Lectures on Market Microstructure
Illiquidity and Asset Pricing

Ingrid M. Werner
Martin and Andrew Murrer Professor of Finance
Fisher College of Business, The Ohio State University
Liquidity and Asset Pricing

- Liquidity and Asset Pricing
  - Liquidity
  - Expected liquidity
  - Liquidity Risk
- Probability of informed trading (PIN)
  - Estimating a sequential trade model
- Information and Asset Pricing
  - PIN and asset prices
Introduction

- Suppose that you buy a stock today and the quotes are $99-$101 => you pay $101 (using a MO).
- If the stock (mid quote) price increases by $5 (5%), and the spread is constant in $ terms, the price at the end of the year is $104-$106.
- If you sell the stock you will receive $104 (using a MO).
- In other words, you return is $(104-101) = 3 < 5 = (106-101)$

- **Key Insight:**
  (II)Liquidity affects asset prices.
Longstaff (2004)
Liquidity and Bond Pricing

• Longstaff (2002) separates default from liquidity premia by comparing U.S. Treasury with Refcorp bonds.
• The principal repayment for Refcorp bonds is fully collateralized by Treasury Bonds, and the Treasury guarantees full payment of the coupons.
• Refcorp bonds are identical to Treasury Bonds aside from any illiquidity issues.
• Average premia of Refcorp compared to Treasury Bonds range from 10 to 16 basis points of yield.
• The premia vary significantly over time.
• Translate into large price spreads for long-maturity bonds => 15%...
Amihud and Mendelson (1986)  
Liquidity and Asset Prices

• The previous example suggests that transactions costs (liquidity) should affect asset prices.

• Model assumes a perpetual per-period dividend, $d$, a required risk adjusted return, $r$, a relative bid ask spread $S$, and an expected trading frequency of $\mu$.

\[
P = \frac{d}{r + \mu S}
\]

• Generalized to account for risk, the expected risk-adjusted return (using the CAPM) is:

\[
E(R) = r_f + \beta (E(R_m) - r_f) + \mu S
\]

where $E(R)$: Expected return on the market portfolio, $r_f$: Risk-free rate, $\beta$: Covariance of market returns and security returns, $\mu$: Mean return on the market portfolio, $S$: Relative bid ask spread.

Compensation for Trading costs => “Liquidity premium”
Amihud and Mendelson (1986)
Liquidity and Asset Prices

• Empirical tests are based on data for NYSE stocks during 1960-1979.
  – Calculate relative bid-ask spread for each stock monthly
  – Estimate Market Model beta for each stock monthly
  – Form portfolios based on spreads and beta, and perform cross-sectional regressions of average monthly portfolio return on the portfolio spread and beta.

\[
\bar{R}_p = 0.36 + 0.672 \cdot \beta_p + 0.211 \cdot S_p + u_p
\]

*Both beta and spreads affect the cross-section of stock returns!*

0.211 => stocks are traded once every five months
If included, \((S_p)^2\) does not have a significant coefficient...
Amihud (2002)
Illiquidity and Asset Prices

• Spread data is not always available for long time series.
• Desirable to develop proxies for trading costs that do not depend on intraday data => ILLIQ.
• The inspiration of ILLIQ comes from Kyle’s lambda.
  – A stock with high liquidity is expected to have a low price impact per unit traded.
  – A stock with low liquidity is expected to have a high price impact per unit traded.
• ILLIQ is computed as the average daily absolute price change, |ΔP|, divided by trading volume, V.
• Use CRSP daily data for 1964-1997 to test the predictions:
  \[ ILLIQ_t = \frac{1}{D} \sum_{d=1}^{D} \frac{|\Delta P_t|}{V_t} \]
  \[ \bar{R}_{pt} = -0.364 + 1.183 \cdot \beta_{pt} + 0.162 \cdot ILLIQ_{pt} + u_{pt} \]
  \[ \bar{R}_{pt} = 1.922 + 0.217 \cdot \beta_{pt} + 0.112 \cdot ILLIQ_{pt} - 0.134 \ln(SIZE)_{pt} + u_{pt} \]

Illiquidity is priced!
Conclusions are robust to controlling for size.
Brennan and Subrahmanyam (1996)  
Price Impact and Asset Prices

- Do privately informed investors impose illiquidity costs on uninformed investors that are significant enough to affect asset prices?
- ISSM data for 1984-1988 to estimate Kyle’s lambda
  - Rely on Glosten & Harris (1988) and Hasbrouck (1991)
- Construct portfolios sorted by estimate price impacts and size $\Rightarrow$ 25 liquidity quintiles, $L_i$.
- Add these to the standard FF (1993) 3-factor model.

$$R_{it} = \alpha + \sum_{i=2}^{25} \lambda_i L_i + \beta_i R_{Mt} + s_i SMB_t + h_i HML_t + e_{it}$$

Where $R_M$ is the return on the market portfolio, SMB is the return on the small minus big (size) portfolios, and HML is the return on the high minus low book-to-market (B/M) portfolios.

The coefficients $\lambda$ measure the incremental return for the liquidity portfolios.

Results show that an additional return of 6.6% per year is required for the Lowest liquidity portfolio compared to the highest liquidity portfolio!
Chordia, Roll and Subrahmanyam (2001)  
Market Liquidity and Trading

  – When market returns are positive, spreads decline, depth increases, and volume increases.
  – When market returns are negative, spreads widen, depth declines, and volume increases.
Commonality in Liquidity
Chordia et al (2000) and Hasbrouck and Seppi (2001)

Table 9
Individual liquidity determinants and industry commonality

Individual stock liquidity measures (levels) are regressed cross-sectionally each trading day on the standard deviation of individual daily returns from the preceding calendar month (STD), the concurrent day’s mean price level (PRICE), the day’s dollar trading volume (DVOL), and an equally-weighted liquidity measure of all stocks in the same industry (INDUSTRY). The INDUSTRY observation corresponding to an individual stock excluded that stock. Natural logarithmic transformations are used for all variables. Cross-sectional coefficients are then averaged across the 254 trading days in the sample and are reported with t-statistics in parentheses. QSPR is the quoted spread. PQSPR is the proportional quoted spread. DEP is quoted depth. ESPR is the effective spread. PESPR is the proportional effective spread. The $R^2$ is adjusted.

<table>
<thead>
<tr>
<th></th>
<th>QSPR</th>
<th>PQSPR</th>
<th>DEP</th>
<th>ESPR</th>
<th>PESPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>STD</td>
<td>0.1268</td>
<td>0.1171</td>
<td>-0.1372</td>
<td>0.1295</td>
<td>0.1218</td>
</tr>
<tr>
<td>$t$</td>
<td>(45.41)</td>
<td>(35.54)</td>
<td>(-17.45)</td>
<td>(32.49)</td>
<td>(27.98)</td>
</tr>
<tr>
<td>PRICE</td>
<td>0.3738</td>
<td>-0.6215</td>
<td>-0.9010</td>
<td>0.3296</td>
<td>-0.6669</td>
</tr>
<tr>
<td>$t$</td>
<td>(108.8)</td>
<td>(-164.8)</td>
<td>(-103.2)</td>
<td>(54.96)</td>
<td>(-101.9)</td>
</tr>
<tr>
<td>DVOL</td>
<td>-0.0669</td>
<td>-0.0670</td>
<td>0.4127</td>
<td>-0.0523</td>
<td>-0.0525</td>
</tr>
<tr>
<td>$t$</td>
<td>(-33.17)</td>
<td>(-33.99)</td>
<td>(129.4)</td>
<td>(-42.06)</td>
<td>(-43.23)</td>
</tr>
<tr>
<td>INDUSTRY</td>
<td>0.3333</td>
<td>0.1871</td>
<td>0.2737</td>
<td>0.2428</td>
<td>0.1413</td>
</tr>
<tr>
<td>$t$</td>
<td>(30.75)</td>
<td>(29.49)</td>
<td>(13.11)</td>
<td>(29.63)</td>
<td>(30.36)</td>
</tr>
<tr>
<td>$R^2$ mean</td>
<td>0.290</td>
<td>0.810</td>
<td>0.432</td>
<td>0.216</td>
<td>0.735</td>
</tr>
<tr>
<td>Median</td>
<td>0.288</td>
<td>0.806</td>
<td>0.422</td>
<td>0.208</td>
<td>0.733</td>
</tr>
</tbody>
</table>

*Note: $t$ denotes t-statistic corrected for first-order auto-correlation.
This is similar to the Fama and MacBeth (1973) method for returns. The eight industry classifications follow Roll (1992) and Chalmers and Kadlec (1998).
Liquidity (Risk) and Asset Prices

- Pastor and Stambaugh (2003) investigate whether market-wide liquidity is a state variable for asset pricing.
- Acharya and Pedersen (2005) develop a theoretical model which delivers both a liquidity premium (expected liquidity) and a liquidity risk-premium (covariance between liquidity and returns).
  - Emphasizes that the permanent price impact component, rather than the transitory cost component, is the priced liquidity factor.
Estimation: Stylized model…

• Define unexpected liquidity as $U_t = L_t - E_{t-1}(L_t)$

• Regress individual stock returns on market returns and the unexpected shock to liquidity. $R_{it}^e = \alpha_t + \beta_i r_{mt} + \delta_i U_t + \epsilon_{it}$

• Expected returns: $E(R_i^e) = \beta_i E(R_m^e) + \delta_i \lambda_L$

• Adding expected liquidity: $E(R_i^e) = \beta_i E(R_m^e) + \delta_i \lambda_L + \mu E(L_i)$

• Pastor and Stambaugh (2003) estimate the following (GMM): $R_{it}^e = \beta_i R_{mt}^e + \delta_i (U_t + \lambda_L) + \epsilon_{it}$

• Acharya and Pedersen (2005) estimate the following cross-sectional regression (second step in the augmented two-step procedure, using 25 portfolios sorted by monthly ILLIQ): $\hat{r}_i = \lambda_m \hat{\beta}_i + \lambda_L \hat{\delta}_i + \mu \overline{L}_i + u_i$
Results

• Pastor and Stambaugh (2003) estimate an expected return differential of 7.5% between the high liquidity-sensitive stock portfolio and the low liquidity-sensitive stock portfolio.
• Sadka (2006) find a liquidity risk premium of 5%-6% per annum, measured as the difference in expected return between a high liquidity-exposure portfolio and a low liquidity-exposure portfolio.
• Acharya and Pedersen (2005) find that there is a difference between the highest and lowest liquidity portfolio return corrected for the other risk factors, of 4.6% per year, of which 3.5% is compensation for expected liquidity and the remaining 1.1% is compensation for liquidity risk.
• Including the expected liquidity makes a difference…
Measurement Issues

• Microstructure would suggest bid-ask spreads and/or some sort of price impact measure (Kyle’s lambda).
  – Short time series
  – Cumbersome

• CRSP data suggest:
  – Roll’s spread
  – Amivest liquidity ratio
  – Amihud’s ILLIQ
  – Pastor and Stambaugh’s reversal measure
  – Liu’s Zero volume LMx

• Hasbrouck (2009)
  – Horse-race: which daily measure best proxies for intradaily data?
  – Gibbs estimate of Roll’s spread wins out!
• Liu (2006) proposes an alternative measure of illiquidity, $LM_x =$ the standardized (to 21 day months) turnover-adjusted (to break ties) number of zero daily trading volumes over the prior $x$ months.

$$LM_x = \left[ \frac{\text{Number of zero daily volumes in prior } x \text{ months}}{\text{Deflator}} + \frac{1}{(x\text{-month turnover})} \right] \times \frac{21x}{\text{NoTD}},$$

(1)

• LM12 is materially different from existing liquidity measures such as turnover, bid-ask spread, etc.
• Captures trading speed…
• Documents a significant and robust liquidity premium over the sample period 1963-2003.
• Distinct from systematic market risk and the FF 3 factor risks.
Illiquidity: References

- DeJong and Rindi, 2009, 6.3.2, Chapter 7.
Illiquidity: References

Introduction

• The empirical work we have discussed up to now estimates “reduced” for equations where regression coefficients have been interpreted in light of MM models.
• However, in a set of papers, Easley and O’Hara illustrate how it is possible to estimate a structural MM model directly.
• Would like to estimate the degree of information asymmetries, or the probability of informed trading.
• The probability of informed trading is intimately linked to liquidity, and liquidity may in return affect asset pricing.
• Of course, the structure has to be quite stylized, so Easley, O’Hara and coauthors rely on a sequential trade model…

Key Insight:
Information-based trading affects asset prices.

• Based on Easley and O’Hara (1992)
• Use both trade and no-trade periods
• Use intraday and interday data
• Key is to use the trade direction (B/S)
  – Many more buys than sells => positive information
  – Many more sells than buys => negative information
• Illustrate the importance of asymmetric information models for asset prices (O’Hara (2003))
Easley, Kiefer and O’Hara (1997)

- The paper shows how to estimate such a sequential trade model using maximum likelihood.
- Truly structural model.
- Figure 1 shows the tree diagram of the trading process.
Figure 1
Tree diagram of the trading process.

$\alpha$ is the probability of an information event, $\delta$ is the probability of a low signal, $\mu$ is the probability that the trade comes from an informed trader, $1/2$ is the probability that an uninformed trader is a seller, and $\epsilon$ is the probability that the uninformed trader will actually trade. Nodes to the left of the dotted line occur only at the beginning of the trading day, nodes to the right are possible at each trading interval.
Easley, Kiefer and O’Hara (1997)

- The specialist knows the structure of the problem, but does not know whether a news event has occurred, and if it has occurred whether it is bad or good news, nor does she know if a particular trader is informed or not.
- Specialist is rational and uses Bayesian updating.
- Quoted prices will be conditional expectations.
- The specialist uses the sequence of buys and sells to try to figure out the value of the security.
Easley, Kiefer and O’Hara (1997)

- The neat thing with this paper is that the authors show how to actually cast the problem as an econometric problem where the econometrician tries to infer the parameters of the problem from the sequence of buys, sells, and no-trades.
- The parameters are in Figure 1 (the probability of news, the probability of good news, the probability that a trader is informed, and the likelihood that an uninformed trader is a buyer).
- These parameters are known to the specialist, but unobservable to the econometrician.
Easley, Kiefer and O’Hara (1997)

• They proceed to estimate this model for Ashland Oil.
  – An information event occurs 75% of the time.
  – About 17% of trades are informed if an information event occurs.
  – Good news and bad news are equally likely.
  – The probability that an uninformed trader trades given the opportunity is 33%.
  – Some parameters depend heavily on the no-trade interval assumed.

• Extension to variable trade size etc have been developed.

• If you can draw the tree, you can estimate it!!!

- The end of day value of the stock can be low or high with given probabilities =>
- Uninformed traders (including the dealer) know the probabilities and the possible values => compute $E(V)$
- With probability $\alpha$, there is an information event
- If there is an information event, informed traders learn the end of day value of the stock
- They buy (sell) if they get a good signal (bad) signal
- Market maker posts regret-free prices...

$$V = \{V_L, V_H\}$$

$$\pi(V_L) = \delta, \pi(V_H) = 1 - \delta$$

$$E(V) = V_L \delta + V_H (1 - \delta)$$

$$\alpha = \pi(\text{information event})$$

$$\mu = \pi(\text{informed})$$

$$\varepsilon = \pi(\text{uninformed trade})$$

$$\Rightarrow \ (\alpha, \delta, \mu, \varepsilon)$$

Parameters are assumed to be known to Market participants, but not to the econometrician…
Easley, Kiefer, and O’Hara (1997)
Estimating Sequential Trade Models

• Using the same approach as G&M (1985), EO(1992), we can compute the dealers’ bid and ask prices as conditional probabilities.

\[
E[V | S_1] = b_1 = \frac{\delta V_L (\alpha \mu + \varepsilon (1/2)(1 - \alpha \mu)) + (1 - \delta) V_H (\varepsilon (1/2)(1 - \alpha \mu))}{\delta \alpha \mu + \varepsilon (1/2)(1 - \alpha \mu)}
\]

\[
E[V | B_1] = a_1 = \frac{\delta V_L (\varepsilon (1/2)(1 - \alpha \mu)) + (1 - \delta) V_H (\alpha \mu + \varepsilon (1/2)(1 - \alpha \mu))}{\delta \alpha \mu + \varepsilon (1/2)(1 - \alpha \mu)}
\]

• At each point in time during the day, the dealer knows the history of past trades (including no-trades) Q in (B,S,N)
• In fact, the number of past buys, sells, and no-trades is a sufficient statistic for the quote process.
• What we see in the trade data is actually a censored sample of the quote process as the dealer may update quotes absent trades.

- The econometric implementation uses the fact that the likelihood function for a single day is proportional to:

\[
\Pr\{B, S, N \mid \alpha, \delta, \mu, \varepsilon\} = \alpha(1-\delta)\left[\mu + (1-\mu)(1/2)\varepsilon\right]^B\left[(1-\mu)(1/2)\varepsilon\right]^S\left[(1-\mu)(1-\varepsilon)\right]^N
\]\n
\[
+ \alpha\delta\left[(1-\mu)(1/2)\varepsilon\right]^B\left[\mu + (1-\mu)(1/2)\varepsilon\right]^S\left[(1-\mu)(1-\varepsilon)\right]^N
\]\n
\[
+ (1-\alpha)\left[(1/2)\varepsilon\right]^{B+S}\left[(1-\varepsilon)\right]^N
\]

- This is a mixture of trinomials (B,S,N).
- Using the assumption that information is independent across days, \(d\), we can model the likelihood function for a sample of \(D\) days as:

\[
\Pr\{(B_d, S_d, N_d)_{d=1}^{D} \mid \alpha, \delta, \mu, \varepsilon\} = \prod_{d=1}^{D} \Pr\{B_d, S_d, N_d \mid \alpha, \delta, \mu, \varepsilon\}
\]
Easley, Kiefer, and O’Hara (1997)  
Estimating Sequential Trade Models

- Use Maximum Likelihood to get estimates for $\alpha$, $\delta$, $\mu$, $\epsilon$.
  - Log transform simplifies the estimation.
- Econometrician is trying to use patterns to estimate parameters
  - High volume days with more buys than sells => Good news
  - High volume days with more sells than buys => Bad news
  - Low volume days => No news
- The very common stock is Ashland Oil, 10/1-11/9/90
  - Trade-off: Time series versus parameter stability
- Use L&R (1991) to classify trades as buys/sells

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of informed</td>
<td>$\mu$</td>
<td>0.172</td>
<td>0.014</td>
</tr>
<tr>
<td>Probability that uninformed trade</td>
<td>$\epsilon$</td>
<td>0.332</td>
<td>0.012</td>
</tr>
<tr>
<td>Probability of information</td>
<td>$\alpha$</td>
<td>0.750</td>
<td>0.103</td>
</tr>
<tr>
<td>Probability of bad state</td>
<td>$\delta$</td>
<td>0.502</td>
<td>0.113</td>
</tr>
</tbody>
</table>

- Previous authors have emphasized that large trades may be more informative than small trades.
- E&O (1987) suggest that traders may optimally pool (all trade small) their trades instead of separating (informed trade large: easy to identify informed).
- The authors also show how to incorporate trade size E&O (1987) into the PIN estimation.
- Additional parameters: $\phi =$ prob. UI trades large, $\omega =$ prob. I trades large

<table>
<thead>
<tr>
<th>Label</th>
<th>Param.</th>
<th>Est. UR</th>
<th>Est. R $\phi = \omega$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of informed</td>
<td>$\mu$</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>Probability of uninformed trade</td>
<td>$\varepsilon$</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td>Probability of information</td>
<td>$\alpha$</td>
<td>0.59</td>
<td>0.59</td>
</tr>
<tr>
<td>Probability of bad state</td>
<td>$\delta$</td>
<td>0.58</td>
<td>0.57</td>
</tr>
<tr>
<td>Prob. Uninformed trades large</td>
<td>$\omega$</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>Prob. Informed trades large</td>
<td>$\phi$</td>
<td>0.28</td>
<td></td>
</tr>
</tbody>
</table>

- $R=U \Rightarrow$ Cannot reject that $\phi = \omega \Rightarrow$ Trade size is not informative!
Introducing PIN

- What is the probability that a trade will be based in information, and how does it vary in the cross-section of stocks?

\[ PIN = \frac{\mu(1-P_n)}{(\mu(1-P_n) + 2\epsilon)} \]

- Initial Spread = \[ \frac{\alpha\mu}{\alpha\mu + 2\epsilon} [V_H - V_L] \]

- Draw a random sample of 30 NYSE stocks from the first, fifth, and eighth decile by trading volume (match on P), Oct 1-Dec. 23, 1990.
  - Aggregate trades w/in 5 seconds at the same price with no intervening quote revisions.
  - Use L&R (1991) to classify trades as B, S.
  - Let \( P_n \) be the probability of no-trade (same interval across stocks)??

<table>
<thead>
<tr>
<th>Label</th>
<th>Parameter</th>
<th>Dec. 1</th>
<th>Dec. 5</th>
<th>Dec. 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of I trade</td>
<td>( \mu )</td>
<td>0.105</td>
<td>0.028</td>
<td>0.014</td>
</tr>
<tr>
<td>Probability of UI trade</td>
<td>( \epsilon )</td>
<td>0.137</td>
<td>0.023</td>
<td>0.009</td>
</tr>
<tr>
<td>Probability of information</td>
<td>( \alpha )</td>
<td>0.478</td>
<td>0.449</td>
<td>0.364</td>
</tr>
<tr>
<td>Probability of bad state</td>
<td>( \delta )</td>
<td>0.360</td>
<td>0.418</td>
<td>0.455</td>
</tr>
<tr>
<td>PIN</td>
<td>( \frac{\mu(1-P_n)}{(\mu(1-P_n) + 2\epsilon)} )</td>
<td>0.154</td>
<td>0.206</td>
<td>0.197</td>
</tr>
</tbody>
</table>
Easley, Kiefer, O’Hara and Paperman (1996)

Introducing PIN

The distribution of PIN => PIN is higher for less liquid stocks

Initial quoted (cent) spread is increasing in Price*PIN and decreasing in volume...

Another way of looking at the adverse Selection component of the spread…
Easley, Hvidkjaer and O’Hara (2002) Does PIN Affect Asset Prices?

- Documented that PIN is related to liquidity as measured by trading volume and by quoted spreads...
- Amihud & Mendelson (1986) showed that spreads affect returns.
  - Wide spreads => high cost of trading => liquidity premium => higher returns are required by investors => lower price...
- Does PIN affect asset prices?
- Generalize model slightly to allow for B, S by UI to be different (\( \varepsilon_B \neq \varepsilon_S \)).
- Estimate annual PIN for all NYSE stocks 1983 – 1998...
  - Some trouble with convergence and corner solutions...

### Table II

**Parameter Summary Statistics**

The table contains time-series averages across years 1983 to 1998 of cross-sectional means, medians, standard deviations, and the median of parameter standard errors from the likelihood estimation. The following are variable descriptions: \( \alpha \) is the probability of an information event, \( \delta \) is the probability of a low signal, \( \mu \) is the rate of informed trade arrival, \( \epsilon_b \) is the arrival rate of uninformed buy orders, and \( \epsilon_s \) is the arrival rate of uninformed sell orders. PIN is the probability of informed trading as given by equation (5).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Median Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>0.253</td>
<td>0.281</td>
<td>0.111</td>
<td>0.035</td>
</tr>
<tr>
<td>( \delta )</td>
<td>0.331</td>
<td>0.309</td>
<td>0.181</td>
<td>0.066</td>
</tr>
<tr>
<td>( \mu )</td>
<td>31.075</td>
<td>21.303</td>
<td>32.076</td>
<td>0.996</td>
</tr>
<tr>
<td>( \epsilon_b )</td>
<td>22.304</td>
<td>11.437</td>
<td>31.519</td>
<td>0.324</td>
</tr>
<tr>
<td>( \epsilon_s )</td>
<td>24.046</td>
<td>13.095</td>
<td>31.427</td>
<td>0.299</td>
</tr>
<tr>
<td>PIN</td>
<td>0.191</td>
<td>0.185</td>
<td>0.057</td>
<td>0.019</td>
</tr>
</tbody>
</table>
Easley, Hvidkjaer and O’Hara (2002) Does PIN Affect Asset Prices?

Table III.

<table>
<thead>
<tr>
<th>Size/PIN</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Excess Returns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>0.148</td>
<td>0.202</td>
<td>0.474</td>
</tr>
<tr>
<td>2</td>
<td>0.462</td>
<td>0.556</td>
<td>0.743</td>
</tr>
<tr>
<td>3</td>
<td>0.647</td>
<td>0.695</td>
<td>0.892</td>
</tr>
<tr>
<td>4</td>
<td>0.873</td>
<td>0.837</td>
<td>0.928</td>
</tr>
<tr>
<td>Large</td>
<td>0.953</td>
<td>1.000</td>
<td>0.643</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Size/PIN</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel B: Number of Stocks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>21.9</td>
<td>60.1</td>
<td>142.5</td>
</tr>
<tr>
<td>2</td>
<td>30.5</td>
<td>79.0</td>
<td>115.7</td>
</tr>
<tr>
<td>3</td>
<td>54.4</td>
<td>95.5</td>
<td>75.1</td>
</tr>
<tr>
<td>4</td>
<td>93.4</td>
<td>99.0</td>
<td>32.7</td>
</tr>
<tr>
<td>Large</td>
<td>174.4</td>
<td>41.6</td>
<td>8.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Size/PIN</th>
<th>Panel C: Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>75.3</td>
</tr>
<tr>
<td>2</td>
<td>257.6</td>
</tr>
<tr>
<td>3</td>
<td>665.2</td>
</tr>
<tr>
<td>4</td>
<td>1,719.1</td>
</tr>
<tr>
<td>Large</td>
<td>9,845.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Size/PIN</th>
<th>Panel D: PIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>0.134</td>
</tr>
<tr>
<td>2</td>
<td>0.138</td>
</tr>
<tr>
<td>3</td>
<td>0.137</td>
</tr>
<tr>
<td>4</td>
<td>0.138</td>
</tr>
<tr>
<td>Large</td>
<td>0.127</td>
</tr>
</tbody>
</table>

Excess returns increase when we move from low to high PIN portfolios within size portfolios => *Suggests that PIN is priced*...
PIN is priced in FF (1993) 3-factor model. Robust to errors in variables (portfolios of PIN instead of stock specific estimates)

Does PIN proxy for some omitted variable? PIN remains significant when spread, Volatility, and turnover (CV of turnover) are added...

**Information based trading has large and significant effects on asset returns!**
Estimating PIN requires that the researcher is able to separate buyer initiated trades from seller initiated trades.

This is notoriously difficult!

Boehmer, Grammig, and Theissen (2007) show that the inaccuracy of trade-classification algorithms lead to downward-biased PIN estimates.

They also show that the bias is related to a security’s trading intensity.

The authors propose a data-based adjustment procedure that substantially reduces the misclassification bias.
Duarte and Young (2007)

- Extends the sample to 1983-2004.
- Finds that PIN is no longer significant as a risk factor when illiquidity (ILLIQ) is included in the Fama-MacBeth analysis.

Table 9 - 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 extended model. PSOS is the probability of symmetric order flow shocks estimated with the extended model. PIN's, AdjPIN's and PSOS's 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.

<table>
<thead>
<tr>
<th>Explanatory Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
<th>(11)</th>
<th>(12)</th>
<th>(13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>0.1745</td>
<td>0.1437</td>
<td>0.1673</td>
<td>0.1744</td>
<td>0.1793</td>
<td>0.1391</td>
<td>0.1486</td>
<td>0.1427</td>
<td>0.1355</td>
<td>0.1439</td>
<td>0.1673</td>
<td>0.1644</td>
<td>0.1427</td>
</tr>
<tr>
<td></td>
<td>(0.494)</td>
<td>(0.398)</td>
<td>(0.480)</td>
<td>(0.469)</td>
<td>(0.508)</td>
<td>(0.369)</td>
<td>(0.425)</td>
<td>(0.414)</td>
<td>(0.368)</td>
<td>(0.413)</td>
<td>(0.480)</td>
<td>(0.439)</td>
<td>(0.414)</td>
</tr>
<tr>
<td>SIZE</td>
<td>0.0430</td>
<td>-0.0010</td>
<td>0.0304</td>
<td>0.0291</td>
<td>0.0392</td>
<td>0.0647</td>
<td>0.0881</td>
<td>0.0835</td>
<td>0.0744</td>
<td>0.0817</td>
<td>0.0304</td>
<td>0.0183</td>
<td>0.0835</td>
</tr>
<tr>
<td></td>
<td>(0.630)</td>
<td>(-0.015)</td>
<td>(0.428)</td>
<td>(0.494)</td>
<td>(0.588)</td>
<td>(1.215)</td>
<td>(1.379)</td>
<td>(1.255)</td>
<td>(1.347)</td>
<td>(1.301)</td>
<td>(0.428)</td>
<td>(0.315)</td>
<td>(1.255)</td>
</tr>
<tr>
<td>BM</td>
<td>0.2675</td>
<td>0.2567</td>
<td>0.2615</td>
<td>0.2677</td>
<td>0.2647</td>
<td>0.2554</td>
<td>0.2542</td>
<td>0.2461</td>
<td>0.2472</td>
<td>0.2452</td>
<td>0.2615</td>
<td>0.2631</td>
<td>0.2461</td>
</tr>
<tr>
<td></td>
<td>(2.956)</td>
<td>(2.853)</td>
<td>(2.901)</td>
<td>(2.963)</td>
<td>(2.959)</td>
<td>(2.802)</td>
<td>(2.824)</td>
<td>(2.750)</td>
<td>(2.755)</td>
<td>(2.757)</td>
<td>(2.901)</td>
<td>(2.888)</td>
<td>(2.750)</td>
</tr>
<tr>
<td>PIN</td>
<td>1.0043</td>
<td>1.1842</td>
<td>0.6859</td>
<td>0.6481</td>
<td>0.5514</td>
<td>0.6164</td>
<td>0.5324</td>
<td>0.5224</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.911)</td>
<td>(1.842)</td>
<td>(1.030)</td>
<td>(1.158)</td>
<td>(1.101)</td>
<td>(0.912)</td>
<td>(0.630)</td>
<td>(0.616)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AdjPIN</td>
<td>0.0248</td>
<td>-0.6519</td>
<td>-0.6519</td>
<td>-0.6519</td>
<td>-0.6519</td>
<td>-0.6519</td>
<td>-0.6519</td>
<td>-0.6519</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(-0.928)</td>
<td>(-0.928)</td>
<td>(-0.928)</td>
<td>(-0.928)</td>
<td>(-0.928)</td>
<td>(-0.928)</td>
<td>(-0.928)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSOS</td>
<td>0.4932</td>
<td>0.2940</td>
<td>0.1956</td>
<td>0.938</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.466)</td>
<td>(-0.193)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILLIQ</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.989)</td>
<td>(2.696)</td>
<td>(2.583)</td>
<td>(2.728)</td>
<td>(2.599)</td>
<td>(2.583)</td>
<td>(2.583)</td>
<td>(2.583)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AdjPIN-PIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Yan and Zhang (2006) and Yan (2009)

- A significant problem with PIN is that the ML estimation does not converge for many stocks.
- For example, PIN could not be estimated for 48% of the stocks in the market in 2001 compared to 21% in 1983 (70% in 2007!).
- The problem has also increased dramatically over time as the frequency of trades has risen.
- Yan and Zhang (2006) show how to improve on the success rate by choosing the initial values for the ML estimation “rationally.”
- Yan (2009) proposes a robust alternative way of estimating PIN that avoids this problem.
- Essentially, the method is to start by identifying event days based on event-study methodology (CRSP) combined with the assumption that there should be more buys (sells) on positive (negative) event days.
PIN: References

• DeJong and Rindi, 2009, 6.4.3, Chapter 7.
• Hasbrouck, 2007, Chapter 6.
• Yan, Y., 2009, A new method to estimate PIN (Probability of Informed Trading), working paper, University of Pennsylvania.
Conclusions

• Bonds with less liquidity trade at lower prices, have higher yields.
• Stocks with more illiquidity, as well as higher illiquidity risk, have higher returns (lower prices).
• Stocks with more informed trading (PIN) have higher returns (lower prices), controlling for competing risk-factors such as Beta, Size, BM.
• PIN captures both illiquidity and informed trading risk, and it appears that the explanatory power comes from the illiquidity component.
• This literature provides a nice link between market microstructure and more traditional fields in finance such as asset pricing.