *PIN* is a significant determinant of expected returns (see Table 1). Specifically, Eom, Kang, and Kwon (2016) report that the problems of the original *PIN* model using US data, as reported in DY, also exist in the Korean stock market, and among five variants of DY's adjusted *PIN* model, the model without any restriction fits the Korean market data best.

In this paper, adopting DY's adjusted *PIN* model, we examine the pricing effects of *PSOS*. We mainly focus on explaining the puzzling "negative" and significant relation between *PSOS* and expected stock returns, which has been reported in the global markets except the US, by investigating the sources of the symmetric order flow shocks in the Korean stock market. This requires performing cross-sectional asset pricing tests with *PIN*, *AdjPIN*, and *PSOS* to examine whether they are priced or not.

DY report that *PSOS* is positively priced in the US markets and that it seems to be a proxy for illiquidity.[2] They document that there can be more than one possible cause for symmetric order flow shocks, but they do not attempt to identify the main source; the goal is to show that their inclusion of

[1] A short list of papers by Easley and O'Hara and their coauthors regarding *PIN* includes Easley, Kiefer, O'Hara and Paperman (1996), Easley, Hvidkjaer, and O'Hara (2002, 2010), and Easley and O'Hara (2004).

[2] This role and interpretation of *PSOS* in the US stock market seems uncomfortable for DY. In a recent working paper, Duarte, Hu, and Young (2015) focus on a possible bias in the theoretical underpinning of the derivation of *PIN*. The debate on the validity of *PIN* has been inconclusive and is ongoing. See Section 4.1 for recent developments in the research on *PIN*.

systematic order flow shock in the original *PIN* model improves the ability to explain the data. Lai, Ng, and Zhang (2014) (hereafter LNZ) examine the pricing effect of informational risk based on the original and adjusted *PIN* models, using international data from 47 countries. They find that *PSOS* is negatively priced, but describe this result as "*puzzling*." Therefore, the role of *PSOS* is important in extension of the *PIN* model, but the literature on the relation between *PSOS* and expected stock returns reports mixed results and fails to explain these results.

The Korean stock market has distinctive features compared to the US stock market. First, the Korean stock market does not have designated market makers as does the US stock market. In the Korean market, buyers and sellers, who can submit both limit and market orders, meet via the Automated Trading System. Second, in the Korean stock market, most of the trading is done by individual investors. According to Choe, Kho, and Stulz (2005), in 1998, 77.43% of the gross value of stock sales was by domestic individual investors, and trading by the government and corporations represents a small fraction of the overall trading. They also note that this feature contrasts with the results of Tesar and Werner (1995) in the sense that in more developed countries, foreign investors are more active traders than local investors. We expect that these differences between the Korean and US stock markets may generate a gap between the performances of the sequential trading model, such as the original and adjusted *PIN* models, in the two markets. More importantly, the dominant shock generating the symmetric order flow can be different, and so the pricing effect of the symmetric order flow component estimated from the model can be different.

To verify the driver of the significant effects of *PSOS*, we pay attention to the characteristic of *PSOS* embedded in its definition, which is the trading caused by symmetric order flow shocks arising from difference of opinion on public news, and to the group of individual traders, who may behave as described in *PSOS*: one subgroup buys and another subgroup simultaneously sells based on the same public news. Our analytical procedures and results are as follows. First, following Diether, Malloy, and Scherbina (2002), we adopt the turnover variable (*TURNOVER*) as a proxy for difference of opinion, which is a potential cause of symmetric order flow shocks,[3] and examine the relation between *PSOS* and difference of opinion.[4] We find that *PSOS* is highly correlated with *TURNOVER*, and higher *PSOS* firms show higher increase of difference of opinion (*TURNOVER*) when there is public news that may lead to differing conclusions among investors. These results indicate that difference of opinion is the cause of the symmetric order flow shocks in the Korean market.

Next, to examine how difference of opinion affects the cross-sectional results of *PSOS*, we perform cross-sectional regressions with *TURNOVER*. We subtract the component related to *TURNOVER* from *PSOS*, then examine whether the residual remains significant. We find that the residual is not significant, which means that the component related to *TURNOVER* drives the negative and significant coefficients of *PSOS*.

Finally, we decompose *TURNOVER* into three components—the trading activity of domestic institutional (*INTA*), domestic individual (*RNTA*), and foreign investors (*FNTA*)—and investigate whether the trading activity of individual investors is the main contributor of the explanatory power of *PSOS*. Among these three variables, we focus on the trading activity of individual investors, because there is a belief that individual investors are uninformed noise traders (Kumar and Lee, 2006; Baber and Odean, 2008; and Baber, Odean, and Zhu, 2009), who may draw different conclusions from public news. We find that individual investors show an increase in the numbers of both buys and sells and *TURNOVER* in response to public news, and this increase is larger for higher *PSOS* firms. These results show that individual investors generate symmetric order flow shocks because of difference of opinion. We subtract the component related to *INTA* from *PSOS* and construct an orthogonalized *PSOS* variable

[3] As DY document, there can be more than one cause for symmetric order flow shocks, and the main cause may affect the cross-sectional performance of *PSOS*.

[4] Turnover is used as a proxy for difference of opinion and is documented to be negatively related to expected returns (Harris and Raviv, 1993; Vega, 2006). Hwang, Lee, Lim, and Park (2013) also find the positive relation between *PSOS* and turnover, but they describe this result as “somewhat unexpected” and explain that this positive relation is because turnover is a noisy proxy of liquidity (footnote 15 on page 156). We interpret turnover as a proxy of difference of opinion following Diether et al. (2002).

from the residuals; we repeat this process using *RNTA* instead of *INTA*.[5] We find that only *RNTA* contributes to the cross-sectional explanatory power of *PSOS*.

We use transaction data with information on the trade initiator and the trader type (domestic individual, domestic institution, or foreigner) from 2001 to 2006 in the Korean stock market. In the related literature, the failure of the Lee and Ready (1991) algorithm to specify the trade initiator is shown to be problematic (Elis, Michaely, and O'Hara, 2000; Asquith, Oman, and Safaya, 2010; Hwang et al., 2013). Hwang et al. (2013) report that the asset pricing results of *PIN*, *AdjPIN*, and *PSOS* are sensitive to this failure of the Lee and Ready algorithm. Our data contain precise information on the trade initiator as in Hwang et al. (2013), thus generating clean empirical asset pricing results.

We show that *PSOS* is negatively priced when the main source of the symmetric order flow shocks is difference of opinion. Our results suggest that the proportion of the trader group generating the symmetric order flow shock in a market can explain the differing performance of *PSOS* across markets. Further research in other markets will be necessary to clarify the effect of the main source of symmetric order flow shocks on the explanatory power of *PSOS* on expected stock returns. Moreover, this paper contributes to the literature on the role of individual investors in asset pricing. Individual investors are generally regarded as noise or uninformed traders, so the question of how they affect stock prices has been actively investigated. Kumar and Lee (2006) report that systematic noise trading of individual investors can affect stock prices, using a measure of individual investor order imbalance. They insist that systematic noise trading causes limits to arbitrage and makes stocks overpriced, and consequently reduces future returns. According to Miller (1977), prices reflect more optimistic valuations if pessimistic investors are constrained by short-sale costs, and this overvaluation will lower future returns. We present evidence that individual investors generate symmetric order flow shocks because of differing opinions on the meaning of public news, and their trading activity leads to lower future returns. The symmetric order flow shocks caused by difference of opinion on news can be interpreted as noise

[5] We find that *INTA* and *RNTA* are negatively related to expected stock returns, while *FNTA* is not. We regard *INTA* and *RNTA* as the components of *TURNOVER* that may contribute to the negative coefficient of *PSOS*, thus we examine the effects of *INTA* and *RNTA* but not *FNTA* on the explanatory power of *PSOS*.

trading, resulting in limits to arbitrage, so our results can be interpreted as consistent with Kumar and Lee (2006).

The remainder of this paper proceeds as follows. Section 2 describes the data and methodology of variable construction and overall analysis. Section 3 briefly introduces the original and adjusted *PIN* models. Section 4 presents the results of the model estimation and asset pricing tests. Section 5 describes our conclusions.

2. Data and Methodology

Our sample includes all stocks in the Korea Composite Stock Price Index (KOSPI). To estimate the structural models on market microstructure, our sample period is restricted to 2001–2006, due to the availability of relevant intraday transaction data with trader type. Our data contain a time-ordered record of every stock transaction on the Korea Exchange (KRX) and information regarding the types and order numbers of buyers and sellers, which enables us to clearly identify the initiator of each trade without employing the Lee and Ready (1991) algorithm, which is reported to have misclassification problems (Odders-White, 2000; Ellis, Michaely, and O'Hara, 2000; Asquith et al., 2010; Chakrabarty, Moulton, and Shkilko, 2012). We define the initiator of a trade as the trader who placed his/her order later than the other among the buyer and seller, following the chronological definition of Odders-White (2000).[6] Using the order submission number of the buyer and seller of each transaction, we can figure out who placed the order later.

[Insert Table 2]

---

[6] Odders-White (2000) describes the initiator of a transaction as the person who caused the transaction to occur, and thus defines an initiator is a trader who placed his or her order last, chronologically. Our data provide the order submission numbers of the buyer and seller of each transaction, so we can figure out who placed the order later. However, most of data sets used in many empirical studies do not provide the order submission numbers of the buyer and seller, so using this method is not available. Odders-White (2000) evaluate the accuracy of the Lee and Ready (1991) algorithm by comparing the initiator determined by the algorithm and the true initiator determined by the order submitted time. The order submission number in our data can be a substitute of the order submitted time in the data of Odders-White to determine the true initiator who placed the order last, thus we document in our paper that we can clearly identify the initiator of each trade without employing the Lee and Ready (1991) algorithm. In the literature, many studies find the true initiator of a transaction in the same way and examine the errors from using the algorithm (Aitken and Frino, 1996; Finucane, 2000; Lee and Radhakrishna, 2000; Hwang et al., 2013).

Table 2 shows a part of our data on January 12th, 2004. Each row provides information about one transaction, including a transaction price, number of shares traded, and the time the transaction completed. The 8th and 9th columns present the order submission number of the buyer and seller of the given transaction, respectively. We determine the initiator of each transaction using this information. For example, the first transaction for the firm code 7160 shows that the buyer's order number is 184 and the seller's order number is 151. It indicates that the seller submitted the order first, and then the buyer completed the transaction by submitting the order after the seller. Thus, we assign the buyer as the initiator of this transaction. In case of the second and third transactions, we can see that the order submission number of sellers in these two transactions are the same as 276. It indicates that his/her sell order is completed with two different buyers, one with the order number 185 and the other one with the order number 260. In both transactions, the seller initiates them.

Hwang et al. (2013) document that the misclassification problem of the Lee and Ready algorithm brings about biases in the estimates of *PIN* such that the empirical tests of *PIN* can be distorted seriously. Our data avoid this problem and provide estimates of *PIN*-related measures free from the misclassification problem.

Our data also provide information about whether a buyer or a seller is an individual, institutional, or foreign investor. Using this trader type information, we compare each group's trading pattern with the structural order flow of the models in Section 4.3. Since information about trader type is not available in general, many studies proxy trader type by trade size (Hvidkjaer, 2008; Barber et al., 2009), but Lee and Radhakrishna (2000) document that there are errors using trade size as a proxy for trader type in the US market. Since our data set provides accurate information, our research is free from the trader type identification issue.

Our analyses also require us to use monthly and daily financial data; these are provided by DataGuide and the Korean Research Data Service (KRDS). To construct market beta, turnover, Amihud (2002) illiquidity measure, and other control variables, we use monthly and daily data, including return and volume provided by DataGuide. The turnover measure (*TURNOVER*) is the annual average of daily turnover. For the Amihud illiquidity measure (*ILLIQ*), we compute the annual average of daily price impact of trading volume. We estimate market beta as follows: first, for each firm–month, we estimate market loadings using 60 months of past data. We then form 10 portfolios based on these pre-ranking factor loadings. Using the returns from these portfolios, we estimate the full period beta for each portfolio and assign this beta to each firm in the portfolio. For firm size and book-to-market ratio, we collect the market value of firms in December and the book value of firms in June for each year from KRDS. In each year, we exclude firms that do not have market capitalization data in December of the previous year.

3. The *PIN* Model and Its Extension

Easley and O'Hara's (1992) sequential trade model provides a measure of the probability of informed trades captured by order imbalance. Their model assumes that informed traders buy (sell) stocks when they receive positive (negative) private information and they do not trade if there is no information. Noise traders do not have any private information and they buy and sell regardless of the existence of private information in the market. Their model has been extended by many studies, and in this paper, we employ the Easley et al. (1996) model as the original *PIN* model. The original *PIN* is defined as follows:

$$PIN = \frac{a \times u}{a \times u + \varepsilon_b + \varepsilon_s} \tag{1}$$

where a is the probability that a private information event occurs at the beginning of a day, u is the daily arrival rate of orders from informed traders, and ε_b and ε_s are daily arrival rates of buy and sell orders from noise traders, respectively.

In the US stock market, DY document that the original *PIN* model does not fit the data well, and thus they suggest the adjusted *PIN* model as a solution to the problems of the original model. For the adjusted *PIN* model, DY extend the original model by adding the symmetric order flow shocks that cause buy and sell trades simultaneously. Put concretely, DY replace the original *PIN* by following two measures:

$$AdjPIN = \frac{a \times (d \times u_b + (1-d) \times u_s)}{a \times (d \times u_b + (1-d) \times u_s) + (\Delta_b + \Delta_s) \times (a \times \theta' + (1-a) \times \theta) + \varepsilon_b + \varepsilon_s} \tag{2}$$

$$PSOS = \frac{(\Delta_b + \Delta_s) \times (a \times \theta' + (1-a) \times \theta)}{a \times (d \times u_b + (1-d) \times u_s) + (\Delta_b + \Delta_s) \times (a \times \theta' + (1-a) \times \theta) + \varepsilon_b + \varepsilon_s} \quad (3)$$

where a is the probability that a private information event occurs at the beginning of a day; u_b and u_s are the daily arrival rate of buy and sell orders from informed traders, respectively; ε_b and ε_s are daily arrival rates of buy and sell orders from noise traders, respectively; and θ and θ' are the probability that a symmetric order flow shock occurs conditional on the arrival of private information and absence of private information, respectively. In the event of a symmetric order flow shock, the daily arrival rates of buys and sells are Δ_b and Δ_s, respectively.

Eom et al. (2016) show that the original *PIN* model exhibits exactly the same problems in the Korean stock market that DY pointed in the US market, and suggest that, among the five variants of DY adjusted *PIN* model, the unrestricted version fits best in the Korean market.[7] We employ this unrestricted version in this paper. We estimate the adjusted *PIN* model by the maximum likelihood method following DY and Eom et al. (2016).[8] The aggregated daily number of buys and sells are used for estimation, and in each firm–year, we maximize the likelihood with 50 different, randomly chosen, starting points. Then, the maximum of these 50 maximization results is chosen as our final results.

[Insert Table 3]

Panel A of Table 3 shows the percentile of summary statistics on the buyer- and seller- initiated trades and Panel B shows the estimation results of the adjusted *PIN* model.[9] Our data include 3,774 firm-year observations, indicating an annual average of 629 firms.[10] Consistent with DY's, the correlation

---

[7] DY employ Model 4 which has a restriction that the probabilities of the symmetric order flow shock conditional on the arrival of private information and absence of private information are the same ($\theta = \theta'$). In this paper, we employ Model 5, the unrestricted model allowing $\theta \neq \theta'$.

[8] The likelihood function of the extended model is

$$L(\theta|B,S) = (1-a)(1-\theta)e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!} + (1-a)\theta e^{-(\varepsilon_b+\Delta_b)}\frac{(\varepsilon_b+\Delta_b)^B}{B!}e^{-(\varepsilon_s+\Delta_s)}\frac{(\varepsilon_s+\Delta_s)^S}{S!} +$$
$$a(1-\theta')(1-d)e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}e^{-(\varepsilon_s+u_s)}\frac{(\varepsilon_s+u_s)^S}{S!} + a\theta'(1-d)e^{-(\varepsilon_b+\Delta_b)}\frac{(\varepsilon_b+\Delta_b)^B}{B!}e^{-(\varepsilon_s+u_s+\Delta_s)}\frac{(\varepsilon_s+u_s+\Delta_s)^S}{S!} +$$
$$a(1-\theta')de^{-(\varepsilon_b+u_b)}\frac{(\varepsilon_b+u_b)^B}{B!}e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!} + a\theta' de^{-(\varepsilon_b+\Delta_b+u_b)}\frac{(\varepsilon_b+\Delta_b+u_b)^B}{B!}e^{-(\varepsilon_s+\Delta_s)}\frac{(\varepsilon_s+\Delta_s)^S}{S!}$$

where B and S are the number of buys and sells for a given day and $\theta = (a, u_b, u_s, \varepsilon_b, \varepsilon_s, d, \theta, \theta', \Delta_b, \Delta_s)$ is the parameter vector. The original *PIN* model can be regarded as a restricted model with restrictions $u_b = u_s$ and $\theta = \theta' = \Delta_b = \Delta_s = 0$.

[9] Panel A (Panel B) of Table 3 in this paper and Panel A of Table 1 (Panel B of Table 3) in Eom et al. (2016) are exactly the same, because both papers use the same market data with the same sample periods.

[10] Since our data include 3,774 firm-year observations, each parameter has a sample of 3,774 estimates. We report the 5th, 25th, 50th (median), 75th, and 95th percentile of 3,774 observations for each parameter in Panel B of Table

between buys and sells is mostly positive, which contradicts the non-positive restriction of the original *PIN* model. Though we do not report the moments of buys and sells implied by the original *PIN* model, the variances of buys and sells computed from the data are also much larger than the implied moments. The median of the implied variance of buys (sells) is 19.88 (29.34) while the median of the variance of buys (sells) is 36.84 (64.35). We observe that the values of parameters determining the arrival rates, such as $u_b, u_s, \varepsilon_b, \varepsilon_s, \Delta_b,$ and Δ_s are much smaller than those in the US markets reported by DY. It is possible that the US market has the higher levels of buy and sell trades than the Korean market.

In cross-sectional regressions, we include market beta (*Beta*), the log of market value of firm equity at the end of the previous year (*log(ME)*), the log of book value divided by market value for the previous year (*log(BM)*)[11], and the Amihud (2002) illiquidity measure (*ILLIQ*), which is the annual average of daily price impact. We report the summary statistics for these variables in Panel C of Table 3. Our sample data are not concentrated in the group with a specific characteristic. The differences in the minimum, maximum, and standard deviation of each variable suggest that there is sufficient cross-correlation variation in our sample. For example, the minimum value of *log(ME)* is 1.80 and the maximum value of it is 18.39. Its mean and standard deviation are 11.07 and 1.62, respectively, suggesting that our data are not concentrated in small firms.

4. Results

4.1 Asset pricing tests of *PIN*, *AdjPIN*, and *PSOS*

In this section, we perform cross-sectional analysis with *PIN*, *AdjPIN*, and *PSOS*. We focus on the pricing effect of *AdjPIN* and *PSOS*, but we also test the pricing effect of *PIN* for comparison.

[Insert Table 4]

Table 4 shows that *PIN* and *AdjPIN* are not significantly priced. Our data provide error-free estimates of the original and adjusted *PIN* models, so we expect that these insignificant results are not derived by

3.

[11] The book-to-market ratio (*log(BM)*) values greater than the 0.95 fractile or less than the 0.05 fractile are set to equal the 0.95 and 0.05 fractile values, respectively.

the error of *PIN* and *AdjPIN*. The insignificant results of *PIN* and *AdjPIN* are consistent with DY, LNZ, and other literature. The relevant literature reports that information risk is diversifiable and can be regarded as an idiosyncratic risk, so it is not priced (Hughes, Liu, and Liu, 2007; Lambert, Leuz, and Verrecchia, 2007). Hughes et al. (2007) suggest that information risk is either diversifiable or embodied in existing risk factors, and Lambert et al. (2007) also report that information risk is diversifiable. On the other hand, the existing literature suggests concerns on the validity of the *PIN* measure. Aktas, de Bodt, Declerck, and Oppens (2007) examine the behavior of *PIN* around merger and acquisition announcements that took place in the Euronext Paris, and report evidence of contradictory behaviors of *PIN*. These results raise the question of whether *PIN* is a valid measure capturing the information risk.[12] Akay, Cyree, Griffiths, and Winters (2012) examine what *PIN* measures in the T-bill market and find that it is possibly related to liquidity-based trading instead of information-based trading. Hence, our insignificant results of *PIN* and *AdjPIN* are in the spirit of the growing debate on the validity of *PIN*.

Most importantly, *PSOS* has significant and negative coefficients. *ILLIQ* has a positive and significant coefficient in Model 6, indicating that there exists an illiquidity premium but it becomes insignificant if *PSOS* is included. The coefficient of *PSOS*, however, remains significant even though *ILLIQ* is included. This indicates that *PSOS* is not simply an illiquidity measure, as DY insist.

Our results that *PIN* and *AdjPIN* are not significant, and that *PSOS* is negatively priced and not affected by the illiquidity measure, are not surprising. LNZ report consistent findings in 47 countries, but they leave these results unexplained. In Sections 4.2 and 4.3, we focus on the meaning of *PSOS* in the Korean market and verify why *PSOS* is negatively priced.

4.2 *PSOS* and difference of opinion

[12] Much research from three distinct directions has raised questions on the credibility of the *PIN* measure. The first is related to the fact that *PIN* contradicts the empirical stylized facts, especially in event studies (e.g., see Duarte and Young, 2009; Petchey, Wee, and Yang, 2016). The second is related to problems such as numerical overflow and so-called time-horizon issues which arise in the estimation of *PIN* (e.g., see Easley, Engle, O'Hara, and Wu, 2008; Tay, Ting, Tse, and Warachka, 2009; Lin and Ke, 2011). The third is related to the bias in the theoretical underpinnings of *PIN* (e.g., see Collin-Dufresne and Fos, 2015; Duarte et al., 2015; Back, Crotty, and Li, 2016).

We employ the turnover measure as a proxy for difference of opinion following Diether et al. (2002), and examine whether *PSOS* is significantly associated with the turnover variable. *PSOS* is defined as the ratio of the expected number of trades caused by the symmetric order flow shock to the total expected number of trades. DY note that there are at least two possible explanations for symmetric order flow shocks. One possible cause is that traders coordinate on trading on a certain day to reduce trading costs as Admati and Pfleiderer (1988) suggest. In addition, DY suggest the occurrence of a public news event whose implications traders disagree as another possible cause. In other words, when heterogeneity among investors increases, symmetric order flow can occur. These two possible explanations provide implications for the relation between symmetric order flow and different aspects of liquidity. The first explanation suggests that symmetric order flow occurs with one aspect of liquidity in terms of the smaller price impact. The other explanation, however, suggests that symmetric order flow occurs with liquidity improvement, which is the larger trading volume but not directly related to the smaller price impact. Thus, we employ stock turnover (*TURNOVER*) to capture the disagreement effect and the Amihud illiquidity measure (*ILLIQ*) to account for the price impact, and then explore the two above explanations by examining the relations between *PSOS* and the two measures.

To explore this issue, we first compute correlations among *PIN, AdjPIN*, *PSOS*, *ILLIQ*, *TURNOVER*, and other related variables.

[Insert Table 5]

In Table 5, contrary to DY's results, *PSOS* has a negative correlation with *ILLIQ*, which indicates that it is not appropriate to interpret *PSOS* as an illiquidity measure in the Korean market, but *PSOS* has a positive and more significant correlation with *TURNOVER*. This suggests that *PSOS* is a proxy for difference of opinion, as expected.

To explore the relation between *PSOS* and difference of opinion, we perform two analyses. First, we sort firms into quintiles based on their *PSOS*, and then investigate changes in the number of buys and sells and *TURNOVER* when public news is announced. We regard the news announcement as the event causing disagreement among investors. Second, we perform cross-sectional analysis with *TURNOVER*. We verify whether *TURNOVER* affects the predictive power of *PSOS*.

In each year t, we sort firms into quintiles based on their *PSOS* estimated in the previous year. For each quintile portfolio, we calculate the average of annual *PIN*-related measures, *PIN*, *AdjPIN*, and *PSOS*, and then average number of buys (*BUYS*) and sells (*SELLS*), average net order imbalance (*IMB*), average value of *ILLIQ*, and average value of *TURNOVER* on event and non-event days, respectively. If the symmetric order flow is caused by difference of opinion, then the high *PSOS* portfolio will show an increase in the number of buys and sells when public news is announced. More directly, the high *PSOS* portfolio will also show increased difference of opinion, which is measured by *TURNOVER*, in response to the news. As public news that may cause disagreement among investors, we use fair disclosure announced on the Korea Exchange.[13] Since the fair disclosure rule was implemented in November 2002, for this analysis, we restrict the sample period to 2003–2006.

[Insert Table 6]

Table 6 presents the trading activity of each *PSOS* quintile on event and non-event days, respectively. As *PSOS* increases, *PIN* does not show a big difference across portfolios, while *AdjPIN* deacreases. These patterns are consistent with the results in Table 5 that *PIN* is weakly correlated with *PSOS* and *AdjPIN* is negatively correlated with *PSOS*. Except for the lowest *PSOS* quintile, all portfolios show an increase in the number of buys and sells on event days. The pattern of *IMB* shows that, as *PSOS* increases, the relative increase of *SELLS* is larger than that of *BUYS*. In Table 3, the estimated values of Δ_s are higher than those of Δ_b, thus the larger increase of *SELLS* in high *PSOS* firms seems to be consistent with the model. The increase of *TURNOVER* in the highest *PSOS* quintile is three times that in the lowest quintile. *TURNOVER* of the lowest *PSOS* quintile changes from 0.012 to 0.014, while that of the highest quintile changes from 0.038 to 0.044. This indicates that the increase in difference of opinion in response to public news is larger in higher *PSOS* firms. Overall results in Table 6 support our hypothesis.

Next, we perform the cross-sectional analysis with *TURNOVER*, and investigate whether the significant coefficient of *PSOS* is attributed to difference of opinion. In the literature, the role of

[13] All fair disclosures can be collected from the Korea Exchange website (httep://kind.krx.co.kr).

heterogeneous beliefs among investors in predicting the cross-section of stock returns has been an important issue. Miller (1977) introduces a theoretical model in which prices reflect more optimistic valuation if pessimistic investors are constrained in trading due to short-sale costs, and this overvaluation produces negative future returns. The higher disagreement among investors produces lower future returns according to Miller's model. Diether et al. (2002) and Boehme, Danielsen, and Sorescu (2006) find empirical evidence of overvaluation of high dispersion stocks under short-sale constraints. If *PSOS* is a proxy of difference of opinion as in our hypothesis, then the negative coefficients will be attributed to the component related to difference of opinion.

[Insert Table 7]

In Table 7, the negative coefficients of *TURNOVER* are significant even after controlling for *PSOS* and *ILLIQ*. By contrast, *PSOS* loses its predictive power if *TURNOVER* is included. These features provide further evidence that the difference in opinion among investors is behind the negative relation between *PSOS* and expected stock returns.

To examine more precisely the effect of *TURNOVER* and to compare it with that of *ILLIQ*, we construct two variables, *AdjPSOS*1 and *AdjPSOS*2. *AdjPSOS*1 (*AdjPSOS*2) is an adjusted *PSOS* measure that is orthogonal to *ILLIQ* (*TURNOVER*). In each month, we regress *PSOS* on *ILLIQ* (*TURNOVER*) and then subtract the component related to *ILLIQ* (*TURNOVER*) from *PSOS*. The sum of the intercept and residual is defined as *AdjPSOS*1 (*AdjPSOS*2). We expect the coefficients of *AdjPSOS1 and AdjPOS2* to become insignificant if the eliminated part plays an essential role to be priced.

Model 4 examines whether *PSOS*, after the illiquidity effect is controlled, has an explanatory power for expected stock returns. The coefficient of *AdjPOS1* in Model 4 is negative and statistically significant at the 1% significance level, which shows that the significant effect of *PSOS* is not driven by the illiquidity component. On the other hand, in Models 5 and 6, examining whether the *PSOS* effect survives after controlling *TURNOVER*, the coefficients of *AdjPSOS2* are not statistically significant even at the 10% significance level. Overall, these results show that *PSOS* is negatively priced due to the component related to the difference of opinion *(TURNOVER)*, not due to the illiquidity component.

As DY document, there can be more than one cause of symmetric order flow shocks. The dominant cause in the market can differ across markets and thus derive different meanings of *PSOS*. In the literature, however, the cause of symmetric order flow shocks is left unexplained. In the Korean market, *PSOS* seems to be a proxy for difference of opinion, which explains the negative relation between *PSOS* and stock returns. Our empirical results shed light on this important issue, but further research on the causes of symmetric order flow in other countries is needed.

4.3 Difference of opinion among individual investors

We investigate whether the trading of individual investors is the main contributor of the predictive power of *PSOS*. Difference of opinion can be regarded as a response by uninformed traders to public news. Among various types of investors – domestic institutional, domestic individual, and foreign investors — we focus on individual traders for two reasons. First, individual investors are regarded as noise traders or uninformed traders in the literature (Kumar and Lee, 2006; Barber and Odean, 2008; Barber et al., 2009). This indicates that they potentially have different implications with regard to public news, because of the lack of accurate information. Second, one of the notable features in the Korean market is that there is a large proportion of individual investors trading in the market (Choe, Kho, and Stulz, 1999; Kang, Kwon, and Sim, 2013). The effect of their trading on the Korean market can be greater than that in other markets. Thus, we hypothesize that individual investors are the main group of traders generating the symmetric order flow because of difference of opinion; thus they are the main contributor of the significant relation between *PSOS* and expected returns.

First, we investigate the correlation of each trader group's buys and sells. In the adjusted *PIN* model, informed traders participate in only one side following their information, if there is an occurrence of private information; thus the correlation of their buys and sells is negative. On the other hand, the symmetric order flow contributes to both sides if there is a shock in the market; thus, the correlation of buys and sells is positive in this case. As we hypothesize, if individual investors are those who have heterogeneous beliefs and generate the symmetric order flow, then their buys and sells will have a positive correlation.

Our intraday data provide information about the types of buyers and sellers—whether they are institutional, individual, or foreign investors. For each day, we aggregate the number (volume) of buys and sells of each investor group, and then compute the time-series average of annual correlations of daily buys and sells.

[Insert Table 8]

In Table 8, the individual investor's buys and sells are highly correlated. In terms of number of buys and sells (*nbuy* and *nsell*), the correlation of individual investors is 0.800, which is notably higher than that of other groups, 0.121 and 0.120, and it is the only significant result. We can see the consistent results in terms of volume (*vbuy2* and *vsell2*). This means that individual investors tend to buy and sell simultaneously on a given day, similar to the symmetric order flow of the adjusted *PIN* model. In other words, these results suggest that symmetric order flow can be closely related to the trading of individual investors.

To investigate further the relation between *PSOS* and trading of individual investors, we construct a measure to capture the level of trades of each investor group on a given stock. By modifying the measures of Kumar and Lee (2006) and Han and Kumar (2013), we define the relative trading activity (or concentration) of a given investor group as a normalized number of shares traded by them (*NTA*) as follows.

$$RNTA_t = \frac{Number\ of\ shares\ traded\ by\ retail\ investors\ during\ a\ day\ t}{Number\ of\ shares\ outstanding} \times 10^3 \tag{4}$$

$$INTA_t = \frac{Number\ of\ shares\ traded\ by\ institutional\ investors\ during\ a\ day\ t}{Number\ of\ shares\ outstanding} \times 10^3 \tag{5}$$

$$FNTA_t = \frac{Number\ of\ shares\ traded\ by\ foreign\ investors\ during\ a\ day\ t}{Number\ of\ shares\ outstanding} \times 10^3 \tag{6}$$

We compute the annual average of those daily *NTA* measures. As *TURNOVER* is the ratio of total trading shares to total number of shares outstanding, these three *NTA* measures as components of *TURNOVER*[14] and figure out the main contributor of the predictive power of *PSOS* by comparing the effects of these three measures on *PSOS*.

[14] In principle, the sum of *RNTA*, *INTA*, and *FNTA* should be *TURNOVER*. As described in Section 3, however, intraday trade data are filtered following DY.

First, we examine the correlations of *NTA* and *PIN*-related measures. There is a belief that foreign and institutional investors are relatively more informed than retail investors, but some of the literature shows evidence inconsistent with this belief. Choe, Kho, and Stulz (2005) document that foreign investors in the Korean stock market are not informationally advantaged compared with domestic investors. Thus, we do not expect high correlation of *PIN* or *AdjPIN* with any specific trader group, but we expect *PSOS* and the individual trader group to be highly correlated. As we suggest, if the trading of individual investors is closely related to the symmetric order flow, then *PSOS* will be highly correlated with *RNTA*. The time-series average of the annual correlation among *NTA* measures and other key variables is reported in Table 9.

[Insert Table 9]

In Table 9, *PIN* and *AdjPIN* do not show the highly positive correlation with any specific trader group. By contrast, *PSOS* appears to be highly correlated with *RNTA* (0.388). *INTA* and *FNTA* also have positive correlations with *PSOS* (0.099 and 0.177, respectively), but the levels and significance of correlations are much smaller than *RNTA*. Overall, we confirm that *PSOS* is highly correlated with the trading of individual investors.

To investigate whether the trading of individual investors is caused by difference of opinion, we revisit the analysis in Section 4.2. In each year *t*, we sort firms into quintiles based on the firms' *PSOS* estimated in the previous year. For each quintile portfolio, we calculate each investor group's average number of buys and sells on event and non-event days, respectively.

[Insert Table 10]

In Table 10, all trader groups show increases in buys and sells on event days in general, but the size of increase in buys and sells across *PSOS* quintiles shows different patterns. Institutional and foreign investors show an almost *U*-shaped pattern in changes in buys and sells, as *PSOS* increases, but individual investors produce an almost monotonically increasing pattern in buys and sells on news days. These results suggest that, among three types of investors, domestic individual investors play an important role in generating the symmetric order flow.

Next, we perform the cross-sectional analysis to examine whether trading of individual investors is the main contributor of the pricing effect of *PSOS*. Among three components of *TURNOVER*, and we expect that *RNTA* is the key component driving the negative pricing effect of *PSOS*.

[Insert Table 11]

In Table 11, Models 1 to 3 examine the predictive power of each investor group's trading, respectively. *INTA* and *RNTA* are negatively priced. In Models 4 to 6, we examine the significance of *PSOS* after controlling for *NTA* measures. The results show that the coefficient of *PSOS* becomes insignificant only if *RNTA* is controlled while *INTA* does not reduce the significance of *PSOS*. This indicates that *RNTA* and *INTA* are negatively priced for different reasons, and the effect of *PSOS* is related only to that of *RNTA*, not *INTA*.

Moreover, we construct *AdjPSOS3* (*AdjPSOS4*) variable that is an adjusted *PSOS* measures orthogonal to *INTA* (*RNTA*) as *AdjPSOS1* and *AdjPSOS2* in Section 4.2. Model 7 shows that the explanatory power of *PSOS*, after controlling for *INTA*, is still negatively significant. This means that *RNTA* is the key component of *TURNOVER* that contributes to the negative relation between *PSOS* and expected returns. Model 8 examines whether *PSOS*, after controlling for *RNTA*, has explanatory power for expected returns. This model shows that the explanatory power of *PSOS* is substantially reduced by removing the component related to *RNTA*. The coefficient of *AdjPSOS4* is significant only under the 10% confidence level.

To summarize, we find that domestic individual investors generate symmetric order flow because of difference of opinion and that their trading is the main contributor of the pricing effect of *PSOS*. As we previously find that *PSOS* is closely related to *TURNOVER*, in this section we decompose *TURNOVER* into three trading activity measures, and show that the trading of individual investors contributes to the negative coefficient of *PSOS*, while others do not.

5. Conclusion

We use transaction data including trader type in the Korean stock market, mainly to examine what drives the significant relation between *PSOS* and expected returns. To accomplish this, we first perform asset pricing tests with *PIN*, *AdjPIN*, and *PSOS*, and find that *PSOS* is significantly priced while *AdjPIN* is not. Focusing on the explanatory power of *PSOS*, we investigate the main cause of symmetric order flow and why *PSOS* is negatively related to expected returns. We show that *PSOS* is not a proxy of illiquidity, but *PSOS* is closely related to *TURNOVER*, which is a proxy for difference of opinion. Then, we decompose *TURNOVER* into three components — the trading activity measures of institutional, individual, and foreign investors — and find that the trading of individual investors significantly contributes to the explanatory power of *PSOS*, while other components do not.

Our results show that *PSOS* can be negatively priced when the dominant cause of symmetric order flow shocks is difference of opinion. DY construct the adjusted *PIN* model and find that *PSOS* is positively priced as a proxy of illiquidity in the US stock market, but they do not clarify the cause of the symmetric order flows. Using data from 47 countries, LNZ find that *PSOS* is negatively priced, but they leave this finding unexplained.

The Korean stock market has distinctive features compared to the US stock market. First, the Korean stock market does not have designated market makers as in the US stock market. In the Korean market, buyers and sellers, who can submit both limit and market orders, meet via the Automated Trading System. Second, in the Korean stock market, most of the trading is done by individual investors. In this paper, we put more weight on the second distinctive feature, which is a different composition of investors in the market, deriving the different pricing effects of *PSOS*. Considering that the Korean stock market and other emerging markets have a larger proportion of the individual investors than the US stock market, our results may imply that in the market with a large proportion of individual investors, or uninformed noise traders, the dominant symmetric order flow shock can be difference of opinion, and thus the symmetric order flow component is negatively priced. Our paper sheds light on this important issue, but requires further research in other markets to clarify the effect of the main cause of symmetric order flow shocks to the explanatory power of *PSOS* on expected stock returns.

References

Admati, A., and Pfleiderer, P., 1988. A theory of intraday patterns: Volume and price variability. *Review of Financial Studies* 1, 3–40.

Aitken, M., and Frino, A., 1996. The accuracy of the tick test: Evidence from the Australian Stock Exchange. *Journal of Banking and Finance* 20, 1715–1729.

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

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

Amihud, Y., 2002. Illiquidity and stock returns: Cross section and time series effect. *Journal of Financial Markets* 5, 31–56.

Asquith, P., Oman, R., and Safaya, C., 2010. Short sales and trade classification algorithms. *Journal of Financial Markets* 13, 157–173.

Back, K., Crotty, K., and Li, T., 2016. Estimating information asymmetry in securities markets. Available at SSRN: https://ssrn.com/abstract=2565216.

Barber, B.M., and Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21, 785–818.

Barber, B.M., Odean, T., and Zhu N., 2009. Do retail trades move markets? *Review of Financial Studies* 22, 151–186.

Boehme, R.D., Danielsen, B.R., and Sorescu S.M., 2006. Short-sale constraints, difference of opinion, and overvaluation. *Journal of Financial and Quantitative Analysis* 41, 455–487.

Chakrabarty, B., Moulton, P.C., and Shkilko, A., 2012. Short sales, long sales, and the Lee-Ready trade classification algorithm revisited. *Journal of Financial Markets* 15, 467–491.

Choe, H., Kho, B.-B., and Stulz, R.M., 1999. Do foreign investors destabilize stock markets? The Korean experience in 1997. *Journal of Financial Economics* 54, 227–264.

Choe, H., Kho, B.-C., and Stulz, R.M., 2005. Do domestic investors have an edge? The trading experience of foreign investors in Korea. *Review of Financial Studies* 18, 798–829.

Choe, H., and Yang, C.-W., 2006. Comparisons of information asymmetry measures in the Korean stock market. *Asia-Pacific Journal of Financial Studies* 35, 1–44.

Choe, H., and Yang, C.-W., 2007. Information risk and asset returns in the Korean stock market. *Asia-Pacific Journal of Financial Studies* 36, 567–620.

Collin-Dufresne, P., and Fos, V., 2015. Do prices reveal the presence of informed trading? *Journal of Finance* 70, 1555–1582.

Diether K.B., Malloy C.J., and Scherbina, A., 2002. Difference of opinion and the cross section of stock returns. *Journal of Finance* 57, 2113–2141.

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

Duarte, J., Hu, E., and Young, L., 2015. What does the PIN model identify as private information? Available at SSRN: https://ssrn.com/abstract=2564369.

Easley, D., Engle, R.F., O'Hara, M., and Wu, L., 2008. Time-varying arrival rates of informed and uninformed trades. *Journal of Financial Economics* 6, 171–207.

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

Easley, D., Hvidkjaer, S., and O'Hara, M., 2010. Factoring information into returns. *Journal of Financial and Quantitative Analysis* 45, 293–309.

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

Easley, D., and O'Hara, M., 1992. Time and the process of security price adjustment. *Journal of Finance* 47, 577–605.

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

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

Eom, K.S., Kang, J., and Kwon, K.Y., 2016. The best *PIN* model in the Korean stock market. *Asian Review of Financial Research* 29, 425–436.

Finucane, T., 2000. A direct test of methods for inferring trade direction from intra-day data. *Journal of Financial and Quantitative Analysis* 35, 553–576.

Han, B., and Kumar, A., 2013. Speculative retail trading and asset prices. *Journal of Financial and Quantitative Analysis* 48, 377–404.

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

Hughes, J.S., Liu, J., and Liu, J., 2007. Information, diversification, and cost of capital. *Accounting Review* 82, 705–729.

Hvidkjaer, S., 2008. Small trades and the cross-section of stock returns. *Review of Financial Studies* 21, 1123–1151.

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

Kang, J., Kwon, K.Y., and Park, H., 2014. Momentum and foreign investors: Evidence from the Korean stock market. *Emerging Markets Finance and Trade* 50, 157–172.

Kang, J., Kwon, K.Y., and Sim, M., 2013. Retail investor sentiment and stock returns. *The Korean Journal of Financial Management* 30, 35–68. (in Korean)

Kumar, A., and Lee, C.M.C., 2006. Retail investor sentiment and return comovements. *Journal of Finance* 61, 2451–2486.

Lai, S., Ng, L., and Zhang, B., 2014. Does *PIN* affect equity prices around the world? *Journal of Financial Economics* 114, 178–195.

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

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

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

Lin, H.W.W., and Ke, W. C., 2011. A computing bias in estimating the probability of informed trading. *Journal of Financial Markets* 14, 625–640.

Miller, E.M., 1977. Risk, uncertainty, and divergence of opinion. *Journal of Finance* 32, 1151–1168.

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

Park, J.-H., and Eom, K.S., 2008. The usefulness of PIN as a measurement for private-information risk in Korean Stock Markets. *Asia-Pacific Journal of Financial Studies* 37, 501–536. (in Korean with long English abstract)

Petchey, J., Wee, M., and Yang, J., 2016. Pinning down an effective measure for probability of informed trading. *Pacific-Basin Finance Journal* 40, 456-475.

Tay, A., Ting, C., Tse, Y., and Warachka, M., 2009. Using high-frequency transaction data to estimate the probability of informed trading. *Journal of Financial Econometrics* 7, 288–311.

Tesar, L.L., and Werner, I.M., 1995. Home bias and high turnover. *Journal of International Money and Finance* 14, 467–492.

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

Table 1. Findings from *PIN* models in the Korean stock markets

This table presents the previous findings on the performance of the *PIN* models in the Korean stock markets.

Authors	Sample period	Findings
Choe and Yang (2006)	January 2002 to March 2002	*PIN* shows little correlations with other measures for information asymmetry while others are closely correlated with each other, and it also shows an insignificant relation with the expected stock returns.
Choe and Yang (2007)	January 1993 to December 2004	*PIN* is not significantly priced while other measures for information risk are significantly priced.
Park and Eom (2007)	January 1997 to December 2005	*AdjPIN* generally shows a significant relation with the expected stock returns while *PIN* does not. The original PIN model does not fit Korean stock market data, but is only useful for detecting the private-information risk that occurs during an extremely short period of time.
Hwang, Lee, Lim, and Park (2013)	January 2000 to December 2004	*AdjPIN* is significantly related to the expected return, which is measured by the implied cost of equity, while *PIN* is not.
Lai, Ng, and Zhang (2014)	January 1997 to December 2010	Both *PIN* and *AdjPIN* are not significantly priced in the international stock markets including the Korean stock market.
Kang, Kwon, and Eom (2016)	January 2001 to December 2006	The original *PIN* model does not fit to the Korean stock market data.

Table 2. Data specification

This table presents a part of the intraday data on January 12th, 2004. Each row provides information about one transaction including a transaction price, number of shares traded, or the time the transaction completed. "B-A spread" indicates the bid-ask spread, and "Buyer (Seller) number" indicates the order submission number of the buyer (seller) of each transaction.

Firm code	Transaction time	Date	Bid price	Ask price	B-A spread	Midpoint	Buyer number	Seller number	Contraction number	Number of shares	Price	Amount
7160	32568.21	2004-01-12	3995	3935	60	3965	184	151	30	100	3995	399500
7160	32979.86	2004-01-12	3935	3915	20	3925	185	276	72	30	3915	117450
7160	32979.86	2004-01-12	3935	3915	20	3925	260	276	73	370	3915	1448550
7160	33402.78	2004-01-12	3925	3920	5	3922.5	341	332	107	2030	3925	7967750
7160	33402.79	2004-01-12	3950	3920	30	3935	341	302	108	100	3950	395000
7160	33402.8	2004-01-12	3970	3920	50	3945	341	271	111	330	3970	1310100
7160	33402.81	2004-01-12	3990	3920	70	3955	341	259	112	800	3990	3192000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 3. Descriptive statistics

This table presents the descriptive statistics of our data. Panel A shows the median and the percentiles of the cross-sectional distribution of a series of statistics on the daily number of buys and sells for each stock in the sample. Panel B shows the cross-sectional distribution of the estimated parameters of the adjusted *PIN* model along with the cross-sectional distribution of the estimated *AdjPIN* and *PSOS* in the adjusted *PIN* model. Panel C shows the summary statistics of monthly control variables. *Beta* is post-ranking beta estimated using 10 portfolios. *log(ME)* is the log of market value of firm equity from December of year *t*-1 and *log(BM)* is the log of book value divided by market value for year *t*-1. *ILLIQ* is the Amihud (2002) illiquidity measure calculated by annual average of daily price impact. The sample period is from 2001 to 2006.

Panel A. Percentile of summary statistics on the number of buys and sells

	95th percentile	75th percentile	Median	25th percentile	5th percentile
Mean buys	9.41	5.09	3.15	1.76	0.74
Mean sells	15.18	8.53	5.37	3.29	1.61
Variance buys	248.26	89.85	36.84	11.97	1.75
Variance sells	398.34	164.31	64.35	20.15	4.26
Correlation between buys and sells	0.69	0.47	0.23	0.03	-0.29

Panel B. Adjusted *PIN* model

	95th percentile	75th percentile	Median	25th percentile	5th percentile
a	1.00	0.54	0.39	0.30	0.17
u_b	30.23	12.42	6.95	1.97	0.00
u_s	32.99	15.68	9.66	4.32	0.22
d	0.84	0.58	0.42	0.28	0.00
ε_b	3.19	1.46	0.82	0.38	0.00
ε_s	5.41	3.11	2.01	1.25	0.24
θ	1.00	0.23	0.10	0.04	0.00
θ'	0.53	0.27	0.18	0.11	0.00
Δ_b	32.19	16.06	8.42	1.48	0.00
Δ_s	42.61	21.64	10.66	2.30	0.00
AdjPIN	0.59	0.44	0.38	0.30	0.19
PSOS	0.44	0.34	0.27	0.21	0.12

Panel C. Summary statistics of control variables

	ILLIQ	*log(ME)*	*log(BM)*	*Beta*
Mean	0.16	11.07	7.58	0.93
Max	17.05	18.39	9.03	1.79
Min	0.00	1.80	4.73	0.32
Std. dev.	0.64	1.62	0.82	0.32
Skewness	9.47	0.80	-0.75	0.70
Kurtosis	130.97	1.06	0.32	0.27

Table 4. *PIN* and the cross-section of expected returns

This table presents time series averages of the monthly cross-sectional regression. *Beta* is post-ranking beta estimated using 10 portfolios. *log(ME)* is the log of market value of firm equity from December of year *t*-1 and *log(BM)* is the log of book value divided by market value for year *t*-1. *ILLIQ* is the Amihud (2002) illiquidity measure calculated by annual average of daily price impact. *PIN*, *AdjPIN*, and *PSOS* are estimated for each calendar year. The sample period is from 2001 to 2006 and *t*-statistics are in parentheses. *, **, and *** indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

	1	2	3	4	5	6	7	8	9
Beta	0.807	0.731	1.028	1.126*	1.003	0.784	1.069	1.167*	1.043
	(1.25)	(1.15)	(1.64)	(1.79)	(1.61)	(1.22)	(1.69)	(1.83)	(1.66)
log(ME)	-0.088	-0.059	-0.063	-0.096	-0.066	-0.025	-0.037	-0.064	-0.040
	(-0.50)	(-0.34)	(-0.36)	(-0.55)	(-0.38)	(-0.14)	(-0.21)	(-0.35)	(-0.22)
log(BM)	0.724***	0.775***	0.725***	0.679***	0.717***	0.785***	0.735***	0.685***	0.728***
	(3.19)	(3.49)	(3.30)	(3.04)	(3.25)	(3.48)	(3.32)	(3.06)	(3.27)
ILLIQ						1.064	0.902	1.001	0.898
						(1.92)	(1.68)	(1.87)	(1.65)
PIN	-0.571			-0.656				-0.939	
	(-0.50)			(-0.54)				(-0.76)	
AdjPIN		0.254			-0.785				-0.831
		(0.27)			(-0.94)				(-0.94)
PSOS			-2.930***	-2.937***	-3.233***		-2.740***	-2.744***	-3.059***
			(-3.04)	(-2.91)	(-3.37)		(-2.93)	(-2.80)	(-3.29)

Table 5. Correlations of *PIN*-related measures

This table shows time series average of correlation between trading activity measures, *PIN*-related measures, and other control variables. *log(ME)* is the logarithm of market value of firm equity from December of year *t*-1, and *log(BM)* is the logarithm of book value divided by market value for year *t*-1. *ILLIQ* is Amihud (2002) illiquidity measure calculated by annual average of daily price impact. *PIN*, *AdjPIN*, and *PSOS* are estimated for each calendar year. *TURNOVER* is calculated by annual average of daily turnover. The sample period is from 2001 to 2006 and *t*-statistics are in parentheses. *, **, and *** indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

	PIN	*AdjPIN*	*PSOS*	*ILLIQ*	*log(ME)*	*log(BM)*	*TURNOVER*
PIN	1	0.522***	0.018	0.037	0.062	-0.006	-0.098
		(6.03)	(0.12)	(0.36)	(0.60)	(-0.19)	(-1.11)
AdjPIN		1	-0.112	-0.034	0.040	0.015	-0.213*
			(-0.46)	(-0.41)	(0.53)	(0.44)	(-1.97)
PSOS			1	-0.179***	0.031	-0.159**	0.399***
				(-3.18)	(0.42)	(-2.11)	(8.50)
ILLIQ				1	-0.204***	0.080	-0.018
					(-5.04)	(1.09)	(-0.88)
log(ME)					1	-0.195	-0.416***
						(-1.38)	(-8.13)
log(BM)						1	-0.056
							(-0.68)
TURNOVER							1

Table 6. Trading activity on days with and without fair disclosure

This table presents information about trading activity of each *PSOS* quintile on event days and non-event days. Event days are days with fair disclosure and non-event days are the remaining days. In each year *t*, we sort firms into quintiles based on the firms' *PSOS* estimated in the previous year. For each quintile portfolio, we calculate the average of annual *PIN*-related measures, and then calculate average numbers of buys (*BUYS*) and sells (*SELLS*), average net order imbalance (*IMB*), average value of Amihud (2002) illiquidity measure (*ILLIQ*), and average value of turnover (*TURNOVER*) on event and non-event days, respectively. The sample period is from 2003 to 2006 and *t*-statistics are in parentheses. *, **, and *** indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

PSOS quintile	*PSOS*	*PIN*	*AdjPIN*	*BUYS*		*SELLS*		*IMB*		*ILLIQ*×100		*TURNOVER*×100	
				non-event	*event*	*non-event*	*event*	*non-event*	*event*	*non-event*	*event*	*non-event*	*event*
1	0.084	0.449***	0.412***	6.2	7.6	9.9	8.9	-3.6	-1.3	0.075	0.030	0.012	0.014
	(1.08)	(3.58)	(2.60)	(0.25)	(0.25)	-(0.14)	(0.12)	(0.26)	(0.50)	(0.47)	-(0.07)	(0.21)	(0.36)
2	0.222***	0.437***	0.393***	10.1	17.9	14.6	19.9	-4.5	-2.0	0.030	0.006	0.014	0.016
	(5.66)	(4.88)	(5.03)	(0.33)	(0.33)	-(0.13)	(0.12)	(0.28)	(0.43)	(0.48)	-(0.05)	(0.33)	(0.28)
3	0.281***	0.415***	0.363***	12.4	21.9	18.6	28.8	-6.2	-6.9	0.022	0.006	0.015	0.014
	(10.03)	(4.78)	(5.64)	(0.35)	(0.37)	-(0.14)	(0.09)	(0.33)	(0.42)	(0.45)	-(0.12)	(0.11)	(0.48)
4	0.337***	0.401***	0.334***	11.9	20.8	19.3	28.7	-7.3	-7.9	0.018	0.005	0.020	0.024
	(13.06)	(4.90)	(6.27)	(0.36)	(0.38)	-(0.19)	(0.08)	(0.32)	(0.42)	(0.46)	-(0.14)	(0.10)	(0.36)
5	0.435***	0.402***	0.288***	12.7	20.3	23.5	32.6	-10.8	-12.3	0.024	0.007	0.038	0.044
	(7.48)	(4.13)	(6.26)	(0.38)	(0.41)	-(0.28)	(0.08)	(0.38)	(0.51)	(0.51)	-(0.25)	(0.22)	(0.48)

Table 7. *TURNOVER* and the cross-section of expected returns

This table presents time series averages of the monthly cross-sectional regression. *Beta* is post-ranking beta estimated using 10 portfolios. *Log(ME)* is the log of market value of firm equity from December of year *t*-1 and *log(BM)* is the log of book value divided by market value for year *t*-1. *ILLIQ* is Amihud (2002) illiquidity measure calculated by annual average of daily price impact. *PSOS* is estimated for each calendar year. *TURNOVER* is calculated by annual average of daily turnover. The *AdjPSOS*1 (*AdjPSOS*2) measure is an adjusted *PSOS* measure that is orthogonal to *ILLIQ* (*TURNOVER*). The sample period is from 2001 to 2006 and *t*-statistics are in parentheses. *, **, and *** indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

	1	2	3	4	5	6
Beta	0.931	1.000	1.029	0.980	0.796	0.865
	(1.48)	(1.60)	(1.63)	(1.57)	(1.26)	(1.36)
log(ME)	-0.207	-0.197	-0.172	-0.068	-0.042	-0.014
	(-1.23)	(-1.18)	(-0.98)	(-0.39)	(-0.24)	(-0.08)
log(BM)	0.631***	0.626***	0.638***	0.727***	0.771***	0.780***
	(3.10)	(3.07)	(3.11)	(3.32)	(3.43)	(3.44)
PSOS		-0.910	-0.883			
		(-1.02)	(-0.99)			
ILLIQ			0.717			1.035*
			(1.44)			(1.87)
TURNOVER	-0.249***	-0.226***	-0.215***			
	(-3.35)	(-2.85)	(-2.79)			
*AdjPSOS*1				-2.717***		
				(-2.93)		
*AdjPSOS*2					-1.087	-0.988
					(-1.19)	(-1.08)

Table 8. Correlations between daily buys and sells of three investor groups

This table shows time series averages of correlation between buys and sells of three investor groups. *nbuyi* (*nselli*) means the number of buys (sells) of an investor group *i* (*i*=1, 2, 3). *vbuyi* (*vselli*) means the share volume of buys (sells) of an investor group *i* (*i*=1, 2, 3). Investor group 1, 2, and 3 indicate domestic institutional investor group, domestic individual investor group, and foreign investor group, respectively. Every day, we compute the correlation of daily buys and sells, then calculate the time series average of the correlation. The sample period is from 2001 to 2006. *t*-statistics are in parentheses. *, **, and *** indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

	nsell1	*nsell2*	*nsell3*	*vsell1*	*vsell2*	*vsell3*
nbuy1	0.121	0.024	0.083			
	(1.00)	(0.37)	(0.79)			
nbuy2	0.048	0.800***	0.076			
	(0.52)	(7.12)	(0.64)			
nbuy3	0.086	0.060	0.120			
	(0.77)	(0.60)	(0.84)			
vbuy1				0.098	0.040	0.060
				(0.65)	(0.37)	(0.52)
vbuy2				0.071	0.722***	0.124
				(0.51)	(4.18)	(0.64)
vbuy3				0.063	0.096	0.105
				(0.53)	(0.61)	(0.63)

Table 9. Correlations of trading activity measures with key variables

This table shows time series averages of correlation between trading activity measures, *PIN*-related measures, and other control variables. *Log(ME)* is the logarithm of market value of firm equity from December of year *t*-1 and *log(BM)* is the logarithm of book value divided by market value for year *t*-1. *ILLIQ* is Amihud (2002) illiquidity measure calculated by annual average of daily price impact. *PIN*, *AdjPIN*, and *PSOS* are estimated for each calendar year. *RNTA*, *INTA*, and *FNTA* are annual average of daily normalized trading activity of retail, institutional, and foreign traders, respectively. *TURNOVER* is calculated by annual average of daily turnover. The sample period is from 2001 to 2006, and *t*-statistics are in parentheses. *, **, and *** indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

	INTA	*RNTA*	*FNTA*	*PIN*	*AdjPIN*	*PSOS*	*ILLIQ*	*log(ME)*	*log(BM)*	*TURNOVER*
INTA	1	0.067	0.060	0.040	-0.007	0.099***	-0.003	0.142	-0.057	0.116
		(0.43)	(0.74)	(0.79)	(-0.10)	(3.41)	(-0.03)	(1.40)	(-1.40)	(0.74)
RNTA		1	0.220	-0.100	-0.205*	0.388***	-0.017	-0.429***	-0.054	0.998***
			(0.67)	(-1.21)	(-1.78)	(7.05)	(-0.76)	(-8.52)	(-0.68)	(14.09)
FNTA			1	-0.070	-0.100	0.177*	-0.106***	0.129	-0.061	0.243
				(-0.76)	(-0.95)	(1.88)	(-2.74)	(0.45)	(-0.72)	(0.73)

Table 10. Trading activity of three investor groups on days with and without fair disclosure

This table presents information about trading activity of three investor groups for each *PSOS* quintiles on event days and non-event days. Event days are days with fair disclosure and non-event days are the remaining days. In each year *t*, we sort firms into quintiles based on the firms' *PSOS* estimated in the previous year. For each quintile portfolio, we calculate each investor group's average numbers of buys (*BUYS*) and sells (*SELLS*) on event and non-event days. 1, 2, and 3 are for institutional, individual, and foreign investor groups, respectively. The sample period is from 2003 to 2006 and *t*-statistics are in parentheses. *, **, and *** indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

PSOS quintile	*BUYS*1		*SELLS*1		*BUYS*2		*SELLS*2		*BUYS*3		*SELLS*3	
	non-event	*event*	*non-event*	*event*	*non-event*	*event*	*non-event*	*event*	*non-event*	*event*	*non-event*	*event*
1	1.224	1.614	1.003	1.240	4.502	5.271	8.338	7.027	0.505	0.699	0.515	0.614
	(0.14)	(0.24)	(0.14)	(0.17)	(0.21)	(0.43)	(0.22)	(0.44)	(0.10)	(0.16)	(0.10)	(0.14)
2	2.986	6.633	2.267	4.679	5.910	9.051	11.136	13.608	1.214	2.180	1.240	1.602
	(0.18)	(0.22)	(0.19)	(0.24)	(0.26)	(0.35)	(0.28)	(0.40)	(0.11)	(0.22)	(0.11)	(0.21)
3	3.237	6.835	2.814	5.929	7.600	11.990	14.208	19.544	1.527	3.071	1.567	3.346
	(0.19)	(0.22)	(0.18)	(0.25)	(0.28)	(0.36)	(0.32)	(0.38)	(0.12)	(0.15)	(0.13)	(0.22)
4	2.576	5.088	2.290	5.235	8.120	13.081	15.688	21.059	1.247	2.647	1.312	2.452
	(0.17)	(0.18)	(0.16)	(0.21)	(0.29)	(0.37)	(0.33)	(0.39)	(0.14)	(0.15)	(0.12)	(0.17)
5	1.063	2.586	1.064	2.676	10.943	16.513	21.584	27.797	0.704	1.151	0.822	2.096
	(0.12)	(0.18)	(0.12)	(0.15)	(0.35)	(0.49)	(0.39)	(0.51)	(0.11)	(0.15)	(0.12)	(0.10)

Table 11. *NTA* measures and the cross-section of expected returns

This table presents time series averages of the monthly cross-sectional regression. *Beta* is post-ranking beta estimated using 10 portfolios. *Log(ME)* is the log of market value of firm equity from December of year *t*-1 and *log(BM)* is the log of book value divided by market value for year *t*-1. *ILLIQ* is Amihud (2002) illiquidity measure calculated by annual average of daily price impact. *PSOS* is estimated for each calendar year. The *AdjPSOS*3 (*AdjPSOS*4) measure is an adjusted *PSOS* measure, which is orthogonal to *INTA* (*RNTA*). The sample period is from 2001 to 2006 and *t*-statistics are in parentheses. *, **, and *** indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

	1	2	3	4	5	6	7	8
Beta	0.729	0.790	0.759	1.054	0.944	1.047	1.061	0.959
	(1.14)	(1.24)	(1.17)	(1.68)	(1.51)	(1.66)	(1.68)	(1.52)
log(ME)	-0.034	-0.209	-0.087	-0.045	-0.172	-0.099	-0.040	0.005
	(-0.20)	(-1.32)	(-0.50)	(-0.26)	(-1.14)	(-0.57)	(-0.22)	(0.03)
log(BM)	0.777***	0.723***	0.729***	0.727***	0.709***	0.683***	0.739***	0.765***
	(3.49)	(3.76)	(3.32)	(3.32)	(3.57)	(3.13)	(3.33)	(3.29)
PSOS				-2.866***	-1.621	-2.496***		
				(-2.94)	(-1.51)	(-2.47)		
ILLIQ							0.903	0.989*
							(1.68)	(1.81)
TURNOVER								
INTA	-4.110*			-3.291				
	(-1.94)			(-1.49)				
RNTA		-0.638***			-0.478*			
		(-2.66)			(-1.82)			
FNTA			-5.315			-3.387		
			(-0.68)			(-0.44)		
*AdjPSOS*3							-2.620***	
							(-2.76)	
*AdjPSOS*4								-1.906*
								(-1.79)