Journal of Banking & Finance 43 (2014) 137–149

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

Quality of PIN estimates and the PIN-return relationship

Yuxing Yan[a,1], Shaojun Zhang[b,*]

[a] Department of Economics and Finance, Canisius College, 2001 Main Street, Buffalo, NY 14208, United States
[b] School of Accounting and Finance, Faculty of Business, Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 31 October 2012
Accepted 5 March 2014
Available online 20 March 2014

JEL classification:
C13
C61
G12
G14

Keywords:
Probability of informed trading
PIN-return relationship
Quality of PIN estimates

ABSTRACT

This paper provides new evidence concerning the probability of informed trading (PIN) and the PIN-return relationship. We take measures to overcome known estimation biases and improve the quality of quarterly PIN estimates. We use the average of a firm's PIN estimates in four consecutive quarters to smooth out the effect of seasonal variation in trading activities. We find that when high-quality PIN estimates are used, the Fama–MacBeth cross-sectional regressions show stronger evidence for the positive PIN-return relationship than documented in the prior literature. This finding is robust to controls for the January, liquidity, and momentum effects.

1. Introduction

A vibrant finance and accounting literature attempts to understand whether and how information asymmetry between investors influences asset prices in financial markets. In an influential paper, Easley et al. (2002) use the probability of informed trading (PIN) measure to quantify the degree of information asymmetry and document a significant positive relationship between the PIN measure and stock returns between 1983 and 1998. The positive PIN-return relationship has been widely cited in many studies.[2]

However, some researchers doubt the existence of a significant cross-sectional PIN-return relationship for several reasons. First, the relationship is ambiguous in theory. Easley and O'Hara (2004) demonstrate that, in a finite economy, uninformed investors demand a risk premium for holding stocks of greater information asymmetry because they cannot diversify the risk of trading against informed investors. On the other hand, Hughes et al. (2007) show that, if taking a large economy limit and holding the total information constant, greater information asymmetry in the aggregate information environment leads to higher market cost of capital, but firm-specific information characteristics do not affect individual firm expected return after controlling for systematic factors.

Second, the PIN measure must be estimated via a maximum likelihood approach using the daily number of buyer-initiated and seller-initiated trades. A few studies point out potential biases that arise in the estimation of PIN. Boehmer et al. (2007) find that misclassification of buyer-initiated and seller-initiated trades leads to a downward bias in PIN estimates. Lin and Ke (2011) prove that the mathematical transformation used by Easley et al. (2002, 2010) to simplify the joint likelihood function generates a bias because of computer floating-point exception. Yan and Zhang (2012) show that boundary solutions can be another source of bias. Duarte and Young (2009) argue that the microstructure model used by Easley et al. (2002) to describe the trading process does not distinguish information asymmetry from illiquidity.[3]

Third, several studies find that empirical evidence on the PIN-return relationship is not robust. Mohanram and Rajgopal (2009) report that the PIN-return relationship is significantly positive only in the period 1984–1988. It is insignificant in two periods, 1989–1993 and 1994–1998, and is negative in the period 1999–2002. Kang (2011) presents evidence for a January effect, that is, PIN and

* Corresponding author. Tel.: +852 3400 3458; fax: +852 2330 9845.
E-mail addresses: yany@canisius.edu (Y. Yan), afszhang@polyu.edu.hk (S. Zhang).
1 Tel.: +1 (716) 888 2604.
2 See, e.g., Francis et al. (2004, 2005), Odders-White and Ready (2006), Chen et al. (2007), Duarte et al. (2008), Brockman and Yan (2009), and Chen and Zhao (2012).

3 Duarte and Young (2009) use an extended model to propose another measure of information asymmetry and call it the adjusted PIN. Because many studies have used the same PIN measure as Easley et al. (2002), our focus is to examine whether the quality of PIN estimates affects the PIN-return relationship.

http://dx.doi.org/10.1016/j.jbankfin.2014.03.006
0378-4266/

stock returns are negatively related in January, but positively related in other months.

This paper provides new evidence on the PIN-return relationship using high-quality PIN estimates for over 170,000 stock-quarter pairs in the 22 years between 1983 and 2004.[4] We apply several methods that are developed to overcome known estimation biases in order to improve the quality of quarterly PIN estimates. We implement the Fama and MacBeth (1973) regression methodology over the period of 276 months between April 1983 and March 2005. It is evident that the PIN-return relationship is significantly positive over the whole period. Moreover, we examine the same four sub-periods, consistent with Mohanram and Rajgopal (2009): 1984–1988, 1989–1993, 1994–1998, and 1999–2002. Contrary to their observation that the coefficient of PIN is only significant in the earliest sub-period 1984–1988, we find that it is significantly positive in both 1984–1988 and 1994–1998, and its magnitude is larger than that found by Mohanram and Rajgopal (2009). This finding is robust after we adjust for time-varying precision in monthly regression estimates with the Litzenberger and Ramaswamy (1979) weighted least-square method, exclude January observations from the analysis, and control for the liquidity and momentum effects in the cross-sectional regressions.

We contribute to the literature in three ways in addition to documenting new evidence on the PIN-return relationship. First, the Lee and Ready (1991) classification algorithm with a five-second time adjustment has been commonly applied to estimate PIN in previous studies (see, e.g., Easley et al., 2002, 2010; Duarte and Young, 2009; Brown et al., 2004; Yan and Zhang, 2012). This paper is the first to document empirical evidence that the five-second time adjustment causes a systematic bias in PIN estimates for a substantial number of actively traded stocks in the years after 2000.

Second, Boehmer et al. (2007) show that trade misclassification can result in a *downward* bias in PIN estimates. They study a microstructure model in which the arrival rates of buy and sell trades are assumed to be the same. Previous studies often use another model that allows the two arrival rates to be different (see, e.g., Easley et al., 2002, 2010; Duarte et al., 2008; Brown et al., 2004; Yan and Zhang, 2012). Our simulations demonstrate that trade misclassification may result in an *upward* bias under both models.

At last, we observe a distinct seasonal pattern, that is, the PIN estimates on average tend to decrease in the first quarter of a year relative to the previous quarter. Our preliminary analysis suggests that this seasonal pattern is related to tax-loss selling activities at year end. This finding prompts us to use the average of a firm's PIN estimates in four consecutive quarters to smooth out the effect of seasonal variation in trading activities.

The remainder of this paper is organized as follows. Section 2 reviews the estimation of PIN and trade classification. Section 3 compares the two sets of quarterly PIN estimates that are obtained with two different trade classification methods, reports our findings from a simulation study, and presents a preliminary analysis of seasonal variation in the probability of informed trading. Section 4 reports empirical evidence on the PIN-return relationship and its robustness. Section 5 concludes the paper.

2. Trade classification and the estimation of PIN

2.1. The estimation of PIN

The PIN measure of information asymmetry is derived from the market microstructure model proposed in Easley and O'Hara (1992) and Easley et al. (1997). Mathematically, the model specifies that on any day i, the likelihood of observing the number of buy trades B_i and the number of sell trades S_i is represented by

$$L(\theta|B_i,S_i)=\alpha(1-\delta)e^{-(\mu+\varepsilon_b)}\frac{(\mu+\varepsilon_b)^{B_i}}{B_i!}e^{-\varepsilon_s}\frac{\varepsilon_s^{S_i}}{S_i!}+\alpha\delta e^{-\varepsilon_b}\times\frac{\varepsilon_b^{B_i}}{B_i!}e^{-(\mu+\varepsilon_s)}\frac{(\mu+\varepsilon_s)^{S_i}}{S_i!}+(1-\alpha)e^{-\varepsilon_b}\frac{\varepsilon_b^{B_i}}{B_i!}e^{-\varepsilon_s}\frac{\varepsilon_s^{S_i}}{S_i!} \quad (1)$$

where $\theta = (\alpha, \delta, \mu, \varepsilon_b, \varepsilon_s)$ represents five structural parameters that describe the trading process in each day. Specifically, α denotes the probability that an information event occurs. If an information event occurs, it can be bad news with the probability δ or good news with the probability $1-\delta$, and informed traders who know the quality of new information submit orders at the daily arrival rate μ. Informed traders would buy at the rate μ if it is good news, and sell at the same rate μ if it is bad news. No matter whether an information event occurs or not, uninformed traders submit buy orders at the daily arrival rate ε_b and sell orders at the daily arrival rate ε_s.

Assuming independence between days, the joint likelihood of observing a series of daily buys and sells over trading days $i = 1, \ldots, I$ is the product of the daily likelihoods,

$$L(\theta|M)=\prod_{i=1}^{I}L(\theta|B_i,S_i) \quad (2)$$

where $M = ((B_1, S_1), \ldots, (B_I, S_I))$ represents the data set. The PIN measure of information asymmetry is defined as

$$\mathrm{PIN}=\frac{\alpha\mu}{\alpha\mu+\varepsilon_B+\varepsilon_S}. \quad (3)$$

Intuitively, PIN equals the fraction of trades in a day that arise from informed trading.

Maximizing the joint likelihood in Eq. (2) over the parameters in θ produces the maximum likelihood estimates of these structural parameters. There is no closed form solution to this maximization problem. A numerical maximization technique must be used to obtain a solution. Easley et al. (2010) use the following factorization of the joint likelihood function to facilitate numerical maximization

$$L((B_i,S_i)_{i=1}^{I}|\theta)=\sum_{i=1}^{I}[-\varepsilon_b-\varepsilon_s+M_i(\ln x_b+\ln x_s)+B_i\ln(\mu+\varepsilon_b)+S_i\ln(\mu+\varepsilon_s)]+\sum_{i=1}^{I}\ln[\alpha(1-\delta)e^{-\mu}x_x^{S_i-M_i}x_b^{-M_i}+\alpha\delta e^{-\mu}x_b^{B_i-M_i}x_s^{-M_i}+(1-\alpha)x_s^{S_i-M_i}x_b^{B_i-M_i}] \quad (4)$$

where $M_i=\min(B_i,S_i)+\max(B_i,S_i)/2, x_s=\frac{\varepsilon_s}{\mu+\varepsilon_s}$, and $x_b=\frac{\varepsilon_b}{\mu+\varepsilon_b}$. Here, $\min(B_i, S_i)$ represents the smaller of B_i and S_i, and $\max(B_i, S_i)$ represents the larger one.

Lin and Ke (2011) point out that when the above factorized likelihood function is used, floating-point exception in computer software narrows the set of feasible solutions, which causes a downward bias in the estimate of PIN. In order to avoid the influence of floating-point exception, they recommend the following factorization of the joint likelihood function

$$L((B_i,S_i)_{i=1}^{I}|\theta)=\sum_{i=1}^{I}[-\varepsilon_b-\varepsilon_s+B_i\ln(\mu+\varepsilon_b)+S_i\ln(\mu+\varepsilon_s)+e_{\max i}]+\sum_{i=1}^{I}\ln[\alpha(1-\delta)\exp(e_{1i}-e_{\max i})+\alpha\delta\exp(e_{2i}-e_{\max i})+(1-\alpha)\exp(e_{3i}-e_{\max i})] \quad (5)$$

[4] We do not extend the estimates beyond 2004 for two reasons. First, the implementation of Regulation NMS in 2005 had substantial impact on trading activities, which increases the difficulty in reliably classifying buyer-initiated and seller-initiated trades. Second, it is financially costly for us to gain access to the intraday trade and quote data and to use a powerful computing platform. Our PIN estimates are available upon request.