Asian Review of Financial Research Vol. 29 No. 3 (August 2016)

The Best *PIN* Model in the Korean Stock Market

Kyong Shik Eom Affiliated Researcher, CRMR, University of California at Berkeley
Jangkoo Kang Professor, College of Business, Korea Advanced Institute of Science and Technology
Kyung Yoon Kwon* Ph.D Candidate, College of Business, Korea Advanced Institute of Science and Technology

Abstract We investigate the effectiveness of the original *PIN* model (Easley, Kiefer, O'Hara, and Paperman, 1996) and five variants of the adjusted *PIN* model (Duarte and Young, 2009) in the Korean stock market. Throughout the series of likelihood-ratio fitness tests, we find that the unrestricted version of the adjusted *PIN* model fits best in the Korean stock market data.

Keywords Adjusted PIN, Information Risk, Likelihood Ratio Test, Monte Carlo Simulation, Korean Stock Market Microstructure

Received 19 Apr. 2016
Accepted 21 Jun. 2016

* **Address:** College of Business, Korea Advanced Institute of Science and Technology, 85 Hoegiro, Dongdaemoon-gu, Seoul, 02455, South Korea; **E-mail:** noldya@business.kaist.ac.kr; **Tel:** 82-2-958-3693.

Ⅰ. Introduction

Easley and O'Hara (1992) 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 trades and seller-initiated trades. DY suggest an adjusted *PIN* model that adds a symmetric order flow shock generated by public information to the original *PIN* model, and show that their model fits the data much better than does the original in the US stock markets. They replace the original *PIN* measure with two components, the adjusted probability of informed trading (*AdjPIN*) and the probability of trading caused by symmetric order flow shocks (*PSOS*) arising from public information.

In this paper, we examine the effectiveness of the original *PIN* model of Easley et al. (1996) and five variants of DY's adjusted *PIN* models in the Korean stock market, and suggest the model that fits the Korean market data best.

We find that the original *PIN* model shows the same problems in the Korean stock market explaining the two stylized facts as in the U.S. stock market, and among five variants of the adjusted *PIN* models, the model without any restriction fits the Korean market data best. Lai, Ng, and Zhang (2014) investigate the performance of the adjusted *PIN* model with the worldwide data including the Korean market data, but they simply adopt the model which DY accept based on the US market data. However, we find that the different version of the adjusted model fits the Korean market better than the model that Lai et al. (2014) adopt. Our results suggest that the choice of the best model should be considered separately in applying models to different countries.

I. The *PIN* Model and Its Extension

Easley and O'Hara (1992) suggest a sequential trade model with informed traders and noise traders. They assume that informed traders buy (sell) stocks only if they receive positive (negative) private information and they do not participate in any trades if there is no private information. Noise traders, on the other hand, always buy and sell without private information. In this paper, we employ the Easley et al. (1996) model as the original *PIN* model. According to their model, the original *PIN* is defined as follows:

$$PIN = \frac{a \times u}{a \times u + \epsilon_b + \epsilon_s} \tag{1}$$

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.

As DY show, the original *PIN* model has two critical problems. First, the correlation of buys and sells on a given day is restricted to be non-positive under the original model, but the data show that most of firms have positive correlations of buys and sells. Second, the variances of buys and sells implied by the model are much lower than those directly calculated from the data. To resolve these problems, DY extend the original *PIN* model by adding in the model symmetric order flow shocks that cause buy and sell trades simultaneously. In their model, the probability of informed trading (*AdjPIN*) and the probability of trading caused by symmetric order flow shocks (*PSOS*) can be defined as follows:

$$AdjPIN = \frac{a\times(d\times u_b+(1-d)\times u_s)}{a\times(d\times u_b+(1-d)\times u_s)+(\triangle_b+\triangle_s)\times(a\times\theta'+(1-a)\times\theta)+\epsilon_b+\epsilon_s} \quad (2)$$

$$PSOS = \frac{(\triangle_b+\triangle_s)\times(a\times\theta'+(1-a)\times\theta}{a\times(d\times u_b+(1-d)\times u_s)+(\triangle_b+\triangle_s)\times(a\times\theta'+(1-a)\times\theta)+\epsilon_b+\epsilon_s} \quad (3)$$

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. 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 rate of buys and sells are Δ_b and Δ_s.

Ⅲ. Data and Methodology

Our sample includes all the stocks belonging to Korea Stock Price Index (KOSPI). To estimate the structural microstructure models, our sample period is restricted to the period from 2001 to 2006 due to the availability of relevant intraday transaction data. Our intraday transaction data contain a time-ordered record of every stock transaction on the Korea Exchange. The data provide information regarding the types and the order numbers of buyers and sellers, which enables us to clearly identify the initiator of each trade without employing Lee and Ready (1991) algorithm, which is reported to have misclassification problems (Odders-White, 2000; Ellis, Michaely, and O'Hara, 2000; Asquith, Oman, and Safaya, 2010; Chakrabarty, Moulton, and Shkilko, 2012). More importantly, Hwang, Lee, Lim, and Park (2013) document that misclassification problem of Lee and Ready algorithm derives the biased estimates of *PIN* in the Korean stock market, thus the empirical tests of this *PIN* can be critically affected. Our data avoid the problem by using the order numbers and defining the initiator of a trade as the one who placed his/her order later than the other among the buyer and seller, thus provide the error-free estimates of *PIN*-related measures.[1)]

We exclude all trades that occur before the open and after the close, or have negative bid-ask spreads.[2)] The parameters of models are estimated by the maximum likelihood method. Every day, we aggregate the number of buys and sells for each firm and use this aggregated number of buys and sells for the estimation. In each firm-year, following DY, we maximize the likelihood with 50 different, randomly chosen, starting points. Then, we finally choose the maximum of these 50 maximization results.

Ⅳ. Results

First, we examine the effectiveness of the original *PIN* model in the

1) Choe and Yang (2007), Park and Eom (2008), and Hwang et al. (2013) estimate the original and/or the adjusted *PIN* model in the Korean stock market, and report that *PIN* and adjusted *PIN* do not seem to properly measure the information risk. None of these studies, however, examine which model fits the data best among the five variants of the adjusted model. All of these papers employ the most restricted version of the model without the fitness test.

2) To avoid possible data errors, we filter intraday data following DY reported in <Appendix A> of their paper.

Korean stock market. DY document that the original *PIN* model cannot account for the positive correlation between buys and sells and the high variances of buys and sells in the US market. We examine whether these problems are also observable in the Korean stock market.

<Table 1> Percentiles of Summary Statistics on the Number of Buyers and Seller Initiated Trades

Panel A presents 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 displays the cross-sectional distribution of the moments of the number of buyer-initiated and seller-initiated trades implied by the original *PIN* model. The sample period is from 2001 to 2006.

	95th percentile	75th percentile	Median	25th percentile	5th percentile
Panel A: Statistics on the number of buyer and seller-initiated trades					
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: Moments of buys and sells implied by the model					
Mean buys	8.77	4.77	2.96	1.72	0.65
Mean sells	13.30	7.77	4.99	3.07	1.46
Variance buys	122.99	45.57	19.88	7.12	1.02
Variance sells	170.52	65.57	29.34	11.19	2.06
Correlation between buys and sells	0.00	-0.07	-0.11	-0.15	-0.28

Panels A and B of <Table 1> show that the original *PIN* model's problems pointed out by DY exist in the Korean stock market. As DY point out, the original *PIN* model requires a non-positive correlation between buys and sells, but Panel A shows that most of the sample firms have positive correlations. The estimated variances of buys and sells in Panel B are too small relative to the observed variances in Panel A.

To overcome the limitations of the original *PIN* model, DY suggest five variants of extension for the model. Model 5 is the unrestricted version of the adjusted *PIN* model. Model 3 and Model 4 restrict the probability of symmetric order flow shocks; In particular, Model 4 restricts the probability of symmetric order flow shocks to be the same on days with and without private information ($\theta' = \theta$) and Model 3 restricts the occurrence of symmetric order flow shocks to days with private information ($\theta' = \theta$). Model 2, in addition to the restriction in Model 3, restricts the arrival intensities of informed buyers and sellers to be the same ($u_b = u_s$). Finally, Model 1, which is the most restricted version of the adjusted model, additionally restricts the arrival intensities of buyer and seller order flow shocks to be the same ($\Delta_b = \Delta_s$). <Table 2> shows the estimation results of these five models.

<Table 2> shows that all five variants of the adjusted *PIN* model successfully generate the positive correlation between buys and sells and high levels of volatilities of buys and sells that the original *PIN* model fails to generate. These variants of the adjusted *PIN* models seem to account for the data much better than the original model.

Panel A of <Table 3> shows the results of the likelihood ratio tests about the restrictions of the models relative to less restricted models. For the test statistics in Tests 1 and 4 in Panel A of <Table 3>, we conduct the Monte Carlo simulations to obtain their distributions as follows. First, we randomly select 100 firm-years and estimate both null and alternative models. Then, based on the estimates, we generate one firm-year of trading days under the null and compute the likelihood ratio statistic with the estimates of both models based on the simulated data. For each firm-year, this process is repeated 500 times. Finally, we compare the likelihood ratio from the data to critical values from the simulated distribution. For the rest of tests, we use χ^2 distribution with one degree of freedom.

<Table 2> Percentiles of the Moments of Buys and Sells Implied by the Adjusted *PIN* Model

This table presents the median and the percentiles of the mean, variance, and correlation of buys and sells implied by the adjusted trading model and calculated using estimated parameters for each stock-year in the sample. Each panel displays the results based on different versions of the adjusted model. The sample period is from 2001 to 2006.

	95th percentile	75th percentile	Median	25th percentile	5th percentile
Panel A: The Adjusted Model 1 ($u_b = u_s,\ \Delta_b = \Delta_s,\ \theta' = 0$)					
Mean buys	9.30	5.05	3.13	1.76	0.64
Mean sells	13.85	7.79	5.01	3.07	1.50
Variance buys	145.24	59.69	24.77	7.67	1.16
Variance sells	169.16	71.23	30.22	10.53	2.11
Correlation between buys and sells	0.87	0.60	0.21	-0.10	-0.27
Panel B: The Adjusted Model 2 ($\Delta_b = \Delta_s,\ \theta' = 0$)					
Mean buys	8.94	4.88	3.01	1.71	0.64
Mean sells	14.10	8.04	5.08	3.10	1.45
Variance buys	132.79	50.61	21.05	7.12	1.04
Variance sells	221.53	85.56	34.51	12.29	2.07
Correlation between buys and sells	0.85	0.56	0.00	-0.12	-0.30
Panel C: The Adjusted Model 3 ($\theta' = 0$)					
Mean buys	8.92	4.79	2.97	1.69	0.59
Mean sells	14.40	8.07	5.14	3.10	1.46
Variance buys	136.34	49.86	21.58	6.95	1.02
Variance sells	222.09	94.59	38.05	13.08	2.13
Correlation between buys and sells	0.82	0.53	0.14	-0.10	-0.27
Panel D: The Adjusted Model 4 ($\theta' = 0$)					
Mean buys	9.37	5.02	3.10	1.68	0.60
Mean sells	14.82	8.39	5.24	3.09	1.45
Variance buys	149.14	56.17	24.54	7.51	0.93
Variance sells	248.36	100.98	43.42	12.57	2.13
Correlation between buys and sells	0.50	0.27	0.08	-0.05	-0.31
Panel E: The Adjusted Model 5 (unrestricted adjusted model)					
Mean buys	9.20	4.99	3.07	1.67	0.60
Mean sells	14.82	8.39	5.24	3.09	1.45
Variance buys	149.48	57.21	24.69	7.92	1.00
Variance sells	250.27	103.92	41.74	13.16	2.05
Correlation between buys and sells	0.50	0.27	0.08	-0.05	-0.31

<Table 3> The Best *PIN* Model

This table presents the rejection frequency of the null hypothesis for a series of likelihood ratio tests (Panel A) and the estimation results of the preferred adjusted model (Panel B). In Panel A, The test statistics in Tests 1 and 4 have non-standard asymptotic distributions and hence, the reported rejection frequencies are estimated using Monte Carlo simulations. Test 1 and 4 are performed on a sample of 100 randomly selected firm-years. The test statistics in Tests 2, 3, and 5 are distributed chi-square with one degree of freedom. Panel B presents the cross-sectional distribution of the estimated parameters of the preferred adjusted model along with the cross-sectional distribution of the estimated probability of information trading in the adjusted model (*AdjPIN*), and the cross-sectional distribution of the estimated probability of symmetric order-flow shock (*PSOS*). The sample period is from 2001 to 2006.

Panel A: Likelihood ratio tests

Test	Null hypothesis	Alternative hypothesis	Frequency of rejection of null (99%)	Frequency of rejection of null (95%)	Frequency of rejection of null (90%)
1	Original *PIN* model	Adjusted Model 1	61%	63%	63%
2	Adjusted Model 1	Adjusted Model 2	53%	56%	57%
3	Adjusted Model 2	Adjusted Model 3	51%	56%	58%
4	Adjusted Model 3	Adjusted Model 5	99%	99%	99%
5	Adjusted Model 4	Adjusted Model 5	23%	24%	24%

Panel B: Estimation of the Best Model

	95^{th} percentile	75^{th} percentile	Median	25^{th} percentile	5^{th} 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

We begin to test the original *PIN* model as null against Model 1. Under the 5% confidence level, 63% of randomly selected firm-years reject the original *PIN* model. Having rejected the original model in favor of Model

1, we test Model 1 as null against Model 2. Likewise, we reject the null model with high probability until test 4 in Panel A of <Table 3>. Finally, we examine Model 4 and 5 as the last test. DY conclude that it is not high enough to reject Model 4 in the last test since it is rejected with the probability of 9%. However, in the Korean market, Model 4 is rejected with the probability of 24% under the 5% confidence level, which is high enough to reject it. Consequently, we choose Model 5 as a preferred model that best fits the Korean stock market.

Panel B of <Table 3> shows the estimation results of the adjusted *PIN* model 5 (Model 5). Not surprisingly, since the overall buys and sells in the Korean market are lower than those in the US markets, u_b, u_s, ϵ_b, ϵ_s, Δ_b, and Δ_s are relative lower than those of the US markets.

To summarize, we verify that the original *PIN* model does not fit well in the Korean stock market and encounters the same problems that DY document in the U.S. stock market. Hence, throughout the series of fitness test among various versions of extension, we find that the unrestricted version of the adjusted *PIN* model fits best. This model appears to solve the problems of the original model and generates the moments that are quite close to those from the Korean real-world data.

References

Asquith, P., R. Oman, and C. Safaya, "Short Sales and Trade Classification Algorithms," *Journal of Financial Markets*, Vol. 13, No. 1 (2010), pp. 157-173.

Chakrabarty, B., P. C. Moulton, and A. Shkilko, "Short Sales, Long Sales, and the Lee-Ready Trade Classification Algorithm Revisited," *Journal of Financial Markets*, Vol. 15, No. 4 (2012), pp. 467-491.

Choe, H. and C.-W. Yang, "Information Risk and Asset Returns in the Korean Stock Market," *Asia-Pacific Journal of Financial Studies*, Vol. 36, No. 4 (2007), pp. 567-620.

Duarte, J. and L. Young, "Why is PIN Priced?," *Journal of Financial Economics*, Vol. 91, No. 2 (2009), pp. 119-138.

Easley, D., N. M. Kiefer, M. O'Hara, and J. Paperman, "Liquidity, Information, and Infrequently Traded Stocks," *Journal of Finance*, Vol. 51, No. 4 (1996), pp. 1405-1436.

Easley, D. and M. O'Hara, "Time and the Process of Security Price Adjustment," *Journal of Finance*, Vol. 47, No. 2 (1992), pp. 577-605.

Ellis, K., R. Michaely, and M. O'Hara, "The Accuracy of Trade Classification Rules: Evidence from Nasdaq," *Journal of Financial and Quantitative Analysis*, Vol. 35, No. 4 (2000), pp. 529-551.

Hwang, L.-S., W.-J. Lee, S.-Y. Lim, and K.-H. Park, "Does Information Risk Affect the Implied Cost of Equity Capital? An Analysis of PIN and Adjusted PIN," *Journal of Accounting and Economics*, Vol 55, No. 2-3 (2013), pp. 148-167.

Lai, S., L. Ng, and B. Zhang, "Does PIN Affect Equity Prices around the World?," *Journal of Financial Economics*, Vol. 114, No. 1 (2014), pp. 178-195.

Lee, C. M. C. and M. J. Ready, "Inferring Trade Direction from Intraday Data," *Journal of Finance*, Vol. 46, No. 2 (1991), pp. 733-746.

Odders-White, E. R., "On the Occurrence and Consequences of Inaccurate Trade Classification," *Journal of Financial Market*, Vol. 3, No. 3 (2000), pp. 259-286.

Park, J.-H. and K. S. Eom, "The Usefulness of PIN as a Measurement for Private-Information Risk in Korean Stock Markets," *Asia-Pacific Journal of Financial Studies*, Vol. 37, No. 3 (2008), pp. 501-536 (in Korean).

박종호, 엄경식, "한국주식시장에서 사정 정보위험 PIN의 유용성 분석: 주식수익률에 미치는 영향을 중심으로," 증권학회지, 제37권 제3호(2008), pp. 501-536.