Appendix

A. Firm-Specific Determinants of *PIN*, *PIN_ G*, and *PIN_ B*

We consider how *PIN* and its good and bad information components depend on the following firm-specific characteristics, several of which have been found to be associated with measures of illiquidity in previous studies:[1]

SIGMA: The standard deviation of daily returns in the previous 12 months (computed monthly with a 12-month rolling window). This measure of return volatility is a proxy for the rate of information arrival and is associated with lower liquidity. Therefore, we conjecture that higher volatility is associated with higher *PIN* measures.

BTM and *BMDUM*: *BTM* is the natural logarithm of *BM* if *BM* is positive [*BM*=(book value)/(market value)], and zero otherwise. *BMDUM* is unity if *BM* is positive, and zero if *BM* is negative or missing. These variables follow Pontiff and Woodgate (2008). We include *BTM* because of the possibility that glamour stocks with low *BTM* may have more cash flow uncertainty and therefore more scope for informed trading (for details about these variables, see Subsection 3.3.).

IH: The natural logarithm of one plus the fraction of stock that is held by insiders. Insider holdings are taken from the Insider Filing Table 1 (Stock Transactions) in the Thomson Reuters database, and the value for any given month is the latest value reported in the database. The data are available from 1986 but are sparse before 1990, so the analysis using this variable is limited to 1990-2010. Chiang and Venkatesh (1988) consider the role of insiders as a determinant of the bid-ask spread, assuming that inside holdings are the channel through which private information is conveyed to the market.

ANA: The natural logarithm of one plus the number of analysts following a firm each month. The data are obtained from IBES. It is natural to expect that increased analyst coverage will reduce the prevalence of private information and hence the *PIN* measures.

[1] See Benston and Hagerman (1974), and Brennan and Subrahmanyam (1995).

Consistent with this, Brennan and Subrahmanyam (1995) find that an increase in the number of analysts reduces the cost of transacting as measured by the Kyle lambda.

SIZE: The natural logarithm of the month-end market value of equity, defined as the stock price times the number of shares outstanding. Small firms are less liquid and there is generally less public information about them.

Since information asymmetry may be more characteristic of some industries than others, we include five industry dummy variables that correspond to the four-digit SIC code classifications given on Kenneth French's website:

CONSM: The industry dummy for consumer durables, nondurables, wholesale, retail, and some services (laundries and repair shops). This industry category is treated as the base case, and the variable is not included in the cross-sectional regressions.

MANUF: The industry dummy for manufacturing, energy, and utilities.

HITEC: The industry dummy for hi-tech businesses such as business equipment, telephone, and television transmission.

HLTH: The industry dummy for healthcare, medical equipment, and drugs.

OTHR: The dummy for other industries, including mines, construction, building materials, transportation, hotels, business services, entertainment, and finance.

To allow for potential endogeneity between the *PIN* measures and other variables, we consider a simultaneous equation system.[2] Following Brennan and Subrahmanyam (1995) who treat the Kyle (1985) lambda, analyst coverage, and share turnover as jointly determined, we assume that *PIN* (and its components) are determined jointly with an-

[2] Aslan, Easley, Hvidkjaer, and O'Hara (2011) also investigate why *PIN* varies in the cross-section, but unlike us, they use a single equation approach, and do not decompose *PIN* into good- and bad-news components.

alyst coverage and turnover. Therefore, we specify the following three-equation system:

$$
\begin{aligned}
Y_{i,t} &= a_1 + b_1 SIGMA_{i,t} + c_1 BTM_{i,t} + d_1 BMDUM_{i,t} + e_1 IH_{i,t} + f_1 ANA_{i,t} \\
&\quad + g_1 SIZE_{i,t} + \sum_{j=1}^{4} \gamma_{j,t} A_{i,j,t} + \nu_{1,i,t}, && \text{(A1)} \\
ANA_{i,t} &= a_2 + b_2 Y_{i,t} + c_2 IO_{i,t} + d_2 TURN_{i,t} + e_2 SIGMA_{i,t} \\
&\quad + f_2 \ln(P)_{i,t} + g_2 SIZE_{i,t} + \sum_{j=1}^{4} \phi_{j,t} A_{i,j,t} + \nu_{2,i,t}, && \text{(A2)} \\
TURN_{i,t} &= a_3 + b_3 Y_{i,t} + c_3 ANA_{i,t} + d_3 \ln(P)_{i,t} + e_3 SIZE_{i,t} \\
&\quad + f_3 SP\_IND_{i,t} + \nu_{3,i,t}, && \text{(A3)}
\end{aligned}
$$

where $Y_{i,t}$ is one of the three informed-trading measures, *PIN*, *PIN_ G*, and *PIN_ B*, for stock i in month t; $IO_{i,t}$ is the natural logarithm of one plus the fraction of shares owned by institutional investors, obtained from the Thomson Reuters Institutional (13f) Holdings database; $TURN_{i,t}$ is share turnover calculated as the monthly share volume divided by the number of shares outstanding as of the month end; $\ln(P)_{i,t}$ is the natural logarithm of the stock price at the end of the month; $SP\_IND_{i,t}$ is the indicator variable for the S&P 500 Index components (i.e., 1 if the stock is included in the Index, and 0 otherwise), obtained from Compustat; and $A_{i,j,t}$ includes the four industry dummy variables (*MANUF*, *HITEC*, *HLTH*, and *OTHR*): the consumer goods industry (*CONSM*) is treated as the base case. The remaining variables are defined above.

Equation (A1) treats the *PIN* variable as a function of return volatility, book-to-market ratio, insider holdings, analyst following, and firm size, as well as the industry dummy variables: the specification is motivated by earlier studies on illiquidity in securities markets (Benston and Hagerman, 1974; Branch and Freed, 1977; Stoll, 1978; and Chiang and Venkatesh, 1988).

In Equation (A2), analyst coverage (*ANA*) is modeled as a function of *PIN* or its component, institutional ownership, share turnover, return volatility, share price, firm size, and the industry dummy variables. Institutional ownership, firm size, share price, and return volatility have been shown to be important determinants of analyst following by Bhushan (1989) as well as by Brennan and Subrahmanyam (1995, 1998). Share turnover is included as a proxy for liquidity, which Brennan and Subrahmanyam (1995, 1998) have shown to be also important for analyst following. The third equation [Equation (A3)] explains share turnover in terms of *PIN* or its component, analyst coverage, price, firm size, and the dummy variable for S&P 500 Index membership.[3]

For each of the three *PIN* measures ($Y_{i,t}$), the system is estimated by two-stage least-squares. Following Fama and MacBeth (1973), we estimate cross-sectional regressions each month, and the final estimator is the time-series average of the monthly coefficients. Since the dependent variables in the system (*PIN* or its components, *ANA*, and *TURN*) are highly persistent, we report heteroskedasticity- and autocorrelation-consistent (HAC) t-statistics computed based on Newey and West (1987), instead of the usual Fama-MacBeth t-statistics.[4]

To save space, in Table A1 we report only the estimates of Equation (A1) for each of the three *PIN* measures. Note that since $PIN = PIN\_G + PIN\_B$ the coefficients of the PIN regression are equal to the sum of the coefficients in the other two regressions. The coefficient of return volatility, *SIGMA*, is negative and highly significant in the *PIN* regression, but this is entirely due to its significance in the *PIN_ B* regression: the coefficient in the *PIN_ G* regression is positive and insignificant. Thus, *SIGMA* is strongly negatively associated with informed trading, but only on *bad* news. We have no explanation for this phenomenon which clearly merits further investigation.

The coefficient of the book-to-market ratio variable, BTM, is insignificant in the PIN regression but is negative and marginally significant in the *PIN_ G* regression, and positive and significant in the *PIN_ B* regression. This implies that value firms experience

[3]Using the rank condition (e.g., Maddala, 1977, 233-235), it is straightforward to verify that all three equations in our system are exactly identified.

[4]As suggested by Newey and West (1987) in choosing bandwidth parameter $N(= L + 1)$ for the Bartlett kernel to compute the standard errors, we let the lag length L be equal to the integer portion of $4(T/100)^{2/9}$, where T is the number of observations in the estimated coefficient series.

less informed trading on good news than do growth firms, but more informed trading on bad news: it is possible that this is because there is less private good news about value firms and more private bad news about growth firms. The coefficient on *BMDUM* is positive and strongly significant in all regressions. This implies that there is *less* informed trading in the shares of firms with missing or negative BM. This may be because shares in these firms are thinly traded and may not provide informed traders with adequate camouflage. The coefficient of insider holdings, IH, is positive but not significant in any of the regressions. This is contrary to our conjecture that insider holdings would be more likely to give rise to sales on bad private information than to purchases on good private information, since sales might be more easily disguised as portfolio adjustments than purchases. Instead, the analysis suggests that informed trading emanates from agents other than corporate insiders.[5]

We note that the coefficients of firm size, $SIZE$, and analyst following, ANA, are negative and highly significant in all three regressions. This is consistent with the finding of Brennan and Subrahmanyam (1995, 1998) that these firm characteristics reduce the adverse selection costs of trading as captured by an empirical proxy for the Kyle (1985) lambda, since the adverse-selection costs arise from trading on private information.[6]

Turning to the effects of the industry dummy variables, we find that relative to the consumer goods industry which is the omitted industry category, firms in other service industries (*OTHR*) are in general less prone to informed trading as captured by any of the three *PIN* measures. On the other hand, firms in hi-tech industries (*HITEC*) are more subject to informed trading on good news but not on bad news. Reeb, Zhang, and Zhao (2012) find evidence of more informed trading in the pharmaceutical industry using different proxies for informed trading. We find that for the broader health care sector,

[5] Vega (2006, p. 105) makes a similar argument. She conjectures that "...*PIN* is not exclusively an insider trading measure as it also captures informed trading by investors who are particularly skillful in analyzing public news." Our results on IH and $SIGMA$ do not accord with the single equation estimates of Aslan, Easley, Hvidkjaer, and O'Hara (AEHO) (2011). The differences might arise due to our use of quarterly PIN's, as opposed to the annual ones used by AEHO.

[6] Brennan and Subrahmanyam (1995) use volume of trading instead of firm size in their regression.

$HLTH$, there is a higher probability (than in the consumer goods sector) of informed trading on good news, but a lower probability of informed trading on bad news.[7]

The significance of the analyst following variable and the fact that BTM and the industry membership dummies, $HITECH$ and $HLTH$, have opposite effects on PIN_B and PIN_G provide further confirmation for the role of these variables as proxies for *informed* trading since analyst following is a determinant of the informational environment, and it is difficult to think of industry specific idiosyncracies in the order flow that would increase PIN_G and reduce PIN_B that are *not* information-related.

[7]The pharmaceutical industry is by no means immune to insider trading on bad news as the ImClone case in 2001 (http://en.wikipedia.org/wiki/ImClone_stock_trading_case) and the Elan case in 2008 (http://www.marketswiki.com/mwiki/Mathew_Martoma) show.

Appendix References

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

Benston, G., and R. Hagerman, 1974, Determinants of bid-asked spreads in the over-the-counter market, *Journal of Financial Economics* 1, 353-364.

Bhushan, R., 1989, Firm size and analyst following, *Journal of Accounting and Economics* 11, 255-274.

Branch, B., and W. Freed, 1977, Bid-asked spreads on the Amex and the Big Board, *Journal of Finance* 32, 159-163.

Brennan, M., and A. Subrahmanyam, 1995, Investment analysis and price formation in securities markets, *Journal of Financial Economics* 38, 361-381.

Brennan, M., and A. Subrahmanyam, 1998, The determinants of average trade size, *Journal of Business* 71, 1-25.

Chiang, R., and P. Venkatesh, 1988, Insider holdings and perceptions of information asymmetry: A note, *Journal of Finance* 43, 1041-1048.

Maddala, G., 1977, *Econometrics*, McGraw-Hill, New York, NY.

Newey, W., and K. West, 1987, A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.

Reeb, D., Y. Zhang, and W. Zhao, 2012, Informed trading in regulated industries, Working paper, Temple University.

Stoll, H., 1978, The pricing of security dealer services: An empirical study of NASDAQ stocks, *Journal of Finance* 33, 1153-1172.

Table A1

Determinants of the *PIN* Measures: Results from the Three-Equation System

This table reports the results of monthly Fama-MacBeth (1973) cross-sectional regressions using *PIN*, *PIN_G*, or *PIN_B* as one of the dependent variables in the three-equation system [see Equations (A1)-(A3) in the Appendix] for NYSE/AMEX stocks over the 252 months (1990:01-2010:12). The three-equation system is estimated simultaneously via the two-stage least-squares (2SLS) with one of the three *PIN* measures being the dependent variable in Equation (A1). The results from Equation (A1) only are reported in this table. The variables are defined as follows. *PIN*: the probability of informed trading defined by $\alpha\mu/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, where α is the probability with which a private information event occurs on a given day, μ is the rate at which orders from informed traders arrive if the information event does occur, ε_b is the rate at which orders from uninformed buyers arrive, and ε_s is the rate at which orders from uninformed sellers arrive; *PIN_G*: the probability of informed trading on good news defined by $\alpha\mu(1-\delta)/(\alpha\mu+\varepsilon_b+\varepsilon_s)$; *PIN_B*: the probability of informed trading on bad news defined by $\alpha\mu\delta/(\alpha\mu+\varepsilon_b+\varepsilon_s)$; *SIGMA*: standard deviation of daily stock returns within the previous 12-month (rolling) period; *BTM*: natural logarithm of the book to market ratio (BM) if BM is positive, and 0 if BM is negative or missing; *BMDUM*: 1 if BM is positive, and 0 if BM is negative or missing; *IH*: natural logarithm of one plus the fraction of total (direct and indirect) insider stock holdings; *ANA*: natural logarithm of one plus the number of analysts following a firm each month; and *SIZE*: natural logarithm of the market value of equity. In addition, we also include five industry dummy variables based on four-digit SIC code classifications reported on Ken French's web site. They are: *CONSM*: the industry dummy (base case not included in the regressions) for consumer durables, nondurables, wholesale, retail, and some services (laundries and repair shops); *MANUF*: the industry dummy for manufacturing, energy, and utilities; *HITEC*: the industry dummy for hi-tech industries such as business equipment, telephone, and television transmission; *HLTH*: the industry dummy for healthcare, medical equipment, and drugs; and *OTHR*: the industry dummy for other industries, including mines, construction, building materials, transportation, hotels, business services, entertainment, and finance. The values in the first row for each explanatory variable are time-series averages of coefficients obtained from the monthly cross-sectional regressions, and the values italicized in the second row of each variable are heteroskedasticity- and autocorrelation-consistent (HAC) *t*-statistics computed using Newey and West (1987) standard errors. *Avg Obs* is the average number of stocks used in the monthly cross-sectional regressions. To survive in the sample for the *PIN*-related measures, stocks must have at least 40 positive-volume days within each quarter. Coefficients significantly different from zero at the significance levels of 1% and 5% are indicated by ** and *, respectively.

Determinants of PIN Measures for NYSE/AMEX Stocks (1990:01-2010:12)						
Explanatory Vars.	PIN		PIN_G		PIN_B	
Intercept	0.313	**	0.161	**	0.152	**
	39.68		*27.04*		*30.59*	
SIGMA	-0.335	**	0.131		-0.466	**
	-4.75		*1.78*		*-10.36*	
BTM	0.002		-0.001		0.002	**
	1.61		*-1.85*		*3.33*	
BMDUM	0.026	**	0.019	**	0.007	**
	10.44		*7.82*		*4.28*	
IH	0.013		0.013		0.000	
	0.93		*1.23*		*0.02*	
ANA	-0.049	**	-0.025	**	-0.024	**
	-14.02		*-11.88*		*-10.08*	
SIZE	-0.013	**	-0.006	**	-0.007	**
	-10.80		*-7.94*		*-7.78*	
MANUF	0.001		0.001		0.000	
	0.92		*1.09*		*0.42*	
HITEC	0.003	*	0.005	**	-0.002	
	1.96		*3.90*		*-1.26*	
HLTH	0.004	*	0.007	**	-0.003	*
	2.05		*5.74*		*-2.40*	
OTHR	-0.004	**	-0.003	**	-0.002	*
	-4.26		*-2.99*		*-2.35*	
Avg Obs	1371.6		1374.0		1374.3	