Негосударственное образовательное учреждение
высшего профессионального образования
«Российская экономическая школа» (институт)

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

по направлению подготовки 080100.68
Программа «Магистр экономики»

MASTER THESIS

Тема: *Насколько точны меры асимметрии информации, основанные на секвенциальных торговых моделях?*

Title: *How accurate do sequential trading models measure information asymmetry?*

Студент/ Student:

Александр Сергеевич Копытов
(Ф.И.О. полностью студента, выполнившего работу)
Alexandr Kopytov
(Name of the student)

Научный руководитель/ Advisor:

PhD, Профессор финансов, Российская Экономическая Школа, Патрик Дж. Келли
(Ученая степень, звание, место работы, Ф.И.О.)
Patrick J. Kelly, PhD, Associate Professor of Finance, New Economic School
(Name, scientific degree, scientific rank, place of work)

Москва 2014

How accurate do sequential trading models measure information asymmetry?

Alexandr Kopytov*

New Economic School

May, 2014

Abstract

Sequential trading models are used to measure information asymmetry basing on order flow imbalance. In this paper I compare the basic PIN(probability of informed trading) model of Easley, Kiefer, O'Hara, and Paperman (1996) and its extension proposed by Duarte and Young (2009). I verify that model implied information asymmetry proxies, PIN and $AdjPIN$(adjusted probability of informed trading), respectively, are related to stocks with high information asymmetry measures. I find some evidence that information asymmetry proxy implied by the extended PIN model is downward biased. I then propose the accuracy test of sequential trading models which allows one to investigate whether models are able to identify information based trading by order flow imbalance correctly. I show that information asymmetry proxies implied by both models capture not only information asymmetry, and the basic PIN model is likely to be less precise than its extended version.

*I would like to thank my research advisors Dmitry Makarov and Oleg Shibanov for their helpful directions and comments. Valuable comments were provided by several students of New Economic School, especially by Vasily Afonin and Elena Zherko. Special thanks go to my research advisor Patrick J. Kelly for his invaluable help, wise guiding and generous sharing of data and data manuals. All mistakes are mine.

1 Introduction

Classic asset pricing models (for example, CAPM) describe the impact of systematic risks on asset prices while the process of price formation is left out. That is to say, microstructural effects on asset prices are not taken into consideration. However, in recent years it was demonstrated that some aspects of trading process may affect asset prices.

One of these aspects is information asymmetry. Classic asset pricing models usually assume that all information is already incorporated into asset prices as a result of investors having common beliefs. At the same time, in real markets new information about assets arrives constantly. It seems reasonable to assume that investors may have access to different information. Therefore, their beliefs about stock returns may not be common. Easley and O'Hara (2004) show that in a rational expectation model with two types of unequally informed investors information asymmetry affect asset prices. In their model uninformed investors are rational agents who demand premia for bearing nondiversifiable risk of buying high information asymmetry stocks.

On the other hand, Hughes, Liu, and Liu (2007) and Lambert, Leuz, and Verrecchia (2007) argue that under assumptions of correlated asset returns and large economy with infinite number of securities and investors information asymmetry is not priced. They claim that information asymmetry risk is idiosyncratic and, hence, is diversifiable. However, if the economy is not infinitely large, the conclusions of these two papers coincide with the result of Easley and O'Hara (2004), because idiosyncratic information asymmetry risk can not be fully eliminated by diversification.

Empirical evidence about information asymmetry impact on asset returns is controversial too. One of the main reasons is that there is no direct measures of information asymmetry. In recent years several order flow imbalance based proxies of information asymmetry were proposed. Typically, these measures are based on the assumption that periods of private information trading can be identified by some parts of order flow imbalance. However, in the literature there is no consensus which part of order flow imbalance correspond to information based trading.

In this paper I compare two concurrent sequential trading models which differ in their assumptions about sources of order flow imbalance. In particular, I investigate famous and widely used basic PIN (probability of informed trading) model of Easley, Kiefer, O'Hara, and Paperman (1996) and the extended PIN model of Duarte and Young (2009). The basic PIN model empirically confirms that information asymmetry is priced. As shown in Easley, Hvidkjaer, and O'Hara (2002), PIN is positively and significantly related to asset returns. Moreover, Easley, Hvidkjaer, and O'Hara (2010) show that PIN factor plays significant role in the presence of three Fama-French factors (see Fama and French, 1993), as well as momentum and liquidity factors. At the same time, Duarte and Young (2009) claim that PIN captures not only information asymmetry but illiquidity as well. They propose a more sophisticated extended PIN model where information asymmetry measure, $AdjPIN$ (adjusted probability of informed trading), is cleaned from irrelevant illiquidity part. Finally, they find that $AdjPIN$ is not priced.

In recent years PIN has been widely used as a proxy for information asymmetry. However, if the extended PIN model is correct then PIN might be a bad measure of information asymmetry. Therefore, it seems interesting to investigate which information asymmetry proxy, PIN or $AdjPIN$, is more precise.

I study my research question in two steps. Firstly, I verify that model implied information asymmetry proxies indeed measure information asymmetry[1]. However, I find some evidence that information asymmetry proxy implied by the extended PIN model is downward biased. Second, I show that information asymmetry proxies implied by both models capture not only information asymmetry, and the basic PIN model is likely to be less precise than its extended version. Particu-

[1]For the basic PIN model this is already done in Aslan, Easley, Hvidkjaer, and O'Hara (2011)

larly, I find that both investigated models (extended one to a less degree) tend to assign order flow imbalance to informed trading periods incorrectly. Before I explain how I arrive at this conclusions, I briefly describe both investigated models.

The basic PIN model assumes that all trades are initiated either by informed investors or by liquidity traders. Each day informed investors initiate buy/sell trade if they receive positive/negative private signal about a stock. Orders from liquidity, or uninformed investors arrive at constant rate, regardless of the day. The only source of order flow imbalance in this model is informed trading. Large number of sell and buy orders occur in different days because each day with new private information informed investors either initiate only buy or only sell orders. Therefore, according to the basic PIN model, correlation between buys and sells is negative.

However, Duarte and Young (2009) note that in the real markets this correlation is significantly positive. Taking this into consideration, they propose the extended PIN model where they allow for symmetric order flow, or liquidity shocks. Around such shocks liquidity traders increase rates of both buy and sell orders. This extension implies two sources of order flow imbalance: trades initiated by informed investors and trades initiated by liquidity investors during the days of symmetric order flow shocks.

As can be seen from the above, the main difference between basic and extended PIN models is symmetric order flow shocks. Duarte and Young (2009) argue that the probability that a trade caused by such a shock, $PSOS$ (probability of symmetric order flow shock), is a proxy for illiquidity and is not connected with information asymmetry. However, the nature of symmetric order flow shocks is not clear. According to Duarte and Young (2009), these shocks may be caused by at least two reasons. The first one is public news event which may lead to disagreement between investors and, hence, to increase of both sell and order flow rates. The second one is investors' coordination to trade on certain days to reduce trading costs. At the same time, Aslan, Easley, Hvidkjaer, and O'Hara

(2011) emphasize that even public news may lead to increase of information asymmetry. Indeed, Kim and Verrecchia (1994) theoretically and Lee, Mucklow, and Ready (1993) empirically show that information asymmetry increase around such public news events as earning announcements. Possibly, some investors are able to do better analysis of public news and acquire more precise information. Different interpretation of the same event may lead to increase of information asymmetry among investors. The source of this new asymmetry is not pure private signals but private analysis of public signals. In this case, symmetric order flow shocks might lead to additional information asymmetry.

It does not mean, however, that the model of Duarte and Young (2009) is absolutely incorrect. Symmetric order flow shocks might lead to increase of rates of buy and sell orders but at least partly this increase is caused by informed investors. In this case, $PSOS$ indeed reflects the information asymmetry to some extent. $AdjPIN$, in its turn, is a downward biased estimator of the true probability of informed trading.

In the first part of this paper I investigate the relationship between PIN, $AdjPIN$, $PSOS$ and different measures of information asymmetry, previously used in the literature. Controlling for illiquidity proxies, I show that not only PIN and $AdjPIN$, but also $PSOS$ are significantly connected with information asymmetry measures. As noted above, the latter result implies that $AdjPIN$ is a downward biased measure of information asymmetry. At the same time, I also find that PIN, $AdjPIN$ and $PSOS$ are all significantly connected to illiquidity. This result gives some evidence that both PIN and $AdjPIN$ are likely to capture not only information asymmetry.

The second part of the paper is devoted to the accuracy tests of basic and extended PIN models. The idea of the accuracy test is following. Sequential trading models assign order flow imbalance to periods of informed trading or to liquidity shock days (in case of the extended PIN model). Put it otherwise, models classify market events leading to order flow shocks based on the value of order flow imbalance. For example, the basic PIN model classifies all order flow shocks into good or bad private information events. The extended PIN model also allows for liquidity events. However, this classification is based on the number of models' assumptions about sources of order flow imbalance and behavior of investors. It these assumptions do not hold then the classification may be incorrect and some market events may be imprecisely described by the models. Therefore, having a sample of previously classified market events (based on some external classification mechanism), one can test whether models' implied classification coincide with the external one. If two classifications do not coincide then it is a sign that sequential trading models are likely to assign order flow imbalance to period of informed trading imprecisely. In this case models implied information asymmetry proxies capture not only information asymmetry.

I conduct the accuracy test using the sample of quarter earning announcements. As mentioned above, quarter earning announcements are good examples of public events that also are likely to increase information asymmetry. I show that model implied abnormal probabilities of good/bad private signals are significantly positive around not only 'good'/'bad' [2] quarter earning announcements, but also around other types of earning announcements. I conclude that some events are incorrectly identified by both models as private information events (although the extended PIN model performs marginally better). Therefore, there is some evidence that model implied information asymmetry measures capture not only information asymmetry.

In this paper I demonstrate that both basic and extended PIN models do capture information asymmetry. Hence, order flow imbalance is connected to informed trading. However, neither widely used basic PIN nor extended PIN model precisely identify parts of order flow imbalance corresponding to informed trading. Hence, one should be careful when interpreting PIN and $AdjPIN$ as information asymmetry proxies.

[2] This classification is based on the sign of earning surprise and is quite common in the literature. See, for example, Lee (1992) and Skinner and Sloan (2002)

The remainder of the paper is organized as follows. Section 2 gives a brief overview of basic PIN and extended PIN models and provides details of PIN, $AdjPIN$ and $PSOS$ estimation. Section 3 describes the data and the sample, discusses the explanatory variables and presents summary statistics. Section 4 investigates the relationship between PIN, $AdjPIN$, $PSOS$ and the explanatory variables. Section 5 presents the results of the accuracy tests. Section 6 concludes.

2 Basic and extended PIN model

2.1 Overview of the basic PIN model

In this section I give a brief overview of the basic PIN model proposed by Easley, Kiefer, O'Hara, and Paperman (1996). According to the model, there are two types of investors. Informed investors trade based on their private signals. Uninformed, or liquidity investors trade for some exogenous reasons. The uninformed liquidity provider sets the bid and ask quotes basing on the number of buy and sell orders. The bid-ask spread is set by the liquidity provider to compensate the risk of trading against informed investors.

The trade process is organized as depicted in the Figure 1. At the beginning of each day the nature decides whether there is new private information. The probability of information events is a. There are two types of signals: high (with probability d) and low. Informed investors initiate buy/sell trades only if they receive high/low signals, otherwise they do not trade at all. Uninformed investors trade regardless of the signals as they are assumed to trade because of some exogenous reasons. Therefore, on the high private signal days buy orders arrive at rate $\mu + \epsilon_b$ and sell orders arrive at rate ϵ_s, where μ is the rate of informed investors' orders, ϵ_b/ϵ_s is the arrival rate of uninformed investors' buy/sell orders. Similarly, on the low private signal days buy orders arrive at rate ϵ_b and sell orders arrive at rate $\mu + \epsilon_s$. On the no information days the buy/sell order flow rate is given by

ϵ_b/ϵ_s. All the order flows are assumed to be independent and have Poisson distribution.

According to the model, expected total order flow rate is given by $a\mu+\epsilon_s+\epsilon_b$ and the information based order flow rate is $a\mu$. Hence, the unconditional probability of a trade to be initiated by an informed investor (PIN) is given by the ratio of the information based order flow rate and the total rate.

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

2.2 Estimation of the basic PIN model

The PIN model can be estimated numerically via Maximum Likelihood. As the buy and sell orders are assumed to arrive according to independent Poisson processes then for each firm year likelihood function can be written as follows.

$$\begin{aligned} L(a, d, \mu, \epsilon_b, \epsilon_s | B, S) = \prod_{i=1}^{N} \Bigg[&(1-a)e^{-\epsilon_b-\epsilon_s} \frac{\epsilon_b^{B_i} \epsilon_s^{S_i}}{S_i! B_i!} + a(1-d)e^{-\epsilon_b-\epsilon_s-\mu} \frac{\epsilon_b^{B_i} (\epsilon_s + \mu)^{S_i}}{S_i! B_i!} + \\ &ade^{-\epsilon_b-\epsilon_s-\mu} \frac{(\epsilon_b + \mu)^{B_i} \epsilon_s^{S_i}}{S_i! B_i!} \Bigg] \end{aligned} \tag{2}$$

where B_i and S_i are the numbers of buyer-initiated and seller-initiated trades at day i; a, d, μ, ϵ_s, ϵ_b are the parameters of estimation, N is the number of trading days in a calendar year.

The direct estimation of this likelihood function is almost impossible, especially for actively traded securities. When B_i of S_i are sufficiently large, calculation of such terms as $B_i!$ or $\epsilon_s^{S_i}$ leads to numerical overflows. Therefore, I use the approach suggested by Easley, Hvidkjaer, and O'Hara (2010) and estimate the following log-likelihood function, which can be straightforwardly derived from Equation (2):

$$\hat{L}(a,d,\mu,\epsilon_b,\epsilon_s|B,S) = \sum_{i=1}^{N}\Bigg\{-\epsilon_b-\epsilon_s+B_i\log(\epsilon_b+\mu)+S_i\log(\epsilon_s+\mu)+M_i\log x_s+M_i\log x_b+ \log\Bigg[\exp[\log(1-a)+(B_i-M_i)\log x_b+(S_i-M_i)\log x_s]+ \exp[\log a(1-d)-\mu+(B_i-M_i)\log x_b-M_i\log x_s]+ \exp[\log ad-\mu+(S_i-M_i)\log x_s-M_i\log x_b]\Bigg]\Bigg\} \tag{3}$$

where $x_s = \dfrac{\epsilon_s}{\epsilon_s+\mu}$, $x_b = \dfrac{\epsilon_b}{\epsilon_b+\mu}$, $M_i = \xi \cdot \min(B_i, S_i)$. The factoring of $x_s^{M_i}$ and $x_b^{M_i}$ is done to prevent direct calculation of $x_s^{S_i} \cdot x_b^{B_i}$. This factoring is especially important for actively traded securities with large S_i and B_i. Parameter ξ is chosen to reduce the number of stocks with extreme values of log likelihood function summands. For each firm year I try $\xi = 1/4, 1/3, 1/2, 1$. Such flexibility in ξ helps to reduce the number of computational overflows.

2.3 Overview of the extended PIN model

In this section I briefly describe the extension of the PIN model proposed by Duarte and Young (2009). Similarly to the basic PIN model, there are two types of investors and uninformed liquidity provider. The main difference from the initial model is symmetric order flow shocks. The trade process is organized as depicted in the Figure 2. At the beginning of each day the nature not only decides whether there is new information (with probability of a) but also decides if there is symmetric order flow shock (with probability of θ). These decisions are assumed to be independent. On no symmetric order flow shocks day all the order arrival rates coincide with corresponding rates of the basic PIN model. Symmetric order flow shock increase both buy and sell order flows of uninformed investors by Δ_b and Δ_s, respectively.

The extended PIN model allows one to distinguish between impact of symmetric order flow shocks and informed trading on order flow. The total order flow rate is $a(d\mu_b + (1-d)\mu_s) + (\Delta_b + \Delta_s)(a + \theta + (1-a)\theta) + \epsilon_s + \epsilon_b$, the information based order flow rate is $a(d\mu_b + (1-d)\mu_s)$ and the order flow caused by symmetric order flow shocks has the rate of $(\Delta_b + \Delta_s)(a\theta + (1-a)\theta)$. Therefore, according to the extended PIN model, the probability of a trade to be initiated by informed investor ($AdjPIN$) is given by the ratio of the information based order flow rate and the total rate:

$$AdjPIN = \frac{a(d\mu_b + (1-d)\mu_s)}{a(d\mu_b + (1-d)\mu_s) + (\Delta_b + \Delta_s)\theta + \epsilon_s + \epsilon_b} \tag{4}$$

Similarly, the probability that a trade is caused by symmetric order flow shock ($PSOS$) is given by the ratio of the order flow rate caused by the symmetric order flow shock and the total rate:

$$PSOS = \frac{(\Delta_b + \Delta_s)\theta}{a(d\mu_b + (1-d)\mu_s) + (\Delta_b + \Delta_s)\theta + \epsilon_s + \epsilon_b} \tag{5}$$

2.4 Estimation of the extended PIN model

Similarly to the basic PIN model, its extension can be estimated via Maximum Likelihood. Again, all the buy and sell order flows are assumed independent Poisson processes. As in Duarte and Young (2009), for each firm year likelihood function can be written as follows.

$$\begin{aligned} L(a, d, \theta, \mu_b, \mu_s, \Delta_s, \Delta_b, \epsilon_b, \epsilon_s | B, S) = \prod_{i=1}^{N} \Bigg\{ & (1-a)(1-\theta)e^{-\epsilon_b}\frac{\epsilon_b^{B_i}}{B_i!}e^{-\epsilon_s}\frac{\epsilon_s^{S_i}}{S_i!} + \\ & (1-a)\theta e^{-\epsilon_b-\Delta_b}\frac{(\epsilon_b+\Delta_b)^{B_i}}{B_i!}e^{-\epsilon_s-\Delta_s}\frac{(\epsilon_s+\Delta_s)^{S_i}}{S_i!} + \\ & a(1-\theta)(1-d)e^{-\epsilon_b}\frac{\epsilon_b^{B_i}}{B_i!}e^{-\epsilon_s-\mu_s}\frac{(\epsilon_s+\mu_s)^{S_i}}{S_i!} + \\ & a(1-\theta)de^{-\epsilon_b-\mu_b}\frac{(\epsilon_b+\mu_b)^{B_i}}{B_i!}e^{-\epsilon_s}\frac{\epsilon_s^{S_i}}{S_i!} + \\ & a\theta(1-d)e^{-\epsilon_b-\Delta_b}\frac{(\epsilon_b+\Delta_b)^{B_i}}{B_i!}e^{-\epsilon_s-\mu_s-\Delta_s}\frac{(\epsilon_s+\mu_s+\Delta_s)^{S_i}}{S_i!} + \\ & a\theta de^{-\epsilon_b-\mu_b-\Delta_b}\frac{(\epsilon_b+\mu_b+\Delta_b)^{B_i}}{B_i!}e^{-\epsilon_s-\Delta_s}\frac{(\epsilon_s+\Delta_s)^{S_i}}{S_i!} \Bigg\} \end{aligned} \tag{6}$$

where B_i and S_i are the numbers of buyer-initiated and seller-initiated trades at day i; a, d, θ, μ_s, μ_b, Δ_s, Δ_b, ϵ_s, ϵ_b are the parameters of estimation, N is the number of trading days in a calendar year.

For the same reasons as for the basic PIN model, the idea of direct estimation of this likelihood function does not seem reasonable. Therefore, I use the approach of Easley, Hvidkjaer, and O'Hara (2010) again. After factoring, dropping some constant terms and rearranging, the log likelihood function can be written as follows.

$$
\begin{aligned}
&\hat{L}(a, d, \theta, \mu_b, \mu_s, \Delta_s, \Delta_b, \epsilon_b, \epsilon_s | B, S) = \\
&\qquad \sum_{i=1}^{N} \Bigg\{ -\epsilon_s - \epsilon_b + B_i \log(\epsilon_b + \mu_b + \Delta_b) + S_i \log(\epsilon_s + \mu_s + \Delta_s) + M_i \log x_b x_s + \\
&\qquad \log \Bigg[\exp[\log(1-a)(1-\theta) + (B_i - M_i) \log x_b + (S_i - M_i) \log x_s] + \\
&\qquad \exp[\log(1-a)\theta - \Delta_b - \Delta_s + B_i \log y_b - M_i \log x_b + S_i \log y_s - M_i \log x_s] + \\
&\qquad \exp[\log a(1-\theta)(1-d) - \mu_s + (B_i - M_i) \log x_b + S_i \log z_s - M_i \log x_s] + \\
&\qquad \exp[\log a(1-\theta)d - \mu_b + B_i \log z_b - M_i \log x_b + (S_i - M_i) \log x_s] + \\
&\qquad \exp[\log a\theta d - \mu_b - \Delta_b - \Delta_s - M_i \log x_b + S_i \log y_s - M_i \log x_s] + \\
&\qquad \exp[\log a\theta(1-d) - \Delta_b - \mu_s - \Delta_s + B_i \log y_b - M_i \log x_b - M_i \log x_s] \Bigg] \Bigg\}
\end{aligned}
\tag{7}
$$

where $x_j = \dfrac{\epsilon_j}{\epsilon_j + \mu_j + \Delta_j}$, $y_j = \dfrac{\epsilon_j + \Delta_j}{\epsilon_j + \mu_j + \Delta_j}$, $z_j = \dfrac{\epsilon_j + \mu_j}{\epsilon_j + \mu_j + \Delta_j}$, $j = b, s$. $M_i = \xi \cdot \min(B_i, S_i)$. Again, for each firm year I try $\xi = 1/4, 1/3, 1/2, 1$. The factoring is done for the same reasons as in the PIN model.

3 Data

Following Easley, Hvidkjaer, and O'Hara (2002), in this paper I use the sample of common New York Stock Exchange (NYSE) and American Exchange (AMEX) stocks excluding those incorporated outside the US, real estate investment funds, ADRs, closed-end funds, for the years from 1993 through 2001. I also exclude stocks with less than 60 days with trade data in a given year. The final sample includes between 1,464 and 1,907 yearly observations and totally 15,443 observations. In the following section I describe explanatory variables I use in my regressions.

3.1 Explanatory variables

In order to investigate whether PIN, $AdjPIN$ and $PSOS$ capture information risk, I use several measures of information environment previously utilized in the literature (see, for example, Verdi (2005) for overview). The description of these measures is below. The data sources are CRSP (trading data), Compustat (accounting data), I/B/E/S (analyst coverage data), Thomson Financial Insider Filing (insider trading data)[3].

1. $Size$ of the firm i is the market value of its equity. Expected correlation between $Size$ and information asymmetry proxies is negative. This measure is widely used in the information risk literature (for example, Verdi, 2005).

2. $Turnover$ is the annual mean of daily turnovers calculated as the ratio of daily share trading volume and number of shares outstanding. The expected connection with information risk is ambiguous. According to Jiang, Lee, and Zhang (2005), $Turnover$ is high for stocks with high information uncertainty. However, Aslan, Easley, Hvidkjaer, and O'Hara (2011) find the negative connection between PIN and $Turnover$.

3. Age is the number of years since a stock was covered by CRSP for the first time. As noted in Barry and Brown (1985), listing period can be used as a proxy for the quantity of information with some limitations. Generally, firms with high Age are supposed to be more attractive to

[3] I thank Patrick Kelly for generous sharing of this data

investors as investors have more information about such firms. Therefore, *Age* is expected to be negatively connected with information risk.

4. *Volatility* is standard deviation of daily returns in year t. This measure is widely used in the information risk literature (see, for example, Jiang, Lee, and Zhang, 2005). High *Volatility* of stock returns may be a signal that investors are uncertain about them. Therefore, expected connection between *Volatility* and information risk is positive.

5. *Coverage* for a firm i is the number of analysts who provide one fiscal year ahead forecasts of this firms' earnings per share. I calculate analyst coverage using I/B/E/S Historical Detailed files. If there is no information about analyst coverage available, I set it equal to zero. The expected connection between *Coverage* and information risk is negative as analysts tend to convert private information into public information (see, for example, Barry and Brown, 1985).

6. *Cash*. Following Aslan, Easley, Hvidkjaer, and O'Hara (2011), I include cash scaled by assets as firms with higher information asymmetry might have problems with raising external funds. Therefore, firms with high information asymmetry are supposed to have higher *Cash*.

7. *Lag* is the annual average number of days between quarter end earnings and quarter earnings report. This variable is used by Aslan, Easley, Hvidkjaer, and O'Hara (2011) to measure the quality of earnings. The expected connection between *Lag* and earnings quality is negative. High earnings quality, in its turn, reduces information risk.

8. $Turnover_{ins}$ is the annual mean of daily insider turnovers calculated as the ratio of daily share volume of open market buy and sell trades made by company higher level insiders and number of shares outstanding. Following Beneish and Vargus (2002), I include transactions made by top five executives (CEO, CFO, COO, president, chairman of the board) as higher level insiders are

likely to possess valuable information[4]. The data is obtained from Thomson Financial Insider Filing Data. If there is no information about insider trading available, I set insider turnover equal to zero. I expect positive connection between information risk and $Turnover_{ins}$.

Duarte and Young (2009) argue that $PSOS$ is a proxy for illiquidity. They show that $PSOS$ is significantly and positively connected with Amihud (2002) measure of illiquidity, $ILLIQ = \frac{1}{N}\sum_{i=1}^{N}\frac{|r_i|}{V_i}$, where r_i is daily return, V_i is daily trading volume and N is the number of trading days in a year. To control for illiquidity, I include $ILLIQ$ in my regressions. I expect that connection between PIN, $PSOS$, $AdjPIN$ and $ILLIQ$ is positive.

3.2 PIN, AdjPIN, PSOS

PIN, $AdjPIN$, $PSOS$ are estimated using the data on number of buyer and seller initiated trades each day for each stock. The data includes buy-sell statistics for 1993-2003 period. However, the final sample includes only data for the years from 1993 through 2001 as for the years 2002 and 2003 I do not manage to get reliable estimates of PIN, $AdjPIN$, $PSOS$ (for around 10% of firms computational algorithm does not converge)[5]. Generally, computational overflows are the most frequent for actively traded stocks of the firms with large market capitalization. Exclusion of these observations is nonrandom and may lead to biased estimates of the regression coefficients. Therefore, I prefer to limit my sample only to 1993-2001 period to avoid possible biases.

3.3 Earning announcement data

For the accuracy tests of the investigated sequential trading models I use sample of quarter earning announcements. The sample covers quarter earning announcements of all the investigated firms for

[4] The results are qualitatively the same if transactions of all insiders are taken into account

[5] Easley, Hvidkjaer, and O'Hara (2010) limit their sample by the year 1999 for the similar reasons. They include years 2000 and 2001 only to do asset pricing tests

1993 through 2001. The data source is Compustat. The sample includes 47,731 announcements.

3.4 Summary statistics and correlations

Summary statistics of the explanatory variables are given in Panel A of Table 1. All the figures are reasonable and standard deviations are rather large. Panel B of Table 1 contains summary statistics of PIN, $AdjPIN$ and $PSOS$. Figure 3 presents the time series of 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles of PIN, $AdjPIN$ and $PSOS$. They are stable over the sample period. $PSOS$ seems to be more variable than PIN and $AdjPIN$. PIN tends to be higher than $AdjPIN$ which is consistent with the extended PIN model as $AdjPIN$ captures the probability of informed trading only partly. $PSOS$ tends to be higher than both PIN and $AdjPIN$. Generally, these results are consistent with those of Duarte and Young (2009).

Panel A of Table 2 presents correlations between the parameters. Following Aslan, Easley, Hvidkjaer, and O'Hara (2011), I use logarithms of $Size$, $Turnover$, $Illiq$, Age, $Volatility$, Lag and $Turnover_{ins}$ to make distributions of the variables more symmetric and leptokurtic. Correlations between the explanatory variables look reasonable. Some correlations are quite large in absolute values; the highest in absolute value correlation is -0.85 (between $Lsize$ and $Lilliq$). This may lead to multicollinearity problem. Panel B of Table 2 contains variance inflation factors (VIFs). Typically, high VIFs (often 10 is suggested as a cutoff value. See, for example, Weisberg, 2005) imply that there is multicollinearity problem. Among the explanatory variables $Lilliq$ has the largest VIF of $6.82 < 10$. This supports that multicollinearity is not that important for these explanatory variables, especially for the measures of information asymmetry. However, for robustness issues I estimate regressions in absence of $Lilliq$ regressor and compare the results with the case of full set of regressors. I discuss the results of the regressions in estimation results subsection below.

Correlation between PIN and $AdjPIN$, as well as between PIN and $PSOS$ is high (0.53 and

0.42, respectively). At the same time, correlation between $AdjPIN$ and $PSOS$ is only 0.05. This is consistent with the hypothesis that $AdjPIN$ and $PSOS$ reflect different aspects of information asymmetry. PIN captures both of these aspects. Correlations between PIN, $AdjPIN$, $PSOS$ and the explanatory variables are generally consistent with my expectations. $Lsize$, $Lage$ and $Lcoverage$ are negatively correlated with PIN, $AdjPIN$, $PSOS$; $Lilliq$, $Lvolatility$, $Cash$, $Llag$ and $Lturnover_{ins}$ are positively correlated with PIN, $AdjPIN$, $PSOS$. $Lturnover$ is correlated negatively with PIN and $AdjPIN$ which is consistent with Aslan, Easley, Hvidkjaer, and O'Hara (2011). However, correlation between $PSOS$ and $Lturnover$ is positive. As noted in Jiang, Lee, and Zhang (2005), high turnover of a stock might be a signal that information uncertainty for this stock is also high. Symmetric order flow shocks caused by public news lead to increase of both buy and sell order flows as public news lead to additional divergence in investors' opinions. So, for high $PSOS$ stocks information uncertainty tends to be high which is reflected in positive correlation between $PSOS$ and $Lturnover$.

4 Relation between PIN, AdjPIN, PSOS and information asymmetry

In this section I investigate the relation between PIN, $AdjPIN$, $PSOS$ and information asymmetry measures. Firstly, I describe my estimation methodology and then discuss the results of estimation.

4.1 Estimation methodology

In this subsection I describe two ways I use to investigate the relation between PIN, $AdjPIN$, $PSOS$ and information asymmetry measures. The sample consists of 15,443 observations for the period 1993-2001. The estimated equation is given below.

$$PIN_{it}/AdjPIN_{it}/PSOS_{it} = \beta_0 + \beta_1 Lsize_{it} + \beta_2 Lturnover_{it} + \beta_3 Lilliq_{it} + \beta_4 Lage_{it} + \beta_5 Lvolatility_{it} + \beta_6 Lcoverage_{it} + \beta_7 Cash_{it} + \beta_8 Llag_{it} + \beta_9 Lturnover_{ins\ it} + \epsilon_{it}, \tag{8}$$

where i and t are firm and time indices, respectively.

The most obvious way of estimation of Equation 8 is simple pooled OLS. Assuming correct specification, no perfect collinearity and exogeneity of errors ϵ_{it}, OLS estimates of the parameters are consistent. Estimation of standard errors is less straightforward. As noted in Cochrane (2005), shocks ϵ_{it} are likely to be correlated both over time and cross sectionally. Hence, non corrected standard errors are likely to be underestimated. To avoid this problem I calculate double clustered standard errors in the sense of Miller, Cameron, and Gelbach (2011). Multi-way clustering technique allows for calculation of standard errors that are robust to correlation of errors within several clusters.

I use time and industry clusters. Time clustered standard errors correct for correlation of shocks over time. Industry clustering corrects for correlation of shocks within one industry. In the literature there is no unified method of assigning stocks to industry. Therefore, for robustness issues I use two industry classifications. They are proposed by Moskowitz and Grinblatt (1999) and Fama and French (1997) and divide all the stocks into 20 and 49 groups, respectively. Both of them are based on Standard Industrial Classification (SIC) codes. The results of pooled OLS estimation and double clustered errors are given in Panel A of Table 3. We can see that standard errors calculated using these two industry classification are very similar.

Although double clustering of standard errors in pooled OLS corrects for correlated shocks, I also estimate Equation 8 using Fama-MacBeth method (see Fama and MacBeth (1973) for details). This method implies two steps. On the first step Equation 8 is estimated for each period of time. The result of the first stage is 10 time series of $\hat{\beta}_{jt},\ j = 0, ..., 9$. The length of each time series is 9 as we have observations over the period from 1993 to 2001. On the second stage each $\hat{\beta}_{jt}$ is regressed on the constant. The estimated constant is simple time series average of corresponding $\hat{\beta}_{jt}$: $\hat{\beta}_j^{FM} = \frac{1}{T}\sum_{t=1}^{T}\hat{\beta}_{jt}$. Finally, standard error of $\hat{\beta}_j^{FM}$ is estimated from the second stage of the Fama-MacBeth procedure.

To correct for serial correlation, I use Newey-West heteroscedasticity and autocorrelation robust errors (Newey and West, 1987). As mentioned in Cochrane (2005), Fama-MacBeth method is robust to cross sectional correlation of shocks. The results of the Fama-MacBeth estimation are given in Panel B of Table 3. We can see that Fama-MacBeth and double clustered standard errors are sometimes quite different. However, significance levels of the regression coefficients estimated by the two methods mostly coincide.

It is worth noting that there is no additional source of error connected with the using of estimated independent variables. Unlike asset pricing models, in this case on the first stage I use directly observable independent variables rather than some preestimated coefficients. Therefore, there should not be additional corrections to Fama-MacBeth standard errors.

4.2 Estimation results

The estimation results are give in Table 3. Panels A and B present the estimates of pooled OLS and Fama-MacBeth regressions, respectively. Generally, the results of both regressions are similar. Following Aslan, Easley, Hvidkjaer, and O'Hara (2011), I focus on the results of Fama-MacBeth regression.

4.2.1 PIN regression

The estimated coefficients of the PIN regression are consistent with the results of Aslan, Easley, Hvidkjaer, and O'Hara (2011). I find that the coefficients of $Lsize$ and $Lturnover$ are insignificant. The expected connection between $Lturnover$ and PIN is negative as stocks with high trading activity are supposed to have stronger order flow from liquidity traders. However, this effect of $Lturnover$ is captured by $Lilliq$ which is positively and significantly connected with PIN (as noted in Panel C of Table 3 in absence of $Lilliq$ the coefficient on $Lturnover$ is significantly negative).

Connection between PIN and different information asymmetry measures confirms that PIN is a proxy for information asymmetry. The coefficient on $Lage$ is highly significant and negative which might signal that liquidity investors prefer not to trade young stocks without enough historical data. The coefficient on $Lvolatility$ is negative but insignificant. The negative sign seems a bit surprising but consistent with Aslan, Easley, Hvidkjaer, and O'Hara (2011). The coefficient on $Lcoverage$ is negative and significant. Indeed, analysts tend to reduce information asymmetry by converting private information into public information. Stocks with active insider trading tend to have high information asymmetry. This is confirmed by significantly positive coefficient of the $Lturnover_{ins}$ coefficient. Finally, low earnings quality measured by $Cash$ and $Llag$ tends to increase information asymmetry.

4.2.2 AdjPIN regression

The estimated coefficients of the $AdjPIN$ regression are quite similar to the coefficients of the PIN regression. According to expectations, $AdjPIN$ has significantly negative connection with $Lsize$, $Lage$ and $Lcoverage$ and significantly positive connection with $Lilliq$, $Cash$, $Llag$ and $Lturnover_{ins}$. Connection between $AdjPIN$ and $Lturnover$ is positive but insignificant. Again, the effect of $Lturnover$ is likely to be captured by $Lilliq$ (see Panel C of Table 3 for more details).

The coefficient on $Lvolatility$ is significant and negative. A possible explanation of the sign is following. According to the PIN model, there is only one source of order flow imbalance. It is caused by informed trading. However, the extended PIN model implies two sources of order flow imbalance, the first is caused by informed trading and the second is caused by symmetric order flow shocks.

Symmetric order flow shocks lead to divergent opinions of liquidity investors who increase both sell and buy order flows. As a result, *Lvolatility* is also increased. At the same time, probability of informed trading, $AdjPIN$, should be lower as liquidity investors are more active and their impact on order flow is more significant. Conversely, $PSOS$ is higher. The PIN model does not distinguish between two sources of order flow imbalance and PIN captures both $AdjPIN$ and $PSOS$ effect of *Lvolatility*. Therefore, the coefficient on *Lvolatility* is significantly negative for $AdjPIN$, insignificant for PIN and, as noted below, significantly positive for $PSOS$.

4.2.3 PSOS regression

The estimated coefficient of the $PSOS$ regression support the hypothesis that symmetric order flow shocks lead to increase of both liquidity and informed investors' activity. The coefficient on *Lvolatility* is significantly positive, supporting the arguments given above. The coefficient on *Lturnover* is significantly positive. There are two reasons for this. First of all, the effect of high trading activity is captured by *Lilliq*, similarly to the PIN and $AdjPIN$ regressions (as noted in Panel C of Table 3 that in absence of *Lilliq* regressor the coefficient on *Lturnover* is smaller, although it is still significantly positive). Second, as noted in Jiang, Lee, and Zhang (2005) high *Lturnover* is a signal that investors have divergent opinions about this stock. Stocks with high divergence in opinions tend to have high $PSOS$. Surprisingly, the coefficient on *Lsize* is significantly positive. The possible explanation is as follows. As noted in Table 2, $PSOS$ and *Lsize* have negative correlation. However, in the regression this negative connection is captured by other independent variables (mainly, by *Lilliq* and *Lcoverage*). Note that in absence of *Lilliq* the coefficient on *Lsize* is insignificant (see Panel C of Table 3 for details).

The coefficients on information asymmetry measures coincide with expectations. $PSOS$ has significantly negative relation with *Lage*, *Lcoverage* and significantly positive relation with *Lage*,

$Cash$ and $Lturnover_{ins}$. The coefficient on $Llag$ is insignificant. Generally, the coefficients on information asymmetry measures are lower in absolute value and less significant than for the PIN and the $AdjPIN$ regressions.

In the whole, the results described above are consistent with the hypothesis that PIN, $AdjPIN$ and $PSOS$ reflect information asymmetry to some extent. I find that for firms with low earnings quality (measured by $Cash$ and $Llag$), for younger firms and firms with smaller analysts' coverage and more active insider trading PIN, $AdjPIN$ and $PSOS$ tend to be higher. However, these facts do not guarantee that PIN, $AdjPIN$ and $PSOS$ measure solely information asymmetry.

First of all, information asymmetry proxies I use in the regressions do not measure information asymmetry directly. For example, as noted in Barry and Brown (1985), Age may not be a good measure of information quantity because "...the period of listing cannot be regarded as powerful measure of information quantity since it is insensitive to the different rates at which information is produced for different securities". The analysts' choice which securities to follow may be based on specific features of securities (such as securities' liquidity. See Roulstone (2003) for the discussion). Therefore, my results support the hypothesis only to extent of validity of regressors as information asymmetry measures.

Second, even if all the three measures capture information asymmetry one can not be sure that they capture *only* information asymmetry. Coefficients on illiquidity proxies are highly significant for PIN, $AdjPIN$ and $PSOS$. Hence, the results of the regressions give some evidence that PIN, $AdjPIN$ and $PSOS$ do measure not only information asymmetry. In the following section I investigate this issue in more details.

5 Accuracy test of sequential trading models

In this section I investigate how model implied probabilities of good and bad news, as well as of symmetric order flow shocks (for the extended PIN model) vary around quarter earning announcements. My analysis is based on the idea of Kelly (2005). I divide the sample of earning announcements into three groups, depending on the sign of earning surprise. Assuming that the sign of earning surprise reflects whether the announcement is 'good', 'bad' or 'neutral', it is possible to test whether model implied probabilities of good and bad private information events correspond to 'good' and 'bad' announcements, respectively.

It is important to note that both investigated sequential trading models assume that only informed investors can understand whether new information is good or bad. In the basic PIN model liquidity investors do not respond to any market events at all. Therefore, their activity around announcement days should be the same as around non announcement days. In the extended PIN model liquidity investors are more sophisticated and may increase their activity around public news events. However, while liquidity investors are likely to increase their activity around announcement days, their reaction on announcements does not depend on type of these announcements. Hence, they are not able to process whether public news event is good or bad and their activity should be the same around all types of earning announcements. Conversely, as informed investors can discriminate between 'good' and 'bad' announcements, they should initiate buy/sell trades around 'good'/'bad' announcements but should not be particularly active around 'neutral' ones.

In this section I test these model predictions. In the following subsections I firstly give formulas to calculate model implied probabilities of order flow shock days, then I describe the mechanism of dividing announcements into three groups in more details and, finally, I discuss the results.

5.1 Model implied probabilities of information and liquidity events

Sequential trading models allow one to calculate ex-post probabilities of each day to be a day with an order flow shock. For instance, in case of the PIN model, one can calculate the probability of each day to be a day with good or bad private news. In addition to that, the extended PIN model allows one to calculate model implied probability of symmetric order flow shock day. For example, having estimated all the parameters of the PIN model, the probability of private news day is defined by the number of buyer and seller initiated trades on this day:

$$\mathbb{P}(News|B,S) = \frac{\mathbb{P}(B,S|News)\mathbb{P}(News)}{\mathbb{P}(B,S|News)\mathbb{P}(News) + \mathbb{P}(B,S|NoNews)\mathbb{P}(NoNews)} \tag{9}$$

Here $\mathbb{P}(News) = a$, $\mathbb{P}(B,S|News) = de^{-\epsilon_b-\mu}\frac{(\epsilon_b+\mu)^B}{B!}e^{-\epsilon_s}\frac{(\epsilon_s)^S}{S!} + (1-d)e^{-\epsilon_b}\frac{(\epsilon_b)^B}{B!}e^{-\epsilon_s-\mu}\frac{(\epsilon_s+\mu)^S}{S!}$. Similar but more cumbersome formulas describe probabilities of good and bad private news, as well as probability of symmetric order flow shock (in case of the extended PIN model).

5.2 Earning announcement classification

My sample includes 47,731 quarter earning announcements. I divide them into three groups in the following manner. For each announcement I calculate the most recent consensus analyst forecast of one quarter ahead forecast of the announcing firm's earning per share. Consensus forecast is estimated as mean analysts' forecast[6]. Then I compare announced earning per share with the most recent consensus forecast. The difference between the two is earning surprise. Following widespread approach used in the literature (see, for example, Lee (1992) and Skinner and Sloan, 2002), I assume that the earning announcement is 'good'/'neutral'/'bad' if corresponding earning surprise is positive/zero/negative. The sample includes 16,594 'bad', 6,888 'neutral' and 24,249 'good' announcements.

This classification implies that, say, 'good' announcements contain new information that is above

[6]The result is very similar if consensus forecast is estimated as median analysts' forecast

analysts' forecasts. According to the considered sequential trading models, only informed investors are able to understand that such an announcement contains positive signal. According to the models, they should initiate buy sells after receiving this positive signal.

Using the basic PIN and the extended PIN models, I calculate abnormal probabilities of good and bad private news event days for 7 days before and after every announcement (announcement window). The extended PIN model also allows to calculate abnormal probabilities of symmetric order flow shock days. Abnormal probabilities of order flow shock day are calculated as the difference between probability in the announcement window and the mean of corresponding probabilities in the non announcement window for the same year.

5.3 Accuracy test results

5.3.1 The basic PIN model

The basic PIN model allows only for one type of order flow shock caused by private signal. In the ideal world of the model the results of the accuracy tests should be as follows. Abnormal probability of good/bad private news should be significantly higher than zero around 'good'/'bad' announcements and insignificantly different from zero around other types of announcements. Around 'neutral' announcements expected activity of informed investors (measured by abnormal probability of private signal event) should be statistically indistinguishable from zero.

The results of the accuracy test of the basic PIN model are described in Table 4 and Figure 4. The results of the accuracy test are rather contradictory. On the one hand, abnormal probabilities of good/bad private signals are significantly different from zero around 'good'/'bad' announcements. The effect is the most significant at the reported days of announcements (day 0 in Table 4 and Figure 4) and persists for the next 3 days. This persistence can be explained at least in two ways. Firstly, it might signal that new information brought by earning announcements is complicated and can not be fully analyzed at once even by informed investors. Therefore, abnormal probability of private signal event is significant not only at the very announcement days. Second, there is evidence (see, for example, Kandel and Pearson, 1995) that dates of earning announcements in Compustat files do not always correspond to dates information about them becomes available for investors. As a result, true and reported announcement days may differ by one. This may lead to broadening of peaks in Figure 4.

On the other hand, abnormal probabilities of good/bad private signals are significantly higher than zero for 'bad'/'good' and 'neutral' announcements, although they are smaller than for 'good'/'bad' announcements. Again, abnormal probabilities are persistent. This result contradicts predictions of the basic PIN model as it implies that informed investors should be buyers/sellers around 'bad'/'good' announcements. Moreover, they are more active even if announcements are not informative ('neutral'). The latter observation is not surprising. Indeed, the basic PIN model implies that liquidity investors do not respond to any market events. Hence, well documented abnormal trading activity around all earning announcements, including, of course, 'neutral', can be explained only by informed investors in the model.

5.3.2 The extended PIN model

The extended PIN model allows for two sources of order flow shocks. The role of informed investors in this model coincide with the case of the basic PIN model. Hence, similarly to the previous case, the model predicts that abnormal probability of good/bad private news should be significantly higher than zero around 'good'/'bad' announcements and insignificantly different from zero around other types of announcements. However, the extended PIN model does not imply that abnormal trading activity around 'neutral' announcements should be zero. In this model liquidity investors can respond to earning announcements. Therefore, I expect that abnormal probability of symmetric order flow shock should be significant for all types of earning announcements.

The results of the accuracy test of the extended PIN model are described in Table 5 and Figure 4. Generally, the extended PIN model describes announcements only marginally better than the basic PIN model. Similarly to the previous case, abnormal probabilities of good/bad private signals are significantly higher than zero around 'good'/'bad' announcements. Abnormal probability of bad news is only marginally significant around 'good' announcements. At the same time, abnormal probability of good news is highly significant around 'bad' announcements. Abnormal probability of both good and bad news are marginally significant around 'neutral' announcements. In line with the predictions of the extended PIN model, abnormal probability of symmetric order flow shocks is significantly positive for all types of the announcements.

Both investigated sequential trading models do not pass the accuracy test ideally (the extended PIN model performs slightly better). What does it mean in terms of accuracy of PIN and $AdjPIN$ as information asymmetry proxies? Both basic and extended PIN models assume that periods of information asymmetry can be identified by some parts order flow imbalance. In case of the basic PIN model all order flow imbalance is supposed to be caused by informed trading. The results of the accuracy test clearly demonstrate that this approach is quite imprecise. Indeed, the fact that abnormal probability of, say, bad private signal is significantly positive for 'good' announcement means that order flow imbalance sometimes does not behave according to the model's predictions. For some announcements order flow imbalance around 'good' events is actually correspond to bad news in the model. Put simply, sometimes around 'good'/'bad' announcements the number of seller/buyer initiated orders is abnormally high for some reasons. As a result, the basic PIN model incorrectly classify these events. This observation gives evidence that the model's assumption of the only source of order flow imbalance is oversimplified.

The extended PIN model seems to solve this problem at least partly. In this model there are two sources of order flow imbalance and informed trading explains only part of order flow imbalance. This approach seems to be more precise, while it does not pass the accuracy test ideally.

The results of the accuracy tests indicate that order flow imbalance (or its part in the extended PIN model) capture not only informed trading. Therefore, $AdjPIN$ and especially PIN are likely to capture not only information asymmetry. This paper does not answer what exactly is captured by PIN and $AdjPIN$ besides information asymmetry. At the same time, it is clear that real market process is much more complicated than oversimplified model implied process, although the extension of Duarte and Young (2009) seems to propose some improvement. However, careful economic interpretation of both PIN and $AdjPIN$ is unclear.

It goes without saying that the results of accuracy test are correct to the extent of validity of its assumptions. Its main assumption is that quarter earning announcement can be classified into 'bad', 'neutral' and 'good' groups using the sign of earning surprise. As noted above, this classification is rather widespread in the literature. However, this approach has at least two limitations. Firstly, it classifies all the announcements into three groups based only on one indicator (earnings per share). Second, the sign of earning surprise may be subject to managers' manipulation (see, for example, Burgstahler and Eames, 2006). In this case, some 'bad' announcements are likely to be incorrectly classified as 'good' or 'neutral'.

6 Conclusion

PIN is the proxy of information asymmetry implied by sequential trading model of Easley, Kiefer, O'Hara, and Paperman (1996). This model assumes that order flow imbalance is caused by informed trading. PIN reflects the value of this imbalance. However, according to Duarte and Young (2009) the PIN model does not reflect some important features of real data, such as high positive correlation between the number of buy and sell trades. To take this into account, Duarte and Young (2009) allow for another source of order flow imbalance caused by symmetric order flow shocks. One possible example of symmetric order flow shock is public news event. $AdjPIN$ and $PSOS$ measure the value of imbalances caused by informed trading and symmetric order flow shocks, respectively.

The goal of this paper is to find out whether model implied measures of information asymmetry, PIN and $AdjPIN$, indeed capture only information asymmetry. Firstly, I show that not only PIN and $AdjPIN$ but also $PSOS$ are positively and significantly connected with different measures of information asymmetry, such as analyst coverage, age, insider trading activity and earnings quality measures. The latter result is in line with Aslan, Easley, Hvidkjaer, and O'Hara (2011) who note that public news (for example, earning announcements) events may lead to additional information asymmetry as only part of investors are able to interpret even public information properly. In this case, $PSOS$ reflects the information asymmetry to some extent and, hence, $AdjPIN$ does not capture all information asymmetry based trades.

Having shown that PIN and $AdjPIN$ do capture information asymmetry, I then demonstrate that PIN and, to a less degree, $AdjPIN$ capture not only it. Assuming that information asymmetry rises around quarter earning announcements (the result confirmed by, for example, Kim and Verrecchia (1994) and Lee, Mucklow, and Ready, 1993), I demonstrate that basic and, to a less degree, extended PIN models incorrectly identify some 'good'/'bad' announcements as low/high information shocks. These results demonstrate that order flow imbalance assigned by the models to periods of informed trading are not really connected with it. I conclude that $AdjPIN$ and especially PIN capture not only information asymmetry. Therefore, careful interpretation of the facts that PIN is priced (Easley and O'Hara, 2004) while $AdjPIN$ is not priced (Duarte and Young, 2009) is unclear.

The results of this paper demonstrate that order flow imbalance is connected to informed trading. However, neither widely used basic PIN nor extended PIN models precisely identify parts of order flow imbalance corresponding to informed trading. Further sophistication of sequential trading models may lead to more precise estimates of information asymmetry but is likely to be subject to computational problems and overparametrization. Search of more realistic and compact sequential trading model seems to be an appealing goal for the future research.

References

Amihud, Y. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56, 2002.

Aslan, H., Easley, D., Hvidkjaer, S., and O'Hara, M. The characteristics of informed trading: Implications for asset pricing. *Journal of Empirical Finance*, 18(5):782–801, 2011.

Barry, C. B. and Brown, S. J. Differential information and security market equilibrium. *Journal of Financial and Quantitative Analysis*, 20(4):407–422, 1985.

Beneish, M. D. and Vargus, M. E. Insider Trading, Earnings Quality, and Accrual Mispricing. *The Accounting Review*, 77(4):755–791, 2002.

Burgstahler, D. and Eames, M. Management of Earnings and Analysts' Forecasts to Achieve Zero and Small Positive Earnings Surprises. *Journal of Business Finance & Accounting*, 33(5-6):633–652, 2006.

Cochrane, J. H. *Asset Pricing*. Princeton University Press, 2005.

Duarte, J. and Young, L. Why is PIN priced? *Journal of Financial Economics*, 91(2):119–138, 2009.

Easley, D. and O'Hara, M. Information and the cost of capital. *The Journal of Finance*, 59(4): 1553–1583, 2004.

Easley, D., Kiefer, N. M., O'Hara, M., and Paperman, J. B. Liquidity, Information, and Infrequently Traded Stocks.pdf. *The Journal of Finance*, 51(4):1405–1436, 1996.

Easley, D., Hvidkjaer, S., and O'Hara, M. Is information risk a determinant of asset returns? *The Journal of Finance*, 57(5):2185–2221, 2002.

Easley, D., Hvidkjaer, S., and O'Hara, M. Factoring information into returns. *Journal of Financial and Quantitative Analysis*, 45(2):293–309, 2010.

Fama, E. F. and French, K. R. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.

Fama, E. F. and French, K. R. Industry costs of equity. *Journal of financial economics*, 43:153–193, 1997.

Fama, E. F. and MacBeth, J. D. Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, 81(3):607–636, 1973.

Hughes, J., Liu, J., and Liu, J. Information asymmetry, diversification, and cost of capital. *The Accounting Review*, 82(3):705–729, 2007.

Jiang, G., Lee, C. M. C., and Zhang, Y. Information Uncertainty and Expected Returns. *Review of Accounting Studies*, 10(2-3):185–221, 2005.

Kandel, E. and Pearson, N. Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy*, 103(4):831–872, 1995.

Kelly, P. Information efficiency and firm-specific return variation. *EFA 2005 Moscow Meetings Paper*, 2005.

Kim, O. and Verrecchia, R. E. Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics*, 17(1-2):41–67, 1994.

Lambert, R., Leuz, C., and Verrecchia, R. E. Accounting information, disclosure, and the cost of capital. *Journal of Accounting Research*, 45(2):385–420, 2007.

Lee, C. Earnings news and small traders: An intraday analysis. *Journal of Accounting and Economics*, 15:265–302, 1992.

Lee, C. M., Mucklow, B., and Ready, M. J. Spreads, depths, and the impact of earnings information: An intraday analysis. *Review of Financial Studies*, 6(2):345–374, 1993.

Miller, D., Cameron, A., and Gelbach, J. Robust inference with multi-way clustering. *Journal of Business & Economic Statistics*, 29(2):238–249, 2011.

Moskowitz, T. J. and Grinblatt, M. Do industries explain momentum? *The Journal of Finance*, 54 (4):1249–1290, 1999.

Newey, W. K. and West, K. D. A simple, positive semi-definite, heteroskedasticity and aurocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.

Roulstone, D. T. Analyst Following and Market Liquidity. *Contemporary Accounting Research*, 20 (3):552–578, 2003.

Skinner, D. and Sloan, R. Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio. *Review of accounting studies*, 7(2-3):289–312, 2002.

Verdi, R. S. Information Environment and the Cost of Equity Capital. *SSRN Electronic Journal*, 2005.

Weisberg, S. *Applied linear regression.* John Wiley & Sons, Inc., 3 edition, 2005.

Table 1: Summary statistics of the parameters. Panel A contains summary statistics of the explanatory variables. $Size$ is market value of equity in $billions in year t. $Turnover$ is the average daily turnover for the year t, multiplied by 10^3. $Illiq$ is the yearly measure of illiquidity proposed by Amihud (2002), multiplied by 10^6. Age is the number of years since the firm i was firstly covered by CRSP. $Volatility$ is yearly standard deviation of the firm's i returns. $Coverage$ for the firm i is the number of analysts who provide one fiscal year ahead forecasts of this firms' earnings per share during the year t. $Cash$ represents the ratio of the firm's i cash holdings and total assets in the year t. Lag is the average over year t number of days between quarter end earnings and quarter earnings report. $Turnover_{ins}$ is the annual mean of daily insider turnovers calculated as the ratio of daily share volume of trades made by company higher level insiders and number of shares outstanding, multiplied by 10^3. **Panel B** contains summary statistics of PIN, $AdjPIN$ and $PSOS$. PIN and $AdjPIN$ are the measures of probability of informed trading proposed by Easley, Kiefer, O'Hara, and Paperman (1996) and Duarte and Young (2009), respectively. $PSOS$ is the probability of symmetric order flow shock proposed by Duarte and Young (2009).

	Mean	Median	Min	Max	St Dev
Panel A					
$Size$	3.53	0.76	0.00	302.21	10.92
$Turnover$	3.69	2.88	0.04	63.37	3.13
$Illiq$	1.60	0.18	0.00	739.58	10.78
Age	22.14	19	0	76	19.41
$Volatility$	0.026	0.022	0.001	0.36	0.016
$Coverage$	9.50	7	0	53	8.85
$Cash$	0.051	0.023	0	0.96	0.075
Lag	28.60	26.25	3.25	588	12.16
$Turnover_{ins}$	0.011	0	0	4.58	0.075
Panel B					
PIN	0.162	0.153	0	0.527	0.055
$AdjPIN$	0.115	0.110	0	0.582	0.049
$PSOS$	0.200	0.192	0	0.741	0.077

Table 2: Correlations and Variance Inflation Factors. *Lsize* is logarithm of market value of equity in \$billions in year t. *Lturnover* is logarithm of the average daily turnover for the year t, multiplied by 10^3. *Lilliq* is logarithm of the yearly measure of illiquidity proposed by Amihud (2002), multiplied by 10^6. *Lage* is logarithm of the number of years since the firm i was firstly covered by CRSP. *Lvolatility* is logarithm of yearly standard deviation of the firm's i returns. *Lcoverage* for the firm i is logarithm of the number of analysts who provide one fiscal year ahead forecasts of this firm's earnings per share during the year t. Cash represents the ratio of the firm's i cash holdings and total assets in the year t. *Llag* is the average over year t number of days between quarter end earnings and quarter earnings report. $Lturnover_{ins}$ is logarithm of the annual mean of daily insider turnovers calculated as the ratio of daily share volume of trades made by company higher level insiders and number of shares outstanding, multiplied by 10^3. PIN is the measure of probability of informed trading proposed by Easley, Kiefer, O'Hara, and Paperman (1996). $AdjPIN$ is the measure of probability of informed trading proposed by Duarte and Young (2009). $PSOS$ is the probability of symmetric order flow shock.

Panel A contains Pearson correlations between the parameters. All the correlations are significantly different from 0 (significance level is 1%).

	$AdjPIN$	$PSOS$	$Lsize$	$Lturnover$	$Lilliq$	$Lage$	$Lvolatility$	$Lcoverage$	$Cash$	$Llag$	$Lturnover_{ins}$
PIN	0.53	0.42	−0.54	−0.23	0.59	−0.25	0.16	−0.47	0.07	0.29	0.07
$AdjPIN$		0.05	−0.43	−0.14	0.46	−0.24	0.12	−0.37	0.06	0.23	0.07
$PSOS$			−0.10	0.20	0.10	−0.20	0.22	−0.12	0.09	0.08	0.09
$Lsize$				0.17	−0.85	0.23	−0.40	0.68	−0.09	−0.46	−0.07
$Lturnover$					−0.47	−0.09	0.39	0.29	0.12	−0.06	0.07
$Lilliq$						−0.25	0.23	−0.71	0.05	0.41	0.06
$Lage$							−0.27	0.22	−0.05	−0.16	−0.10
$Lvolatility$								−0.24	0.17	0.29	0.08
$Lcoverage$									−0.10	−0.37	−0.06
$Cash$										0.02	0.07
$Llag$											0.06

Panel B contains variance inflation factors of the explanatory variables.

	$Lsize$	$Lturnover$	$Lilliq$	$Lage$	$Lvolatility$	$Lcoverage$	$Cash$	$Llag$	$Lturnover_{ins}$
VIF	5.22	2.43	6.82	1.19	1.76	2.18	1.05	1.31	1.03

Table 3: Regression results. The table presents results of different regressions of PIN, $AdjPIN$ and $PSOS$ (multiplied by 100) on information asymmetry measures and controls. PIN and $AdjPIN$ are the measures of probability of informed trading proposed by Easley, Kiefer, O'Hara, and Paperman (1996) and Duarte and Young (2009), respectively. $PSOS$ is the probability of symmetric order flow shock proposed by Duarte and Young (2009). $Lsize$ is logarithm of market value of equity in $billions. $Lturnover$ is logarithm of the average daily turnover, multiplied by 10^3. $Lilliq$ is logarithm of illiquidity measure proposed by Amihud (2002), multiplied by 10^6. $Lage$ is logarithm of the number of years since the firm was firstly covered by CRSP. $Lvolatility$ is logarithm of yearly standard deviation of the firm's returns. $Lcoverage$ is logarithm of the number of analysts who provide one fiscal year ahead forecasts of the firm's earnings per share. $Cash$ represents the ratio of the firm's cash holdings and total assets. $Llag$ is the yearly average of number of days between quarter end earnings and quarter earnings report. $Lturnover_{ins}$ is logarithm of the annual mean of daily insider turnovers calculated as the ratio of daily share volume of trades made by company higher level insiders and number of shares outstanding, multiplied by 10^3.

Panel A. Pooled OLS regression. 20 industry classification of Moskowitz and Grinblatt (1999) (round parentheses) and 49 industry classification of Fama and French (1997) (square parentheses) are used to calculate industry-year double clustered standard errors. ***, **, * mark coefficients that are significant on 1%, 5% and 10% levels, respectively.

	PIN	$AdjPIN$	$PSOS$
$Intercept$	21.91***	15.43***	11.88***
	(2.03)	(1.77)	(2.94)
	[2.19]	[1.64]	[2.96]
$Lsize$	−0.37***	−0.49***	1.23***
	(0.08)	(0.08)	(0.21)
	[0.09]	[0.08]	[0.22]
$Lilliq$	1.11***	0.72***	1.41***
	(0.06)	(0.10)	(0.11)
	[0.06]	[0.09]	[0.11]
$Lturnover$	−0.09	0.29	2.94***
	(0.25)	(0.26)	(0.33)
	[0.26]	[0.26]	[0.28]
$Lcoverage$	−0.38***	−0.20**	−0.86***
	(0.09)	(0.10)	(0.26)
	[0.09]	[0.10]	[0.25]
$Lvolatility$	−0.41	−1.08***	1.36**
	(0.55)	(0.37)	(0.55)
	[0.56]	[0.37]	[0.57]
$Lage$	−0.53***	−0.58***	−0.68***
	(0.08)	(0.05)	(0.11)
	[0.08]	[0.04]	[0.10]
$Cash$	2.55**	1.82***	3.70***
	(1.13)	(0.33)	(1.20)
	[0.84]	[0.47]	[1.20]
$Lturnover_{ins}$	4.16***	3.92***	7.46***
	(0.96)	(0.71)	(1.74)
	[1.01]	[0.69]	[1.89]
$Llag$	0.46***	0.36***	−0.24
	(0.15)	(0.08)	(0.33)
	[0.15]	[0.11]	[0.34]
$Adjusted\ R^2$	0.38	0.26	0.13

Panels B and C. Fama-MacBeth regression. This table present the results of Fama-MacBeth regressions. Coefficients are estimated each year from 1993 to 2001 and then averaged across time. **Panel B** reports regression results with full set of regressors. **Panel C** reports regression results in absence of *Lilliq* regressor. Heteroscedasticity and autocorrelation robust standard errors are reported between parentheses. ***, **, * mark coefficients that are significant on 1%, 5% and 10% levels, respectively.

	Panel B			Panel C		
	PIN	$AdjPIN$	$PSOS$	PIN	$AdjPIN$	$PSOS$
Intercept	20.91***	14.77***	10.03***	34.12***	23.68***	26.18***
	(1.43)	(1.09)	(2.49)	(1.74)	(1.41)	(2.71)
Lsize	−0.19	−0.29***	1.26***	−1.18***	−0.96***	0.05
	(0.13)	(0.04)	(0.21)	(0.15)	(0.06)	(0.22)
Lilliq	1.27***	0.87***	1.56***			
	(0.07)	(0.07)	(0.10)			
Lturnover	0.06	0.42	3.03***	−1.10**	−0.38	1.60***
	(0.32)	(0.31)	(0.34)	(0.39)	(0.38)	(0.40)
Lcoverage	−0.35***	−0.24***	−0.66***	−0.58***	−0.39***	−0.92***
	(0.05)	(0.03)	(0.11)	(0.05)	(0.03)	(0.11)
Lvolatility	−0.16	−0.65***	1.04***	−0.06	−0.59***	1.15***
	(0.15)	(0.16)	(0.20)	(0.17)	(0.14)	(0.23)
Lage	−0.51***	−0.53***	−0.71***	−0.73***	−0.68***	−0.97***
	(0.03)	(0.03)	(0.06)	(0.03)	(0.03)	(0.06)
Cash	1.99***	1.13**	3.84**	2.18***	1.26**	4.06***
	(0.36)	(0.37)	(1.31)	(0.16)	(0.42)	(1.03)
$Lturnover_{ins}$	4.15**	3.99***	7.48**	5.48*	4.86***	9.16**
	(1.34)	(0.35)	(3.24)	(1.63)	(0.26)	(3.65)
Llag	0.42***	0.29***	−0.16	0.61***	0.43***	0.08
	(0.06)	(0.07)	(0.20)	(0.05)	(0.06)	(0.18)
Adjusted R^2	0.41	0.29	0.13	0.38	0.27	0.11

Table 4: The basic PIN model implied abnormal probabilities. The table presents abnormal probabilities (in %) of private signals (Ab.News for all news, Ab.Good for good news and Ab.Bad for bad news) around 'bad', 'neutral' and 'good' quarter earning announcements implied by the basic PIN model of Easley, Kiefer, O'Hara, and Paperman (1996). Abnormal probabilities are calculated for 7 days before and after each announcement. Announcements are classified as 'good'/'neutral'/'bad' if the value of corresponding earning surprise is positive/zero/negative. Standard errors are reported between parentheses. ***, **, * mark coefficients that are significant on 1%, 5% and 10% levels, respectively.

The basic PIN model

Days	'Bad' announcement			'Neutral' announcement			'Good' announcement		
	Abn. Good	Abn. Bad	Abn. News	Abn. Good	Abn. Bad	Abn. News	Abn. Good	Abn. Bad	Abn. News
−7	−0.6	−0.9	−1.4	−0.8	−0.9	−1.7	−1.2	−1.2	−2.3
	(1.4)	(1.2)	(2.0)	(2.4)	(1.4)	(2.4)	(1.8)	(0.7)	(1.9)
−6	0.2	−1.1	−0.9	−0.6	−0.3	−0.9	−1.6	−1.2	−2.7
	(1.5)	(0.8)	(1.9)	(1.8)	(1.4)	(1.4)	(1.9)	(0.9)	(1.7)
−5	0.0	−0.8	−0.8	−0.8	−0.5	−1.3	−0.8	−1.2	−2.0
	(2.0)	(1.0)	(2.2)	(1.4)	(1.4)	(1.1)	(2.1)	(0.7)	(2.4)
−4	0.2	−1.3	−1.1	−0.6	−1.2	−1.8	−0.6	−1.4	−2.0
	(2.1)	(1.3)	(2.8)	(1.7)	(1.4)	(1.4)	(2.1)	(0.9)	(1.8)
−3	0.0	−1.4	−1.4	−1.0	−1.0	−2.0	−0.8	−1.3	−2.2
	(2.3)	(0.6)	(2.1)	(2.2)	(1.1)	(2.4)	(2.4)	(0.6)	(2.4)
−2	0.3	−0.6	−0.3	−0.6	−1.2	−1.8	0.0	−0.8	−0.8
	(1.7)	(1.1)	(2.0)	(1.8)	(0.7)	(1.6)	(2.5)	(1.4)	(3.2)
−1	2.3	2.4	4.7	3.1*	1.1	4.2	4.1	1.9	6.0
	(2.6)	(2.6)	(4.5)	(1.6)	(2.8)	(3.9)	(3.0)	(2.8)	(5.3)
0	8.3***	10.5***	18.8***	8.7***	7.4***	16.1***	13.0***	8.8**	21.8***
	(2.3)	(3.5)	(3.0)	(2.7)	(2.0)	(2.0)	(2.1)	(3.8)	(3.5)
1	8.2***	8.3***	16.5***	8.4***	6.5***	14.9***	10.2***	7.2***	17.3***
	(2.0)	(2.0)	(2.3)	(2.3)	(2.2)	(2.6)	(3.2)	(2.7)	(1.9)
2	5.2***	3.6***	8.8***	4.4**	2.6*	7.0***	5.7***	3.6***	9.3***
	(1.7)	(1.3)	(2.0)	(2.1)	(1.5)	(3.0)	(2.3)	(1.1)	(2.4)
3	3.3**	2.2*	5.5**	1.9	2.2	4.1	3.1**	2.3	5.4***
	(1.6)	(1.2)	(2.5)	(1.9)	(1.5)	(2.5)	(1.5)	(1.4)	(2.2)
4	2.1	1.5	3.5	3.5	−0.9	2.6	2.1	1.3	3.4
	(2.0)	(1.1)	(2.7)	(6.1)	(4.0)	(4.1)	(1.5)	(1.5)	(2.0)
5	2.1	1.1	3.1	1.0	−0.6	0.5	1.8	0.4	2.2
	(1.8)	(1.3)	(2.1)	(1.8)	(2.0)	(2.9)	(1.6)	(1.3)	(2.0)
6	1.3	0.2	1.4	0.6	−0.5	0.0	1.1	0.1	1.2
	(1.6)	(1.5)	(2.4)	(2.0)	(3.2)	(3.5)	(1.6)	(1.1)	(2.1)
7	0.8	0.0	0.7	0.3	−1.3	−1.0	0.8	−0.3	0.4
	(1.4)	(1.0)	(2.1)	(1.3)	(1.3)	(1.8)	(0.9)	(1.3)	(1.7)

Table 5: The extended PIN model implied abnormal probabilities. The table presents abnormal probabilities (in %) of private signals (Abn.News for all news, Abn.Good for good news and Abn.Bad for bad news) and symmetric order flow shocks (Abn. SOF) around 'bad', 'neutral' and 'good' quarter earning announcements implied by the extended PIN model of Duarte and Young (2009). Abnormal probabilities are calculated for 7 days before and after each announcement. Announcements are classified as 'good'/'neutral'/'bad' if the value of corresponding earning surprise is positive/zero/negative. Standard errors are reported between parentheses. ***, **, * mark coefficients that are significant on 1%, 5% and 10% levels, respectively.

The extended PIN model

	'Bad' announcement				'Neutral' announcement				'Good' announcement			
Days	Abn. Good	Abn. Bad	Abn. News	Abn. SOF	Abn. Good	Abn. Bad	Abn. News	Abn. SOF	Abn. Good	Abn. Bad	Abn. News	Abn. SOF
-7	0.1	−1.1	−0.9	−1.0	−0.8	−0.8	−1.6	−2.5	−0.6	−1.0	−1.6	−2.1
	(1.3)	(0.9)	(1.3)	(1.6)	(2.2)	(1.7)	(2.3)	(1.9)	(1.4)	(0.8)	(1.0)	(1.4)
-6	0.4	−1.2	−0.8	−0.7	−0.5	−0.6	−1.1	−1.2	−0.6	−0.9	−1.5	−2.2
	(1.3)	(0.8)	(1.3)	(2.1)	(2.6)	(1.7)	(1.4)	(1.9)	(1.3)	(0.6)	(1.0)	(1.5)
-5	0.6	−0.9	−0.3	−0.4	0.5	−0.8	−0.2	−1.4	−0.4	−0.8	−1.2	−1.8
	(1.1)	(1.0)	(1.4)	(2.2)	(1.1)	(1.4)	(1.4)	(1.9)	(1.3)	(1.0)	(1.2)	(2.0)
-4	0.4	−1.3	−0.9	−0.7	−0.1	−0.9	−1.0	−1.9	−0.4	−0.9	−1.3	−1.6
	(1.4)	(1.1)	(1.5)	(2.3)	(1.3)	(1.5)	(1.3)	(1.5)	(2.1)	(1.3)	(1.2)	(1.7)
-3	0.2	−1.2	−0.9	−1.1	−0.6	−1.0	−1.6	−1.1	0.0	−1.2	−1.2	−2.0
	(1.7)	(1.0)	(1.9)	(2.3)	(1.5)	(1.1)	(1.6)	(1.9)	(1.8)	(1.0)	(1.2)	(2.3)
-2	0.1	−0.5	−0.3	−0.2	0.1	−0.8	−0.7	−0.7	0.6	−0.5	0.1	−0.7
	(1.5)	(0.9)	(1.7)	(2.5)	(2.4)	(1.3)	(1.6)	(1.5)	(1.5)	(0.9)	(1.5)	(3.1)
-1	1.6	1.6	3.3	4.5	2.0*	0.4	2.4	4.2	3.0*	0.7	3.7	6.0
	(1.9)	(2.0)	(2.8)	(4.6)	(1.1)	(1.9)	(2.4)	(3.5)	(1.8)	(2.0)	(3.0)	(5.1)
0	4.0**	7.4***	11.4***	18.2***	4.3*	4.7**	9.0***	16.3***	7.6***	4.9*	12.5***	21.6***
	(1.9)	(2.6)	(1.4)	(2.9)	(2.2)	(2.2)	(1.2)	(2.4)	(1.9)	(2.7)	(2.0)	(3.2)
1	3.9***	6.0***	9.9***	15.6***	4.5*	4.2*	8.7***	14.6***	6.0***	4.5*	10.5***	16.6***
	(1.6)	(1.9)	(1.8)	(2.4)	(2.3)	(2.1)	(2.0)	(2.8)	(1.9)	(2.3)	(1.5)	(2.1)
2	3.1***	2.5***	5.6***	8.1***	2.5*	1.9	4.4**	6.6***	4.0***	2.4*	6.4***	7.8***
	(1.2)	(1.0)	(1.5)	(2.2)	(1.5)	(1.4)	(2.1)	(2.5)	(1.6)	(1.3)	(1.1)	(2.2)
3	2.1**	1.1	3.2**	4.7*	2.1	1.3	3.4	2.4	2.2*	1.6	3.8***	4.5***
	(1.0)	(1.2)	(1.4)	(2.4)	(1.9)	(1.6)	(2.4)	(1.9)	(1.3)	(1.0)	(1.4)	(1.9)
4	2.4	0.3	2.7	2.7	1.7	−0.3	1.4	1.0	1.7	0.8	2.5*	2.8*
	(1.7)	(1.0)	(1.8)	(2.7)	(1.3)	(0.8)	(1.5)	(2.5)	(1.3)	(1.4)	(1.3)	(1.6)
5	2.0	0.4	2.4	2.6	1.6	−0.2	1.4	0.3	1.9*	0.4	2.4**	2.0
	(1.7)	(0.8)	(1.6)	(1.7)	(1.3)	(1.1)	(1.9)	(1.8)	(1.0)	(1.1)	(1.1)	(1.8)
6	1.4	−0.3	1.1	1.3	0.4	−0.9	−0.6	0.2	1.1	0.1	1.2	1.2
	(1.7)	(1.2)	(1.5)	(2.9)	(1.3)	(1.5)	(2.1)	(1.5)	(1.4)	(1.0)	(1.4)	(1.9)
7	1.4	−0.6	0.7	0.3	0.6	−1.3	−0.7	−0.9	1.6	−0.4	1.1	0.3
	(1.0)	(0.8)	(1.5)	(2.2)	(1.3)	(1.0)	(1.7)	(1.7)	(1.2)	(1.0)	(1.0)	(2.1)

Figure 1: The tree of the trading process in the PIN model. Each trading day informed investors receive private signal with probability a. The probability of high signal is d. If the signal is high/low then informed investors send buy/sell orders at rate μ. Uninformed investors send sell orders at rate ϵ_s and buy orders at rate ϵ_b.

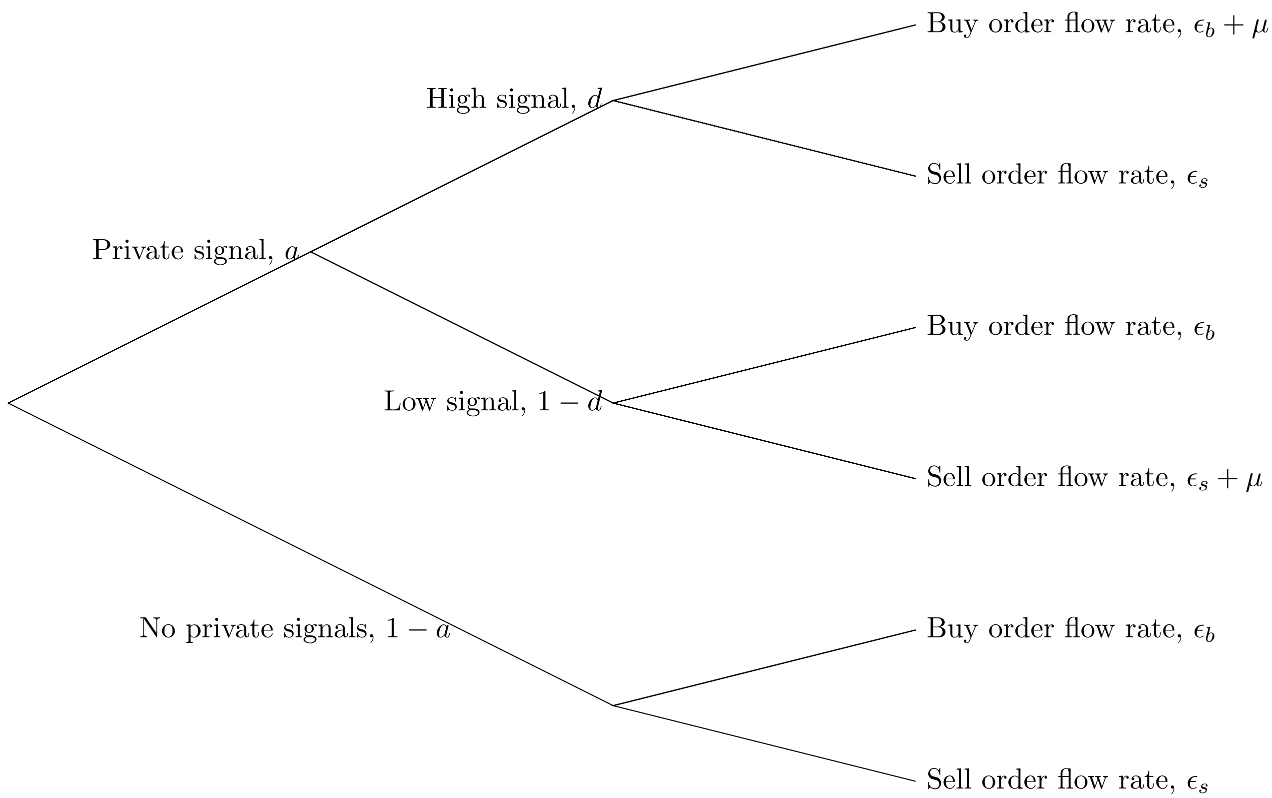

Figure 2: The tree of the trading process in the extended PIN model. Each trading day informed investors receive private signal with probability a. The probability of high signal is d. If the signal is high/low then informed investors send buy (sell) orders at rate μ_b/μ_s. The probability of symmetric order flow (SOF) is θ. If there is (no) SOF then uninformed investors send sell orders at rate $\epsilon_s + \Delta_s$ (ϵ_s) and buy orders at rate $\epsilon_b + \Delta_b$ (ϵ_b).

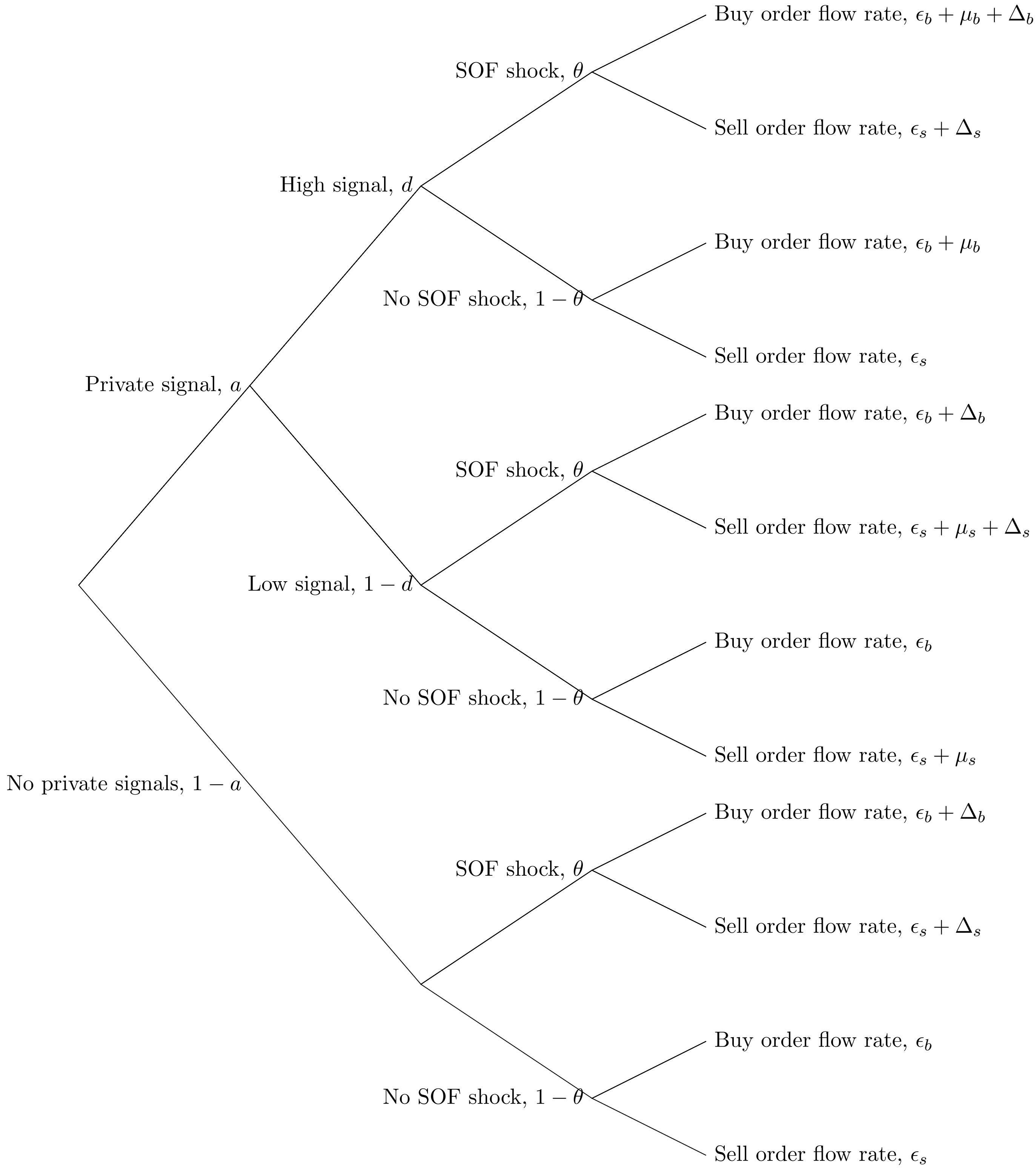

Figure 3: Time series plots of PIN, AdjPIN and PSOS. These figures represents 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles of PIN, $AdjPIN$ and $PSOS$. PIN is the measure of probability of informed trading proposed by Easley, Kiefer, O'Hara, and Paperman (1996). $AdjPIN$ is the measure of probability of informed trading proposed by Duarte and Young (2009). $PSOS$ is the probability of symmetric order flow shock.

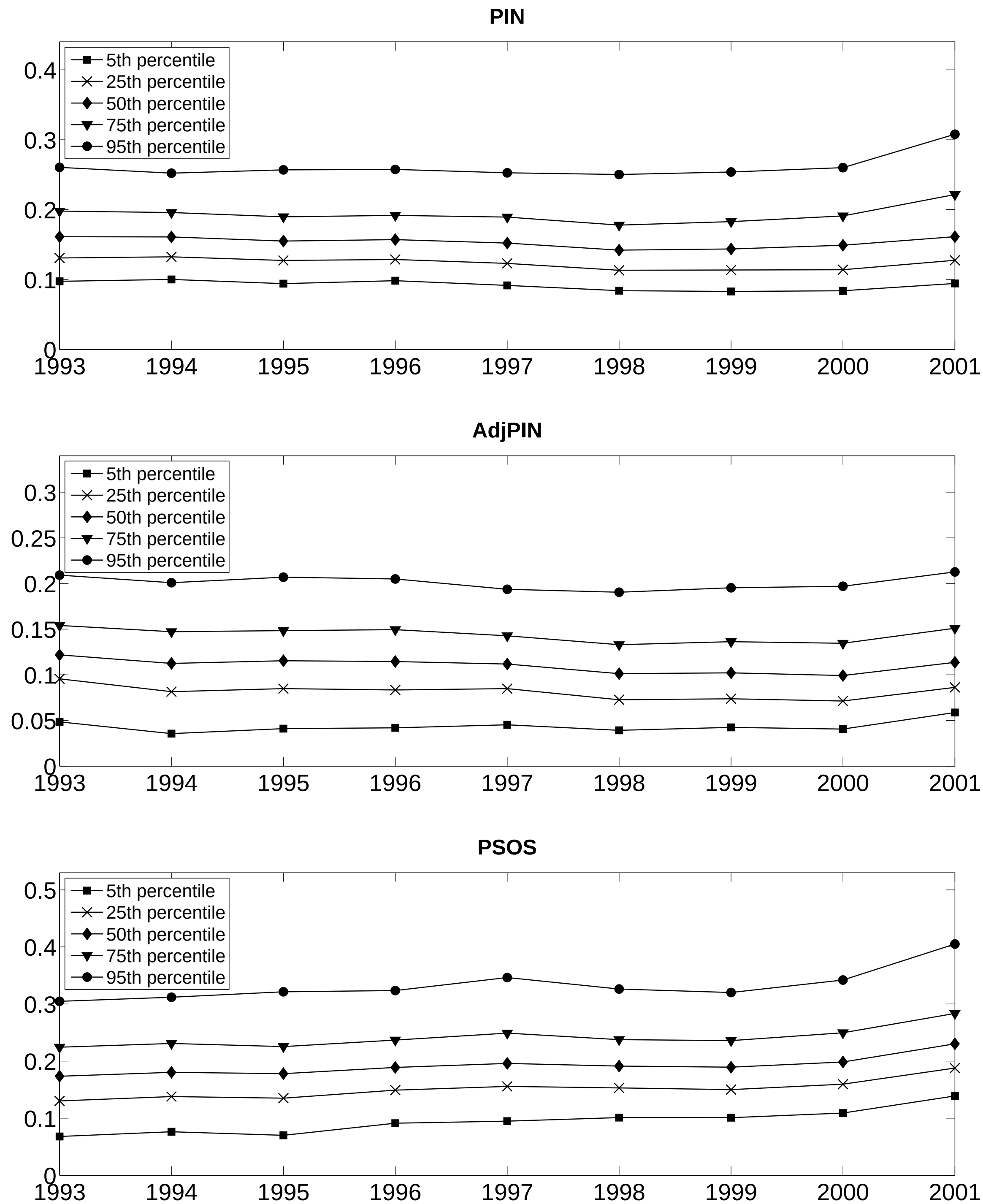

Figure 4: Abnormal probabilities of order flow shocks days. The figures depict abnormal probabilities of good, bad news and liquidity shocks around days of earning announcements. Announcements are classified as 'good'/'neutral'/'bad' if the value of corresponding earning surprise is positive/zero/negative. 'Good', 'neutral' and 'bad' announcements are marked by red, blue and black colors, respectively.