Why is PIN priced?

Jefferson Duarte and Lance Young[*]

August 31, 2007

Abstract

Recent empirical work suggests that a proxy for the probability of informed trading (PIN) is an important determinant of the cross-section of average returns. This paper examines whether PIN is priced because of information asymmetry or because of other liquidity effects that are unrelated to information asymmetry. Our starting point is a model that decomposes PIN into two components, one related to asymmetric information and one related to illiquidity. In a two-pass Fama-MacBeth regression, we show that the PIN component related to asymmetric information is not priced, while the PIN component related to illiquidity is priced. We conclude, therefore, that liquidity effects unrelated to information asymmetry explain the relation between PIN and the cross-section of expected returns.

JEL classification: G12, G14
Keywords: Liquidity; Information Asymmetry

1 Introduction

Market microstructure's impact on asset prices has captured substantial attention in the finance literature in recent years. Easley and O'Hara (2004) argue that stocks with more information asymmetry have higher expected returns. They construct a rational expectations asset pricing model with asymmetric information and find that everything else held constant, uninformed investors demand a premium to hold shares in firms with higher information asymmetry. In this model, the effects of information asymmetry are undiversifiable since the uninformed expect to lose to the informed and therefore demand to be compensated for this expected loss. On the other hand, Hughes, Liu and Liu (2005) and Lambert, Leuz and Verrecchia (2005) show that in a large economy the effect of asymmetric information on expected returns is diversifiable. They argue that asymmetric information is priced in Easley and O'Hara (2004) because the number of assets in their model is finite and hence asymmetric information risk cannot be diversified away. In spite of the fact that private information should be diversifiable in a large economy, empirically a proxy for information asymmetry, PIN, is positively and significantly related to average stock returns. Specifically, Easley, Hvidkjaer and O'Hara (2002) show that a ten percent difference in the $PINs$ of two stocks results in a 250 basis point difference in their annual expected returns. The empirical evidence therefore does not agree with theories that predict that asymmetric information is diversifiable and thus raises the question: Why is PIN priced? Our results indicate that PIN is priced because it is a proxy for illiquidity unrelated to asymmetric information. We arrive at this conclusion in three steps.

First, we examine the Easley, Kiefer, O'Hara and Paperman (1996) structural microstructure model (the PIN model hereafter). Our empirical examination shows that the original PIN model cannot match the pervasive positive correlation between buyer and seller initiated order flow (hereafter buy order flow and sell order flow) or the variances of buy and sell order flow. Our results suggest that the PIN model cannot match these moments because the PIN model specifies only two possible motives for trades, information and exogenous liquidity needs. All of the trades that are not initiated by informed traders are considered liquidity trades. Information related trading happens on days in which informed traders receive a private signal, which induces a larger number of buy orders if the private signal is positive and a larger number of sell orders on days with negative private signals. As a result of this, large numbers of buys and sells arrive on different days, creating a negative correlation between buys and sells. However, we find that for more than 95% of stocks, buys and sells are positively correlated. Furthermore, with only two motives for trade, the PIN model cannot match the relatively large variances of buys and sells.

Our second step is an extension of the PIN model which accommodates the positive correlation between buys and sells and generates variances closer to those observed in the data. Our extension of the PIN model accomplishes this by allowing for simultaneous positive shocks to both buy and sell order flow. The extended model also allows us to compute a measure of asymmetric information, $AdjPIN$, which like the PIN measure, is identified by periods of abnormal order flow imbalance as motivated by sequential trade models such as Glosten and Milgrom (1985) and as assumed in Easley, Kiefer, O'Hara and Paperman (1996).

In contrast to the PIN measure, however, $AdjPIN$ is consistent with the high variances of buys and sells and the positive correlation between buys and sells. Consequently, a Fama-MacBeth regression that includes $AdjPIN$ instead of PIN as a proxy for asymmetric information allow us to examine whether the relation between expected returns and PIN obtains because PIN is a proxy for information asymmetry or because of the original PIN model's inability to match some of the characteristics of the order flow data. Our results indicate that $AdjPIN$ is orthogonal to expected returns. Thus, the evidence suggests that PIN is not priced because it is a proxy for information asymmetry.

Our third step is to explore the relation between PIN and expected returns using our extended model to develop a measure of illiquidity unrelated to information asymmetry. We call this measure $PSOS$ (probability of symmetric order flow shock.) The $PSOS$ is the probability that any given trade happens during a shock to both the number of buyer initiated and seller initiated trades. To connect $PSOS$ to illiquidity, we show that stocks with a high $PSOS$ match the usual trading activity patterns of illiquid stocks, namely, very low trading activity on most days with sudden spikes in volume associated with the release of public information. We also show that firms in the highest $PSOS$ decile have Amihud illiquidity measures nearly thirty times that of the median $PSOS$ firm. In addition to being related to illiquidity, $PSOS$ is strongly correlated with PIN while the correlation between $PSOS$ and $AdjPIN$ is relatively low, indicating that $PSOS$ is the component of PIN that proxies for illiquidity unrelated to asymmetric information. Consequently, the relation between expected returns and $PSOS$ reveals the extent to which PIN is priced because it

3

is a proxy for illiquidity effects unrelated to information asymmetry. The estimated relation between expected returns and $PSOS$ is strong, indicating that the relation between PIN and expected returns is due to the fact that PIN is also a proxy of illiquidity not related to private information. Naturally, our evidence does not address the open question in the literature of why illiquidity would matter for the cross-section of expected returns. (See, for instance, Constantinides, 1986.) Our evidence does suggest however, that the relation between expected returns and illiquidity cannot be explained by information asymmetry effects.

It is important to note that our interpretation of the results is conditional on the assumption that periods of private information can be identified by abnormal order flow imbalance. Both PIN and $AdjPIN$ identify the arrival of private information as periods of abnormal order flow imbalance. Despite the evidence that the original PIN model captures information asymmetry,[1] it is possible that the abnormal order flow imbalances used to identify PIN and $AdjPIN$ are not the result of informed trade, but instead reflect liquidity shocks or the effect of the changes in the demand for immediacy in the sense of Grossman and Miller (1988). In this case, the separation of illiquidity into asymmetry of information and inventory concerns implied by the original PIN model is not correct. Even though we recognize that this is an important caveat, this paper's objective is not to question the notion that informed trade can be identified by abnormal order flow imbalance. Instead, adopting the same identification methodology as Easley, Hvidkjaer and O'Hara (2002), our objective is to explore the reasons why PIN is priced.

---

[1] See Easley, Kiefer and O'Hara (1997), Easley, Kiefer, O'Hara and Paperman (1996) and Vega (2006).

This paper adds to the growing literature on the analysis of order flow data and market microstructure's effect on asset prices. Studies closely related to this paper include Easley, Kiefer, O'Hara and Paperman (1996), Easley, Kiefer and O'Hara (1997), Easley, Hvidkjaer and O'Hara (2002), Vega (2006), and Duarte, Harford, Han and Young (2006). All of these studies use order flow data and the PIN model to extract information from the trade process that is important to asset prices. Our paper complements these papers by allowing for simultaneous positive shocks to both buy and sell order flow. Our paper relates to the recent papers by Venter and de Jongh (2004) and Boehmer, Grammig and Theissen (2007) that examine statistical issues related to PIN. Our focus, however, is different because we explore a large cross-section of stocks to focus on the economic effect of asymmetric information on expected returns.

The remainder of the paper is outlined as follows. In section 2 we outline the data we use for our empirical results. In section 3 we outline our extension of the Easley, Kiefer, O'Hara and Paperman (1996) structural microstructure model and present empirical evidence that the model matches important characteristics of the data. In section 4, we present our empirical results relating $AdjPIN$ to the cross-section of average returns. Section 5 concludes.

2 Data

Estimation of the structural microstructure models requires data on the number of buyer and seller initiated trades for each firm-day. To compute the number of buyer and seller initiated trades each day, we use the intra-day data from the Institute for the Study of Securities Markets (ISSM) (1983-1992) and the NYSE Trade and Quote (TAQ) (1993-2005)

databases. We exclude all trades and quotes that occur before the open and at the open as well as those at the close and after the close to avoid including trades that occurred during the opening and closing auctions. Furthermore, we exclude all trades with non-typical settlement conditions because these trades may have been made under special arrangements from which the models abstract. We exclude all quotes with zero bid or ask prices, quotes for which the bid-ask spread is greater than 50 percent of the price, and trades with zero prices to eliminate possible data errors. We employ the Lee and Ready (1991) algorithm to sign the trades. That is, trades above the mid-point between the bid and ask prices are considered buyer initiated and trades below the mid-point of the bid and ask prices are considered seller initiated. Trades that occur at the mid-point of the bid and ask prices are classified as buyer or seller initiated according to a tick test.[2] In addition, if there are no quotes posted during the trading day, we use the tick test to sign any trades made during the day. Each firm-day, we compute the number of trades classified as buyer-initiated and the number of trades classified as seller-initiated. The full sample includes all NYSE and Amex stocks. As in Easley, Hvidkjaer and O'Hara (2002), we exclude American Depository Receipts (ADRs), closed end funds and firm-years with less than 60 days of trading data. The Appendix provides more detail on the intra-day data.

Table 1 presents the median, first, and third quartile as well as the fifth and 95th percentiles of the mean, variance of both buys and sells and correlation between buys and sells. Interestingly, buys appear more volatile than sells. Furthermore, the median correlation

---

[2]The tick test classifies a trade as buyer initiated if the price was above that of the previous trade, and a sell if the price was below that of the previous trade.

between buys and sells is in excess of .50. In fact, even the fifth percentile is positive for every year in the sample. Indeed on average, 50 percent of firms had correlations of buys and sells between 0.34 and 0.66, and more than 95 percent of firms had buy and sell correlations above 0.15.[3] Figure 1 presents graphs of the number of buys and sells for the firm at the 25th percentile, the median and the 75th percentile. The positive correlation is easily visible in the graphs. As we will see in Section 3, the PIN model cannot reproduce the positive correlations between buys and sells or the high variances observed in the data.

To understand the relation between symmetric order flow shocks and illiquidity, we also use a comprehensive database of news events.[4] The database contains a random sample of firms, which represents between 20 and 25 percent of the firms in the CRSP database. The database records each day each firm in the sample is mentioned in publications that are covered in the Dow Jones Interactive Publications Library. The database is described in detail in Chan (2003).

The asset pricing tests require daily and monthly return data which we gather from CRSP. The tests also require book-value data which we gather from COMPUSTAT, specifically data item 60. We compute monthly turnover (TURNOVER) as the logarithm of the monthly volume divided by shares outstanding. Following Amihud (2002), we compute the Amihud measure (ILLIQ) as the average ratio of daily absolute price change and daily (dollar) volume for the year $t-1$. Our asset pricing regressions also require us to compute the market βs. We estimate market βs as follows: First, for each firm-month, we estimate the market

---

[3]We have also computed the correlation between buyer and seller initiated order flow at the hourly level and find that the correlations are similar to daily correlations.

[4]We thank Wesley Chan for graciously providing us with his database.

loadings using 60 months of past data. We then form portfolios based on these pre-ranking factor loadings. Using the returns from these portfolios, we estimate the full period β for each portfolio and assign this β to each firm in the portfolio.

3 The PIN Model and its Extension

This section briefly discusses the PIN model, then extends the model and presents estimates of various parameterizations of the extension as well as tests of these parameterizations.

3.1 The PIN model

Before considering our extension, we will briefly review the Easley, Kiefer, O'Hara and Paperman (1996) model. This model is based on the Glosten and Milgrom (1985) and Easley and O'Hara (1987) sequential trade models. The model contains both informed traders who trade for speculative purposes based on private information, and noise traders whose reasons for trading are exogenous. It also posits the existence of a uninformed liquidity provider who sets the bid and ask quotes by observing the flow of buy and sell orders, and assessing the probability that the orders come from informed traders. The bid-ask spread compensates the liquidity provider for the possibility of trading with the informed traders. At the beginning of each day, nature decides whether a private information event will occur. The probability that a private information event will occur on a given day is a. If a private information event occurs on a particular day, informed traders receive a private signal which is positive with probability d. If the signal is positive, buy order flow for that day arrives according to a Poisson distribution with intensity parameter $u + \varepsilon_b$ and sell order flow arrives

according to a Poisson distribution with intensity parameter ε_s. The intuition is that on days with positive private information, both informed traders and noise traders arrive in the market as buyers. The total buy order flow for the day therefore consists of arrivals of both noise traders, who arrive at rate ε_b, and informed traders who arrive at rate u. On the other hand, only noise traders arrive to sell, so the arrival rate of sell order flow is ε_s. If the signal is negative, buy orders consist only of noise traders with intensity parameter ε_b, and sell order flow arrives according to a Poisson distribution with intensity parameter $\varepsilon_s + u$ to reflect both the arrivals of noise sellers and of informed sellers. If there is no private signal, only noise traders will arrive in the market, so buy and sell order flow arrives by Poisson distributions with intensity parameters ε_b and ε_s, respectively. A tree outlining the information structure in the model can be found in Figure 2. The PIN is computed (suppressing the firm subscript) as:

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

The intuition behind the formula for PIN is that the probability of an informed trade is the ratio of expected informed order flow to expected total order flow.

3.1.1 The PIN Model Estimation

We estimate the PIN model numerically via maximum likelihood for each firm year from 1983 to 2005. The likelihood function of the Easley, Kiefer, O'Hara and Paperman (1996)

model is:

$$L(\theta|B,S) = (1-a)e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!}$$
$$+ade^{-(u+\varepsilon_b)}\frac{(u+\varepsilon_b)^B}{B!}e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!} \qquad (2)$$
$$+a(1-d)e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}e^{-(u+\varepsilon_s)}\frac{(u+\varepsilon_s)^S}{S!}$$

where B and S are the number of buys and sells for a given day and $\theta = (a, u, \varepsilon_b, \varepsilon_s, d)$ is the parameter vector. The likelihood equation shows that at each node of the tree in Figure 2 buys and sells arrive according to independent Poisson distributions, with the intensity parameters differing according to the node of the tree. For instance, conditional on positive information, buys arrive at a rate $\varepsilon_b + u$ and sells arrive at a rate of ε_s. The direct computation of this likelihood function when B is large may result in numerical overflow since ε_b^B or $B!$ become very large numbers. To avoid this problem we compute $e^{-\varepsilon_b}\varepsilon_b^B/B!$ as:

$$e^{[-\varepsilon_b + B\ln(\varepsilon_b) - \sum_{i=1}^{B}\ln(i)]} \qquad (3)$$

We compute the ratios $e^{-\varepsilon_s}\varepsilon_s^S/S!$, $e^{-(u+\varepsilon_b)}(u+\varepsilon_b)^B/B!$ and $e^{-(u+\varepsilon_s)}(u+\varepsilon_s)^S/S!$ similarly. To account for the fact that a particular numerical optimization may arrive at a local maximum, we run the likelihood optimization ten times for each firm year, each with randomly selected starting points, and we then select the maximum of these ten optimizations. Because the optimization procedure is computationally intensive[5], we divide the optimization across more than 100 PCs. The estimation procedure converges for virtually all firm-years.

Panel A of Table 2 presents quartiles, median, fifth and 95th percentile of the parameters

---

[5]The estimation of the original and of the extended models for all 48,611 firm years entails 2,916,660 optimizations.

of the PIN model across all firm-years. The median firm-year has a PIN of 0.2, with a first quartile of 0.15 and a third quartile of 0.28. Panel A also shows the median as well as the first and third quartiles of the t-statistics for the hypothesis that $PIN = 0$. The t-statistics range between 4.44 and 15.01 indicating that the PIN is relatively precisely estimated. In general, our PIN estimates are consistent with the ones in Easley, Hvidkjaer and O'Hara (2002).

3.1.2   Empirical and PIN Model Implied Moments

Figure 3 presents graphs of data simulated under the PIN model for the same firm-years as in Figure 1. The parameters used in these simulations are the same as those estimated from the samples displayed in Figure 1.[6] In Figure 3, sell order flow appears on the vertical axis and buy order flow appears on the horizontal axis. In each of the graphs, days with negative information appear in the upper left hand corner of the graph and days with positive private information appear in the lower right hand corner of the graph. The non-information days are represented by the cluster of points in the left bottom corner.

These graphs in Figure 3 show how PIN is identified in the data. Days with negative or positive private information produce large order flow imbalances relative to the normal level of trade. Another interesting feature of the graphs is that buy and sell order flow are negatively correlated under the PIN model. In fact, this is not peculiar to these firm-years, it is a general feature of the PIN model, in which contemporaneous covariance between the

---

[6]The parameter values are $a = 0.225$, $\varepsilon_b = 2.568$, $\varepsilon_s = 2.749$, $u = 7.395$, and $d = 0.452$ in the first panel; $a = 0.175$, $\varepsilon_b = 4.004$, $\varepsilon_s = 7.179$, $u = 15.236$, and $d = 0.723$ in the second panel; and $a = 0.410$, $\varepsilon_b = 37.827$, $\varepsilon_s = 32.206$, $u = 45.870$, and $d = 0.369$ in the third panel.

number of buys and sells in the PIN model is:

$$cov[B, S] = (au)^2 (d - 1)d \leq 0 \qquad (4)$$

This result is central to the intuition behind the model. Informed trade can inflate the arrival rate of buy or sell order flow, but not on the same day. Thus, when buy order flow is inflated, sell order flow is expected to be below its mean and vice-versa. Panel B of Table 2 shows the correlations between buys and sells computed using the estimated parameters of the model. As is clear from Equation 4, all of the correlations are below or equal to zero. Note that the negative correlation between buys and sells in the original PIN model is not consistent with the data. Recall from Table 1 that even the fifth percentile of firms has a positive correlation. Indeed, Table 1 indicates that more than 95 percent of the firm-years in the sample have correlation between buys and sells above 0.15.

Panel B in Table 2 also presents the implied means and variances of buys and sells under the PIN model, calculated for each firm using the estimated parameters. (See Appendix B for the formulas of these means and variances.) Comparison with the corresponding sample means and variances in Table 1 shows that while the model closely matches the buy's and sell's means, the model produces buy and sell variances that are smaller than the actual variances in the data. For about 50% of the sample, the actual variances of buys and sells are nearly twice as large as the implied variance from the model. This problem is most pronounced for firms with large buy and sell variances.

Thus, it appears that some aspect of the data generating process that creates a large positive correlation between buys and sells and high levels of buy and sell volatility is missing

from the original PIN model. In the next section, we extend the PIN model to better match the positive correlations between buys and sells as well as the high volatility in buy and sell order flow.

3.2 Model Extension

We extend the PIN model to account for the large buy and sell volatility and pervasive positive correlation between buys and sells by allowing both buy and sell order flow to increase on certain days. Naturally, there are many possible different extensions of the PIN model that match the data well. We however consider only extensions that are mixtures of a finite number of Poisson random variables. We focus on these extensions because they are parsimonious and imply measures, such as PIN, which have economic interpretations. In this section we present the unrestricted extended model, while in Section 3.2.1, we search for a more parsimonious version of the extended model.

The unrestricted extended model's information structure is in Figure 4. The first small difference between the extension and the PIN model is that the extension allows for the arrival rate of informed buyers, u_b to be different from the arrival rate of informed sellers, u_s. The motivation for this is to better allow the model to account for the fact that in the data, buy order flow has a larger variance than sell order flow for almost all firms. More importantly, the model allows for increased buy and sell variation and a positive correlation between buys and sells because each day an event can occur that causes both buy and sell order flow to increase. We call this event a symmetric order flow shock. The probability of such an event in the model, conditional on the absence of private information is represented

by θ. The probability of such an event, conditional on the arrival of private information is θ'. In the event of symmetric order flow shock, the additional arrival rate of buys is Δ_b and sells is Δ_s.

There are at least two possible explanations for symmetric order flow shocks. One possible cause of a symmetric order flow shock is the occurrence of a public news event about whose implications traders disagree. This disagreement causes both buy and sell trades to arrive at higher rates. The notion that disagreement about public news events is an important source of volume has a long history in the literature dating back to Bachelier (1900). As Kandel and Pearson (1995) note, agents observing identical signals can disagree about how to interpret these signals if they disagree about the model that should be used to interpret them. A number of studies have found empirical support for this hypothesis including, Harris and Raviv (1993), Kandel and Pearson (1995), Bessembinder, Chan and Seguin (1996) and Bamber, Barron and Stober (1999). Another potential cause for symmetric order flow shocks is that traders simply coordinate on trading on certain days to reduce trading costs, see Admati and Pfleiderer (1988). Even though there is more than one possible cause for symmetric order flow shocks, our purpose here is not to distinguish between the possible causes of such shocks. Instead, our objective is to present evidence that their inclusion in the model adds to its ability to explain the data, and has important consequences for the relation between information asymmetry and average returns.

The addition of symmetric order flow shocks to the model allows the correlation of buys

and sells to be positive. Under the extended model, the covariance of buys and sells is:

$$
\begin{aligned}
cov[B,S] \;=\; & a \times d \times u_b \times (a \times (d-1) \times u_s - (1-a) \times \Delta_s \times (\theta - \theta')) \\
& - \Delta_b \times (\Delta_s \times (\theta \times (\theta - 1) + a \times (2 \times \theta' \times \theta - 2 \times \theta^2 + (\theta - \theta')) \qquad (5) \\
& + a^2(\theta'^2 - \theta \times \theta')) + (-1 + a) \times a \times (-1 + d) \times (\theta - \theta') \times u_s),
\end{aligned}
$$

Unlike the PIN model, this expression can be positive or negative. To see this, note that when $\theta = \theta'$, the formula above becomes

$$
cov[B,S] = a^2 \times u_b \times u_s \times (d-1) \times d - \Delta_b \times \Delta_s \times (\theta - 1) \times \theta \qquad (6)
$$

As with the original PIN model, the probability of informed trade in the adjusted model is the ratio of the expected informed order to the total expected order flow:

$$
AdjPIN = \frac{a \times (d \times u_b + (1-d) \times u_s)}{a \times (d \times u_b + (1-d) \times u_s) + (\Delta_b + \Delta_s) \times \left(a \times \theta' + (1-a) \times \theta\right) + \varepsilon_s + \varepsilon_b}. \qquad (7)
$$

The extended model also allows us to introduce a related probability, the probability of a symmetric order flow shock ($PSOS$). The $PSOS$ is the unconditional probability that a given trade will come from a shock to both buy and the sell order flows. The $PSOS$ is given by:

$$
PSOS = \frac{(\Delta_b + \Delta_s) \times \left(a \times \theta' + (1-a) \times \theta\right)}{a \times (d \times u_b + (1-d) \times u_s) + (\Delta_b + \Delta_s) \times \left(a \times \theta' + (1-a) \times \theta\right) + \varepsilon_s + \varepsilon_b}. \qquad (8)
$$

As we show in Section 3.2.3, high $PSOS$ firms tend to be firms with low volume, relative to other firms, on most days but who experience large increases in both buy and sell order flow on days associated with public news events. For instance, a high $PSOS$ firm has

15

about one fifth the volume of the median firm on days without news and nearly twice as much volume as the median firm on days with news. As we will see later, this tendency of high $PSOS$ firms to have very low volume on most days and very high volume on a few days means that these firms tend to be very illiquid relative to other firms. To see this note that an investor who wishes to trade shares in a high $PSOS$ firm faces a trade off between waiting, perhaps for some time, for a period of high volume when the stock will be more liquid than on a typical day and when his trade can presumably be crossed with another trade, or trading immediately and involving a market maker or liquidity provider at additional cost. Thus, these firms' shares tend to among the least liquid firms in the market, as measured by their Amihud measures. Based on this we argue that $PSOS$ is effectively a proxy for illiquidity.

3.2.1 Model Selection and Estimation of Extended Model

As with the PIN model, we estimate the extended model numerically via maximum likelihood for each available firm-year. The likelihood function of the extended model is:

$$
\begin{aligned}
L(\theta|B,S) &= (1-a)(1-\theta)e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!} \\
&+ (1-a)\theta e^{-(\varepsilon_b+\Delta_b)}\frac{(\varepsilon_b+\Delta_b)^B}{B!}e^{-(\varepsilon_s+\Delta_s)}\frac{(\varepsilon_s+\Delta_s)^S}{S!} \\
&+ a(1-\theta')(1-d)e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}e^{-(u_s+\varepsilon_s)}\frac{(u_s+\varepsilon_s)^S}{S!} \\
&+ a\theta'(1-d)e^{-(\varepsilon_b+\Delta_b)}\frac{(\varepsilon_b+\Delta_b)^B}{B!}e^{-(u_s+\varepsilon_s+\Delta_s)}\frac{(u_s+\varepsilon_s+\Delta_s)^S}{S!} \\
&+ a(1-\theta')de^{-(u_b+\varepsilon_b)}\frac{(u_b+\varepsilon_b)^B}{B!}e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!} \\
&+ a\theta de^{-(u_b+\varepsilon_b+\Delta_b)}\frac{(u_b+\varepsilon_b+\Delta_b)^B}{B!}e^{-(\varepsilon_s+\Delta_s)}\frac{(\varepsilon_s+\Delta_s)^S}{S!}
\end{aligned}
\tag{9}
$$

16

where B and S are the number of buys and sells for a given day and $\theta = (a, u_b, u_s, \varepsilon_b, \varepsilon_s, d, \theta, \theta',$ $\Delta_b, \Delta_s)$ is the parameter vector. The estimation procedure is similar to the one described in Section 3.1.1 and it converges in virtually all cases. Analogous to the original PIN model, each term in the likelihood function above corresponds to a branch in the trading tree in Figure 4.

Because the general model contains twice as many parameters as the PIN model, before continuing, we seek a parsimonious parameterization that matches the characteristics of the data. Table 3 presents implied means, and variances of buys and sells as well as the correlation between buys and sells for four restricted versions of the general model, as well as the unrestricted version. These implied means, variances and covariances are calculated with the formulas in Appendix B. These models are meant to represent a variety of potential parameter restrictions. Specifically, in the first three, we allow symmetric order flow shocks only on days with no private information, $(\theta' = 0)$. Moreover, in Model 1 we restrict the buyer initiated and seller initiated information related trades to have the same intensities $(u_b = u_s)$, as well as the buyer and the seller symmetric order flow shocks to have the intensities $(\Delta_b = \Delta_s)$. In Model 2, we keep the restriction $u_s = u_b$ and relax the restriction $\Delta_b = \Delta_s$. Model 3 relaxes the restrictions $u_s = u_b$ and $\Delta_b = \Delta_s$. Model 4 allows symmetric order flow shocks on all days, but restrict the probability of symmetric order flow arrival to be the same on days with and without private information $(\theta = \theta')$. Model 5 is the unrestricted general model.

Each of the models produces positive correlations between buys and sells because, as

opposed to the original PIN model, all the estimated extended models do not restrict the probability of asymmetric order flow shocks to be zero. Furthermore, the buy and sell correlations are close to those found in the data. The median correlation between buys and sells ranges between .38 and .45, while in the data the median correlation is .50. Furthermore, each extended model is able to produce higher buy and sell variances than the PIN model, while closely matching the means of buys and sells. The first three models, which assume that symmetric order flow shocks arrive only on days without private information, produce somewhat higher correlations than the last two models, which allow for symmetric order flow shocks on private information days. The final two models, however, produce somewhat higher buy and sell variances, particularly for the high variance firms.

In order to choose the model which best fits the data, we run a series of likelihood ratio tests. We start by testing the original PIN model as null against Model 1 in Panel A of Table 3, which restricts $u_b = u_s$, $\Delta_b = \Delta_s$ and $\theta' = 0$. This amounts to a test of the null that $\theta' = \theta = 0$ or $\Delta_b = \Delta_s = 0$. Note that the standard regularity conditions that assure that the likelihood ratio statistic is asymptotically χ^2 do not apply in this case because in the PIN model $\theta' = \theta = 0$ and thus, the parameters lie on the frontier of the parameter space under the null hypothesis. In order to obtain the distribution of the test statistics, we conduct Monte Carlo simulations. For 100 randomly selected firm-years, we estimate the parameters of both models. We then generate one firm-year of trading days under the null using the estimated parameters and estimate both models using the simulated data to compute the likelihood ratio statistic. We repeat this process 500 times for each firm-

18

year. We then compare the likelihood ratio from the data to critical values of the simulated distribution. Table 4 shows that the PIN model is rejected at the five percent level in favor of Model 1 in 99% of the firm-years in the randomly generated sample.

Having rejected the PIN model in favor of Model 1, we then test Model 1 as null against Model 2. This amounts to a test of the null hypothesis that $\Delta_s = \Delta_b$. As the parameters do not lie on the frontier of the parameter space under the null, the standard asymptotic results apply and the likelihood ratio is distributed asymptotically χ^2 with one degree of freedom. This being the case, we can apply this test to the entire sample. The results in Table 4 show that the null is rejected at the five percent level in 53% of the firm years. We interpret this as evidence that Model 2 is preferable to Model 1 and therefore that the restriction that $\Delta_s = \Delta_b$ should not be enforced.

We then conduct a similar experiment and test Model 2 as null against Model 3. This amounts to a test of $u_b = u_s$, and, therefore, the likelihood ratio is distributed asymptotically χ^2 with one degree of freedom. The results indicate that the null is rejected in favor of the alternative 23% of the time. While this is not as high a rejection rate as we see above, it is higher than the five to ten percent rejection rate we would expect by pure chance. Therefore, we conclude that the restriction $u_b = u_s$ should not be enforced and reject Model 2 in favor of Model 3.

Finally, we test the last two models, which involve restrictions on the parameter θ'. First, we test Model 3 against Model 5, the unrestricted model. This amounts to a test of the null that $\theta' = 0$. As in the first likelihood ratio test, the likelihood ratio statistic is no

longer distributed χ^2. Therefore, we conduct Monte Carlo simulations, as described above, to obtain the likelihood ratio statistics' critical values under the null. Table 4 shows that the null is rejected in 79% of the randomly selected firm-years. We therefore conclude that the restriction that $\theta' = 0$ should not be imposed. Lastly, we test Model 4 against the unrestricted alternative, Model 5. This is a test of $\theta' = \theta$ and therefore the likelihood ratio is distributed χ^2 with one degree of freedom. The null is rejected in nine percent of the firm years. As this is close to what we would expect by chance, we conclude that the restriction $\theta' = \theta$ is innocuous, and choose Model 4 as our preferred model. Henceforth, we refer to this model as the $AdjPIN$ model. In what follows, all results use estimates from $AdjPIN$ model. However, all of the results that follow are robust to any of the other four alternative specifications. The preferred extended model has nine free parameters, which is four more parameters than the original PIN model.

3.2.2 The Preferred Extended Model

Figure 5 presents graphs of buys and sells simulated under the preferred model, for the firm-years featured in Figure 1 and Figure 3.[7] Here, there are six groups of points, one for each branch of the tree in Figure 4: The group of points in the lower left show the normal level of noise trade in the stock. The cluster of points in the upper right reflect days with symmetric order flow shocks. The graphs show that the preferred model better matches the data in Figure 1 than the PIN model. It is clear from the graphs that symmetric order

---

[7]The parameter values used in these simulations are $a = 0.585$, $\varepsilon_b = 1.553$, $\varepsilon_s = 1.624$, $u_b = 4.075$, $u_s = 2.631$, $\Delta_b = 5.151$, $\Delta_s = 6.112$, $d = 0.377$, and $\theta = 0.173$ in the first panel; $a = 0.426$, $\varepsilon_b = 2.067$, $\varepsilon_s = 5.810$, $u_b = 4.665$, $u_s = 14.143$, $\Delta_b = 12.711$, $\Delta_s = 7.492$, $d = 0.867$, and $\theta = 0.171$ in the second panel; and $a = 0.498$, $\varepsilon_b = 19.749$, $\varepsilon_s = 23.274$, $u_b = 24.261$, $u_s = 39.520$, $\Delta_b = 41.558$, $\Delta_s = 34.536$, $d = 0.659$, and $\theta = 0.410$ in the third panel.

flow shocks allow the preferred model to generate positive correlations between buys and sells. Furthermore, the graphs indicate that buys and sells have higher variances under the preferred model than under the PIN model, particularly for the firm-year representing the third quartile. This is because the preferred model has a larger number of possible reasons for trade than the original PIN model, which is reflected in a tree in Figure 4 with more branches and consequently more variation on the number of buys and sells.

Table 5 presents the first quartile, mean and median of the parameters of the $AdjPIN$ model for all firm-years. The median and both quartiles of $AdjPIN$ are lower than PIN, ranging between a fifth percentile of 0.08 and a 95th percentile of 0.37. The $t-$statistics indicate that the $AdjPIN$ estimates are large relative to their standard errors, with even the first quartile being over three. The median and both quartiles of $PSOS$ are also displayed in Table 5. They range between a fifth percentile of 0.12 and a 95th percentile of 0.66. The $t-$statistics indicate that the $PSOS$ estimates are also large relative to their standard errors, with even the first quartile being over two. Figure 6 shows the time series of the first quartile, median, and third quartile of PIN, $AdjPIN$ and $PSOS$. The graph indicates that PIN and $AdjPIN$ are stable over time. Furthermore, the graph indicates that $AdjPIN's$ quartiles and its median are consistently lower than $PIN's$ quartiles and median The median $PSOS$ is also stable over time at about 0.2, though it appears that $PSOS$ is somewhat more variable than PIN and $AdjPIN$, particularly in the third quartile.

The results in Table 5 show that the probability of trade due to symmetric order-flow shocks is non-trivial. The median estimated θ is 0.25, indicating that one in four days

includes simultaneously elevated buy and sell order flow. In fact, the probability of a symmetric order flow shock day is similar in magnitude to the probability of an information day (a) for the median firm as well as the first and third quartiles. The median, first, and third quartiles of $PSOS$ are also larger than the median, first and third quartiles of $AdjPIN$. Thus, for most firms, trade related to positive order flow shocks is more likely than trade related to private information on any given day.

3.2.3 PIN, $PSOS$ and Illiquidity

Having examined the models' relative ability to match the data, we now contrast PIN, $AdjPIN$ and $PSOS$. Table 6 presents the correlations between PIN, $AdjPIN$, and $PSOS$ and a number of other variables employed in this study. The correlation between PIN and $AdjPIN$ is high, at around 0.71. The correlation between PIN and $PSOS$ is also around 0.71. The correlation between $AdjPIN$ and $PSOS$ is not as high, at around 0.34. This reflects the fact that the $AdjPIN$ model explicitly accounts for the possibility of symmetric order flow shocks, so the model is better able to discriminate between days with symmetric order flow shocks and days that involve only abnormal buy or sell volume, which we identify as informed trade. The correlations between $AdjPIN$ and size, and opening spreads (SPREAD) are consistent with the idea that $AdjPIN$ is a proxy for the asymmetry of information. The correlation between $PSOS$ and Amihud's illiquidity (ILLIQ) is also consistent with the interpretation that $PSOS$ is a measure of illiquidity.

We explore the relation between $PSOS$ and illiquidity further in Table 7. First, in each year t, we sort firms into deciles based on the firms' $PSOS$ estimated in year $t - 1$. Using

22

the previously mentioned database of news events, we then compute the average number of buys, sells, the average product of buys and sells, the average volume (number of trades), the average net order flow imbalance and the average value of the Amihud measure separately on news and non-news days. The results indicate that for the low $PSOS$ deciles the average number of buys and sells are similar between news and non-news days. By contrast, for the large $PSOS$ deciles, the average number of buys and sells as well as their average product are much larger on news days than on non-news days. For the largest $PSOS$ decile, the number of buys and sells on news days is an order of magnitude bigger than on non-news days. Furthermore, relative to the other $PSOS$ deciles, the high $PSOS$ firms experience much smaller volumes on non-news days than the lower $PSOS$ deciles; only around 14 buys and 12 sells per day as opposed to upwards of 100 of each for lower $PSOS$ deciles. This evidence indicates that the expected number of both buys and sells increase on news days for high $PSOS$ firms. The evidence in Table 7 also indicates that high $PSOS$ firms have much higher Amihud measures than low $PSOS$ firms. This is particularly true on non-news days. However, despite the fact that high $PSOS$ firms have nearly twice the volume on news days as the median firm, their Amihud measures on news days are still an order of magnitude larger than the median firm's Amihud measure. It is also worth noting that news days are much more infrequent for high $PSOS$ firms than for low $PSOS$ firms. The average high $PSOS$ firm has news on less than five percent of days, while a low $PSOS$ firm has news on nearly 20 percent of days. Thus, the high $PSOS$ firms tend to have much lower volume than the median firm on most days, punctuated occasionally by news days where the high $PSOS$

23

firms have higher volume of both buys and sells than the median firm. Not surprisingly, these firms tend to be much less liquid, in terms of their Amihud measures, than low $PSOS$ firms, since, on most days, there is very little trade.

The results in Table 7 also shed light on the effect of symmetric order flow shocks on PIN. While high $PSOS$ firms tend to have much larger $PINs$ than other firms, they have only slightly larger $AdjPINs$. The reason for this lies with the symmetric order flow shocks. For high $PSOS$ firms, the periodic shocks associated with increased buy and sell volume also are associated with more volatile buy and sell volume. Thus, the increase in both buy and sell volume on days with symmetric order flow shocks also leads to a higher probability of much larger imbalances than on a typical day. The PIN model if forced to identify these days as private information days while the $AdjPIN$ model is not. Therefore, high $PSOS$ firms have only slightly higher information asymmetry than low $PSOS$ firms, as evidenced by their $AdjPIN$ measures, but they have much larger estimated $PINs$. At the same time, the high $PSOS$ firms also tend to be very illiquid, despite the fact that they have similar $AdjPINs$ to other firms. These firms experience low levels of trade, compared to the median firm, on most days and consequently are much less liquid than other firms. Furthermore, even on news days, the high $PSOS$ firms remain more illiquid than a typical firm. Therefore, $PSOS$ appears to be strongly related to liquidity, but for reasons apparently unrelated to information asymmetry and adverse selection.

Table 8 presents additional evidence that $PSOS$ is a proxy for illiquidity. Table 8 presents coefficient estimates from Fama-MacBeth regressions of $PSOS$ on various com-

24

mon proxies for illiquidity, including the natural logarithm of size, the natural logarithm of turnover, the Amihud measure, and the natural logarithm of the coefficient of variation of turnover and opening spreads. We calculate size in year t as the market value of the firm's equity at the end of December of $t-1$. We calculate turnover in year t as the average monthly turnover in year $t-1$. We calculated opening spreads in year t as the average, daily, opening, relative spread in year $t-1$. Size and turnover are both measures of liquidity, so we expect that if $PSOS$ is a proxy for illiquidity, the coefficients on size and turnover will be negative. The Amihud measure, the coefficient of variation of turnover and bid ask spreads are measures of illiquidity, so we expect that they will be positively related to $PSOS$. The coefficient estimates agree with these expectations in each case. $PSOS$ is negatively related to size and turnover and positively related to the Amihud measure, the coefficient of variation of turnover and spreads. Furthermore, $PSOS$ is non-linearly related to each of these variables. Regression 2 in Table 8 shows the coefficients on each of the liquidity proxies squared. In each case, the squared liquidity proxies are significantly related to $PSOS$. Based on the results in Table 8 and Table 7, we conclude that $PSOS$ is strongly and non-linearly related to illiquidity.

Our discussion so far suggests that symmetric order flow shocks creates differences between $AdjPIN$ and PIN. Table 9 presents further evidence that suggests that the difference between $AdjPIN$ and PIN is not random noise, but is related to liquidity. We explore cross-sectional differences between PIN and $AdjPIN$, by running regressions of $AdjPIN$ on PIN and regressions of the difference $(PIN - AdjPIN)$ on variables related to liquidity.

Table 9 presents the average difference between PIN and $AdjPIN$ as well as the time-series average of the estimated coefficients from cross-sectional regressions of $AdjPIN$ on PIN. This regression is useful in this context because if PIN is an unbiased estimate of $AdjPIN$, then the intercept in a regression of $AdjPIN$ on PIN is zero, and the coefficient on PIN is equal to one. The average difference between PIN and $AdjPIN$ is significantly positive, at around 0.04. Furthermore, the hypothesis that PIN is an unbiased estimator of $AdjPIN$ is strongly rejected by the data. The regression's intercept is positive and highly significant. In addition, the coefficient on PIN is 0.52 which, while significantly different from zero, is also statistically different from one. The t-statistic for the test that the linear coefficient is equal to one is -46.

Table 9 also presents evidence that the difference between PIN and $AdjPIN$ is related to liquidity. Table 9 presents averages of coefficient estimates from cross-sectional regressions of the difference between PIN and $AdjPIN$ on a number of variables related to liquidity, including $PSOS$. The results suggest that the bias in PIN is strongly related to size, turnover, $PSOS$, and the Amihud measure. Furthermore, since both PIN, $AdjPIN$ and $PSOS$ are non-linear functions of order flow, the bias in PIN could be non-linearly related to liquidity. The results in Table 9 indicate that this is indeed the case. The coefficients on squared size, turnover and $PSOS$ are large and highly significant. These results demonstrate that the difference between PIN and $AdjPIN$ is systematically related to variables that proxy for liquidity. Thus, despite the strong positive correlation between the two measures of the probability of informed trade, positive shocks to both buy and sell order flow render

PIN a biased measure of $AdjPIN$, and this bias is related to liquidity. In the following sections we show that this bias is important in interpreting the strong positive relation between PIN and average returns.

4 Asset Pricing Results

In the previous sections, we showed that the possibility of symmetric order flow shocks creates a bias in the PIN measure and that despite having about average $AdjPINs$, high $PSOS$ firms tend to have high estimated $PINs$. At the same time, high $PSOS$ firms tend to be very illiquid. All of this suggests that cross-sectional variation in PIN may, in part, capture liquidity effects that are unrelated to information asymmetry. Given that a number of recent papers such as Amihud (2002), Acharya and Pedersen (2005), Sadka (2006) and Pastor and Stambaugh (2003) document that liquidity is related to average stock returns, the above evidence suggests PIN may be related to average returns not because it is a proxy for information asymmetry per se, but because it is a proxy for illiquidity unrelated to information asymmetry. In this section, we investigate this possibility by replicating the asset pricing tests in Easley, Hvidkjaer and O'Hara (2002) to examine whether, after accounting for the possibility of simultaneous shocks to both buy and sell order flow, the positive relation between informed trade and average stock returns remains.

To this end, Table 10 presents Fama-MacBeth estimates of regressions of returns on β, size, book-to-market ratios (BM) and PIN as well as regressions that include $AdjPIN$ in place of PIN. These regressions are similar to those found in Easley, Hvidkjaer and

O'Hara (2002).[8]   Discussion of the computation of β, size and book-to-market can be found in the Data section above. As documented in Easley, Hvidkjaer and O'Hara (2002), the coefficient on PIN is large, at about 100 basis points per month, and is significant at conventional levels. However, when $AdjPIN$ is substituted for PIN, the result changes dramatically. The coefficient on $AdjPIN$ is economically small, at only 13 basis points per month, and is insignificant. Furthermore, when PIN and $AdjPIN$ are included in the same regression, the coefficient on PIN remains large, positive and significant, while the coefficient on $AdjPIN$ is insignificant and negative. On the other hand, when $PSOS$ is included in the regression alone it has a positive and significant coefficient. However, when both $PSOS$ and PIN are included in the regression, the coefficients on both measures drop considerably and both are rendered insignificant. Table 10 also shows that when the difference between $AdjPIN$ and PIN is included in the regression, it has a negative and significant coefficient. However, when the difference is in the presence of PIN, the coefficients on both drop dramatically and are no longer significant. This indicates that the bias in PIN is important for expected returns, and that this bias, rather than information asymmetry, is behind the relation between PIN and average returns.

In order to further examine the possibility that PIN is priced for liquidity reasons unrelated to the adverse selection problem created by informed trade, Table 10 also presents Fama-MacBeth regressions that include the Amihud measure. The Amihud measure is significantly positively related to average returns. However, in the presence of the Amihud

---

[8]Note that Easley, Hvidkjaer and O'Hara (2002) also present coefficient estimates from regressions that include additional control variables such as volume, the coefficient of variation of volume, spreads and return volatility. Our results are robust to the inclusion of these variables.

measure, the coefficient on PIN drops substantially and is insignificantly different from zero. Similarly, when $PSOS$ is included in the regression along with the Amihud measure the coefficient on $PSOS$ is reduced considerably and rendered insignificant. Overall, the evidence described above suggests that PIN is priced for reasons unrelated to informed trade, such as market makers' aversion to holding inventories. Naturally, the evidence in Table 10 does not address the open question in the literature of why liquidity concerns would matter for the cross-section of expected returns. (See, for instance, Constantinides, 1986.) Instead, the results in Table 10 show that if private information is identified by extreme order imbalances, then PIN is priced because it is a proxy for liquidity concerns unrelated to private information.

5   Conclusion

This study presents an extended version of the Easley, Kiefer, O'Hara and Paperman (1996) structural microstructure model that allows for the possibility of symmetric order flow shocks. We show that the extended model is better able to capture the variances of buys and sells as well as the large positive contemporaneous correlation between buys and sells observed in the data. We further show that the presence of symmetric order flow shocks creates differences between the PIN measure and the analogous measure from the extended model, $AdjPIN$. While the two measures are correlated, PIN is a biased measure of $AdjPIN$ and the bias is related to the tendency of a stock to have simultaneous positive shocks to both buy and sell order flow. Moreover, we show that high $PSOS$ firms have high $PINs$, but only slightly above average levels of information asymmetry as measured by $AdjPIN$. At the

29

same time, high $PSOS$ firms are also very illiquid, with very little trading volume, relative to the median $PSOS$ firm, on most days. Most importantly, we demonstrate that while PIN appears to be related to average returns, the $AdjPIN$ is not. We present evidence that the cross-sectional variation in PIN that is important for average returns relates to $PSOS$ as well as to the Amihud measure.

Assuming that periods of private information can be identified by periods of abnormal order flow imbalance as motivated by sequential trade models such as Glosten and Milgrom (1985) and as assumed in empirical work such as Easley, Kiefer, O'Hara and Paperman (1996), we conclude from our evidence that information based trading does not affect expected stock returns. Instead we interpret our evidence as indicating that while information asymmetry does not appear to be related to average returns, microstructure and liquidity effects unrelated to information asymmetry are still important for expected returns. It is worth noting that this interpretation relies on the assumption that periods of asymmetric information can be identified as periods with abnormal order flow imbalances. It is possible the relation between private information and order flow is more complex than the one implied by sequential trade models, in which case private information could indeed be related to expected returns. However, in this case, both PIN and $AdjPIN$ would be inappropriate proxies for information asymmetry. We leave further exploration of this matter to future research.

Appendix A - Database Construction

Quotes are limited from 8:30 am to 3:58 pm. We only consider quotes with non-zero bid price, non-zero ask price, and with ask price greater than bid price. We exclude quotes with large differences in bid and ask price. More specifically, we exclude a quote if the mid-point of the bid-ask spread falls within five and 50 dollars and the ratio of the bid-ask spread to the sum of the bid and ask price is greater than 0.25. We also exclude a quote if the mid-point of the bid-ask spread is greater than 50 dollars and the ratio of the bid-ask spread to the sum of the bid and ask price is greater than 0.1.

We only consider trades with positive price. We only consider regular trades that were not corrected, changed, or signified as cancelled or in error. We only consider trades made without any stated conditions. In particular, trades under the following stated conditions are excluded: split trade, pre- and post-market close trades, average price trades, sold sale, crossing session, etc. The trades and quotes data are matched together by symbol and time. To match the quotes with trades, we use the five-second quote rule, i.e., compare the trade with quotes that are at least five seconds older. The stock symbol from ISSM/TAQ is then matched with the CRSP symbol to get CRSP PERMNO for each stock. For a CRSP observation where the trading volume is zero, we assign a buys and sells a value of zero on that firm day. We only consider common stocks. We exclude ADRs, REITs, and closed-end funds.

Appendix B - Expected Value, Variance and Correlations of Buys and Sells

In this appendix, we sketch the proof of the formulas used to calculate the expected value of buys and sells, their variances, and covariances. The expected value of the number of buys in the original PIN model is equal to the sum of the probabilities of each branch of the tree in Figure 2 multiplied by the intensity of the buy trades arrival in the branch:

$$E[B] = (1 - a) \times \varepsilon_b + a \times (1 - d) \times \varepsilon_b + a \times d \times (\varepsilon_b + u). \tag{10}$$

Analogously, the expected value of the seller initiated trades in the original PIN model is:

$$E[S] = (1 - a) \times \varepsilon_s + a \times d \times \varepsilon_s + a \times (1 - d) \times (\varepsilon_s + u) \tag{11}$$

The variances of the number of buys is $\sigma^2[B] = E[B^2] - E[B]^2$, calculating $E[B^2]$ in the same way that we calculate $E[B]$ above and using the expression above for $E[B]$ we get:

$$\sigma^2[B] = \varepsilon_b^2 + a \times d \times u \times (1 + u) + \varepsilon_b \times (1 + 2 \times a \times d \times u) - \tag{12}$$

$$((1 - a) \times \varepsilon_b + a \times (1 - d) \times \varepsilon_b + a \times d \times (\varepsilon_b + u))^2$$

The variance of the seller initiated traded can be calculated in the same way and it is:

$$\sigma^2[S] = \varepsilon_s^2 - a \times (-1 + d) \times u \times (1 + u) + \varepsilon_s \times (1 - 2 \times a \times (-1 + d) \times u) - \tag{13}$$

$$((1 - a) \times \varepsilon_s + a \times d \times \varepsilon_s + a \times (1 - d) \times (\varepsilon_s + u))^2$$

The covariance between buys and sells is $cov[B, S] = E[B \times S] - E[B] \times E[S]$, calculating the expected value $E[B \times S]$ in the same way as above and using the above expressions for

$E[B]$ and $E[S]$ we get:

$$cov[B, S] = (au)^2 \times (d - 1) \times d \tag{14}$$

The method to calculate the expected value of buys and sells, their variances and co-variances is analogous to the one used in the original PIN model above. These expected values, variances and covariances in the unrestricted model are:

$$E[B] = \varepsilon_b + \Delta_b \times (\theta - a \times \theta + a \times \theta') + a \times d \times u_b \tag{15}$$

$$E[S] = \varepsilon_s + \Delta_s \times (\theta - a \times \theta + a \times \theta') - a \times (-1 + d) \times u_s \tag{16}$$

$$
\begin{aligned}
\sigma^2[B] \quad = \quad & \varepsilon_b - \Delta_b^2 \times ((-1 + a)^2 \times \theta^2 + a \times \theta' \times (-1 + a \times \theta') \\
& -(-1 + a) \times \theta \times (-1 + 2 \times a \times \theta')) \\
& +a \times d \times u_b \times (1 + u_b - a \times d \times u_b) \\
& +\Delta_b \times (a \times \theta' \times (1 - 2 \times (-1 + a) \times d \times u_b) \\
& +(-1 + a) \times \theta \times (-1 + 2 \times a \times d \times u_b))
\end{aligned} \tag{17}
$$

$$
\begin{aligned}
\sigma^2[S] \quad = \quad & \varepsilon_s - \Delta_s^2 \times ((-1 + a)^2 \times \theta^2 + a \times \theta' \times (-1 + a \times \theta') \\
& -(-1 + a) \times \theta \times (-1 + 2 \times a \times \theta')) \\
& -a \times (-1 + d) \times u_s \times (1 + (1 + a \times (-1 + d)) \times u_s) \\
& -\Delta_s \times ((-1 + a) \times \theta \times (1 + 2 \times a \times (-1 + d) \times u_s) \\
& +a \times \theta' \times (-1 + 2 \times (-1 + a + d - a \times d) \times u_s))
\end{aligned} \tag{18}
$$

$$cov[B,S] = (-a) \times d \times u_b \times ((-(-1+a)) \times \Delta_s \times (\theta - \theta'))$$

$$-a \times (-1+d) \times u_s)$$

$$-\Delta_b \times (\Delta_s \times ((-1+a)^2 \times \theta^2 + a \times \theta' \times (-1+a \times \theta') \qquad (19)$$

$$-(-1+a) \times \theta \times (-1+2 \times a \times \theta'))$$

$$+(-1+a) \times a \times (-1+d) \times (\theta - \theta') \times u_s)$$

The expected values of the buys and sells, their variances, and covariances in the restricted extended models follow by substituting the restrictions in the expressions above. For instance, substituting the restriction $\theta = \theta'$ in the expressions above, we arrive at the formulas for the expected values, variances and covariances of the buys and sells in the preferred extended model:

$$E[B] = \varepsilon_b + \Delta_b \times \theta + a \times d \times u_b \qquad (20)$$

$$E[S] = \varepsilon_s + \Delta_s \times \theta - a \times (-1+d) \times u_s \qquad (21)$$

$$\sigma^2[B] = \varepsilon_b + \Delta_b \times \theta - \Delta_b^2 \times (-1+\theta) \times \theta + a \times d \times u_b \times (1 + u_b - a \times d \times u_b) \qquad (22)$$

$$\sigma^2[S] = \varepsilon_s + \Delta_s \times \theta - \Delta_s^2 \times (-1+\theta) \times \theta - a \times (-1+d) \times u_s \times (1+u_s - a \times u_s + a \times d \times u_s) \qquad (23)$$

$$cov[B,S] = (-\Delta_b) \times \Delta_s \times (-1+\theta) \times \theta + a^2 \times (-1+d) \times d \times u_b \times u_s \qquad (24)$$

References

[1] Acharya, Viral V., and Lasse H. Pedersen, 2005, Asset pricing with liquidity risk, *Journal of Financial Economics* 77, 375-410.

[2] Admati, Anat, R., and Paul Pfleiderer, 1986, A theory of intraday patterns: volume and price variability, *Review of Financial Studies* 1, 3-40.

[3] Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31-56.

[4] Bachelier, Louis, 1900, Theorie de la Speculation (Thesis), Annales Scientifiques de l'École Normale Superieure III -17, 21-86. (English Translation in Cootner, P. H., 1964, *Random Character of Stock Market Prices*, MIT Press, 17-78).

[5] Bamber, Linda Smith, Orie E. Barron, and Thomas L. Stober, 1999, Differential interpretations and trading volume, *Journal of Financial and Quantitative Analysis* 34, 369-385.

[6] Bessembinder, Hendrik, Kalok Chan, and Paul J. Seguin, 1996, An empirical examination of information, differences of opinion, and trading activity, *Journal of Financial Economics* 40, 105-134.

[7] Boehmer, Ekkehart, Joachim Grammig and Erik Theissen, 2007, Estimating the probability of informed trading–does trade misclassification matter? *Journal of Financial Markets*, 10, 26-47.

[8] Chan, Wesley, 2003, Stock price reaction to news and no-news: drift and reversal after headlines, *Journal of Financial Economics* 70, 223-260.

[9] Constantinides, George M., 1986, Capital market equilibrium with transaction costs, *Journal of Political Economy* 94, 842-862.

[10] Duarte, Jefferson, Xi Han, Jarrad Harford, and Lance Young, 2006, Information Asymmetry, Information Dissemination and the Effect of Regulation FD on the Cost of Capital, *Journal of Financial Economics,* forthcoming.

[11] Easley, David and Maureen O'Hara, 1987, Price, trade size, and information in securities markets, *Journal of Financial Economics* 19, 69-90.

[12] Easley, David, and Maureen O'Hara, 2004, Information and the cost of capital, *Journal of Finance* 59, 1553-1583.

[13] Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, 1997, One day in the life of a very common stock, *Review of Financial Studies* 10, 805-835.

[14] Easley, David, Nicholas M. Kiefer, Maureen O'Hara, and Joseph B. Paperman, 1996, Liquidity, information, and infrequently traded stocks, *Journal of Finance* 51, 1405-1436.

[15] Easley, David, Soren Hvidkjaer, and Maureen O'Hara, 2002, Is Information Risk a Determinant of Asset Returns?, *Journal of Finance* 57, 2185-2221.

[16] Fama, Eugene F., and James D. MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Public Economy* 81, 607-636.

[17] Glosten, L. and P. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 13, 71-100.

[18] Harris, Milton and Artur Raviv, 1993, Differences of opinion make a horse race, *Review of Financial Studies* 6, 473-506.

[19] Hughes, John, Jing Liu, and Jun Liu, 2005, Private information, diversification, and asset pricing, *The Accounting Review,* forthcoming .

[20] Kandel, Eugene, and Neil D. Pearson, 1995, Differential interpretation of public signals and trade in speculative markets, *Journal of Political Economy* 103, 831-872.

[21] Lambert, Richard, Christian Leuz, and Robert E. Verrecchia, 2005, Accounting information, disclosure, and the cost of capital, University of Pennsylvania working paper.

[22] Lee, Charles M.C. and Mark J. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733-746.

[23] O'Hara, Maureen, 2003, Presidential address: liquidity and price discovery, *Journal of Finance* 58, 1335-1354.

[24] Pástor, Luboš, and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642-685.

[25] Sadka, Ronnie, 2005, Momentum and Post-Earnings-Announcement Drift Anomalies: The role of liquidity risk, *Journal of Financial Economics,* forthcoming.

[26] Vega, Clara, 2006, Stock price reaction to public and private information, *Journal of Financial Economics* 82, 103-133.

[27] Venter, Johannes H. and Dawid C.J. de Jongh, 2005, Extending the EKOP model to estimate the probability of informed trading, North-West University Center for Business Mathematics working paper.

Table 1 - Percentiles of summary statistics on the number of buyer and seller initiated trades. This table presents the median and the percentiles of the cross-sectional distribution of a series of statistics on the number of buys and sells of each stock in the sample. The number of buys and sells in a day for each stock is estimated with the Lee and Ready (1991) algorithm applied to intraday transactions and quotes data. The means, variances and the correlations between buys and sells of each stock are estimated within each calendar year.

	95th Percentile	75th Percentile	Median	25th Percentile	5th Percentile
Mean buys	322	37	10	3	0.4
Mean sells	266	35	10	3	0.3
Variance buys	14263	513	71	10	0.9
Variance sells	9805	379	54	8	0.7
Correlation between buys and sells	0.84	0.66	0.50	0.34	0.15

Table 2 - Summary of estimation results. Panel A presents the cross-sectional distribution of the estimated parameters of the original PIN model along with the cross-sectional distribution of the estimated probability of information trading (PIN). The t-statistics of the PINs are calculated with delta method based on the asymptotic covariance matrix of the estimated model parameters. 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.

A - Distribution of the estimated parameters of the original PIN model

	95th Percentile	75th Percentile	Median	25th Percentile	5th Percentile
a	0.46	0.35	0.26	0.18	0.08
u	193.84	44.78	17.53	7.04	2.13
d	0.98	0.82	0.69	0.54	0.26
ε_b	270.13	28.58	7.47	1.79	0.13
ε_s	245.15	31.07	8.96	2.49	0.19
PIN	0.51	0.28	0.20	0.15	0.10
t-statistic	15.05	11.65	9.14	6.96	4.44

B - Distribution of the moments implied by the Original PIN model

	95th Percentile	75th Percentile	Median	25th Percentile	5th Percentile
Mean buys	321	37	10	3	0.4
Mean sells	267	35	10	3	0.4
Variance buys	6581	297	48	8	0.8
Variance sells	3303	162	29	6	0.7
Correlation between buys and sells	0	-0.04	-0.08	-0.13	-0.22

Table 3 - The percentiles of the moments of the buys and sells implied by the extended PIN model. This table presents the median and the percentiles of the cross-sectional distribution of a series of moments of the buys and sells of each stock in the sample implied by the parameters estimated for the extended trading model. The implied moments are estimated within each calendar year. Each panel displays the results based on different versions of the extended model.

A - Model 1 ($u_b = u_s$, $\Delta_b = \Delta_s$, $\theta' = 0$)

	95th Percentile	75th Percentile	Median	25th Percentile	5th Percentile
Mean buys	322	37	10	3	0.4
Mean sells	270	35	11	3	0.4
Variance buys	8123	339	53	9	0.9
Variance sells	7740	320	49	8	0.9
Correlation between buys and sells	0.79	0.61	0.46	0.31	0.10

B - Model 2 ($u_b = u_s$, $\theta' = 0$)

	95th Percentile	75th Percentile	Median	25th Percentile	5th Percentile
Mean buys	320	37	10	3	0.4
Mean sells	266	35	10	3	0.3
Variance buys	9636	381	57	9	0.8
Variance sells	6401	268	41	7	0.7
Correlation between buys and sells	0.71	0.53	0.39	0.26	0.12

C - Model 3 ($\theta' = 0$)

	95th Percentile	75th Percentile	Median	25th Percentile	5th Percentile
Mean buys	321	37	10	3	0.4
Mean sells	266	35	10	3	0.3
Variance buys	9684	382	57	9	0.8
Variance sells	6337	269	42	7	0.7
Correlation between buys and sells	0.80	0.61	0.45	0.30	0.10

D - Model 4 ($\theta' = \theta$)

	95th Percentile	75th Percentile	Median	25th Percentile	5th Percentile
Mean buys	321	37	10	3	0.4
Mean sells	266	35	10	3	0.3
Variance buys	10973	393	57	9	0.8
Variance sells	7005	283	43	7	0.7
Correlation between buys and sells	0.67	0.50	0.37	0.25	0.11

E - Model 5 (unrestricted extended model)

	95th Percentile	75th Percentile	Median	25th Percentile	5th Percentile
Mean buys	321	37	10	3	0.4
Mean sells	266	35	10	3	0.3
Variance buys	10350	391	58	9	0.9
Variance sells	6758	281	43	7	0.7
Correlation between buys and sells	0.66	0.50	0.37	0.25	0.10

Table 4 - Likelihood ratio tests. This table presents the frequency that the null hypothesis of a series of likelihood ratio tests is rejected at the 5% significance level. The test statistics of tests number 2, 3 and 5 are chi-square distributed with one degree of freedom. The test statistics of tests number 1 and 4 do not have standard assymptotic distribution and hence the distribution of the tests are estimated with Monte Carlo simulations. Tests 2, 3 and 5 are done over the entire sample. Tests 1 and 4 are performed on a sample of 100 randomly selected stocks.

Test Number	Null Hypothesis	Alternative Hypothesis	Frequency of Rejection of Null
1	Original PIN model	Extended Model with the restrictions $u_b = u_s$, $\Delta_b = \Delta_s$, $\theta' = 0$	0.99
2	Extended model with the restrictions $u_b = u_s$, $\Delta_b = \Delta_s$, $\theta' = 0$	Extended Model with the restrictions $u_b = u_s$, $\theta' = 0$	0.53
3	Extended model with the restrictions $u_b = u_s$, $\theta' = 0$	Extended Model with the restriction $\theta' = 0$	0.23
4	Extended model with the restriction $\theta' = 0$	Unrestricted extended model	0.79
5	Extended model with the restriction $\theta' = \theta$ (Preferred Model)	Unrestricted extended model	0.09

Table 5 - Summary of results of estimation of preferred extended model. This table presents the cross-sectional distribution of the estimated parameters of the preferred extended model along with the cross-sectional distribution of the estimated probability of information trading in the extended model (AdjPIN) and the cross-sectional distribution of the estimated probability of symmetric order flow shock (PSOS). The preferred extended model is the extended model with the constraint $\theta'=\theta$. The t-statistics of the PINs and PSOSs are calculated with delta method based on the asymptotic covariance matrix of the estimated model parameters.

	95th Percentile	75th Percentile	Median	25th Percentile	5th Percentile
a	0.62	0.48	0.39	0.28	0.08
u_b	153	32	13	5	1.0
u_s	148	33	13	4	0.8
d	0.98	0.78	0.56	0.32	0.03
ε_b	212	21	5	1	0
ε_s	185	23	6	1	0
θ	0.55	0.36	0.25	0.14	0.04
Δ_b	188	39	13	4	1.2
Δ_s	156	31	11	4	1.1
AdjPIN	0.37	0.23	0.17	0.13	0.08
t-statistic	15.86	11.85	8.47	5.70	3.03
PSOS	0.66	0.38	0.25	0.18	0.12
t-statistic	16.24	10.41	7.51	5.24	2.50

Table 6 - Correlation between variables of interest. The displayed correlation matrix is calculated with the entire cross-section of NYSE and AMEX stocks. PIN is a measure of probability of information-related trading estimated with the Easley, Hvidkjaer and O'hara (2002) trading model. AdjPIN is the probability of information-related trading estimated with the preferred extended model. PSOS is the probability of symmetric order flow shocks. PINs, AdjPINs and PSOSs are estimated for each calendar year from 1983 to 2004. SPREAD is the average of the daily absolute open bid-ask spread of each stock calculated over the entire calendar year. SIZE is the log of market value of firm equity in millions for each December. ILLIQ is the Amihud (2002) illiquidity measure for year t-1. P-values are between parentheses.

	PIN	AdjPIN	PSOS	SPREAD	SIZE	BM	ILLIQ
PIN	1.0000	0.709	0.711	0.154	-0.659	0.269	0.188
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
AdjPIN		1.0000	0.341	0.117	-0.582	0.251	0.106
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
PSOS			1.0000	0.151	-0.578	0.182	0.178
				(0.000)	(0.000)	(0.000)	(0.000)
SPREAD				1.0000	-0.214	0.100	0.190
					(0.000)	(0.000)	(0.000)
SIZE					1.0000	-0.441	-0.221
						(0.000)	(0.000)
BM						1.0000	0.129
							(0.000)
ILLIQ							1.0000

Table 7 - Trading activity on days with news and on days without news. This table presents the means and standard deviations of a series of variables related to trading activity on days with news and on days without news. PSOS is the probability of symmetric order flow shocks. PIN is a measure of probability of information-related trading estimated with the Easley, Hvidkjaer and O'hara (2002) trading model. AdjPIN is the probability of information-related trading estimated with the preferred extended model. PINs, AdjPINs and PSOSs are estimated for each calendar year from 1983 to 2004. BUYS is the number of buyer initiated trades on a given day. SELLS is the number of seller initiated trades. IMBALANCE is the difference between BUYS and SELLS. TRADES is total number of trades in a day. ILLIQ is the Amihud (2002) illiquidity measure. Days with news are days in which a given firm is mentioned at least once in the publications covered by Dow Jones Interactive Publications Library. Stocks are sorted on their PSOS and the means and standard deviations of the trading related variables are calculated for each PSOS decile. Standard deviations are between parentheses.

PSOS Decile	PSOS	PIN	Adj.PIN	BUYS	SELLS	BUYS x SELLS	IMBALANCE	TRADES	ILLIQ x 100
						DAYS WITH NEWS			
1	0.11	0.15	0.16	94	86	38234	8	181	0.002
	(0.02)	(0.06)	(0.08)	(194)	(162)	(2454143)	(61)	(352)	(0.128)
2	0.16	0.15	0.15	96	86	24877	10	182	0.001
	(0.01)	(0.05)	(0.07)	(145)	(122)	(117774)	(53)	(263)	(0.021)
3	0.18	0.16	0.15	79	70	20116	9	149	0.001
	(0.01)	(0.06)	(0.07)	(139)	(111)	(104017)	(48)	(246)	(0.012)
4	0.21	0.17	0.16	67	60	15015	7	128	0.002
	(0.01)	(0.05)	(0.06)	(118)	(98)	(343047)	(42)	(214)	(0.059)
5	0.24	0.20	0.18	76	68	130111	9	144	0.006
	(0.01)	(0.07)	(0.09)	(407)	(315)	(5360897)	(124)	(718)	(0.219)
6	0.27	0.21	0.19	107	91	290159	16	198	0.003
	(0.01)	(0.07)	(0.08)	(617)	(468)	(7033388)	(198)	(1077)	(0.071)
7	0.32	0.24	0.20	31	27	5338	4	59	0.008
	(0.02)	(0.07)	(0.09)	(76)	(63)	(92762)	(27)	(137)	(0.122)
8	0.39	0.30	0.25	44	37	15836	7	80	0.021
	(0.02)	(0.12)	(0.12)	(146)	(105)	(252815)	(62)	(247)	(0.395)
9	0.50	0.37	0.28	98	83	304485	15	181	0.058
	(0.04)	(0.13)	(0.11)	(578)	(520)	(6280817)	(105)	(1094)	(0.945)
10	0.73	0.41	0.23	127	106	71118	21	233	0.089
	(0.12)	(0.13)	(0.09)	(268)	(220)	(254637)	(70)	(485)	(1.637)
						DAYS WITHOUT NEWS			
1	0.11	0.15	0.16	115	102	76432	13	217	0.005
	(0.02)	(0.06)	(0.08)	(273)	(242)	(362186)	(61)	(512)	(0.155)
2	0.16	0.15	0.15	142	124	79774	18	267	0.003
	(0.01)	(0.05)	(0.07)	(268)	(238)	(301207)	(67)	(503)	(0.056)
3	0.18	0.16	0.15	132	114	70847	19	246	0.002
	(0.01)	(0.06)	(0.07)	(257)	(224)	(438455)	(66)	(477)	(0.054)
4	0.21	0.17	0.16	104	89	98685	15	192	0.007
	(0.01)	(0.05)	(0.06)	(321)	(283)	(3142570)	(67)	(602)	(0.436)
5	0.24	0.20	0.18	64	54	35845	9	118	0.013
	(0.01)	(0.07)	(0.09)	(199)	(168)	(971293)	(58)	(365)	(0.600)
6	0.27	0.21	0.19	69	61	135712	8	130	0.011
	(0.01)	(0.07)	(0.08)	(388)	(347)	(4654363)	(88)	(731)	(0.230)
7	0.32	0.24	0.20	35	30	10088	5	66	0.025
	(0.02)	(0.07)	(0.09)	(106)	(89)	(141160)	(33)	(193)	(0.562)
8	0.39	0.30	0.25	25	22	95261	3	47	0.066
	(0.02)	(0.12)	(0.12)	(334)	(285)	(6595427)	(60)	(618)	(1.456)
9	0.50	0.37	0.28	15	13	17074	2	28	0.111
	(0.04)	(0.13)	(0.11)	(139)	(123)	(1072208)	(28)	(261)	(2.098)
10	0.73	0.41	0.23	14	12	10537	2	26	0.267
	(0.12)	(0.13)	(0.09)	(109)	(96)	(150114)	(22)	(205)	(5.448)

Table 8 - PSOS and Liquidity. This table contains monthly time-series averages of the estimated coefficients in cross-sectional regressions of PSOS on a series of stock characteristics related to liquidity. PSOS is the probability of symmetric order-flow shocks estimated with the preferred extended model. PSOSs are estimated for each calendar year from 1983 to 2004. SIZE is the logarithm of the December market equity for year t-1. TURNOVER is the logarithm of the monthly average turnover. ILLIQ is the Amihud (2002) illiquidity measure. CV_TURNOVER is the logarithm of the monthly coefficient of variation of turnover. SPREAD is the average of the daily absolute open bid-ask spread of each stock calculated over the entire calendar year. The t-statistics are between parentheses.

Explanatory Variables	(1)	(2)
Intercept	0.5779	2.0343
	(9.771)	(15.854)
SIZE	-0.0355	-0.2228
	(-9.839)	(-16.507)
TURNOVER	-0.0457	-0.0585
	(-19.196)	(-19.717)
ILLIQ X 100	0.0014	0.0017
	(2.536)	(2.210)
CV_TURNOVER	0.0291	-0.0856
	(6.881)	(-3.112)
SPREAD	0.0072	-0.0001
	(4.421)	(-0.076)
$SIZE^2$		0.0072
		(14.458)
$TURNOVER^2$		0.0174
		(10.637)
$ILLIQ^2 X 10^6$		-0.0023
		(-2.758)
$CV\_TURNOVER^2$		0.0125
		(4.553)
$SPREAD^2$		0.0001
		(0.941)

Table 9 - Factors affecting the difference between PIN and AdjPIN. This table contains monthly time-series averages of the estimated coefficients in cross-sectional regressions of PIN-AdjPIN and AdjPIN on a series of stock characteristics. PIN is a measure of probability of information-related trade estimated using the Easley, Hvidkjaer and O'Hara (2002) trading model. AdjPIN is the probability of information-related trade estimated with the preferred extended model. PSOS is the probability of symmetric order-flow shocks estimated with the preferred extended model. PINs, AdjPINs and PSOSs are estimated for each calendar year from 1983 to 2004.   SIZE is the logarithm of the December market equity for year t-1. TURNOVER is the logarithm of the average monthly turnover. The t-statistics are between parentheses.

Explanatory Variables	Dependent Variables			
	PIN-AdjPIN	AdjPIN	PIN-AdjPIN	
Intercept	0.0412	0.0698	-0.0313	0.0965
	(35.202)	(19.184)	(-25.168)	(3.584)
PIN		0.5205		
		(49.891)		
PSOS			0.2769	-0.0894
			(41.511)	(-5.402)
SIZE				-0.0116
				(-3.110)
TURNOVER				-0.0149
				(-11.486)
ILLIQ X 100				0.0002
				(2.163)
$PSOS^2$				0.4311
				(23.982)
$SIZE^2$				0.0004
				(3.288)
$TURNOVER^2$				0.0073
				(11.393)

Table 10 - PIN and the cross-section of expected returns. This table contains monthly time-series averages of the estimated coefficients in cross-sectional asset-pricing tests. The dependent variable is the stock monthly return. Betas are post-ranked betas estimated using 40 portfolios. SIZE is the logarithm of the December market equity for year t-1, BM is the logarithm of book value divided by market value for year t-1. PIN is a measure of probability of information-related trade estimated using the Easley, Hvidkjaer and O'Hara (2002) trading model. AdjPIN is the probability of information-related trade estimated with the preferred extended model. PSOS is the probability of symmetric order flow shocks estimated with the preferred extended model.   PINs, AdjPINs and PSOSs are estimated for each calendar year from 1983 to 2004.   ILLIQ is the Amihud(2002) illiquidity measure for year t-1. The t-statistics are inside parentheses.

Explanatory Variables	Fama-MacBeth regressions												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Beta	0.1745	0.1397	0.1775	0.1929	0.1926	0.1391	0.1486	0.1529	0.1532	0.1547	0.1775	0.1809	0.1427
	(0.494)	(0.388)	(0.506)	(0.520)	(0.546)	(0.369)	(0.425)	(0.441)	(0.418)	(0.444)	(0.506)	(0.483)	(0.414)
SIZE	0.0430	0.0043	0.0354	0.0372	0.0487	0.0647	0.0881	0.0883	0.0822	0.0902	0.0354	0.0186	0.0835
	(0.630)	(0.064)	(0.501)	(0.639)	(0.733)	(1.216)	(1.379)	(1.337)	(1.497)	(1.436)	(0.501)	(0.324)	(1.255)
BM	0.2675	0.2558	0.2581	0.2661	0.2645	0.2554	0.2542	0.2443	0.2457	0.2451	0.2581	0.2613	0.2461
	(2.956)	(2.824)	(2.869)	(2.940)	(2.953)	(2.802)	(2.824)	(2.732)	(2.737)	(2.758)	(2.869)	(2.869)	(2.750)
PIN	1.0043		1.3650		0.7455		0.6481	0.6277		0.6453	0.6108		0.5224
	(1.911)		(2.875)		(1.141)		(1.168)	(1.266)		(0.957)	(0.785)		(0.616)
AdjPIN		0.1324	-0.7541					0.0147					
		(0.182)	(-1.092)					(0.024)					
PSOS				0.6448	0.3731				0.2455	0.0199			
				(2.587)	(1.215)				(0.998)	(0.071)			
ILLIQ						0.0003	0.0003	0.0003	0.0003	0.0003			0.0003
						(2.989)	(2.696)	(2.543)	(2.738)	(2.581)			(2.583)
AdjPIN-PIN											-0.7541	-1.1302	-0.0290
											(-1.092)	(-2.650)	(-0.047)

Figure 1 - Example of number of buyer and seller initiated trades in the sample. This figure diplays the plots of the daily number of buyer and seller initiated trades for three different stock-years in the sample. The number of buyer and seller initiated trades are calculated by Lee and Ready (1991) algorithm. The stock-year in the top panel is the one that has an estimated correlation between buys and sells equal to the 25th percentile of the correlations in the sample. The stock-year in the second panel has an estimated correlation between buys and sells equal to median correlation and the stock-year in the third panel has a correlation between buys and sells equal to the 75th percentile.
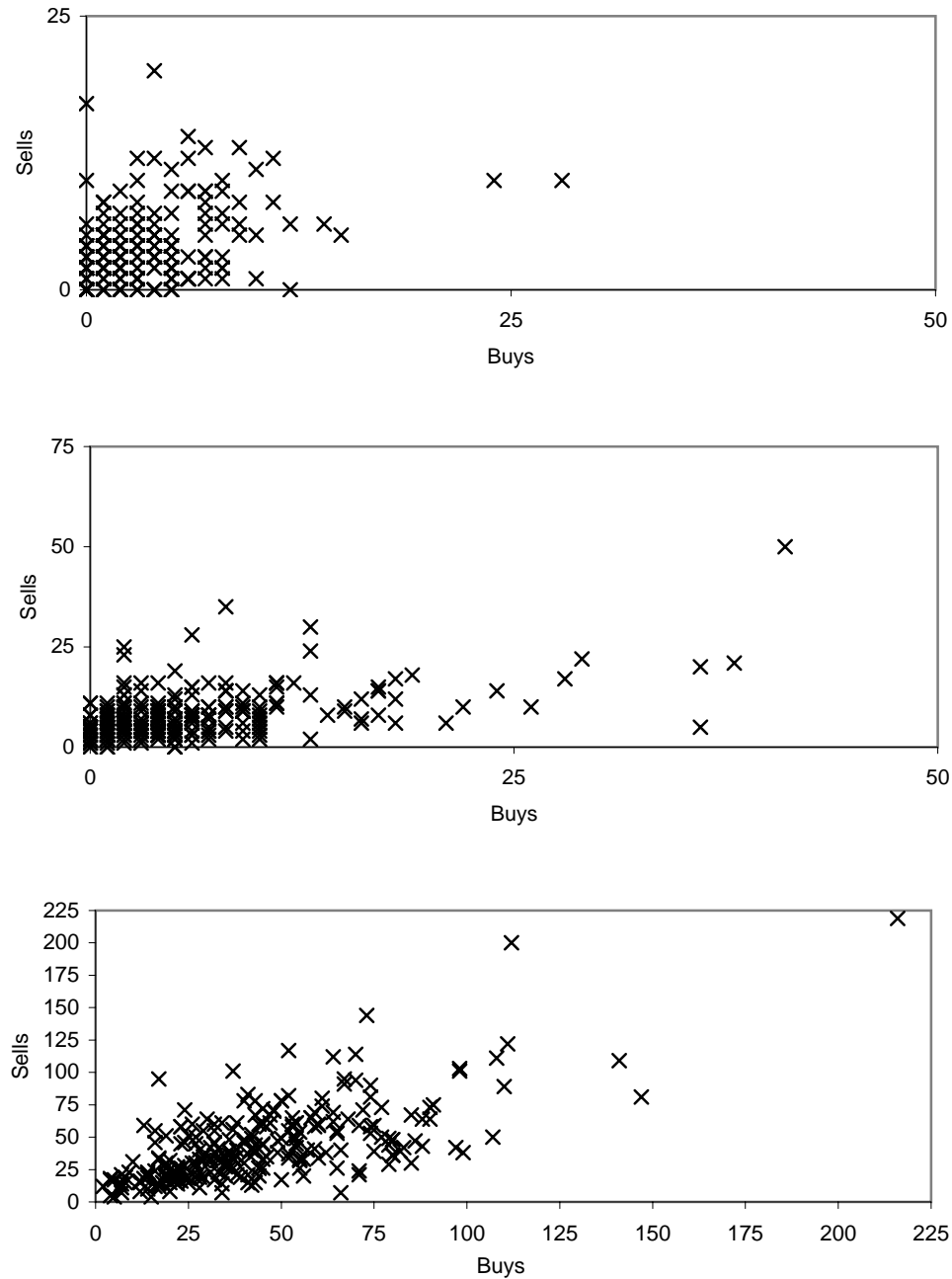
Figure 2 - Tree of the original PIN trading model. This tree represents the Easley, Hvidkjaer and O'hara (2002) trading model. In any given trading day, private information arrives with probability "a". When there is no private information, the number of buyer initiated trades is Poisson distributed with intensity ε_b and the number of seller initated trades is Poisson distributed with intensity ε_s. Positive private information induces an increase in the number of buyer initiated trades, as a result the number of buyer initiated trades is Poisson distributed with intensity $\varepsilon_b + u$ when there is positive private information. Analagously, the number of seller initiated trades is Poisson distributed with intensity $\varepsilon_s + u$ when there is negative private information.
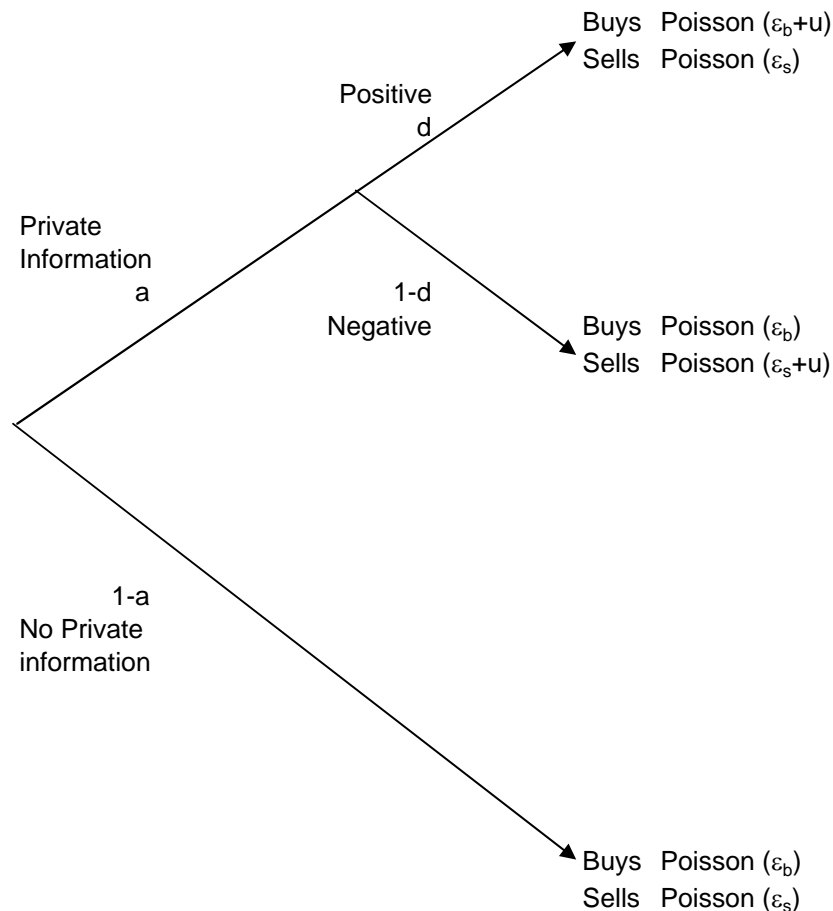

Positive
d

Buys Poisson (ε_b+u)
Sells Poisson (ε_s)

Private
Information
a

1-d
Negative

Buys Poisson (ε_b)
Sells Poisson (ε_s+u)

1-a
No Private
information

Buys Poisson (ε_b)
Sells Poisson (ε_s)

Figure 3 - Simulation of the original PIN trading model. The plots of the daily number of buys and sells on this figure are the result of simulated daily observations from the original PIN trading model. Days with positive private information are represented by a positive sign (see the cloud of observations on the right bottom corner of each plot). Days with negative private information are represented by negative signs (see the cloud of observations on the left top corner of the each plot). Days with only noise trading are represented by circles (see the cloud of observations on the left bottom corner of each plot). The parameters used in the simulation displayed in the first graph are estimated from the sample displayed in the first panel of Figure 1, the parameters used in the simulation in the second panel are estimated from the sample displayed in the second panel of Figure 1, and the parameters used in the third panel are estimated from the sample displayed in the third panel of Figure1.
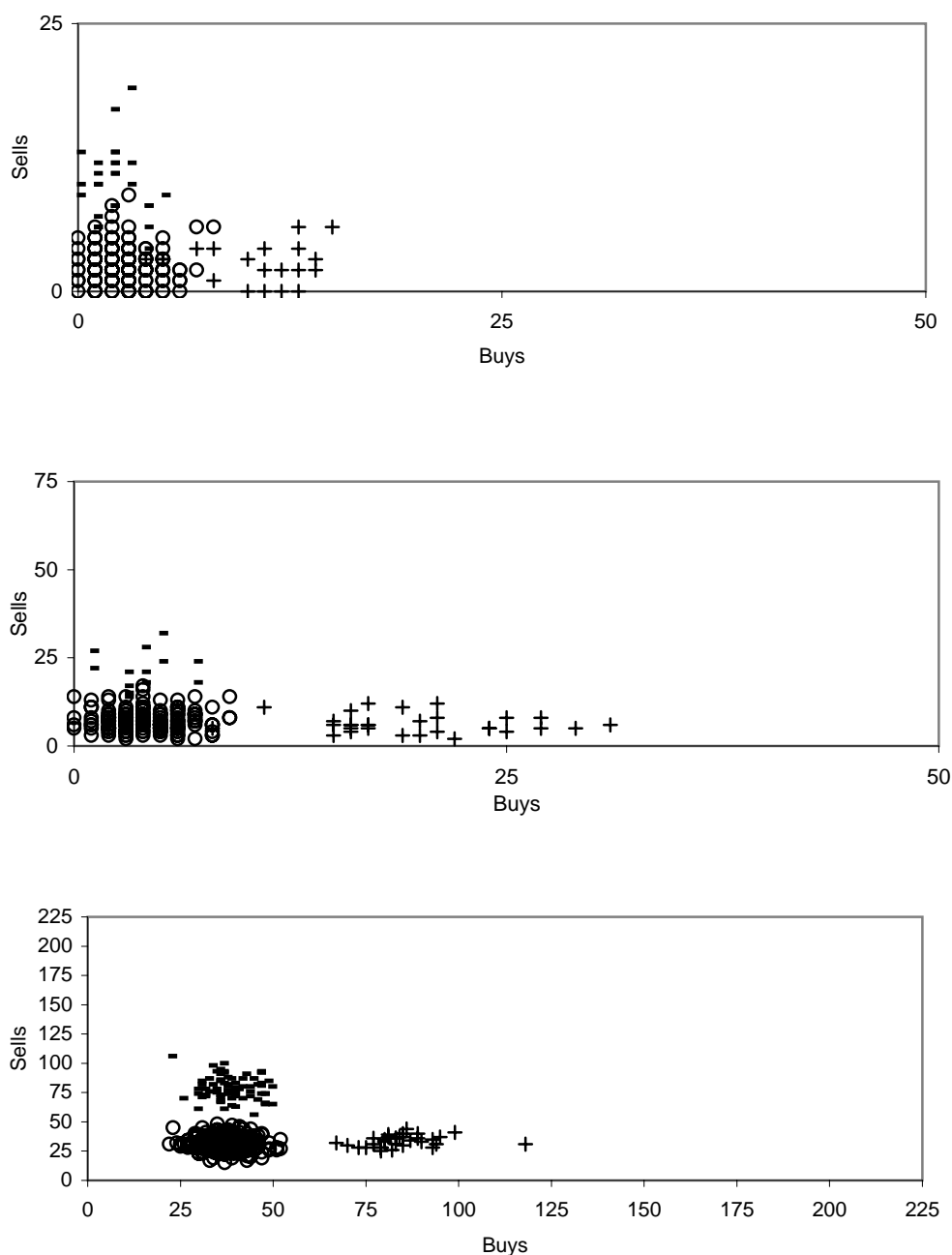
Figure 4 - Tree of the unrestricted extended trading model. The tree represents the extended trading model. The difference between this model and the Easley, Hvidkjaer and O'hara (2002) trading model is twofold. First, the extended model has additional branches for days in which both the number of buys and sells increase (symmetric order flow shocks). These days happen with probability θ' on the days with private information and with probability θ on the days without private information. Second, the number of buyer initiated informed trades has a different distribution from the number of seller initiated trades.
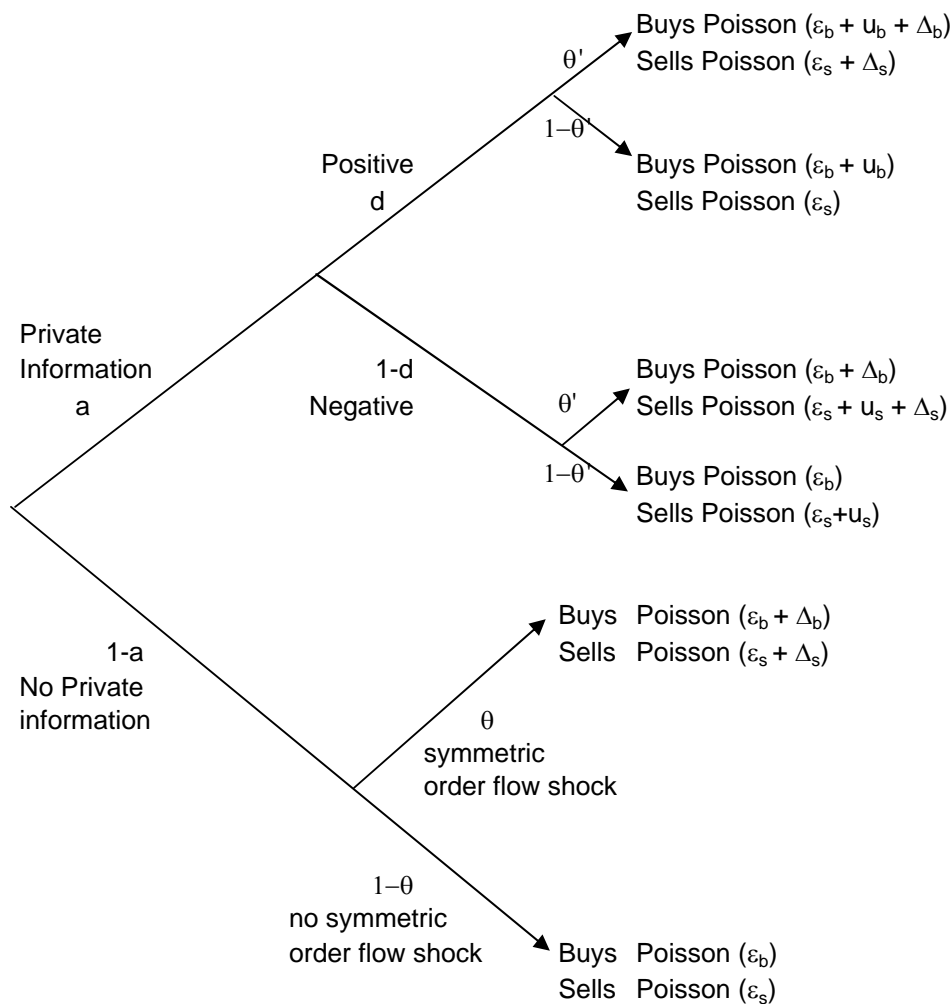
Private Information
a

1-a
No Private information

Positive
d

1-d
Negative

θ'

$1-\theta'$

θ'

$1-\theta'$

Buys Poisson $(\varepsilon_b + u_b + \Delta_b)$
Sells Poisson $(\varepsilon_s + \Delta_s)$

Buys Poisson $(\varepsilon_b + u_b)$
Sells Poisson (ε_s)

Buys Poisson $(\varepsilon_b + \Delta_b)$
Sells Poisson $(\varepsilon_s + u_s + \Delta_s)$

Buys Poisson (ε_b)
Sells Poisson $(\varepsilon_s + u_s)$

θ
symmetric
order flow shock

$1-\theta$
no symmetric
order flow shock

Buys Poisson $(\varepsilon_b + \Delta_b)$
Sells Poisson $(\varepsilon_s + \Delta_s)$

Buys Poisson (ε_b)
Sells Poisson (ε_s)

Figure 5 - Simulation of the preferred extended trading model. The plots of the daily number of buys and sells on this figure are the result of simulated daily observations from the preferred extended trading model. Days with positive private information are represented by a positive sign (see the cloud of observations on the right bottom corner of each plot). Days with positive private information and symmetric order flow shocks are represented by squares. Days with negative private information are represented by negative signs (see the cloud of observations on the left top corner of the each plot). Days with negative private information and symmetric order flow shocks are represented by triangles. Days with symmetric order flow shocks and no private information are represented by asteristics. Days with only noise trading are represented by circles (see the cloud of observations on the left bottom corner of each plot). The parameters used in the simulations are estimated from the samples displayed in Figure 1.
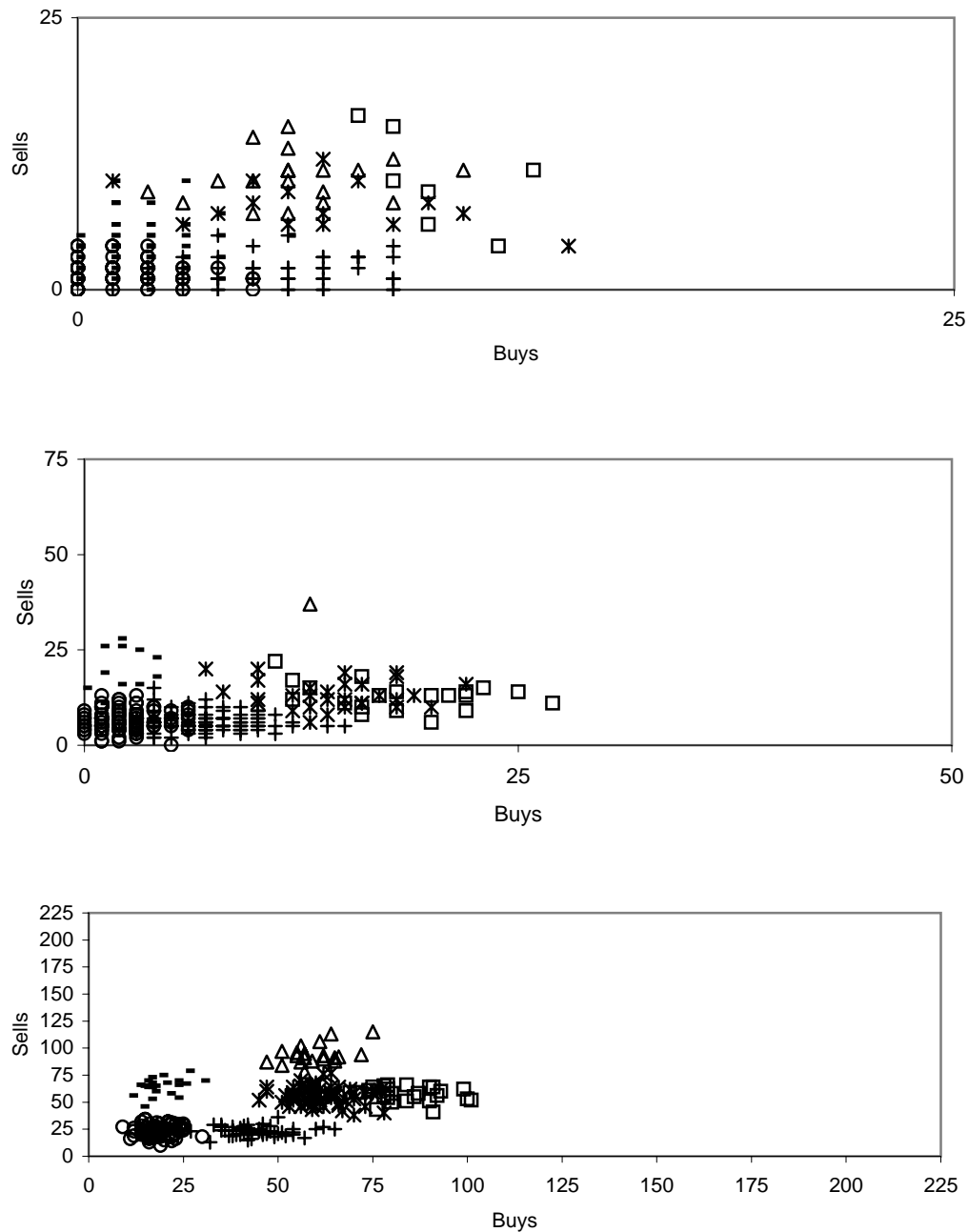
Figure 6 - Time series plots of PIN and AdjPIN. These figures plot some of the properties of the cross-sectional distribution of PIN and AjdPIN. PIN is the probability of private information related trade from the Easley, Hvidkjaer and O'hara (2002) model. AdjPIN is the probability of private information related trade from the preferred extended model. PSOS is the probability that a given trade happens during a symmetric order flow shock.