# Informed Trading, Information Asymmetry and Pricing of Information Risk: Empirical Evidence from the NYSE

Bardong, Florian and Bartram, Söhnke M. and Yadav, Pradeep K.

Lancaster University Management School, Price College of Business, Oklahoma University

7 July 2005

# Informed Trading, Information Asymmetry and Pricing of Information Risk: Empirical Evidence from the NYSE

by

Florian Bardong, Söhnke M. Bartram, and Pradeep K. Yadav[#]

**Abstract**

We analyze commonality in informed trading across stocks, and how informed trading varies with the structural and trading characteristics of a firm. We thereby isolate the residual level of informed trading that is unrelated to commonality, trading characteristics, and structural characteristics and analyze this measure with respect to its characteristics and pricing relevance. We find evidence of commonality in informed trading, and a systematic dependence of the level of informed trading on firm characteristics, such as, tick size, the existence of options, and the size of the ownership stake of outside parties. Most importantly, we find that the residual level of informed trading is the component of informed trading most strongly related to required returns. This indicates that an important part of the information risk premium is related to the inability to differentiate between price fluctuations that are caused by changes in fundamental value from random price moves.

**Keywords:** Market microstructure, Common factors, Risk factors, Asymmetric information
**JEL Classification:** G10, G12, G14
**First version:** July 07, 2005
**This version:** October 10, 2008

[#] Florian Bardong (f.bardong@lancaster.ac.uk), and Söhnke Bartram (s.m.bartram@lancaster.ac.uk) are at the Department of Accounting and Finance, Lancaster University Management School, Lancaster, LA1 4YX, U.K., tel. +44-1524-65201. Pradeep Yadav (pyadav@ou.edu) is at the Price College of Business, Oklahoma University, 307 W. Brooks Suite 205A, Norman, OK 73019, U.S., tel. 405 325 6640, fax 405-325-2096. The authors would like to thank Aslı Aşçıoğlu, Chris Brooks, and Alexander Stremme for detailed discussions. We are grateful to Martin Evans, Chitru Fernando, Scott Linn, Narayan Naik, John O'Hanlon, Michael Pagano, Grzegorz Pawlina, Janya Portniaguina, Jesus Salas, Brian Stanhouse, Stephen Taylor, and seminar participants at the European Finance Association Annual Meetings 2006 in Zürich, Indian School of Business, Lancaster University, University of Oklahoma, the EMG/ESRC Workshop "Microstructure Issues of Financial Markets" at Cass Business School, the DRM-CEREG Dauphine Workshop "Financial Market Quality" at Euronext, Paris, and at the Western Finance Conference Association Annual Meetings 2008 in Waikoloa, Hawaii for helpful comments and suggestions. Pradeep Yadav thanks the *Scottish Institute for Research in Investment and Finance (SIRIF)* at the University of Strathclyde and the *Center for Financial Research* at the University of Cologne for data support.

# 1 Introduction and Motivation

Recent corporate scandals (e.g., Enron, WorldCom and others) have focused the attention of regulators and market participants on the extent and market impact of private information and the resulting information asymmetry between investors. It is fairly challenging empirically, however, to numerically capture such fuzzy a concept as the information environment in general and private information in particular. At least a subset of such private information is periodically revealed to the market through trading by investors with access to private value-relevant information. Hence, an important way to infer the level of private information that informed investors have and uninformed investors do not is through the use of a suitable variable that can measure the profits from informed trading. This paper is an empirical investigation of the level of informed trading that is inferred using methodologies from the market microstructure literature.

In this paper, the level of informed trading is estimated in two ways. First, the probability of informed trading, popularly known as PIN, is used, which is developed, tested and used in, for instance, Easley and O'Hara (1992), Easley, Kiefer, and O'Hara (1997a, b), and Easley, Hvidkjaer, and O'Hara (2002). Second, to improve the validity of the analysis and the degree the results presented in this paper can be generalized, informed trading is also approximated by the spread revenues lost, on average, by liquidity suppliers in the short-term to liquidity takers, a group that arguably includes informed investors (Harris (2003, p. 226)). Such a measure has been extensively used in the literature in a different context to proxy for informed trading (see, e.g., Hansch, Naik, and Viswanathan (1999); Naik and Yadav (2003b)).

A large body of academic literature has modeled the role of private information in asset markets.[1] The existence of private value-relevant information that is known by only a few investors implies the existence of information asymmetry, i.e., an unequal distribution of value-relevant information across investors. The inter-relationship between private information that exploits information asymmetry and trading behavior has also been extensively modeled.[2] And, very importantly, the empirical results of Easley, et al. (2002) show that information asymmetry is priced in the required rate of return. Clearly, the underlying information environment is of critical importance in explaining both observed trading behavior (Chordia, Roll, and Subrahmanyam (2000)) and the cost of capital (Easley and O'Hara (2004)). Yet, even though the extant empirical market microstructure literature addresses a wide range of issues that are broadly relevant in this context,[3] surprisingly little is known about informed trading observed in the financial markets, such as the extent of commonality in informed trading across stocks or the degree of co-

---

1 Akerlof (1970) relates the occurrence of private information to asset characteristics and trading mechanisms. Incentives to conceal negative information make companies in Brealey, Leland, and Pyle (1977) actively create private information. Informed traders that exploit private information acquire such information as long as the marginal benefit, realized via profitable trading, compensates for the marginal costs of information acquisition, which implies that there always remains some pieces of un-revealed private information (Grossman and Stiglitz (1980)). Uninformed investors will try to infer additional information from similar, correlated firms, which – if communicating information is costly – makes firms to disclose less (Admati and Pfleiderer (2000)). Incomplete information makes uninformed investors require a risk premium (Merton (1987)), which also applies if some of the missing information is known by only a few, informed, investors (Easley and O'Hara (2004)).

2 Early models include the following papers. In Kyle (1985), private information gets revealed through the price-impact of order-imbalances. To minimize the price-impact of their transactions, informed investors split their trades strategically over time, which results in private information being only successively impounded into prices (Kyle (1985)). Glosten and Milgrom (1985) suggest that liquidity suppliers recover their losses from informed trades by charging higher bid-ask spreads. Their sequential trade model is developed further in Easley and O'Hara (1987), who add the trade sequence to order-imbalance as conditioning information used by the market maker. This feature helps explain the positive relationship between the price-impact of a trade and trade size. Some stylized facts of empirical market microstructure, such as intra-day seasonality in turnover, are explained by the model by Admati and Pfleiderer (1988) where liquidity considerations lead to a clustering of trades by informed and uninformed investors.

3 For example, in addition to the references cited in the text, (Brennan and Subrahmanyam (1996)) find that liquidity is priced. Breen, Hodrick, and Korajczyk (2002) show that stock-level trading characteristics and cross-sectional characteristics are related to microstructure phenomena. Stock-level order-imbalances have been found to be related to returns (Chordia and Subrahmanyam (2004)). In particular, the relationship of lagged order-imbalance, firm size, and daily stock returns suggests – in the spirit of Kyle (1985) – that informed investors in large firms break up their trades into smaller ones that are successively executed stretching out the price adjustment to new information over time.

variation of informed trading across stock-level and market-wide characteristics.[4] This paper aims to address this gap in the literature.

Information asymmetry and thus informed trading can arise not just from insider information that corporate managers and their affiliates have. Investors unconnected to a particular firm may invest resources to acquire price-relevant private information (Grossman and Stiglitz (1980)) that can be firm-specific (Kim and Verrecchia (1994, 1997)), applicable to many firms (Chordia, et al. (2000); Subrahmanyam (1991)), or that can be related to the trading environment (Madrigal (1996)).[5]

[4] There has been some research focusing on links between information asymmetry and single firm-specific characteristics, such as the relationship of information asymmetry and options volume (Easley, O'Hara, and Srinivas (1998)), institutional ownership (Dennis and Weston (2001)), or credit ratings (Odders-White and Ready (2006)).

[5] Subrahmanyam (1991) shows that trading index certificates instead of the individual underlying securities reduces uninformed investors' adverse selection costs related to private information about systematic return factors. Modelling private information about market-wide systematic return factors – next to firm-specific insider information – is also an important element in Admati (1984) and Hughes, Liu, and Liu (2005). Theoretically, information about the systematic return component centers around mapping the factor realizations into returns via factor-mimicking portfolio returns (see, e.g., Ferson (2003)). If the mapping is known for sure in markets with no private information, prices should adjust instantaneously to new factor realizations. In reality, this adjustment process is stretched out over time, however. Fleming and Remolona (1999), for instance, attribute to private information the period of price volatility that follows the almost transaction-free price adjustment in the Treasury market to economic news. In this situation, the economic factor realization itself is already known to the public, this implies that private information contains knowledge about the mapping of the factor realization into prices. This makes private information about market-wide phenomena the type of private information investors are exposed to in the Treasury market (see, e.g., Boni and Leach (2002)). Private information generated from financial reporting data is suspected to be amongst the drivers of informed trades in the stock market by some theoretical models (see, e.g., Kim and Verrecchia (1994, 1997)). Thus, public data can be a valuable source of private information.

Gilson, Healy, Noe, and Palepu (2001) find some analysts to consistently issue superior forecasts about the financial performance of firms. Some market participants are therefore particularly skilled in processing public data derived from, e.g., financial reports or analyses of the current economic situation, in private information that underlies such superior forecasts. Price-relevant private information is only generated as long as the trading profit derived from this private information equals marginal costs (Grossman and Stiglitz (1980)). This implies that the private information that financial analysts generate at a cost is exploited by informed traders, which is evidenced by the fact that analysts' information is reflected in prices (Bhushan (1989a)). This implies that there are two types of informed trades: Trades that are based on insider information from investors that just happen to acquire a piece of price-relevant private information, e.g., tip-offs from corporate executives, and informed trades that are based on the analysis of publicly available data.

As analysts differ significantly in their area of expertise (Gilson, et al. (2001)), private information is likely to be about individual firms (Kim and Verrecchia (1994)), about sectors (Chordia, et al. (2000)), or about market-wide return factors (Subrahmanyam (1991)), depending on analysts' skill and expertise. Public information that applies to many firms and that is used to generate private information should lead to common movements in informed trading of those stocks. In fact, commonality in informed trading based on public information is likely to be amongst the causes for commonality in trading activity. Hasbrouck and Seppi (2001), for instance, observe commonality in trading activity, which they attribute to investors with macroeconomic information that trade the same basket of stocks. Likewise, Chordia, et al. (2000) find commonality in quoted and effective spreads and note that informed trading

The level of informed trading should also depend on the structural characteristics of a firm. For example, as small firms are traded less frequently and have a lower analysts following Bhushan (1989b), they should expose public investors to more information asymmetry as larger firms. Lakonishok and Lee (2001), for instance, find that investors in small firms that have a low analyst following are particularly exposed to trades by corporate insiders. Easley, et al. (1998) document a relationship between information asymmetry and options volume, Dennis and Weston (2001) find a relationship between information asymmetry and institutional ownership, and Odders-White and Ready (2006) study the relationship between information asymmetry and credit ratings.

One major contribution of this paper is to comprehensively investigate, for the first time, the extent of commonality in informed trading and the influence of stock-level trading characteristics and firm-specific structural characteristics on the level of informed trading. Given the relationship between informed trading and common market-wide factors, and the relationship between informed trading and the structural and trading characteristics of a firm, it is possible to calculate the expected level of informed trading. The unexplained part of the observed level of informed trading is what this paper refers to as *Residual Asymmetric Information* – labeled *RAIN* for brevity and ease of exposition. *RAIN* is interpreted as a measure that captures the aggregate net economic effect of the abnormal, unexpected level of informed trading which, as suggested in footnote 5, also captures trades by insiders.

This paper makes another important contribution to market microstructure research by examining which component of informed trading is priced in the cross-section. Easley, et al. (2002)

---

could be one explanation. If informed traders exploit information on the stock-level trading environment (Madrigal (1996)), this co-variation could be across the entire market or could be associated with the market environment of individual stocks. Informed trades based on insider information, by contrast, should not show commonality. Thus, one should be able to decompose informed trading into trading on private information based on public data and insider information-related trading that is unrelated to common movements.

and Easley and O'Hara (2004) show that exposure to informed trading is priced in the cross-section of stock returns. However, if commonality is present in informed trading, it is an empirical question whether the resultant predictability of informed trading activity reduces overall information risk – by improving the information set of the public investor – or whether the resultant reduction in the ability to diversify exposure to informed trading increases overall information risk. To address this issue, the return relevance of the systematic common-factor dependent component of informed trading and the return relevance of RAIN, the residual unexplained part of informed trading is tested.

The empirical analysis of this paper is based on all reasonably liquid stocks traded on the New York Stock Exchange (hereafter NYSE) covering the eleven-year period between January 1995 and December 2005. Results show strong evidence of commonality in informed trading.[6] Market volatility, market trading volume, market-level bid-ask spreads, and market-level order-imbalance are all significantly related to the level of informed trading. Evidence of individual firm-specific components in informed trading is also found. Further investigations reveal a common, market-wide component in information asymmetry related to skilled information analysts who, consistent with Kim and Verrecchia (1994, 1997), generate private information from public data. It further appears that trades by this category of investors successively impound private information into prices, which is consistent with the strategic cost-minimizing trading behavior of

[6] In reality, one may figure market-wide informed trading as trading, done by large institutional investors that act in concert to similar information, thereby generating the same orderflow. Information that may be relevant to these investors may be based on economic data or announcements of economic and monetary policy decision that are released on a particular point in time and apply to many firms in a similar fashion. Alternatively, Admati and Pfleiderer (2000) show that new data that applies to a small set of dominant firms in a particular industry may be used by investors to enhance their information-set of related firms in the same industry, potentially causing the same trades across many stocks at once. To implement these trades, traders may need to make use of electronic devices. Some types of investors, e.g., hedge funds, may be particularly strong users of this type of information as its implementation calls for a high degree of technical expertise while the associated turnover-considerations, in particular trading costs, are less important. Malkiel and Saha (2005) report that a hedge funds are considered to account for up to half of the trading volume of the NYSE, showing that these traders, which are typically associated with informed traders, are likely to be amongst those that implement high turnover strategies that call for simultaneous trades of many stocks at once. (See also Khandani and Lo (2007) for hedge-fund strategies and their turnover implications.)

informed investors in Kyle (1985) and the empirical findings in Chordia and Subrahmanyam (2004). In addition, informed trading turns out to decrease in firm size and, consistent with Dennis and Weston (2001), in the size of the ownership stake of outside parties. The results are also in line with Easley, et al. (1998) as the level of informed trades seems to be lower if there are exchange-traded options on the stock. This indicates that some informed traders migrate to the options market, which reduces the exposure of equity investors to informed trades.

About forty-five percent of explained variation in informed trading can be attributed to market-wide commonality, while the rest is attributed to the firm-level environment captured by firm-specific structural characteristics and the stock-level trading environment. Differences in the relative importance of market-wide commonality, the stock-level trading environment, and firm-specific structural characteristics across firm size highlight the relevance of disclosure that takes into account characteristics of the firm as shown in the model by Admati and Pfleiderer (2000).

Finally, the unexplained part of informed trading, *RAIN*, is priced in the asset's required rate of return, and its effect on pricing is stronger and more robust than that of total informed trading. This shows that the price relevance of information risk is not just the result of the inability to diversify away exposure to informed trades as is commonly believed. Results indicate that the inability to forecast the level of informed trading, and hence the inability to differentiate between random price fluctuations and changes in fundamental value, is an important driver of the information risk premium.[7] Thus, changes in the component of informed trading that are unrelated to firm characteristics, the trading environment, and the market as a whole contribute significantly to the information risk premium.

[7] Bhushan (1989b) shows that investors can free ride on the costly information acquisition process of information analysts by observing the stock price and thereby, at least partially, infer the information that the informed investor has acquired himself at a cost (see also Admati and Pfleiderer (2000)). The weaker the link between the information content of price innovations and the environment is, however, the less useful as a source of information prices are, as it becomes increasingly difficult to differentiate random price fluctuations from price changes induced by trades on private information.

To summaries, this paper contributes to the literature related to the information environment of a firm in two important ways. First, it is shown how time-series and cross-sectional factors influence informed trading, and how these can be accounted for to infer abnormal information risk. Firms and traders could potentially use the empirical results presented here to limit exposure to informed trading. Second, this paper addresses the question of the price-relevance of the information environment of a firm and thereby contributes to academic market microstructure research that relates the information environment to asset prices (see, e.g., Easley, et al. (2002); Easley and O'Hara (2004)).

The remainder of this paper is structured as follows. Section 2 develops the relevant hypotheses, Section 3 presents the methodology, and Section 4 describes the data. Section 5 documents the empirical results, and Section 6 provides concluding remarks.

# 2 Development of Hypotheses

One source of variation in information asymmetry could be attributed to common factors next to firm-specific ones as is commonly suggested.[8] Variation in trading behavior has been attributed to firm-specific trading characteristics (e.g., see Chordia and Subrahmanyam (2004)) or cross-sectional firm-level structural characteristics (e.g., see Breen, Hodrick, and Korajczyk (2002)). As observed trading behavior is considered to be related to its information content (see, for instance, Kyle (1985)), we expect that information asymmetry measured in financial markets to be related to market-wide commonality factors, and firm-level trading and structural characteristics. So far, the specific nature of the systematic factors in information asymmetry received little em-

[8] Hasbrouck and Seppi (2001) explain commonality in volume by investors who use a similar set of information to trade a similar portfolio of stocks. Kim and Verrecchia (1994) show that this set of information may include private information made out of what is publicly known. Commonality factors that underlie market-wide changes in observed trading behavior are therefore likely to include market-wide changes in trades that exploit private information on market-wide changes in expected returns. As private information can therefore refer to systematic as well as idiosyncratic return components (Hughes, Liu, and Liu (2005)), one would expect that variation in observed information asymmetry is the result of changes in market-wide and firm-level factors.

pirical attention, however. We address this gap in the literature and derive a set of testable hypotheses based on empirical implications of microstructure models and hypothesized relationships between information asymmetry and observable data found in empirical studies.

Order-imbalance, for instance, is commonly attributed to trading pressure created by informed trades (e.g., Kyle (1985)), which also applies to observed return-volatility (French and Roll (1986)). As prices converge faster to their underlying value with lower minimum tick-size (Chordia, Roll, and Subrahmanyam (2005)), it is likely that smaller minimum price-increments in general lead to informationally more efficient prices. This suggests that the lower the tick-size, the more information is impounded into prices over a given time-interval, which results in a higher level of information asymmetry measured in the financial markets. Alternatively, as prices are more efficient the lower the tick-size (Chordia et al. (2005)), one could also see a decline of information asymmetry in tick-size, which makes the association of tick-size and realized information asymmetry an empirical issue.

The most common way liquidity providers react to informed trades is by increasing the bid-ask spread to recover their expected losses to informed traders (Glosten and Milgrom (1985)). As informed traders act strategically in timing their trades (Kyle (1985)), they are likely to postpone their trading activity until market depth recovers. Higher trading activity could lead to this recovery, as it is considered being mainly the result of liquidity providers joining the market (Glosten and Milgrom (1985)). Alternatively, higher volume results in relatively noisy prices, which increases the benefit of private information (Bhushan (1989a)). As informed traders hide among liquidity traders (Admati and Pfleiderer (1988)), this could also lead to more informed trades with higher trading volume. If some traders have private information about systematic return components, one would expect their trading to affect aggregate market-volume, as fluctuations in random uninformed trading activity should cancel out on a market level. This implies that higher market-level trading activity is likely the result of informed traders joining the market (as Admati and Pfleiderer (1988) show for intra-day trade-clustering) while increases in trading activity on an individual stock-level is likely to be due to uninformed traders.

We therefore expect to find commonality factors in information asymmetry that are positively related to market-wide changes in bid-ask spreads, trading volume, volatility, and order-imbalance. The observed market-wide variation should be the result of informed trades acting on common, market-wide signals.

Most of the literature on information asymmetry considers the interaction of informed trades and firm-level phenomena, whereby firm-specific trading characteristics receive particular attention by market-microstructure research. Based on earlier studies mentioned previously, we hypothesize that stock-level volatility, order-imbalance, and bid-ask spreads all have a positive association with the level of information asymmetry. Stock-level trading volume and tick-size, however, are expected to have a negative association with the level of information asymmetry. Similarly, the strategic trading-behavior of informed traders (Kyle (1985)) should be reflected in a negative association between the level of information asymmetry and large drops in liquidity, such as large increases in bid-ask spreads.

Chalmers and Kadlec (1998) find that the asset characteristics of a firm are also important in determining how the stock of a firm is traded. Therefore, firm-specific structural characteristics, reflecting asset characteristics and operating conditions constitute another set of factors that potentially influence the information environment and hence the level of information asymmetry public investors face. The more the public knows about a firm, the less scope there is for information asymmetry to arise. One rough measure of public exposure is firm size, which has already been shown to be negatively related to the informativeness of stock-trades (Hasbrouck (1991a, 1991b)). As investors in large firms have more sources of information that get more frequently updated than the information that investors in small firms have (Bushan (1989a)) and small firms tend to suffer from insider trading (Lakonishok and Lee (2001)), one would expect information asymmetry to decrease in firm size.

The difficulty to derive a firm's risk and expected return from its assets is another factor that determines the scope better informed investors have in exploiting their informational advantage. There has been some empirical evidence that intangible assets, for instance, have a highly uncertain payoff (Kothari, Laguerre, and Leone (2002)) and are difficult to value for company-outsiders (Cotter and Richardson (2002)). This suggests that the more intangible assets a firm has, the more incentives there are to acquire additional, private information and to exploit the resulting information asymmetry in the financial markets. This suggests that one should expect

information asymmetry to increase in the relative level of intangible assets.

Firms could reduce information asymmetry by providing guidance on how their results should be interpreted. There is only a thin line, however, between making investors more knowledgeable and managing perceptions, for instance in order to avoid negative earnings surprises. Rather than giving an objective interpretation of its situation, a firm may create information asymmetry by managing investors' perception into a desired direction. Those investors that are able to identify such window-dressing may therefore have a more objective set of information and therefore be at an information advantage to the general public. It has been shown empirically that this window-dressing is most common at small, profitable firms with growth options (Matsumoto (2002)). In addition, growth firms that manage the perception of their investors tend to have a high book-to-market ratio (Bhattacharaya, Black, Christensen, and Mergenthaler (2004)).

Company-insiders (Aboody, Hughes, and Liu (2004); Lakonishok and Lee (2001)) and company-outsiders (Maug (2002)) seem to exploit their information advantage. In addition, the larger the ownership stake, the higher the level of information asymmetry should be as this class of informed investors may get additional price-relevant information, by, for instance, being represented on a company's board of directors (Maug (2002)). On the other hand, outside shareholders may improve the communication with financial markets (Bushee and Noe (2000)), which therefore makes the role of the ownership structure in determining the level of information asymmetry investors face in the financial markets an empirical issue. Finally, it has been shown that investors exploit information asymmetry by means of options (Easley et al. (1998)). The higher leverage of options relative to stocks could make informed traders prefer these instruments. The realized level of information asymmetry in the stock market should therefore be lower if options are traded as some of the informed traders move to the options market.

In sum, we hypothesize that the level of information asymmetry measured in the stock market is lower the larger the firm, the more measurable (tangible) its assets, the lower its propensity to window-dress, and if there are options written on its stock.

Whether investors care about the existence of information asymmetry can ultimately be looked into when testing for whether exposure to information asymmetry is a priced risk factor. Typically, the justification for information asymmetry to be a priced risk-factor is based on the difficulty to diversify it away (Aboody, Hughes, and Liu (2004)) as it is one-sided – the uninformed al-ways loses out to the informed trader. If non-diversifiability is the only factor that gives rise to the information risk-premium, this effect should be more pronounced the stronger the common variation in information asymmetry as a rising level in information asymmetry in one stock would not be offset by reductions in another.

The informativeness of prices can be increased – and hence the information risk-premium be reduced – by trades of informed investors who impound their information into prices. Observed security prices may be used by public investors to complement their information set (e.g., see Habib, Johnsen, and Naik (1997) and Admati and Pfleiderer (2000) for theoretical models where the public investor uses security prices to improve her information set). This implies that informed trading, which helps impounding previously private information into prices, makes this information public and therefore help the uniformed update their information set (Easley and O'Hara (2004)). This implies that the stronger common variation is in information asymmetry, the better one can assess the actual level of information asymmetry. This allows the public investor to better differentiate between true changes in fundamental value and random price fluctuations. Therefore, it is also possible that the information risk-premium declines the stronger common factors are. As both could effects co-exist, it is an empirical issue whether the most important source of the information risk-premium is rather due to the inability to diversify or to the inability to forecast the information environment. Specifically, the return relevance of the systematic common-factor dependent component of informed trading and the return relevance of the residual asymmetric information, *RAIN*, is tested. These hypotheses are summarized in Table 1.

# 3 Methodology

## 3.1 Measuring Informed Trading

One of the measures used is the popular *PIN* (labeled $PIN_1$ hereafter) developed and empirically tested in, for instance, Easley, et al. (1997b). Assuming a unit trade size and uninformed trades to occur randomly, the trading session is modeled as a hazard process of informed and uninformed trades. Private information may be revealed to the informed investor at the beginning of each trading session only. The basic assumption is that informed trades reflect this private information by being either exclusively buys or sales depending on the context of the private information. The trade direction of uninformed investors is assumed to fluctuate randomly between buys and sells. The resulting daily number of buys and sells serves as the empirical input of the model. The estimated coefficients of the theoretical model are used to calculate $PIN_I$ over a set of $T$ trading days.[9] Many empirical studies use $PIN_I$ (see, e.g., Easley, et al. (2002); Odders-White and Ready (2006); Vega (2006)), which shows that this measure is well behaved and yields intuitive results. One of the disadvantages of this measure is, however, that it is estimated over many trading sessions, which makes it less suited for identifying short-lived changes in the information environment. One could, however, adjust the estimation of $PIN_I$ such that $PIN_I$ is available on a higher frequency. The computational difficulties encountered when empirically estimating $PIN_I$, however, makes the use of the publicly available $PIN_I$ data estimated over a one-year horizon more appear most reliable.[10]

To make up for this shortcoming of $PIN_I$, informed trading is alternatively estimated by the daily average loss of liquidity suppliers to traders demanding liquidity. As this loss should, on average, be zero in the absence of information asymmetry, this direct measure of adverse selec-

[9] More specifically, the daily number of buy transactions, $B_t$, and sell transactions, $S_t$, on day $t$ are used as input to the daily likelihood function, $L_t$:

$$L_t\left(\theta \middle| B_t, S_t\right) = \left(1-\alpha\right)e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!} + \alpha\delta e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}e^{-(\mu+\varepsilon_s)}\frac{\left(\mu+\varepsilon_s\right)^S}{S!} + \alpha\left(1-\delta\right)e^{-(\mu+\varepsilon_b)}\frac{\left(\mu+\varepsilon_b\right)^B}{B!}e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!},$$

where $\theta$ is the parameter vector defined as $\theta = (\alpha, \mu, \varepsilon_b, \varepsilon_s, \delta)$, $\alpha$ denotes the probability of an information event to occur and indicating a drop in asset value with probability $\delta$. The resulting information-based orderflow is denoted by $\mu$, uninformed sell transactions are denoted by $\varepsilon_s$ and uninformed buy transactions are captured by $\varepsilon_b$. Maximum likelihood estimate of the individual coefficients that are part of parameter vector $\theta$ are calculated via the product of $L_t$ over $T$ days. Subsequently, the following relationship is used to derive $PIN_I$:

$$PIN_I = \frac{\alpha\mu}{\alpha\mu + \varepsilon_s + \varepsilon_b}.$$

[10] These problems have also been documented by Easley, Engle, O'Hara, and Wu (2001), Easley, Hvidkjaer, and O'Hara (2004), and Vega (2006), which is essentially a truncation error that arises from the fact that the software used for the maximum likelihood estimation reaches its numerical limit. Re-arranging the likelihood function (results not shown to preserve space but available on request from the authors) show that these problems occur when the number of transactions is fairly high or if the estimated parameter $\mu$, which is the imputed level of informed orderflow, is large relative to the uninformed orderflow, $\varepsilon$.

tion results in a numerical estimate of the level of informed trading during a trading session. This variable, labeled $IA_1$, captures the total loss to informed traders by liquidity suppliers, who, on average, can be presumed to have all publicly available information. Huang and Stoll (1996) capture this loss by the difference between the quote mid-point, defined as the sum of the bid and ask quotes divided by two, at the time of the transaction and the re-corded transaction price a fixed time later. Bessembinder and Kaufman (1997) use the difference of the transaction price and the quote mid-point a fixed time interval later divided by the quote mid-point at the time of the transaction. Naik and Yadav (2003b) replace transaction prices by the quote mid-point and thereby address problems related to the bid-ask bounce and unequally spaced transaction times (see, e.g., Lease, Masulis, and Page (1991)). Hasbrouck and Sofianos (1993) also use a similar measure. Daily averages of this variable are robust and provide estimates at a sufficiently high frequency. At a transaction level, $IA_1$ is defined as:

$$IA_{1,t} = D_t \left( M_T - M_t \right) / M_t \,, \tag{1}$$

where $D_t$ is a trade direction indicator taking a value of +1 for a buy transaction and –1 for a sell transaction, $M_t$ and $M_T$ are the quote mid-points at the time of the transaction, $t$, and some time, $T$, later. To account for variation in the time horizon new information is impounded into prices, $IA_1$ is estimated over 15 minutes, 30 minutes, 60 minutes, and over one day.[11] Taking daily value-weighted averages allows the inclusion of trade size as an additional variable that accounts for private information.[12] Following Bessembinder and Kaufman (1997), quotes posted at least five

---

[11] $IA_1$ over one day is calculated using $M_T$ in effect at the exact same time of the day than $M_t$ but one trading day later. One could calculate $IA_1$ also over longer horizons to capture information that takes even longer than one day to be reflected in stock prices. The link between the information contained in a particular transaction at time $t$ and the asset value some time later strongly weakens the more trading sessions one considers as new additional information gets revealed that may not be known to the trader that submitted her transaction at time $t$.

[12] Employing the Lee and Ready (1991)-algorithm to infer the trade-direction used in the calculation of $IA_1$ be problematic as measurement errors induced by the Lee and Ready (1991)-algorithm may in some way be correlated with the error of $IA_1$ to quantify the level of informed trading. As equity markets data are used, one still needs to infer the trade-direction to calculate alternative measures of informed trading, such as orderflow (see Ellis, Michaely, and

seconds before the reference trade are used.[13] The comparison of estimates of this information asymmetry variable with similarly defined variables used in previous studies and some simple time-series diagnostics show that the $IA_1$ values have reasonable characteristics.[14]

A third group of well-established measures of informed trading use the serial covariance of quote mid-points (see Huang and Stoll (1997) for an overview of these models within a unifying framework).[15] The advantages of the covariance-based measures of informed trading do not seem to offset their empirical shortcomings, however, due to the following methodological weaknesses. First, the estimation procedure of informed trading is fairly resource intensive.[16] Second,

---

O'Hara (2000) for a discussion of commonly used trade-direction algorithms). In addition, Froot and Ramadorai (2005) show that orderflow does not exclusively measure informed trades but that it affects prices via uninformed portfolio-balance and liquidity effects. This paper tries to account for the private information contained in orderflow. First, daily averages of $IA_1$ are weighted by the U.S. dollar trading volume associated with each transaction. Second, $PIN_1$ is based on orderflow. Both should give orderflow, which Evans and Lyons (2004), for instance, identify to be an important vehicle of value-relevant information in the foreign exchange market, considerable weight in the empirical set-up.

[13] Outliers are cleaned by excluding the first half-hour of the trading day, $IA_1$ observations larger than ten percent (Huang and Stoll (1996)), and those eight standard deviations away from the daily stock-level average. In addition, the daily top and bottom 0.1 percentiles are deleted to minimize the influence of extreme observations.

[14] The 29.7 basis points reported by Bessembinder and Kaufman (1997) for an equally-weighted average over 24 hours is close to the 22.8 basis points of the average IA1 calculated the same way. The values reported in Huang and Stoll (1996) and Bessembinder (2003a, b) are also very similar if appropriate adjustments, such as dividing by the average quote mid-point, are made. Finally, average auto-correlation of $IA_1$ is only significant for first order auto-correlation, strongly decreasing in the time interval between $M_t$ and $M_T$ and in firm size. The only significant AR(1) coefficients are for the time intervals of 15 minutes with medians (*t*-values) of 0.08 (1.90).

[15] Building on Glosten and Milgrom (1985), these models specify that market makers adjust quotes in response to changes in fundamental values, to expectations of future adverse selection losses, and to changes in the costs of carrying inventory. Researchers define changes in fundamental values to follow a white noise process and infer informed trading as the fraction of quote changes unexplained by a simple inventory model. The inventory model, typically based on Ho and Stoll (1981), assumes market makers to passively respond to orderflow by adjusting bid quotes downward and ask quotes upward after an incoming transaction that is seller or buyer initiated, respectively. These models typically infer the effective spread as twice the square root of minus the auto-covariance of transaction price changes Roll (1984). Glosten (1987) is among the first models to decompose the spread into inventory management and adverse selection components. This model is developed further in George, Kaul, and Nimalendran (1991), for instance, who adjust stock-price changes by estimated changes in the expected return.

[16] Huang and Stoll (1997) use 20 stocks during the year 1991, a year where the amount of data to be processed was likely much lower than today (see, e.g., Easley, et al. (2004) for a discussion of the increase in intra-day activity over time and the associated computational problems). The empirically simpler model by George, et al. (1991) has its own practical difficulties. As large orders are often broken up into successively executed smaller ones, positive auto-correlation in stock returns poses a significant problem (Harris (1990)). Neal and Wheatley (1998) report the level of in-formed trading to monotonically increase in the data sampling frequency when the George, et al. (1991)-measure is used. Our own calculations (unreported but available on request from the authors) lead to similar conclusions. This evidence make the George, et al. (1991)-measure not sufficiently reliable to be used here. By contrast, the data set used in this paper contains daily stock-level observations of around 1,500 individual stocks covering eleven years of

the underlying inventory model assumes that the only statutory obligation of market makers consists of accommodating incoming trades, which does not reflect the situation of the market-maker on the NYSE, the stock market of interest.[17] Third, additional factors are likely to influence market makers' trading and quote-setting behavior.[18] This evidence suggests that the quote-setting behavior of market makers on the NYSE is more complex than suggested by the simple inventory model that underlies the covariance-based measures of informed trading. By contrast, $IA_1$ is independent of the potential trade motivations and market makers' trade motivation as it only relies on the average profit generated from a transaction over a fixed time-interval. Given the comparatively high costs of estimating covariance-based measures of informed trading, the set of information measures is limited to $PIN_1$ and $IA_1$ as these variables are less costly to acquire calculate or have simpler and more robust structural assumptions as compared to the covariance-based measures.

## 3.2 Systematic Variation of Informed Trading

To capture variation of informed trading related to market-wide commonality, the stock-level trading environment, and to firm-specific structural characteristics, the level of informed trading is regressed on the set of explanatory variables as discussed in Section 2. The regression analysis is divided into three parts as this allows examining how the step-wise removal of explained variance of informed trading affects excess returns.[19] For this purpose, informed trading alternatively

---

a time period that is likely to contain more individual trade observations than the data set of the studies mentioned above.

[17] Market makers quote-setting behavior on the NYSE has to account for additional obligations, however, such as price-continuity (NYSE (2006)). This commitment seems to significantly affect market makers' quote-setting behavior and profitability (Panayides (2004)).

[18] These influences include competition for orderflow with regional exchanges, the risk-preferences, capitalization, and costs of capital of market markers, and cross-subsidization of income generated from active stocks to support losses made on inactive ones (Cao, Choe, and Hatheway (1997)). Market makers may even adjust their quotes to exploit their private information derived from the order book (Ready (1999)).

[19] Alternatively, one could relate informed trading to its explanatory variables in one single regression. This procedure does not allow, however, examining how the step-wise removal of explained variation of informed trading af-

captured by $PIN_1$ or $IA_1$, is first regressed on the set of market-wide commonality variables and variation explained by these variables is subsequently removed.[20] This procedure is repeated using stock-level trading characteristics and firm-level structural characteristics until the three sources of common variation of informed trading are taken out of informed trading. Each regression has a firm-specific intercept. Accordingly, the first regression, used to investigate the presence of common market-wide components in informed trading, is specified as:

$$InfoTrade_{1,i,t} = \beta_{i,0} + \beta_1 MBA + \beta_2 MVOL + \beta_3 MVLA + \beta_4 MOIB + \varepsilon_{i,t}, \tag{2}$$

where $InfoTrade_{1,i,t}$ alternatively denotes $PIN_1$ or $IA_1$ of stock *i* on day *t* and the variables *MBA*, *MVOL*, *MVLA*, and *MOIB* are the daily market-level bid-ask spread, U.S. dollar trading volume,

---

fects excess returns. Computations are done both ways. The method used does not greatly affect the sign, magnitude, and significance of the estimated regression coefficients.

[20] The order in which regressions (2) to (6) are estimated is arguably arbitrary. The main theme behind the order chosen in the current set-up is that we go from the general (market-wide variation) to the specific (stock-level trading characteristics or firm-level structural characteristics) and from time-series level variation captured by the trading environment (market-wide and stock-level trading characteristics) to cross-sectional variation captured by the firm-level structural characteristics. One weakness of this approach (instead of putting all variables into the same regression) is that the estimates may suffer from an omitted variable bias (at least to a larger degree than if the regression would be estimated in one go including all of our right-hand side variables). Estimating these regressions in one go hardly affects the sign and the significance of the estimates, but does not allow showing the change in loading of information asymmetry when used in the regression set-up discussed in Section 5.5. We therefore use the current three-step format.

Nevertheless, we attempt to empirically verify whether the order of the regressions as spelled out in this section is supported by the data, i.e., that it reflects roughly the explanatory power of the variables. We therefore estimate a step-wise regression where for each number of right-hand side variables (i.e., from one to eighteen variables), those who best explain information asymmetry as measured by the maximum $R^2$ are retained. For this exercise to be most meaningful in terms of capturing variation in information asymmetry and extending its validity also to Section 5.5, information asymmetry is measured by daily $IA_1$. The step-wise regression is estimated on daily $IA_1$ measured over 15 minutes, 30 minutes, 60 minutes, and 1 day by size decile with a general intercept (i.e., no firm-specific one) to focus on the explanatory power of the individual right-hand side variables. Results show that the variables capturing information asymmetry best are the market-wide and stock-level trading characteristics. Firm size only occasionally features as the third or fourth most important variable across all time-horizons and size deciles. From this exercise, it generally emerges that bid-ask spreads, volatility, and – for the one-day horizon – order-imbalance are the most important variables (which is consistent with what is reported in Section 5.6). For smaller firms, the stock-level measures are most important, whereas for larger firms, market-level bid-ask spreads and volume tend to be the most important variables. This shows that the order in which the regression are run may roughly correspond to the empirical importance of the variables, which implies that potential effects from the omitted variable bias resulting from estimating the regression in three steps rather than in one go is limited. Results also show, however, that stock-level trading characteristics are at least as important as market-wide trading characteristics in explaining observed information asymmetry. One could therefore first regress information asymmetry on stock-level trading characteristics, then on market-wide commonality factors, and then on firm-level structural characteristics. The estimated regression coefficients of these regressions are hardly different from what is reported in Table 5 (see also footnote 35).

volatility, and order-imbalance, respectively. The market-wide component is subsequently taken out of $InfoTrade_{1,i,t}$ by calculating $InfoTrade_{2,i,t}$ defined as:

$$InfoTrade_{2,i,t} = InfoTrade_{1,t} - \left(\hat{\beta}_1 MBA + \hat{\beta}_2 MVOL + \hat{\beta}_3 MVLA + \hat{\beta}_4 MOIB\right). \tag{3}$$

$InfoTrade_{2,i,t}$ is subsequently related to firm-level variables. The relationship between stock-level trading characteristics and informed trading is investigated by the following regression:

$$\begin{aligned} InfoTrade_{2,i,t} = \gamma_{i,0} + \gamma_1 VLA_{i,t} + \gamma_2 BA_{i,t} + \gamma_3 OIB_{i,t} + \gamma_4 TIC_{i,t} + \\ \gamma_5 UEDSpread_{i,t} + \gamma_6 VOL_{i,t} + \eta_{i,t}, \end{aligned} \tag{4}$$

where *VLA*, *BA*, *OIB*, *TIC*, *UEDSpread*, and *VOL* are stock-level volatility, bid-ask spread, order-imbalance, tick size, unexpected changes in the bid-ask spread, and trading volume, respectively. Variation attributable to stock-level trading characteristics is taken out by calculating:

$$\begin{aligned} InfoTrade_{3,i,t} = InfoTrade_{2,i,t} - \big(\hat{\gamma}_1 VLA_{i,t} + \hat{\gamma}_2 BA_{i,t} + \hat{\gamma}_3 OIB_{i,t} + \\ \hat{\gamma}_4 TIC_{i,t} + \hat{\gamma}_5 UEDSpread_{i,t} + \hat{\gamma}_6 VOL_{i,t}\big). \end{aligned} \tag{5}$$

$InfoTrade_{3,i,t}$ is used to investigate the relationship between structural characteristics and information environment:

$$\begin{aligned} InfoTrade_{3,i,t} = \delta_{i,0} + \delta_1 Insider_{i,t} + \delta_2 Outsider_{i,t} + \delta_3 Capex_{i,t} + \delta_4 R\,\&\,D_{i,t} + \\ \delta_5 BTM_{i,t} + \delta_6 Profit_{i,t} + \delta_7 Options_{i,t} + \delta_8 Size_{i,t} + \xi_{i,t}, \end{aligned} \tag{6}$$

where *Insider*, *Outsider*, *Capex*, *R&D*, *BTM*, *Profit*, *Options*, and *Size* denote the fraction of common stocks held by corporate insiders and outsiders, capital expenditures, R&D expenses, the book to market ratio, the profit margin, and an indicator for the availability of exchange-traded options on stock *i*. To investigate whether the absence of a relationship between informed trading with variables that capture the environment of a firm is what specifically exposes the investor to priced information risk, *Residual Asymmetric Information*, *RAIN* is calculated according to:

$$RAIN_{i,t} = InfoTrade_{3,i,t} - \Big(\hat{\delta}_1 Insider_{i,t} + \hat{\delta}_2 Outsider_{i,t} + \hat{\delta}_3 Capex_{i,t} + \hat{\delta}_4 R\,\&\,D_{i,t} + \hat{\delta}_5 BTM_{i,t} + \hat{\delta}_6 Profit_{i,t} + \hat{\delta}_7 Options_{i,t} + \hat{\delta}_8 Size_{i,t}\Big), \tag{7}$$

where $RAIN_{i,t}$ denotes the unexplained residual part of informed trading of firm $i$ on day $t$. It represents the observed level of informed trading that deviates from what the public investor expects given the market environment, the stock-level trading environment, and features that characterize each particular firm. As informed trades based on insider information should not show co-variation across stocks, *RAIN* is likely to capture informed trading that is most strongly associated to insider trades.

# 4 Data Sources

Except for $PIN_1$, which is in yearly frequency, and return data, which are kept in monthly frequency, all other variables are expressed in daily frequency to improve variable-synchronicity. Intra-day data are from TAQ and cover the period between the 2nd of January 1995 and the 30th of December 2005. From the data sample excluded are REITs, ADRs, ADSs, closed-end funds, convertibles, preference shares, multiple classes of shares, warrants, rights issues, certificates, and stocks with less than 60 days of quotes or trades per calendar year. Stocks with a price below one U.S. dollar at the end of a calendar year are also excluded to ensure a minimum level of liquidity and avoid undue influence of the discrete price grid. Trades at the market open, trades out of sequence, trades with special settlement conditions, trades outside the market opening times, or trades that have been corrected are all purged, as are quotes that are posted during the market open, quotes that are negative, or quotes that lead to a bid-ask spread that is either negative, above five U.S. dollars, or larger than 40 percent of the transaction price.[21] The sample is confined to stocks traded on the NYSE as primary market. Data from regional exchanges can be un-

[21] These cleaning procedures are common for these data (see, e.g., Chordia, et al. (2000)).

reliable for stocks that have their primary listing on the NYSE (Odders-White and Ready (2006)). Therefore, daily time-weighted averages of the best bid and offer (hereafter BBO) quotes are calculated using NYSE data only. The trade direction is inferred using the Lee and Ready (1991)-algorithm, which matches trades with quotes posted at least five seconds before this trade is executed. This data item is used to calculate, for instance, raw order-imbalance as the net of U.S. dollar volume bought and sold each day.

Monthly and daily stock returns, closing stock prices, value-weighted market-returns, the number of shares outstanding, four-digit SIC codes, and the daily share volume are retrieved from CRSP. Using these data, stock-level volatility is calculated as the squared daily return,[22] and tick size is defined as the inverse of the closing stock price.[23] U.S. dollar volume is calculated as the product of the closing stock price and share volume, and firm size is defined as the daily product of the closing stock price and the number of shares outstanding. The Fama and French (1995)-factors SMB and HML and the one-month Treasury bill rate are from the Fama-French data-base on WRDS. The Blockholders data set by Dlugosz, Fahlenbrach, Gompers, and Metrick (2006) on WRDS is used to calculate corporate insider ownership as the sum of the percentage of common stock held by executives, directors, and affiliated entities. Ownership by corporate outsiders is defined as the fraction of common stock held by anyone who is neither affiliated nor employed by the respective firm. Values of $PIN_I$ are from Soeren Hvidkjaer's homepage.[24] The data to calculate profit margins, the ratios of book value to market value, R&D expenses to sales,

[22] An alternative volatility proxy is the sum of the squared intra-day returns. Using this definition may potentially lead to methodological problems in our case as our sample includes some stocks that are not traded sufficiently frequent.

[23] Tick size measures the size of the minimum price-increment and therefore the resolution of the price-grid. For instance, a stock priced at $5 has a minimum possible price-increment of 1 cent (or $1/8 prior to the decimalization of NYSE prices) corresponding to 0.002 (0.025) of its price. A stock priced at $200 has a minimum possible price increment corresponding to 0.00005 (0.000625) of its price, which suggests that the price-grid of the stock priced at $200 is finer than the price-grid of a stock priced at $5.

[24] We would like to thank Soeren Hvidkjaer for making the $PIN_I$ data available on his website: http://www.smith.umd.edu/faculty/hvidkjaer/data.htm.

and capital expenditures to sales are from COMPUSTAT, whereby COMPUSTAT data are winsorized at the first top and bottom percentile.[25] The options-availability indicator, based on the Ivy DB database of Option Metrics, is one if options on the common stock of a firm are registered on an exchange in the U.S. on a particular day and zero otherwise.

Order-imbalance is defined as the sum of the intercept and the residual of a regression of the ratio of absolute daily raw U.S. dollar imbalance to daily U.S. dollar volume on U.S. dollar trading volume. The volume data that are further used in the empirical analysis are defined as the residuals of a regression of changes in U.S. dollar volume on market volume, stock-level volatility, and market returns (see Chordia, et al. (2000) for a similar set-up):

$$\$Volume_t = \vartheta_0 + \vartheta_1 MVolme_t + \vartheta_2 MVolume_{t-1} + \vartheta_3 r_{m,t} + \vartheta_4 r_{m,t-1} + \vartheta_5 r_{i,t}^2 + \tau_t, \tag{8}$$

where $\$Volume_t$ is the percentage change in U.S. dollar volume from the previous trading day to day $t$, $r_{m,t}$ and $r_{i,t}$ are the return on the market and on stock $i$ over the same period, and $MVolume_t$ is the equally-weighted market average of stock-level percentage changes in U.S. dollar volume from the previous day to day $t$. Defining volume this way improves the comparability of volume across stocks and removes the time-trend in U.S. dollar volume. Similarly, regression (8) is estimated by replacing *Volume* and *MVolume* by the daily percentage changes in bid-ask percentage spreads and the market average of changes therein. The residual of this regression is labeled *UEDSpread*, and represents the unexpected change in bid-ask spreads, which serves as our proxy for large changes in bid-ask spreads.[26]

[25] These variables are operating profits (item 13), total sales (item 12), the necessary items to calculate book value excluding preference shares (items 60, 74, and 208 less items 56, 175, and 130), R&D expenses (item 46), and capital expenditures (item 128). Firms with negative book values are excluded. Missing values from COMPUSTAT and the Blockholders database are set to zero. Results are insensitive to these data cleaning procedures.

[26] The estimated coefficients of this regression are very close to what Chordia, et al. (2000) report with and without a lead-term. The intercepts are insignificant with average p-values of 0.23 and 0.38 for volume and bid-ask spreads, respectively.

Market-level bid-ask spread, trading volume, and order-imbalance are defined as the value-weighted averages of the stock-level values. The new methodology VIX index is used to measure market-volatility.[27] The merged data set contains 2,407 individual firms that have valid observations for all data-items, with each year having between 1,287 and 1,641 individual firms.

Summary statistics of the data are shown in Table 4. Mean values of unexpected changes in bid-ask spreads are around zero, which is as expected given the definition of this variable as a regression residual. The mean book-to-market ratio is very close to what other studies, such as Easley, et al. (2002), report. Most of the variables exhibit some degree of skewness and a strong size-effect. The value-weighted market average of bid-ask spreads, for instance, is much smaller than the simple mean of the stock-level equivalent, showing that small firms have a much larger bid-ask spread than large firms. Many other empirical market microstructure studies rescale the input data to improve their distributional characteristics. Rescaling the variables has the additional benefit of making cross-sectional comparisons more meaningful Naik and Yadav (2003a). Therefore, the variables that are meant to capture market-wide commonality, stock-level trading characteristics, and firm-specific structural characteristics are rescaled using the non-parametric method of normal scores, which replaces the respective variable value by its ranking scaled by some factor.[28] Thereby, data referring to market-wide commonality and stock-level trading char-

[27] The new methodology VIX index is downloaded from the website of the CBOE:
http://www.cboe.com/micro/vix/introduction.aspx.
It is the implied volatility of near-to-expiry options on the S&P 500 index and thereby provides an *ex ante* forecast of expected future volatility. Our intention is to capture the markets' concurrent information environment, which makes the VIX measure preferable to the backward looking past realization of volatility. Blair, Poon, and Taylor (2001) show that the VIX index has statistical properties equivalent to the actually realized market volatility, which provides confidence that VIX is a statistically sound measure. In addition, as these data is readily available, one avoids errors induced by calculating market-volatility from raw return data.

[28] Odders-White and Ready (2006), for instance, rescale their data using quartiles, and Brennan, Chordia, and Subrahmanyam (1998) normalize their data to a mean of zero and variance of one. Llorente, Michaely, Saar, and Wang (2001) use a non-parametric method to rescale their explanatory variables based on the relative ranking of each data point within the data-series considered. Taking the logarithm of the variables instead of rescaling is probably more common in empirical studies but requires the data to be strictly larger than zero. In addition, it implies a functional relationship (an exponential one) between informed trades and the explanatory variables, which cannot be

acteristics are rescaled within each time-series individually and the time-specific structural variables are rescaled daily across the cross-section.[29] Table 2 summarizes the definitions of the empirical measurements used in this study.

# 5 Discussion of Results

## 5.1 Univariate Analysis

The correlation matrix presented in Table 3 gives a first impression of the co-variation between informed trading and the set of explanatory variables. The association of $PIN_1$ with the explanatory variables is mostly consistent in sign with the association $IA_1$ shows, suggesting that both measures of informed trading capture the same phenomenon.

The variables that show the hypothesized relationship with informed trading are bid-ask spreads, volatility, order-imbalance, firm size, insider ownership, and option availability. The univariate association of informed trading with market-level trading volume, asset measurability (R&D and capital expenditures), and one of the proxies for active management of investor's perception (profit margin) is opposite to the hypotheses shown in Table 1.

Amongst the associations that are considered empirical issues, higher unexpected bid-ask spreads and higher stock-level trading volume appear to be associated with a higher level of informed trading. This implies that market makers are relatively sensitive to changes in the information environment as higher levels of informed trading result in high bid-ask spreads. Informed traders seem to cluster when daily trading volume is high, probably to minimize trading costs,

motivated a priori in this context. A non-parametric method, such as the one used by Llorente, et al. (2001) therefore seems to us more appropriate for this study as it is explicitly free from any postulated relationship between informed trading and the explanatory variables. Re-scaling these data to a mean of zero and a standard deviation of one does not significantly alter the results.

[29] As a result, the mean, median, and interquartile range of the market-wide commonality variables are respectively between –0.18 and 0.16, around zero, and between 0.36 and 0.91. The mean, median, and interquartile range of the rescaled stock-level trading characteristics are all close to zero. The mean, median, and interquartile range of the rescaled firm-specific structural characteristics are between –0.17 and 0.07 between –0.79 and –0.05, and between 0.24 and 1.45, respectively.

which extends the intra-day results of Admati and Pfleiderer (1988) to a daily horizon. The association of tick size implies that lower tick size, and a resulting more efficient price discovery process, leads to a lower level of informed trading. Finally, a higher level of outside ownership is associated with a lower level of informed trading, suggesting that outside parties that own large stakes in a firm improve the information flow to the financial markets as suggested by Bushee and Noe (2000). The strong size-effect in most of the explanatory variables makes one cautious, however, to attach too much importance to the univariate results, however. A multivariate analysis is potentially more fruitful, which is implemented in the next section.

## 5.2 Multivariate Analysis

Using the methodological set-up discussed in Section 3.2, $IA_1$ and $PIN_1$ are regressed in three steps on the set of explanatory variables. To ensure comparability of the associations of $IA_1$ and $PIN_1$, both measures need to be in the same frequency. For that purpose, daily $IA_1$ is expressed as yearly average, resulting in the same frequency than $PIN_1$. Comparing $IA_1$ with the widely used $PIN_1$ serves as a consistency check to verify whether both variables capture the same phenomenon of informed trading. Based on this evidence, one can assess the validity of extending the empirical analysis of informed trading on data observed with daily frequency. Analyzing informed trading on daily frequency, which only $IA_1$ allows, potentially yields richer results than the yearly frequency for variables that capture time-series variation on the market-level or the stock-level. Yearly averages of $IA_1$ and the explanatory variables and implement the methodological set-up as discussed in Section 3.2. Results presented in Table 5 show consistent results between $PIN_1$ (see Panels A and B of Table 5) and $IA_1$ (see Panels A and C of Table 5). In what follows, it is specifically referred to either measure of informed trading in the discussion of Table 5 only if the empirical associations of both measures with any of the explanatory variables differ from each other.

### 5.2.1 Commonality in Informed Trading

Results in Table 5 show a strong presence of commonality in informed trading, as all coefficients are highly significant.[30] As hypothesized, higher levels of informed trading are associated with higher bid-ask spreads. Consistent with French and Roll (1986), high volatility is associated with a high level of informed trading. This suggests that during periods of price uncertainty, informed investors particularly exploit their knowledge. Order-imbalance has a positive association with informed trading, which is consistent with the univariate results. The association with volume that is negative for $PIN_1$ and positive for $IA_1$ likely reflects a size-effect, which is picked up by $PIN_1$ but to a much lesser degree by $IA_1$, as $PIN_1$ has a stronger positive relationship with size than $IA_1$ (see Table 3). Estimating this regression on a daily horizon likely helps addressing this issue as the higher observational frequency allows better controlling for size while giving more room for time-series variation to be reflected in the regression coefficients of this panel regression.

Nevertheless, these results show a strong degree of commonality in informed trading. The findings support the argument put forward by Hasbrouck and Seppi (2001) that commonality in trading may be caused by informed investors acting on the same market-wide information. This implies that some of the observed level of informed trading is related to public, market-wide signals. For public information to be related to informed trading, it needs to be processed into something that is not yet known. Commonality in informed trades is therefore likely a reflection of information analysts mentioned by Kim and Verrecchia (1994, 1997), who generate private infor-

[30] The regressions are estimated for $IA_1$ estimated over 15, 30, and 60 minutes and 24 hours. As the results are very similar, present only the results for the 60 minutes horizon.

mation, and hence cause informed trading, by interpreting publicly available data quicker and more effectively than the public investor. How the stock-level trading environment is related to informed trading is looked at next.

### 5.2.2 Stock-level Trading Environment

In a second step, informed trading where variation explained by market-wide variables has been taken out is regressed on stock-level trading characteristics. Results in Table 5 are consistent with the discussion in Section 5.1: Higher levels of volatility and bid-ask spreads are associated with higher levels of informed trading. The two bid-ask spread measures and trading volume make the strongest contribution to explain informed trading across all measures of information asymmetry whereby the strong association of $PIN_1$ with order-imbalance likely reflects the fact that $PIN_1$ is based on these data. The association of unexpected changes in bid-ask spreads and informed trading measured by $PIN_1$ is opposite to what is found when $IA_1$ is used. This may also mirror the way both measures of informed trading are calculated. $PIN_1$ values are estimated over a full year and therefore seem to pick up the long-term effects that express the strategic cost-minimizing trading behavior of informed traders (see, e.g., Kyle (1985)) as the level of informed trading is lower when spreads are unexpectedly high. $IA_1$, by contrast, is originally estimated over a much shorter horizon and may thus pick up the short-term relationship between bid-ask spreads and informed trading that is also found in the univariate setting: higher levels of informed trading cause market makers to widen the quoted bid-ask spreads, as suggested by Glosten and Milgrom (1985).

The association of firm-level order-imbalance with informed trading measured by $IA_1$ is similar to the one found at a market-level (not shown): if the horizon over which informed trading is measured, is getting longer, one observes the coefficient of order-imbalance to become larger.[31] The positive coefficient of tick size suggests that prices are more efficient with a smaller the minimum price increment as there seems to be less informed trading the smaller the size of the minimum tick. This is consistent with the improvement in price efficiency brought by the switch to decimal pricing on the NYSE that Chordia, Roll, and Subrahmanyam (2005) report. The results do not support the hypothesis by Glosten and Milgrom (1985) related to volume as higher volume seems to be associated with a higher level of informed trading, consistent with the empirical findings of Bhushan (1989a) and Llorente, et al. (2001).

### 5.2.3 Firm-specific Structural Characteristics

Variation explained by stock-level trading characteristics is taken out of informed trading, which is subsequently regressed on firm-specific structural characteristics. Results, reported in Table 5, confirm most of the findings from the univariate analysis. Opposite to the hypotheses listed in Table 1 our expectation is the negative association between informed trading and R&D expenses, which may be explained by the relatively strong size effect in this variable (see Table 3).

As anticipated, however, the level of informed trading declines with firm size. Looking at the association of ownership structure and informed trading, it turns out that informed trading increases in the relative size of the ownership stake of insiders. Outsiders, however, seem to improve the information environment, which is in line with the findings of Bushee and Noe (2000). The negative sign of the coefficient of the book-to-market ratio confirms the hypothesis that growth firms expose investors to a higher level of informed trading. This is consistent with Matsumoto (2002), who find that growth firms bias the information communicated to the public more than others. Another finding by Matsumoto (2002) is that loss-making firms are less likely

[31] Untabulated results show the regression coefficient of stock-level order-imbalance to be insignificant if $IA_1$ is estimated over two days and significantly negative if $IA_1$ is estimated over three to five days. This reflects findings by Chordia, Roll, and Subrahmanyam (2002) that information contained in order-imbalance is impounded into prices within few days and suggests that the link between the information environment in the stock market and stock prices quickly weakens when extending the measurement horizon beyond one day.

to manage the perceptions of their investors. This implies a negative relationship between the profit margin and the level of informed trading, which the results also support. Finally, the availability of options is related to a lower level of informed trading. This suggests that some informed traders prefer trading in the options market. Consequently, uninformed investors in the market for common stocks with options are less exposed to informed trades than investors in comparable stocks without options.

In sum, most of the hypothesized relationships between informed trading and the set of explanatory variables are supported by the data, although some of the results suggest a need for more precise proxies to better capture the relationship of, say, asset tangibility and the information environment. To improve the analysis of potential causes for time-series variation in informed trading, this analysis is repeated using higher frequency data. $PIN_1$ is in yearly frequency. Thus, only $IA_1$ can be used. As both measures of informed trading show very similar associations with the set of explanatory variables, it appears save to assume that $IA_1$ captures the information environment sufficiently well to allow drawing general conclusions about informed trading from the following analysis.

## 5.3 Daily Analysis of Informed Trading

This section reports the results of re-estimating the regressions specified in Section 3.2 on daily frequency. To account for the strong size effect in the data, the regressions are estimated individually by firm size deciles. Results are presented in Table 6.[32]

The sign of the regression coefficients capturing the relationship between informed trading and market-wide commonality on a daily level is consistent with what is found on the yearly level (see Panel A of Table 6). The magnitude of the relationship between informed trading and

[32] As before, only results where $IA_1$ is estimated over 60 minutes are displayed. Results for $IA_1$ estimated over different intraday or daily horizons are hardly different.

market-level bid-ask spread and volume decreases monotonically in firm size. This implies that changes in the information environment due to innovations in the market environment are relatively more important for small firms than for large ones. As small firms are generally less liquid, this implies that informed traders are fairly sensitive to stock market liquidity, which illustrates their strategic trading behavior postulated by Kyle (1985). Similarly, the size of the regression coefficients of market-level order-imbalance and stock-level order-imbalance (see Panel B of Table 6) decreases in firm size. This suggests that smaller, less actively traded stocks need more time than the rest of the stock market to fully incorporate new information that triggered market-wide changes in order-imbalance. This makes sense given that stock prices of small firms are typically less efficient in incorporating new information (Hasbrouck (1991)). The positive association of informed trading and market-level volatility shows that informed traders are more advantaged in periods of uncertainty.

Another result in Table 6 is that the explanatory power of the regression monotonically increases in firm size if $IA_1$ is estimated over short time horizons. Institutional investors tend to concentrate their investment in large firms rather than spread their investment across many smaller ones (Chordia and Subrahmanyam (2004); Farrar and Girton (1981)). Also, these investors tend to be more sophisticated than, say, retail investors (Lee, Shleifer, and Thaler (1991)). It therefore seems that market-wide common variation in informed trading that likely originates from the Kim and Verrecchia (1994, 1997)-type informed analyst investors is the result of institutional traders who exploit their superior skill in analyzing public information by trading stocks of large firms. This is also consistent with Lakonishok and Lee (2001), who find that most of insider trading, i.e., trading that arguably co-varies little across stocks as it typically relates to private information about firm-level idiosyncratic issues, is strongest and carries most information about small stocks.

The sign of the coefficient of tick size is negative (see Panel B of Table 6), which is opposite to the findings using yearly data. This may be explained by the difference in frequency, as more efficient prices imply a quicker price-response to new information, which the daily frequency accounts for. The lower level of informed trading that results from more efficient prices is what information asymmetry measured on a yearly level captures. The smallest ten percent of the sample, however, do not show this relationship, however, which – given that smaller firms tend to be less price efficient and have a lower level of liquidity (Hasbrouck (1991)) – implies that small tick size may not be beneficial to the smallest, least liquid firms. Kairys, Kruza, and Kumpins (2000) find that the quality of the price-discovery process deteriorates for the least liquid stocks in their sample as the trading process becomes more efficient via a switch from a daily batch auction to continuous pricing. The coefficient of tick size of the smallest ten percent of the sample could pick up a similar deterioration in the quality of the price discovery process, as the amount of information impounded into prices per time unit declines the lower the tick size. This relationship between price efficiency, informed trading, and liquidity could provide interesting insights to the discussion about the optimal minimum tick size.

As some of the firm-specific structural variables are observed on a yearly frequency only, the associations between daily informed trading and firm-specific structural characteristics are considered to be of indicative nature only. Nevertheless, some variables show interesting associations. In particular, outside ownership, which on a yearly frequency is associated with a lower level of informed trading, now shows the opposite relationship. This could be interpreted as showing that large outside investors may exploit their preferential access to insider information, as Maug (2002), suggest. Given the findings on a yearly frequency, however, this seems unlikely as outside investors generally seem to improve investor's communication (Bushee and Noe (2000)). Rather, these results are likely to show that large outside investors, which tend to be professional investment firms, are more skilled in analyzing public information than the average market participant. This is consistent with the empirical findings by Yan and Zhang (2007), who find that, what they call "short-term institutions", exploit their information advantage, whereas "long-term" institutions do not. The results using yearly $IA_1$ may therefore pick up the relationship between informed trading and long-term investors and the relationship found for daily $IA_1$ likely captures the relationship between informed trading and short-term investors.

More generally, the results so far suggest a market-wide commonality component in informed trading. Stock-level characteristics that capture co-movements in informed trading over time and across stocks turn out to be related to the trading environment and structural characteristics. The following section discusses the characteristics of the common component in informed trading.

## 5.4 Analysis of Explained and Unexplained Informed Trading

To further explore the characteristics of the information environment, it is calculated how much of the explained variation of informed trading is captured by market-wide commonality, stock-level trading characteristics, or firm-level structural characteristics and whether the explanatory power is dependent on general features of the firm. Institutions are more likely to invest in large firms (Chordia and Subrahmanyam (2004); Farrar and Girton (1981)), which – following the argument of the sophisticated, information generating investor – implies that informed trades by institutions should increase in relative importance with firm size. Therefore, market-wide commonality is likely to be relatively more important to explain informed trading for larger firms.

Lower trading costs associated with more liquid stocks should enable Madrigal (1996)-type traders who exploit their understanding of the trading environment, to publicly trade on their private information that is based on innovations of the stock-level trading environment. Therefore, the importance of the trading environment in informed traders is expected to increase with liquidity. However, as more liquid stocks are also more price-efficient, there should also be less scope for informed traders to exploit private information. Thus, trading characteristics could also be less important to informed traders of larger firms.

Investors of small firms have less public information at their disposal (Bhushan (1989a)) and, as small firms are likely to have less diverse operations than large corporations (Agmon and Lessard (1997)), these investors are exposed to a higher level of risk. The economic prospects of smaller firms may thus be more strongly related to individual asset characteristics. Private information about the value of these assets should be particularly valuable. Therefore, firm-level structural characteristics should be more important to explain informed trading when looking at small firms.

Table 7, which shows the results of regressing $IA_1$ on all variables plus an intercept on the stock-level, confirms most of these priors. About forty percent of explained variation of informed trading is associated with the market-wide component, slightly more than fifty percent of explained variation is attributed to the stock-level trading environment, and slightly less than ten percent of explained variation is attributed to firm-specific structural characteristics. The relatively high level of importance of market-wide commonality and stock-level trading characteristics seems to be the result of the use of daily $IA_1$ data for this particular exercise, where time-series variation likely dominates cross-sectional variation. Thus, more than half of the explained variation of informed trading is related to factors other than asset characteristics and a large proportion of informed trading is explained by market-wide commonality. This is an interesting finding as empirical studies typically consider informed trading to be a firm-specific phenomenon (see, for instance, the references in footnote 4).[33]

Moving from larger to smaller firms, unexplained variation of informed trading becomes more important (see Table 7). This reflects the particular importance of insider trades found for small firms (Lakonishok and Lee (2001)) that are unlikely to be functionally related to the environment of investors, as argued previously. Ranking the relative explanatory power by other characteristics (not reported), such as the book-to-market ratio, the effective spread, or bid-ask spread exhibits very little variation in relative explanatory power across ranks. This implies that there is no systematic, functional relationship of these characteristics with the relative explanatory power of market-wide commonality, the stock-level trading environment, and firm-specific structural characteristics.

The analysis so far has documented common variation in the level of informed trading next to unexplained residual variation, consistent with extant theory, intuition and, where available, prior empirical evidence. It still remains to be shown how much investors care about all this, i.e., whether and if so how informed trading, its explained part, and its unexplained component affect returns. This is looked at in the following section.

## 5.5 Informed Trading and Stock Returns

Easley, et al. (2002) show that exposure to informed trading is priced in the cross-section. This section investigate whether the presence of common variation in informed trading reduces the ability to diversify away exposure to informed trades and thereby increases the price-relevance of

[33] Alternatively, the regression is estimated across the entire data panel by firm-size decile including all explanatory variables. Looking at the same horizon $IA_1$ is estimated over than in Table 6, market-wide commonality factors capture between 14 (for the smallest decile) and 65 (for the largest decile) percent of the variance explained by the explanatory variables. Stock-level trading characteristics account for between 55 (for the smallest decile) and 19 (for the largest decile) percent of the explained variance of $IA_1$. As firm-level structural characteristics account for the remainder of the explained variance, we find numbers that are fairly close to the one reported in Table 6.

information risk. Alternatively, a strong relationship between the level of informed trading and variables that capture the information environment of investors could reduce information risk as price fluctuations could be more readily interpreted as being either random or a reflection of changes in fundamental values. This could be valuable to investors, as these seem to use prices as additional piece of information to derive the fundamental value (Admati and Pfleiderer (2000); Bhushan (1989a)). The former hypothesis implies that the total level of informed trading should have a stronger relationship with stock returns than unexplained level of informed trading, *RAIN*. The latter hypothesis implies the reverse. To improve the strength of the inferences drawn from this exercise, a variable that captures *Explained Informed Trading*, *EXIT*, is defined. This variable, calculated as the difference between total informed trading and *RAIN*, reflects informed trading explained by market-wide commonality, stock-level trading characteristics, and firm-specific structural characteristics.

We test the relationship between returns and informed trading in the cross-section more formally by using a Fama and MacBeth (1973)-type set-up. The time-period covered by the data includes the period after 2000, which is characterized by a prolonged period of negative market returns. According to Potential and Sundaram (1995), negative excess market returns make the estimated loading on the beta-coefficient insignificant unless negative and positive market returns are separately considered in the cross-sectional regression. For this purpose, up-market (down-market) market betas are defined as being equal to the stock-level beta if the realized market return in excess of the risk-free rate is positive (negative) and zero otherwise. To implement the regression, fifty portfolios are formed based on the average level of informed trading of the previous month. Stock returns in excess of the risk-free rate are average within each portfolio and regressed on the portfolio averages of beta, of the logarithm of the book-to-market ratio, and of the logarithm of firm size. The book-to-market ratio, firm size, and beta are measured as of the previous year. The relevant value of informed trading is the level during the previous calendar year if $PIN_1$ is used and the average level during the previous month if $IA_1$ is used as measure of informed trading. This results in the following regression that is estimated every month:

$$R_p^e = \kappa_0 + \kappa_1 BETA_{up,p} + \kappa_2 BETA_{down,p} + \kappa_3 Size_p + \kappa_4 BTM_{up} + \kappa_5 InfoTrade_{k,p} + \zeta_t, \qquad (9)$$

where $R_p^e$ is the average portfolio return in excess of the risk-free rate of portfolio $p$. The variables $BETA_{up,p}$ and $BETA_{down,p}$ are up-market and down-market betas of portfolio $p$, $Size_p$ is the logarithm of firm size of portfolio $p$, and $BTM_p$ is the logarithm of the book-to-market ratio of portfolio $p$, respectively. $InfoTrade_{k,p}$ refers to the average level of informed trading-measure $k$ of portfolio $p$. The monthly effective spread used as an alternative to $InfoTrade_k$ to verify whether the relationship between informed trading and stock returns does not pick up liquidity effects that, as Brennan and Subrahmanyam (1996) find, may also be priced. Results are presented in Table 8 using Litzenberger and Ramaswamy (1979)-adjusted $t$-statistics.[34]

Intercepts are sometimes insignificant, showing that the model specified in regression (9) captures cross-sectional returns fairly well. Firm size is hardly significant though the negative regression coefficients reveal the size effect in stock returns, discussed in Fama and French (1992). The lack of statistical significance reflects the results of Kim (1995), who also finds firm size to be of little importance in explaining cross-sectional returns if data that cover time more recent time periods are used. The inclusion of the up-market and down-market betas turns out to be useful as the loadings are significant and signed as Potential and Sundaram (1995) suggest, while the use of one single beta variable (unreported) leads to insignificant coefficients. Consistent with Easley, et al. (2002), the total unadjusted level of informed trading is priced in the cross-section. Most importantly, however, the association between excess returns and informa-

[34] The Litzenberger and Ramaswamy (1979)-adjustment is essentially a weighed least square estimate using the parameter precision as weight. Results are robust to the choice of the horizon over which $IA_1$ is estimated.

tion risk gets economically and statistically stronger as one moves from total information asymmetry to *RAIN*. This relationship weakens, however, the longer the horizon $IA_1$ is estimated over. Loadings on *EXIT* are hardly statistically significant. The result is robust to whether *RAIN* is estimated in three steps as outlined in equations (2) to (7) or whether it is estimated in one go (referred to as *RAIN2* in Table 8). This indicates that prices are an important source of information to investors as suggested by Admati and Pfleiderer (2000) and Bhushan (1989a). Therefore, investors seem to require a higher return the more they are exposed to an environment that does not allow discriminating between changes in fundamental value and random price fluctuations.

To ascertain these results and to give more consideration to the relationship between returns and informed trading over time, a second asset-pricing test is conducted. Cochrane (2001) discusses the trade-off between the empirical robustness of ordinarily least-square and the statistical efficiency of generalized least-square (henceforth referred to as GLS) in asset pricing. Testing the return relevance of the components of the information environment in a GLS framework could improve the validity of the results, as it is the most efficient way to adjust Fama-MacBeth regressions for biases and the errors-in-variables problem Ferson and Harvey (1999). Following the approach by Brennan and Subrahmanyam (1996), a random intercept GLS regression model is used. In particular, individual stock returns are sorted into five portfolios based on firm size of the previous year. These portfolios are then individually sorted into five groups based on the average level of informed trading. Alternatively, a three-way sort on size, book-to-market, and informed trading is used. Equally-weighted average portfolio excess returns are regressed on market returns in excess of the risk-free rate, the Fama and French (1993)-factors SMB and HML, and the rank of each portfolio regarding its level of informed trading. The GLS regressions are run on the full data panel, whereby every portfolio has individual factor loadings for the market factor, the HML factor, and the SMB factor. The coefficient on the level of informed trading, however, is estimated across all portfolios. This procedure results in a large number of explanatory variables, which therefore necessitates many time-series observations for a statistically valid estimation. The low sampling frequency of $PIN_1$ results in relatively few time-series observations of this variable. Therefore, this particular asset-pricing test is done using monthly observations of $IA_1$ only.

The positive and significant slope coefficients of *RAIN* presented in Table 9 show that returns are higher if exposure to informed trading that is unrelated to the environment of investors increases. In addition, the loadings on the ranking variable mostly increase in economic terms as one successively removes market-wide commonality, stock-level trading characteristics, and firm-specific structural characteristics of informed trading. *EXIT* is hardly related to returns, confirming the associations of the cross-sectional set-up presented in Table 8.

These results show that exposure to informed trading is priced. In addition, it appears to be exposure to residual asymmetric information, *RAIN*, rather than total information asymmetry that captures the largest part of priced information risk. Thus, the reduced predictability of the information environment seems to be important to investors that they even attach a risk premium to the inability to use prices as additional source of information.

## 5.6 Robustness Checks

Several robustness checks are carried out to test the validity of the results. Regressions (2) to (6) on daily $IA_1$ data are estimated within firm size quintiles and including day-of-the-week dummies. As informed trading is expressed by order-imbalance in many theories about the information flow in financial markets (see, e.g., Easley, et al. (2002); Kyle (1985); Lyons (2001)), several alternative specifications of order-imbalance are included, too, such as lagged order-imbalance, dummies that account for the sign of order imbalance, and dummies that classify each daily stock-level order-imbalance observation into one decile group based on the size relative to the rest of the cross-section. To investigate the stability of the coefficient estimates, regressions (2) to (6) are re-estimated within sub-periods of the sample by slicing the time-series into two, three, and four partitions.

Using firm size quintiles instead of deciles results in the same sign of the associations between the set of explanatory variables and informed trading. While the inclusion of day-of-the-week dummies does not affect the sign or significance of the other explanatory variables, informed trading appears to be higher at the beginning of the week. As Chordia, et al. (2002) find that the highest level of trading activity is at the beginning of the week, which therefore seems to be related to informed trading. Order-imbalance lagged by one day is positively related to informed trading if $IA_1$ is estimated over horizons shorter than one day. The coefficient of lagged order-imbalance is negative for larger firms and positive but insignificant for smaller ones if $IA_1$ is estimated over one day. This suggests that the information contained in order-imbalance is impounded into prices within about one day. Running the regression in sub-periods shows estimates largely consistent with Table 6. However, regression coefficients of market-level volatility are mostly negative between 1995 and 1997 as are the coefficients of stock-level bid-ask spreads for the regressions estimated between 2003 and 2005. Both variables exhibit a consistent time-trend during these sub-periods, which the sub-period regression estimates seem to pick-up. Results of estimating regressions (2) to (6) in a different order is shown in Panel A of Table 10, which reveals that the resulting regression coefficients similar different from what is reported in Table 6.[35]

[35] When reversing the order of the regression, we intend to minimize the potential for omitted variable bias that results from the particular order the regressions are estimated. In particular, we select an alternative order where the variables that may potentially explain more of the observed variation in $IA_1$ than the market-wide trading characteristics are come first, followed by the second most important set of variables and the remaining variables. Step-wise regressions reveal that stock-level trading characteristics could potentially be more important than market-wide trading characteristics in explaining observed variation in $IA_1$ (see also footnote 20). This alternative estimation (i.e., using stock-level trading characteristics first, then market-wide trading characteristics, followed by firm-level struc-

Finally, putting all explanatory variables into the same regression[36] shows that the coefficients are hardly affected by the order of inclusion into the regression (unexpected changes in bid-ask spreads and some of the firm-level structural characteristics sometimes have different coefficients but the results appear to be fairly robust).

The average information content of order-imbalance seems also to differ by trade direction. Buying pressure tends to be positively related to informed trading, while selling pressure is related to a lower level of informed trading. According to Aboody, Hughes, and Liu (2005) and Lakonishok and Lee (2001), this asymmetry could be due to buys involving a deliberate choice, potentially based on private information. Sales, however, additionally contain liquidity trades by employees that intend to divest stocks that are part of their compensation package (Aboody, et al. (2005); Lakonishok and Lee (2001)). Another potential reason for sales containing less information than buys are mutual funds that liquidate stock-positions resulting from mergers or stock-financed acquisitions by firms they are invested in (Harris (2003, p. 332)) and thereby became larger than their mandated maximum position. Consistent with theory (see, e.g., Kyle (1985)), the size of the daily order-imbalance position has a positive relationship with the level of informed trading.

Further, it is tested whether $IA_1$ captures the level of informed trading rather than simply daily returns. The main feature that distinguishes $IA_1$ from simple returns is that it measures returns conditional on a trade. A test of the validity of associating $IA_1$ with informed trading – and not just returns – should therefore ascertain whether accounting for the timing of a transaction, as $IA_1$ does, is useful in capturing the information environment. Thus, a return measure is con-

---

tural characteristics) shows very similar results except for the a few stock-level variables (some of the coefficients of order-imbalance are now insignificant and negative and some of the coefficients of bid-ask spreads now contain some negative observations for larger firms), market-wide variables (market-wide volatility now contains some negative observations), and the firm-level structural characteristics.

[36] Results not shown to preserve brevity but available on request.

structed that accounts for the empirical characteristics of informed trading with its information environment but that excludes the information contained in the timing of a transaction. In particular, a hypothetical trader observes the daily realization of the explanatory variables. If this realization deviates from the historical (or cross-sectional) average, the trader takes a position of unit size that has the same direction than the deviation of the explanatory variable from its mean value and keeps this position for one day. Alternatively, the amplitude of the deviation from the historical or cross-sectional average to be reflected in the size of the position is considered.[37] The resulting stock-level return-series is subsequently regressed on the explanatory variables using the same methodology and sectioning into size groups. Regardless of how this alterative return series is specified, the explanatory power of the regressions is very low (the R-square is between 0.004 and 0.022). The associations with the explanatory variables are, although consistent in sign across size deciles, often insignificant.

Another concern related to the use of $IA_1$ is how strongly $IA_1$ is influenced by return momentum. Therefore a time-series of daily 15-minutes unconditional quote returns is constructed. This return variable has the same set-up than $IA_1$ except for explicitly not accounting for the information content of the timing of a trade. This variable should therefore mechanically pick up return momentum. If the relationship between this quote-return variable and the set of explanatory variables is weaker than what is found for $IA_1$, one could conclude that $IA_1$ contains more information than simply daily returns. Results shown in Panel B of Table 10 reveal that a regres-

[37] The first method multiplies the daily stock return from day $t$ to the following day by the sum of the daily stock-level indicator variables observed at the end of the previous day. Each indicator variable is defined as being equal to one if the de-meaned explanatory variable the indicator variable is referring to is larger than zero. It is equal to minus one if the respective de-meaned explanatory variable is negative. Market-wide commonality variables are de-meaned across time, stock-level trading characteristics are de-meaned across time for each stock individually, and firm-specific structural variables are de-meaned every day across all stocks in the sample. The second method adds up all explanatory variables, which are standardized to a mean of zero and a standard deviation of one, for every stock and every day. The sum of the standardized explanatory variables of day $t–1$ is then multiplied by the daily stock return from day $t$ to the following day.

sion of the quote-return series on the set of explanatory variables has a very low explanatory power (the average R-square across all size deciles is 3.2 percent). In addition, signs and the significance of some explanatory variables are not always consistent across size decile. These tests show that $IA_1$ cannot be replicated either using environmental variables observed in the recent past or mechanically by supplying returns over fixed intervals. Therefore, $IA_1$ seems to reflect something that is not yet contained in prices: private information that is transmitted to the market by means of informed trades. We further test whether *RAIN* really constitutes a residual as is claimed in the analysis. For this purpose, we do a principal component analysis on the complete set of stock-level *RAIN* series.[38] If *RAIN* truly represents a residual, one would expect not to find a principal component that captures much of the variance, e.g., in the vicinity of 30 percent, across the individual stocks. As $PIN_1$ has only eleven time-series observations, its time-series variation does not have meaningful variation in the time-series to qualify for this test, which is why this test is only conducted using $IA_1$. Results shown in Panel C of Table 10 confirm this prior. The first principal component does not capture more than 7 percent of the observed variance and the first three principal components together capture less than 12 percent of the observed variance. Thus, *RAIN* seems to be mainly driven by firm-specific events and that the filtering approach presented in this chapter has been successful in removing commonality in variation across stocks.

# 6 Summary and Conclusions

This paper analyses informed trading in about 1,500 reasonably liquid stocks traded on the NYSE between January 1995 and December 2005. One major contribution of this paper is to comprehensively investigate, for the first time, the relationship between informed trading and market-

[38] To ensure a sufficiently large time-series, only stocks that have observations at least 97.5 percent of time during the sample period have been used.

wide commonality, stock-level trading characteristics, and firm-specific structural characteristics. Results show strong evidence of market-wide commonality. Market-level volatility, trading volume, bid-ask spreads, and order-imbalance are all significantly related to the level of informed trading. These results, consistent with Kim and Verrecchia (1994, 1997), indicate a common, market-wide component in informed trading related to skilled information analysts who generate private information from public data. In addition, strong evidence of individual firm-specific factors in informed trading are reported. More than half of observed informed trading can be attributed to commonality and the stock-level trading environment. Most of the relationships between informed trading and our set of explanatory variables are consistent with economic theory, intuition, and, where available, prior empirical evidence obtained using other measures and other empirical approaches.

Given the systematic influence of market-wide commonality and structural and trading characteristics of a firm, the level of informed trading unrelated to the environment an individual investor faces is calculated and referred to as *Residual Asymmetric Information*, or *RAIN*, associated with that firm at that time. We use this new measure to make another important contribution. Easley, et al. (2002) show that exposure to informed trading is priced in the cross-section of stock returns. The systematic relationship of informed trading with investors' environment could increase priced information risk as it lowers the ability to diversify away exposure to informed trades. Alternatively, a higher level of or whether a higher predictability of the level of informed trading helps uninformed investors in acquiring price-relevant information, thus lowering the required return of a stock. Results show that *RAIN* is priced in the asset's required rate of return and its effect on returns is stronger and more robust than that of unfiltered informed trading. These results indicate that the inability to predict the information environment, and therefore the inability to distinguish information-induced price innovations from random price fluctuations, contributes significantly to the information risk premium.

To address the inherent difficulty of attaching a numerical value to the information environment by using two different measures of information asymmetry, which are fairly different in methodology, underlying theoretical assumptions, and sampling frequency. These empirical measures may be imprecise, however, which may be one of the major limitations of this study. In addition, eleven years of data constitute a fairly short time horizon from an asset pricing point of view, which may weaken the conclusions drawn from the respective part of this paper. The robustness checks shown in Section 5.6, however, supply some confidence that the results are sufficiently robust to provide useful insights into the nature and the pricing relevance of the information environment.

# References


Aboody, David, John Hughes, and Jing Liu, 2005, *Earnings quality, insider trading, and costs of capital*, Journal of Accounting Research 43, 651-673.

Admati, Anat R., 1984, *A Noisy Rational Expectations Equilibrium for Multi-asset Securities Markets*, Econometrica 53, 629-658.

Admati, Anat R., and Paul Pfleiderer, 1988, *A theory of intraday patterns: Volume and price variability*, Review of Financial Studies 1, 3-40.

Admati, Anat R., and Paul Pfleiderer, 2000, *Forcing firms to talk: Financial disclosure regulation and externalities*, Review of Financial Studies 13, 479-519.

Agmon, Tamir, and Donald R. Lessard, 1997, *Investor Recognition of Corporate International Diversification*, Journal of Finance 32, 1049-1055.

Akerlof, George A., 1970, *The market for "lemons": Quality uncertainty and market mechanism*, The Quarterly Journal of Economics 84, 488-500.

Bessembinder, Hendrik, 2003a, *Issues in assessing trade execution costs*, Journal of Financial Markets 6, 233-257.

Bessembinder, Hendrik, 2003b, *Trade execution costs and market quality after decimalization*, Journal of Financial & Quantitative Analysis 38, 747-777.

Bessembinder, Hendrik, and Herbert Kaufman, M., 1997, *A cross-exchange comparison of execution costs and information flow for NYSE-listed stocks*, Journal of Financial Economics 46, 293-319.

Bhattacharya, Nilabhra, Ervin L. Black, Theodore E. Christensen, and Richard D. Mergenthaler, 2004, *Empirical evidence on recent trends in pro forma reporting*, Accounting Horizons 18, 27-43.

Bhushan, Ravi, 1989a, *Collection of Information about Publicly Traded Firms*, Journal of Accounting and Economics 11, 183-206.

Bhushan, Ravi, 1989b, *Firm Characteristics and Analyst Following*, Journal of Accounting and Economics 11, 255-274.

Blair, Bevan J., Ser-Huang Poon, and Stephen J. Taylor, 2001, *Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns*, Journal of Econometrics 105, 5-26.

Boni, Leslie, and J. Chris Leach, 2002, *Supply Contraction and Trading Protocol: An Examination of Recent Changes in the U.S. Treasury Market*, Journal of Money, Credit & Banking 34, 740-762.

Brealey, Richard, Hayne E. Leland, and David H. Pyle, 1977*, Informational asymmetries, financial structure, and financial intermediation*, Journal of Finance 32, 371-387.

Breen, William J., Laurie Simon Hodrick, and Robert A. Korajczyk, 2002, *Predicting Equity Liquidity*, Management Science 48, 470-483.

Brennan, Michael J., Tarun Chordia, and Avanidhar Subrahmanyam, 1998, *Alternative factor specifications, security characteristics, and the cross-section of expected stock returns*, Journal of Financial Economics 49, 345-373.

Brennan, Michael J., and Avanidhar Subrahmanyam, 1996, *Market microstructure and asset pricing: On the compensation for illiquidity in stock returns*, Journal of Financial Economics 41, 441-464.

Bushee, Brian J., and Christopher F. Noe, 2000, *Corporate disclosure practices, institutional investors, and stock return volatility*, Journal of Accounting Research 38, 171-202.

Cao, Charles, Hyuk Choe, and Frank Hatheway, 1997, *Does the Specialist Matter? Differential Execution Costs and Intersecurity Subsidization on the New York Stock Exchange*, Journal of Finance 52, 1615-1640.

Chalmers, John M. R. and Gregory B. Kadlec, 1998, *An Empirical Examination of the Amortized Spread*, Journal of Financial Economics 48, 159-188.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2000, *Commonality in liquidity*, Journal of Financial Economics 56, 3-28.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2002, *Order imbalance, liquidity, and market returns*, Journal of Financial Economics 65, 111-130.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2005, *Evidence on the speed of convergence to market efficiency*, Journal of Financial Economics 76, 271-292.

Chordia, Tarun, and Avanidhar Subrahmanyam, 2004, *Order imbalance and individual stock returns: Theory and evidence*, Journal of Financial Economics 72, 485-518.

Cochrane, John H., 2001. Asset Pricing (Princeton University Press, Princeton, NJ).

Cotter, Julie, and Scott Richardson, 2002, *Reliability of asset revaluations: The impact of appraiser independence*, Review of Accounting Studies 7, 435-457.

Dennis, Patrick J., and James P. Weston, 2001, *Who's informed? An analysis of stock ownership and informed trading*, Working Paper, Rice University.

Dlugosz, Jennifer, Rüdiger Fahlenbrach, Paul Gompers, and Andrew Metrick, 2006, *Large blocks of stock: Prevalence, size, and measurement*, Journal of Corporate Finance 12, 594-618.

Easley, David, Robert F. Engle, Maureen O'Hara, and Liuren Wu, 2001, *Time varying arrival rates of informed and uninformed trades*, Working Paper, Cornell University.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2002, *Is information risk a determinant of asset returns?*, Journal of Finance 57, 2185-2221.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2004, *Factoring information into returns*, Working Paper, Cornell University.

Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, 1997a, *The information content of the trading process*, Journal of Empirical Finance 4, 159-186.

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

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

Easley, David, and Maureen O'Hara, 1992, *Time process of security price adjustment*, Journal of Finance 47, 577-605.

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

Easley, David, Maureen O'Hara, and P. S. Srinivas, 1998, *Option volume and stock prices: Evidence on where informed traders trade*, Journal of Finance 53, 431-465.

Ellis, Katrina, Roni Michaely, and Maureen O'Hara, 2000, *The Accuracy of Trade Classification Rules: Evidence from Nasdaq*, Journal of Financial and Quantitative Analysis 35, 529-551.

Evans, Martin D. D., and Richard K. Lyons, 2004, *A New Micro Model of Exchange Rate Dynamics*, Working Paper, NBER.

Fama, Eugene F., and Kenneth F. French, 1995, *Size and book-to-market factors in earnings and returns*, Journal of Finance 50, 131-155.

Fama, Eugene F., and Kenneth R. French, 1992, The Cross-Section of Expected Stock Returns, Journal of Finance 47, 427-465.

Fama, Eugene F., and Kenneth R. French, 1993, *Common risk factors in the returns on stocks and bonds*, Journal of Financial Economics 33, 3-56.

Fama, Eugene F., and Kenneth R. French, 1997, *Industry costs of equity*, Journal of Financial Economics 43, 153-193.

Fama, Eugene F., and James D. MacBeth, 1973, *Risk, Return, and Equilibrium: Empirical Tests*, Journal of Political Economy 81, 607-636.

Farrar, Donald E., and Lance Girton, 1981, *Institutional Investors and Concentration of Financial Power: Berle and Means Revisited*, Journal of Finance 36, 369-381.

Ferson, Wayne E., 2003, *Tests of multifactor pricing models, volatility bounds and portfolio performance*, Working Paper, NBER.

Ferson, Wayne E., and Campbell R. Harvey, 1999, *Conditioning variables and the cross section of stocks returns*, Journal of Finance 54, 1325-1360.

Fleming, Michael J., and Eli M. Remolona, 1999, *Price formation and liquidity in the U.S. Treasury market: The response to public information*, Journal of Finance 54, 1901-1945.

French, Kenneth R., and Richard Roll, 1986, *Stock return variances: The arrival of information and the reaction of traders*, Journal of Financial Economics 17, 5-26.

Froot, Kenneth A., and Tarun Ramadorai, 2005, *Currency Returns, Intrinsic Value, and Institutional-Investor Flows*, Journal of Finance 60, 1535-1566.

George, Thomas J., Gautam Kaul, and M. Nimalendran, 1991, *Estimation of the bid-ask spread and its components: A new approach*, Review of Financial Studies 4, 623-656.

Gilson, Stuart C., Paul M. Healy, Christopher F. Noe, and Krishna G. Palepu, 2001, *Analyst Specialization and Conglomerate Stock Breakups*, Journal of Accounting Research 39, 565-582.

Glosten, Lawrence R., 1987, *Components of the bid-ask spread and the statistical properties of transaction prices*, Journal of Finance 42, 1293-1307.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, *Bid, ask and transaction prices in a specialist market with heterogeneously informed traders*, Journal of Financial Economics 14, 71-100.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980, *On the impossibility of informationally efficient markets*, American Economic Review 70, 393-408.

Habib, Michael A., Johnsen, D. Bruce, and Narayan Y. Naik, 1997, *Spinoffs and information*, Journal of Financial Intermediation *6*, 153-176.

Hansch, Oliver, Narayan Y. Naik, and S. Viswanathan, 1999, *Preferencing, internalization, best execution, and dealer profits*, Journal of Finance 54, 1799-1828.

Harris, Lawrence, 1990, *Statistical properties of the Roll serial covariance bid/ask spread estimator*, Journal of Finance 45, 579-590.

Harris, Lawrence, 2003. *Trading and Exchanges* (Oxford University Press, New York, NY).

Hasbrouck, Joel, 1991, *Measuring the information content of stock trades*, Journal of Finance 46, 179-207.

Hasbrouck, Joel, and Duane J. Seppi, 2001, *Common factors in prices, order flows, and liquidity*, Journal of Financial Economics 59, 383-411.

Hasbrouck, Joel, and George Sofianos, 1993, *The trades of market makers: An empirical analysis of NYSE specialists*, Journal of Finance 48, 1565-1593.

Ho, Thomas, and Hans R. Stoll, 1981, *Optimal dealer pricing under transactions and return uncertainty*, Journal of Financial Economics 9, 47-73.

Huang, Roger D., and Hans R. Stoll, 1996, *Dealer versus auction markets: A paired comparison of execution costs on NASDAQ and the NYSE*, Journal of Financial Economics 41, 313-357.

Huang, Roger D., and Hans R. Stoll, 1997, *The components of the bid-ask spread: a general approach*, Review of Financial Studies 10, 995-1034.

Hughes, John, Jing Liu, and Jun Liu, 2005, *Private information, diversification, and asset pricing*, Working Paper, University of California in Los Angeles.

Kairys, Jr., Joseph P., Raimonds Kruza, and Ritvars Kumpins, 2000, *Winners and losers from the introduction of continuous variable price trading: Evidence from the Riga Stock Ex-change*, Journal of Banking & Finance 24, 603-624.

Khandani, Amir E., and Andrew W. Lo, *What happened to the Quants in August 2007?*, Working Paper, MIT Sloan School of Management.

Kim, Dongcheol, 1995, *The errors in the variables problem in the cross-section of expected stock returns*, Journal of Finance 50, 1605-1634.

Kim, Oliver, and Robert E. Verrecchia, 1994, *Market liquidity and volume around earnings announcements*, Journal of Accounting and Economics 17, 41-67.

Kim, Oliver, and Robert E. Verrecchia, 1997, *Pre-announcement and Event-period Private Information*, Journal of Accounting and Economics 24, 395-419.

Kothari, S. P. , Ted E. Laguerre, and Andrew J. Leone, 2002, *Capitalization versus expensing: Evidence on the uncertainty of future earnings from capital expenditures versus R&D outlays*, Review of Accounting Studies 7, 355-382.

Kyle, Albert S., 1985, *Continuous auctions and insider trading*, Econometrica 53, 1315-1335.

Lakonishok, Josef, and Inmoo Lee, 2001, *Are insider trades informative?*, Review of Financial Studies 14, 79-111.

Lease, Ronald C., Ronald W. Masulis, and John R. Page, 1991, *An Investigation of Market Microstructure Impacts on Event Study Returns*, Journal of Finance 46, 1523-1536.

Lee, Charles M. C., and Mark J. Ready, 1991, *Inferring Trade Direction from Intraday Data*, Journal of Finance 46, 733-746.

Lee, Charles M.C., Andrei Shleifer, and Richard H. Thaler, 1991, *Investor Sentiment and the Closed-End Fund Puzzle*, Journal of Finance 46, 75-109.

Litzenberger, Robert H., and Krishna Ramaswamy, 1979, *The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence*, Journal of Financial Economics 7, 163-195.

Llorente, Guillermo, Roni Michaely, Gideon Saar, and Jiang Wang, 2001, *Dynamic Volume-Return Relation of Individual Stocks*, Review of Financial Studies 15, 1005-10047.

Lyons, Richard K., 2001. *The Microstructure Approach to Exchange Rates* (The MIT Press, Cambridge, MA).

Madrigal, Vicente, 1996, *Non-Fundamental Speculation*, Journal of Finance 51, 553-578.

Malkiel, Burton G., and Atanu Saha, 2005, *Hedge Funds: Risk and Return*, Financial Analyst Journal 61, 80-88.

Matsumoto, Dawn A., 2002, *Management's incentives to avoid negative earnings surprises*, Accounting Review 77, 483-514.

Maug, Ernst, 2002, *Insider trading legislation and corporate governance*, European Economic Review 46, 1569-1597.

Merton, Robert C., 1987, *A simple model of capital market equilibrium with incomplete information*, Journal of Finance 42, 483-510.

Naik, Narayan Y., and Pradeep K. Yadav, 2003a, *Do dealer firms manage inventory on a stock-by-stock or a portfolio basis?*, Journal of Financial Economics 69, 325-353.

Naik, Narayan Y., and Pradeep K. Yadav, 2003b, *Trading costs of public investors with obligatory and voluntary market-making: Evidence from market reforms*, Working Paper, London Business School.

Neal, Robert, and Simon M. Wheatley, 1998, *Adverse selection and bid-ask spreads: Evidence from closed-end funds*, Journal of Financial Markets 1, 121-149.

NYSE, 2006, NYSE Regulation: *Dealings and Settlements (Rules 45-299C)*, (New York Stock Exchange: http://rules.nyse.com/nysetools/Exchangeviewer.asp?SelectedNode=chp_1_2&manual=/nyse/nyse_rules/nyse-rules/).

Odders-White, Elizabeth R., and Mark J. Ready, 2006, *Credit ratings and stock liquidity*, Review of Financial Studies 19, 119-157.

Panayides, Marios, 2004, *The Specialist's Participation in Quoted Prices and the NYSE's Price Continuity Rule*, Working Paper, Yale School of Management.

Potential, Glenn M., and Sridhar Sundaram, 1995, *The conditional relation between beta and returns*, Journal of Financial and Quantitative Analysis 30, 101-116.

Ready, Mark J., 1999, *The Specialist's Discretion: Stopped Orders and Price Improvement*, Review of Financial Studies 12, 1015-1112.

Roll, Richard, 1984, *A simple implicit measure of the effective bid-ask spread in an efficient market*, Journal of Finance 39, 1127-1139.

Subrahmanyam, Avanidhar, 1991, *A theory of trading in stock index futures*, Review of Financial Studies 4, 17-51.

Sydsaeter, Knut, Arne Strøm, and Peter Berck, 2000. *Economists' Mathematical Manual* (Springer, Heidelberg).

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

Yan, Xuemin Sterling, and Zhe Zhang, 2007, *Institutional Investors and Equity Returns: Are Short-term Institutions Better Informed?*, Review of Financial Studies *forthcoming*.

## Table 1 – Hypothesized Relationship of Variables with Informed Trading

This table shows the hypothesized associations between the explanatory variables and informed trading. The columns Measure and Variable Name indicate which measure is thought to interact with the level of informed trading and the name of the variable used to quantify this effect. The columns Sign and Hypothesis list the expected sign of the association of the respective variable with informed trading and the hypothesis the expectation is based on, respectively. The symbols "+" and "–" in the column Sign indicate that increases in the particular variable are expected to be associated with a higher or lower level of informed trading, respectively.

| | Measure | Variable Name | Hypothesis | Sign |
|---|---|---|---|---|
| Market-wide commonality | Bid-ask spread | MBA | Bid-ask spreads are higher the higher the level of informed trades (Glosten and Milgrom (1985)). | + |
| | Trading volume | MVOL | Higher trading volume means trades are relatively less informed (Glosten and Milgrom (1985)). Higher volume results in relatively noisy prices. This increases the benefit of private information (Bhushan (1989a)) that could also lead to more informed trades as informed traders hide among liquidity traders (Admati and Pfleiderer (1988)). | +/– |
| | Volatility | MVLA | Private information drives return volatility (French and Roll (1986)). | + |
| | Order imbalance | MOIB | Private information is expressed via order-imbalance (Kyle (1985)). | + |
| Stock-level trading characteristics | Bid-ask spread | BA | Bid-ask spreads are higher the higher the level of informed trades (Kyle (1985)). | + |
| | Unexpected changes in bid-ask spread | UEDSpread | Unexpectedly high bid-ask spreads imply an unanticipated drop in liquidity. As informed traders are strategic, the lower the liquidity, the lower the level of informed trades (Kyle (1985)). Alternatively, bid-ask spreads are unexpectedly high, so is informed trading as market makers increase spreads to protect themselves from informed trades (Glosten and Milgrom (1985)). | +/– |
| | Trading volume | VOL | A higher trading volume means that trades contain less private information (Glosten and Milgrom (1985)). Higher volume results in relatively noisy prices, which increases the benefit of private information (Bhushan (1989a)), however, that could also lead to more informed trades as informed traders hide among liquidity traders (Admati and Pfleiderer (1988)). | +/– |
| | Volatility | VLA | Private information drives return volatility (French and Roll (1986)). | + |
| | Tick size | TIC | Prices converge quicker to fundamental values the lower the minimum tick size (Chordia, et al. (2005)). The amount of information revealed per time-interval should therefore increase the smaller the tick size. Alternatively, as a smaller minimum tick size makes prices more efficient (Chordia, et al. (2005)), one could also see a decline in the level of informed trading with tick size. | +/– |
| | Order imbalance | OIB | Private information is expressed via order-imbalance (Kyle (1985)). | + |
| Firm-specific structural characteristics | Public exposure | Size | Public investors in large firms have more sources of information that get more frequently updated than investors of smaller firms (Bhushan (1989a)). Small firms suffer more from insider trading (Lakonishok and Lee (2001)). | – |
| | Active management of investor's perception | Profit<br>BTM | Small firms with growth options (i.e., a lower book-to-market ratio, or BTM) bias investor communication more, and loss-making firms less (Matsumoto (2002)). If public information is biased, informed investors are advantaged. Therefore, profitability (Profit) and growth options are positively associated with informed trading. The growth firms that manage the perception of their investors are growth companies with a high BTM (Bhattacharya, Black, Christensen, and Mergenthaler (2004)). | +<br>+/– |
| | Asset tangibility | R&D<br>Capex | Intangibles are difficult to value for outsiders (Cotter and Richardson (2002)) and their economic benefit is uncertain (Kothari, Laguerre, and Leone (2002)). Therefore, higher asset tangibility, as measured by the ratio of R&D-expenses to sales and capital expenses over sales, or Capex, should be associated with a lower level of informed trading. | +<br>– |
| | Ownership structure | Insider<br>Outsider | Corporate insiders (Aboody, et al. (2005); Lakonishok and Lee (2001)) and outsiders (Maug (2002)) exploit their information advantage. Outside ownership by institutional investors may also be associated with better investor communication (Bushee and Noe (2000)). | +<br>+/– |
| | Alternative trading instrument | Options | Informed traders also use options (Easley, et al. (1998)). This could lead to a lower level of informed trading as some informed traders exploit their information via the options market. | – |

# Table 2 – Variable Definitions

This table lists the names of the variables used in this paper in column *Variable Name* and the definition used to construct the respective variable in the column *Definition*.

| Variable Name | Definition |
|---|---|
| $IA_1$ | The daily trade size-weighted average of the difference between the quote mid-point right before a transaction and the quote mid-point 15 minutes, 30 minutes, 60 minutes, or one day later scaled by the first quote mid-point. |
| $IA_2$ | This variable is defined as $IA_1$ less market-wide commonality. |
| $IA_3$ | This variable is defined as $IA_2$ less stock-level trading characteristics. |
| $IA_{RAIN}$ | This variable is defined as $IA_3$ less firm-specific structural characteristics. |
| $IA_{EXIT}$ | This variable is defined as the difference between $IA_1$ and $IA_{RAIN}$. |
| $PIN_1$ | The probability of information-based trades provided on a yearly frequency on Soeren Hvidkjaer's website. |
| $PIN_2$ | This variable is defined as $PIN_1$ purged from market-wide commonality. |
| $PIN_3$ | This variable is defined as $PIN_2$ purged from stock-level trading characteristics. |
| $PIN_{RAIN}$ | This variable is defined as $PIN_3$ less firms-level structural characteristics. |
| $PIN_{EXIT}$ | This variable is defined as the difference between $PIN_1$ and $PIN_{RAIN}$. |
| *Market-level* | The value-weighted daily average of the stock-level variables (except for volatility). |
| *Volatility* | Market-level volatility is measured by the VIX index and by squared daily returns for individual stocks. |
| *Bid-ask spread* | The time-weighted daily average of the individual BBO percentage spread. |
| *Order-imbalance* | The absolute daily dollar-imbalance scaled by dollar trading volume orthogonalized to dollar trading volume. |
| *Unexpected changes in bid-ask spread* | The residual of a market-model fitted to stock-level quoted percentage bid-ask spreads. |
| *Unexpected changes in trading volume* | The residuals of a market-model fitted to stock-level dollar volume |
| *Tick size* | The inverse of the stock price. |
| *Firm size* | The stock-market capitalization. If used in return regressions, the natural logarithm of the last observation in the previous calendar year is used. |
| *Book value-to-market value* | The firm-level book-value divided by firm size. If used in return regressions, the natural logarithm of this measure is used. |
| *Operating profit margin* | The ratio of operating profit to sales. |
| *Research and develop-ment-to-sales* | The ratio of research and development expenses to sales. |
| *Capital expenditures-to-sales* | The ratio of capital expenditures to sales. |
| *Block-ownership - insider* | The total fraction of large stakes in common stock held by corporate insiders. |
| *Block-ownership - out-sider* | The total fraction of large stakes in common stock held by corporate outsiders. |
| *Options availability* | This variable is equal to one if the respective firm has exchange traded options on its common stock and zero other-wise. |
| *Excess returns on the market* | The monthly returns on the market in excess of the risk-free rate. |
| *SMB* | The monthly returns on the Fama and French (1995)-factor portfolio SMB. |
| *HML* | The monthly returns on the Fama and French (1995)-factor portfolio HML. |
| *Excess stock returns* | The individual monthly stock returns in excess of the risk-free rate. |
| *Beta* | The stock-level beta coefficient calculated as in Fama and French (1992). |
| *Effective spread* | The trade size-weighted daily average of the difference between the transaction price and the mid-point of the con-current primary market BBO quotes. |
| $BETA_{up}$ | This variable is an up-market beta based on the definition of Pattengill and Sundaram (1995) and is equal to the estimated beta if excess market returns are positive and zero otherwise. |
| $BETA_{down}$ | This variable is a down-market beta based on the definition of Pattengill and Sundaram (1995) and is equal to the estimated beta if excess market returns are negative and zero otherwise. |

# Table 3 – Correlation of Informed Trading with Explanatory Variables

This table reports the correlation coefficients of the variables that that are used to capture the information environment (see Table 2 for variable definitions).

| | Informed Trading | | | | | Market-wide Commonality | | | | Stock-level Trading Characteristics | | | | | | Firm-specific Structural Characteristics | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $IA_1$ | | | | | Bid-ask | Trading | Vola- | Order | Bid-ask | UED | Trad. | Vola- | Order | Tick | Firm | Profit | Book/ | | | Ownership | |
| | 15 min | 30 min | 60 min | 1 day | $PIN_1$ | Spread | Volume | tility | Imbal. | Spread | Spread | Vol. | tility | Imbal. | Size | Size | Margin | Market | R&D | Capex | Ins. | Outs. |
| $IA_1$ over 30 minutes | 0.92 | | | | | | | | | | | | | | | | | | | | | |
| $IA_1$ over 60 minutes | 0.83 | 0.90 | | | | | | | | | | | | | | | | | | | | |
| $IA_1$ over 1 day | 0.33 | 0.36 | 0.40 | | | | | | | | | | | | | | | | | | | |
| $PIN_1$ | 0.27 | 0.26 | 0.24 | 0.11 | | | | | | | | | | | | | | | | | | |
| Market bid-ask spread | 0.23 | 0.22 | 0.20 | 0.10 | 0.10 | | | | | | | | | | | | | | | | | |
| Market trading volume | -0.08 | -0.08 | -0.07 | -0.04 | -0.15 | -0.63 | | | | | | | | | | | | | | | | |
| Market volatility | 0.07 | 0.06 | 0.06 | 0.02 | -0.06 | -0.06 | 0.30 | | | | | | | | | | | | | | | |
| Market order-imbalance | 0.01 | 0.02 | 0.02 | 0.01 | 0.04 | 0.14 | -0.14 | -0.19 | | | | | | | | | | | | | | |
| Stock bid-ask spread | 0.56 | 0.56 | 0.51 | 0.23 | 0.38 | 0.34 | -0.16 | 0.04 | 0.03 | | | | | | | | | | | | | |
| Unexp. changes in B/A | 0.24 | 0.24 | 0.21 | 0.08 | -0.03 | 0.01 | 0.08 | 0.07 | -0.03 | 0.53 | | | | | | | | | | | | |
| Stock trading volume | 0.08 | 0.07 | 0.07 | 0.03 | 0.01 | 0.00 | -0.02 | -0.02 | 0.00 | 0.00 | -0.02 | | | | | | | | | | | |
| Stock volatility | 0.07 | 0.07 | 0.08 | 0.06 | 0.01 | 0.00 | 0.02 | 0.03 | 0.00 | 0.05 | 0.00 | 0.00 | | | | | | | | | | |
| Stock order-imbalance | -0.01 | 0.00 | 0.00 | 0.00 | 0.04 | -0.03 | 0.02 | -0.01 | 0.02 | -0.01 | 0.00 | 0.01 | 0.00 | | | | | | | | | |
| Stock tick size | 0.42 | 0.41 | 0.37 | 0.16 | 0.26 | 0.00 | 0.04 | 0.07 | -0.03 | 0.74 | 0.37 | 0.00 | 0.06 | -0.01 | | | | | | | | |
| Firm size | -0.12 | -0.11 | -0.11 | -0.05 | -0.27 | -0.07 | 0.07 | 0.01 | -0.01 | -0.16 | -0.01 | 0.00 | -0.01 | -0.01 | -0.13 | | | | | | | |
| Profit margin | -0.12 | -0.12 | -0.11 | -0.05 | -0.15 | -0.05 | 0.05 | 0.02 | -0.01 | -0.19 | -0.02 | 0.00 | -0.01 | 0.00 | -0.19 | 0.11 | | | | | | |
| Book-to-market | 0.14 | 0.15 | 0.13 | 0.06 | 0.21 | -0.08 | 0.08 | 0.05 | -0.04 | 0.24 | 0.11 | 0.00 | 0.01 | -0.01 | 0.27 | -0.17 | -0.08 | | | | | |
| Research and development | -0.04 | -0.04 | -0.04 | -0.01 | -0.12 | -0.04 | 0.03 | 0.00 | -0.01 | -0.07 | 0.00 | 0.00 | 0.01 | 0.00 | -0.01 | 0.14 | -0.05 | -0.15 | | | | |
| Capital expenditures | 0.02 | 0.02 | 0.01 | 0.01 | 0.02 | 0.02 | 0.01 | 0.04 | -0.01 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | -0.02 | 0.40 | 0.06 | -0.01 | | | |
| Ownership - insider | 0.01 | 0.01 | 0.01 | 0.00 | 0.06 | 0.02 | -0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | -0.01 | -0.03 | -0.06 | -0.01 | -0.05 | -0.04 | | |
| Ownership - outsider | -0.03 | -0.03 | -0.03 | -0.01 | -0.02 | 0.01 | 0.00 | 0.02 | 0.00 | -0.06 | 0.01 | 0.00 | 0.00 | -0.01 | -0.03 | -0.10 | -0.07 | 0.01 | 0.04 | 0.00 | 0.01 | |
| Option availability | -0.16 | -0.17 | -0.15 | -0.07 | -0.46 | 0.00 | 0.01 | 0.03 | -0.01 | -0.27 | 0.00 | -0.01 | 0.00 | 0.00 | -0.16 | 0.14 | 0.07 | -0.20 | 0.14 | 0.04 | 0.05 | 0.17 |

# Table 4 – Summary Statistics

This table reports the summary statistics of the firm-level means of the variables listed in column *Name* calculated over the period between January 1995 and December 2005. The column *Observations* shows the total number of daily observations and the columns *Mean*, *Q1*, *Median*, *Q3*, and *IQ Range* report the mean, the first quartile, the median, the third quartile, and the difference between *Q3* and *Q1*. The unit of measurement is given by the symbols *bp*, *%*, and *$*, which refer to basis points, percentages, and dollar values respectively (see Table 2 for variable definitions).

| | Measure | Observations | Mean | Q1 | Median | Q3 | IQ Range |
|---|---|---|---|---|---|---|---|
| Informed trading | $IA_1$ over 15 minutes (bp) | 3,829,045 | 24.01 | 10.74 | 17.93 | 29.99 | 19.25 |
| | $IA_1$ over 30 minutes (bp) | 3,828,167 | 24.12 | 10.79 | 18.22 | 30.09 | 19.30 |
| | $IA_1$ over 60 minutes (bp) | 3,827,750 | 24.18 | 10.79 | 18.16 | 30.14 | 19.35 |
| | $IA_1$ over 1 day (bp) | 3,543,002 | 28.05 | 12.15 | 20.72 | 35.11 | 22.95 |
| | $PIN_1$ (%) | 22,164 | 16.39 | 12.34 | 15.58 | 19.67 | 7.34 |
| Market-wide commonality | Market-level bid-ask spread (bp) | 2,770 | 22.28 | 8.74 | 25.02 | 30.43 | 21.69 |
| | Market-level trading volume (millions of $) | 2,770 | 186.16 | 97.80 | 200.54 | 250.17 | 152.37 |
| | Market-level volatility | 2,770 | 20.70 | 15.40 | 20.11 | 24.44 | 9.04 |
| | Market-level order-imbalance (%) | 2,770 | 26.81 | 22.64 | 25.70 | 29.47 | 6.83 |
| Stock-level trading characteristics | Company-level bid-ask spread (bp) | 3,829,045 | 84.50 | 31.02 | 53.80 | 99.99 | 68.97 |
| | Unexpected changes in bid-ask spread (bp) | 3,829,045 | 0.17 | -0.07 | 0.00 | 0.02 | 0.10 |
| | Stock-level trading volume (bp) | 3,829,045 | 85.27 | -2.38 | 0.00 | 47.88 | 50.26 |
| | Stock-level volatility (bp) | 3,829,045 | 8.75 | 3.67 | 5.84 | 10.06 | 6.38 |
| | Stock-level order-imbalance (%) | 3,829,045 | 53.40 | 19.93 | 24.86 | 33.30 | 13.36 |
| | Stock-level tick size (%) | 3,829,045 | 7.10 | 3.06 | 4.62 | 7.83 | 4.77 |
| Firm-specific structural characteristics | Firm size (10 millions of $) | 3,829,045 | 401.19 | 37.17 | 92.87 | 258.48 | 221.31 |
| | Operating profit margin (%) | 3,829,045 | 17.00 | 7.19 | 13.72 | 24.07 | 16.88 |
| | Book value-to-market value (%) | 3,829,045 | 54.35 | 26.03 | 47.83 | 74.99 | 48.96 |
| | Research and development-to-sales (%) | 3,829,045 | 1.15 | 0.00 | 0.00 | 0.55 | 0.55 |
| | Capital expenditures-to-sales (%) | 3,829,045 | 8.32 | 1.22 | 3.72 | 7.64 | 6.42 |
| | Block-ownership - insiders (%) | 3,829,045 | 3.90 | 0.00 | 0.00 | 1.22 | 1.22 |
| | Block-ownership - outsiders (%) | 3,829,045 | 9.69 | 0.00 | 0.52 | 17.11 | 17.11 |
| | Option availability (%) | 3,829,045 | 62.48 | 0.00 | 100.00 | 100.00 | 100.00 |
| Data used in return regressions | Excess return on the market (%) | 132 | 0.70 | -2.25 | 1.53 | 3.77 | 6.01 |
| | SMB (%) | 132 | 0.22 | -2.47 | -0.14 | 2.63 | 5.10 |
| | HML (%) | 132 | 0.43 | -1.65 | 0.45 | 2.20 | 3.85 |
| | Excess stock returns (%) | 174,611 | 1.07 | 0.47 | 1.09 | 1.84 | 1.37 |
| | Beta | 3,629,560 | 0.97 | 0.76 | 0.91 | 1.15 | 0.39 |
| | Effective spread (%) | 3,829,045 | 0.68 | 0.26 | 0.44 | 0.79 | 0.53 |

# Table 5 – Decomposition of Yearly Informed Trading

This table shows the results of regressing yearly informed trading, $InfoTrade_1$, on a firm-specific intercept and a set of explanatory variables. The results of estimating the following regression at once is shown in Panel A:

$$\begin{aligned} InfoTrade_{1,i,t} = {} & \beta_{i,0} + \beta_1 MBA + \beta_2 MVOL + \beta_3 MVLA + \beta_4 MOIB \\ & + \gamma_1 VLA_{i,t} + \gamma_2 BA_{i,t} + \gamma_3 OIB_{i,t} + \gamma_4 TIC_{i,t} + \gamma_5 UEDSpread_{i,t} + \gamma_6 VOL_{i,t} \\ & + \delta_1 Insider_{i,t} + \delta_2 Outsider_{i,t} + \delta_3 Capex_{i,t} + \delta_4 R\&D_{i,t} + \delta_5 BTM_{i,t} + \delta_6 Profit_{i,t} + \delta_7 Options_{i,t} + \delta_8 Size_{i,t} + \varsigma_{i,t}, \end{aligned}$$

whereas Panels B and C show the same regression estimated in three steps according to:

$$\begin{aligned} InfoTrade_{1,i,t} &= \beta_{i,0} + \beta_1 MBA + \beta_2 MVOL + \beta_3 MVLA + \beta_4 MOIB + \varepsilon_{i,t}, \\ InfoTrade_{2,i,t} &= \gamma_{i,0} + \gamma_1 VLA_{i,t} + \gamma_2 BA_{i,t} + \gamma_3 OIB_{i,t} + \gamma_4 TIC_{i,t} + \gamma_5 UEDSpread_{i,t} + \gamma_6 VOL_{i,t} + \eta_{i,t}, \\ InfoTrade_{3,i,t} &= \delta_{i,0} + \delta_1 Insider_{i,t} + \delta_2 Outsider_{i,t} + \delta_3 Capex_{i,t} + \delta_4 R\&D_{i,t} + \delta_5 BTM_{i,t} + \delta_6 Profit_{i,t} + \delta_7 Options_{i,t} + \delta_8 Size_{i,t} + \xi_{i,t}, \end{aligned}$$

where $InfoTrade_1$ is alternatively represented by $PIN_1$ or yearly averages of daily $IA_1$. The hypotheses associated with the individual explanatory variables are shown in Table 1 and the variable definitions are shown in Table 2. The regression coefficients associated with $PIN_1$ are in percentages and the coefficients associated with $IA_1$ are in basis points. *P*-values of a two-sided *t*-test of the coefficient being equal to zero are below (Panel A) or to the right (Panels B and C) of the respective coefficients in parentheses. The $R^2$, based on the derivation by Nagelkerke (1991), is in percentages. The variables in Panels B and C are presented in decreasing order of their contribution to the $R^2$.

## Panel A – One-step Decomposition of the Information Environment

| | Variables Capturing Features of the Information Environment of Investors | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Information | Market-wide Commonality | | | | Stock-level Trading Characteristics | | | | | | Firm-specific Structural Characteristics | | | | | | | | |
| Asymmetry | BA | | | | BA | | | | Tick | UED | Ownership Struc. | | Asset Tangibility | | Window-dressing | | Options as | Firm- | |
| Captured by | Spread | VOL | Vola | OIB | Spread | VOL | Vola | OIB | Size | Spread | Insider | Outsider | Capex | R&D | BTM | Profit | Alternative | size | $R^2$ |
| $PIN_1$ | 1.01 | -1.35 | -0.10 | 0.33 | 0.80 | 2.10 | 50.42 | 0.63 | 0.33 | -0.51 | 0.24 | -0.73 | -0.11 | -0.89 | -0.24 | -0.03 | -1.50 | -1.37 | 46.5 |
| | (0.00) | (0.00) | (0.85) | (0.00) | (0.00) | (0.00) | (0.10) | (0.00) | (0.00) | (0.00) | (0.02) | (0.00) | (0.14) | (0.00) | (0.00) | (0.68) | (0.00) | (0.00) | |
| $IA_{1,15\ min}$ | 9.62 | 2.13 | 1.81 | 2.49 | 2.52 | 25.70 | 22.00 | 1.30 | 0.71 | 1.16 | -0.17 | -0.63 | 0.67 | -1.03 | -0.24 | -0.35 | -1.42 | -10.55 | 44.3 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.68) | (0.56) | (0.00) | (0.55) | (0.00) | (0.00) | (0.05) | (0.12) | (0.05) | (0.00) | (0.00) | |
| $IA_{1,30\ min}$ | 8.74 | 1.91 | 1.89 | 2.56 | 2.54 | 25.17 | 24.57 | 0.83 | 1.43 | 1.12 | -0.24 | -0.62 | 0.65 | -1.14 | 0.18 | -0.30 | -1.31 | -10.62 | 44.1 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.90) | (0.24) | (0.00) | (0.42) | (0.00) | (0.00) | (0.03) | (0.23) | (0.10) | (0.00) | (0.00) | |
| $IA_{1,60\ min}$ | 9.32 | 1.83 | 1.90 | 2.60 | 2.46 | 21.97 | 23.68 | 0.92 | 1.32 | 1.07 | -0.13 | -0.54 | 0.61 | -1.31 | 0.15 | -0.44 | -1.22 | -10.34 | 48.6 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.12) | (0.25) | (0.00) | (0.63) | (0.01) | (0.00) | (0.01) | (0.31) | (0.01) | (0.00) | (0.00) | |
| $IA_{1,1\ day}$ | 10.60 | 1.27 | 1.81 | 3.64 | 5.86 | 0.99 | 36.30 | 1.08 | 0.32 | 1.11 | 0.44 | 0.10 | 0.57 | -1.88 | -0.22 | -1.55 | -1.25 | -10.94 | 31.9 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.64) | (0.00) | (0.05) | (0.13) | (0.00) | (0.39) | (0.78) | (0.11) | (0.04) | (0.42) | (0.00) | (0.00) | (0.00) | |

*(continued)*

## Table 5 – Decomposition of Yearly Informed Trading (continued)

### Panel B – Three-step Decomposition of the Information Environment Captured by $PIN_1$

| | Variable | Coeff | p-value | $R^2$ |
|---|---|---|---|---|
| Market-wide commonality | Volume | -1.31 | (0.00) | 39.7 |
| | Order-imbalance | 0.54 | (0.00) | |
| | Volatility | 0.14 | (0.00) | |
| | Bid-ask spread | 0.37 | (0.00) | |
| Stock-level trading characteristics | Order-imbalance | 0.34 | (0.00) | 43.5 |
| | UED Spread | -0.28 | (0.00) | |
| | Volume | 1.87 | (0.00) | |
| | Bid-ask spread | 0.13 | (0.00) | |
| | Volatility | 54.00 | (0.08) | |
| | Tick size | 0.04 | (0.35) | |
| Firm-specific structural characteristics | Firm size | -1.15 | (0.00) | 34.8 |
| | Ownership - corporate outsiders | -0.77 | (0.00) | |
| | Research and development | -0.92 | (0.00) | |
| | Option availability | -2.47 | (0.00) | |
| | Book-to-market | -0.29 | (0.00) | |
| | Ownership - corporate insiders | 0.27 | (0.01) | |
| | Capital expenditures | -0.14 | (0.07) | |
| | Profit margin | -0.05 | (0.42) | |

### Panel C – Three-step Decomposition of the Information Environment Captured by $IA_1$

| Variable | | Coeff | p-value | $R^2$ |
|---|---|---|---|---|
| Market-wide commonality | Bid-ask spread | 9.01 | (0.00) | 16.0 |
| | Volatility | 2.31 | (0.00) | |
| | Volume | 3.97 | (0.00) | |
| | Order-imbalance | 2.49 | (0.00) | |
| Stock-level trading characteristics | Volatility | 29.47 | (0.00) | 30.7 |
| | UED Spread | 2.98 | (0.00) | |
| | Bid-ask spread | 1.83 | (0.00) | |
| | Volume | 26.50 | (0.00) | |
| | Tick size | 2.61 | (0.00) | |
| | Order-imbalance | 1.24 | (0.00) | |
| Firm-specific structural characteristics | Firm size | -9.27 | (0.00) | 21.2 |
| | Profit margin | -0.64 | (0.00) | |
| | Option availability | -1.15 | (0.00) | |
| | Capital expenditures | 0.50 | (0.02) | |
| | Research and development | -1.29 | (0.01) | |
| | Ownership - corporate outsiders | -0.21 | (0.34) | |
| | Book-to-market | -0.14 | (0.35) | |
| | Ownership - corporate insiders | -0.06 | (0.63) | |

## Table 6 – Time-series and Cross-sectional Associations in Daily Informed Trading

This table shows the results of regressing daily values of informed trading measured by $IA_1$ on a firm-specific intercept and a set of explanatory variables by firm size decile according to:

$$IA_{1,i,t} = \beta_{i,0} + \beta_1 MBA + \beta_2 MVOL + \beta_3 MVLA + \beta_4 MOIB + \varepsilon_{i,t},$$
$$IA_{2,i,t} = \gamma_{i,0} + \gamma_1 VLA_{i,t} + \gamma_2 BA_{i,t} + \gamma_3 OIB_{i,t} + \gamma_4 TIC_{i,t} + \gamma_5 UEDSpread_{i,t} + \gamma_6 VOL_{i,t} + \eta_{i,t},$$
$$IA_{3,i,t} = \delta_{i,0} + \delta_1 Insider_{i,t} + \delta_2 Outsider_{i,t} + \delta_3 Capex_{i,t} + \delta_4 R\,\&\,D_{i,t} + \delta_5 BTM_{i,t} + \delta_6 Profit_{i,t} + \delta_7 Options_{i,t} + \delta_8 Size_{i,t} + \xi_{i,t},$$

where $IA_1$ is estimated over 60 minutes. The hypotheses associated with the individual explanatory variables are shown in Table 1 and the variable definitions are shown in Table 2. Panel A presents the results of regressing $IA_1$ on variables that capture market-wide commonality. Panel B presents the results of regressing $IA_2$ on stock-level trading characteristics, and Panel C shows the results of regressing $IA_3$ on firm-specific structural characteristics. In the table below, the column *Decile* shows the size group with *Decile 10* referring to the largest size group. Regression coefficients shown in column *Coeff* are in basis points, p-values are to the right of the respective coefficients in parentheses and the $R^2$ is in percentages (based on the derivation by Nagelkerke (1991)). The variables are presented in decreasing order of their contribution to the average explanatory power going from the left to the right.

### Panel A – Market-wide Commonality in Daily Informed Trading

| | Bid-ask spread | | Volume | | Volatility | | Order-imbalance | | |
|---|---|---|---|---|---|---|---|---|---|
| Decile | Coeff | p-value | Coeff | p-value | Coeff | p-value | Coeff | p-value | $R^2$ |
| 1 | 14.08 | (0.00) | 5.99 | (0.00) | 4.13 | (0.00) | 0.16 | (0.32) | 37.2 |
| 2 | 11.26 | (0.00) | 2.81 | (0.00) | 2.75 | (0.00) | 0.43 | (0.00) | 40.7 |
| 3 | 10.38 | (0.00) | 2.64 | (0.00) | 1.65 | (0.00) | 0.36 | (0.00) | 41.3 |
| 4 | 8.76 | (0.00) | 1.73 | (0.00) | 1.27 | (0.00) | 0.69 | (0.00) | 40.6 |
| 5 | 7.80 | (0.00) | 1.46 | (0.00) | 1.13 | (0.00) | 0.92 | (0.00) | 40.1 |
| 6 | 7.02 | (0.00) | 1.40 | (0.00) | 0.91 | (0.00) | 1.18 | (0.00) | 39.7 |
| 7 | 6.05 | (0.00) | 0.97 | (0.00) | 0.61 | (0.00) | 1.12 | (0.00) | 39.2 |
| 8 | 5.26 | (0.00) | 0.80 | (0.00) | 0.59 | (0.00) | 1.21 | (0.00) | 38.5 |
| 9 | 4.26 | (0.00) | 0.55 | (0.00) | 0.32 | (0.00) | 1.25 | (0.00) | 37.8 |
| 10 | 3.24 | (0.00) | 0.50 | (0.00) | 0.39 | (0.00) | 1.34 | (0.00) | 38.1 |

### Panel B – Informed Trading and Stock-level Trading Characteristics

| | Volatility | | UED Spread | | Bid-ask spread | | Volume | | Tick size | | Order-imbalance | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Decile | Coeff | p-value | Coeff | p-value | Coeff | p-value | Coeff | p-value | Coeff | p-value | Coeff | p-value | $R^2$ |
| 1 | 5.09 | (0.00) | 3.08 | (0.00) | 12.60 | (0.00) | 7.16 | (0.00) | 3.09 | (0.00) | 4.06 | (0.00) | 39.1 |
| 2 | 5.81 | (0.00) | 3.12 | (0.00) | 4.98 | (0.00) | 3.41 | (0.00) | -1.30 | (0.00) | 3.00 | (0.00) | 39.6 |
| 3 | 5.64 | (0.00) | 2.90 | (0.00) | 3.56 | (0.00) | 2.44 | (0.00) | -1.06 | (0.00) | 2.73 | (0.00) | 39.4 |
| 4 | 7.47 | (0.00) | 2.82 | (0.00) | 2.50 | (0.00) | 1.89 | (0.00) | -1.04 | (0.00) | 2.38 | (0.00) | 39.2 |
| 5 | 4.30 | (0.00) | 2.34 | (0.00) | 2.02 | (0.00) | 1.68 | (0.00) | -0.62 | (0.00) | 2.08 | (0.00) | 37.8 |
| 6 | 6.60 | (0.00) | 2.16 | (0.00) | 1.54 | (0.00) | 1.44 | (0.00) | -0.65 | (0.00) | 1.87 | (0.00) | 37.3 |
| 7 | 5.49 | (0.00) | 1.56 | (0.00) | 1.17 | (0.00) | 1.15 | (0.00) | -0.28 | (0.00) | 1.71 | (0.00) | 36.3 |
| 8 | 8.84 | (0.00) | 1.43 | (0.00) | 0.69 | (0.00) | 1.03 | (0.00) | -0.28 | (0.00) | 1.68 | (0.00) | 36.2 |
| 9 | 8.08 | (0.00) | 1.12 | (0.00) | 0.44 | (0.00) | 0.88 | (0.00) | -0.23 | (0.00) | 1.64 | (0.00) | 35.2 |
| 10 | 12.32 | (0.00) | 0.91 | (0.00) | 0.11 | (0.00) | 0.81 | (0.00) | -0.04 | (0.04) | 1.68 | (0.00) | 37.4 |

*(continued)*

# Table 6 – Time-series and Cross-sectional Associations in Daily Informed Trading (continued)

## Panel C – Informed Trading and Firm-specific Structural Characteristics

| | Firm Size | | Profit | | BTM | | Outsider | | Insider | | R&D | | Capex | | Options | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Decile | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | $R^2$ |
| 1 | -21.67 | (0.00) | -2.24 | (0.00) | -1.11 | (0.00) | 9.00 | (0.00) | 6.57 | (0.00) | -2.24 | (0.01) | 1.26 | (0.00) | 5.46 | (0.00) | 32.9 |
| 2 | -18.95 | (0.00) | -2.53 | (0.00) | -0.31 | (0.05) | 1.19 | (0.00) | -1.79 | (0.00) | -2.74 | (0.00) | -0.01 | (0.98) | 4.50 | (0.00) | 37.9 |
| 3 | -9.83 | (0.00) | -2.26 | (0.00) | -0.03 | (0.82) | 0.58 | (0.01) | 0.12 | (0.67) | -1.65 | (0.00) | -1.29 | (0.00) | 1.79 | (0.00) | 37.9 |
| 4 | -9.23 | (0.00) | -1.31 | (0.00) | 0.55 | (0.00) | 0.51 | (0.00) | 1.73 | (0.00) | 0.13 | (0.73) | -0.79 | (0.00) | 1.11 | (0.00) | 35.9 |
| 5 | -4.35 | (0.00) | -1.39 | (0.00) | 0.13 | (0.27) | 0.52 | (0.00) | 1.38 | (0.00) | 0.60 | (0.17) | -0.02 | (0.91) | -0.31 | (0.00) | 35.7 |
| 6 | -6.20 | (0.00) | -0.87 | (0.00) | 0.04 | (0.68) | 1.03 | (0.00) | 0.30 | (0.07) | 0.43 | (0.23) | -1.11 | (0.00) | -1.66 | (0.00) | 34.6 |
| 7 | -3.87 | (0.00) | -0.89 | (0.00) | 0.51 | (0.00) | 0.86 | (0.00) | 0.35 | (0.01) | -0.11 | (0.65) | -0.21 | (0.05) | -0.87 | (0.00) | 33.9 |
| 8 | -2.82 | (0.00) | -0.40 | (0.00) | 0.91 | (0.00) | 0.25 | (0.00) | -0.19 | (0.10) | -1.55 | (0.00) | 0.20 | (0.05) | -1.09 | (0.00) | 32.3 |
| 9 | -2.26 | (0.00) | -0.35 | (0.00) | 0.27 | (0.00) | -0.17 | (0.00) | -0.66 | (0.00) | -0.61 | (0.00) | -0.13 | (0.13) | -0.62 | (0.00) | 31.3 |
| 10 | -1.31 | (0.00) | 0.16 | (0.00) | 0.38 | (0.00) | 0.14 | (0.00) | 0.35 | (0.00) | 0.66 | (0.00) | -0.23 | (0.00) | -0.31 | (0.00) | 30.7 |

## Table 7 – Analysis of Common Variation and Idiosyncratic Informed Trading

The table below shows a decomposition of explained variation in informed trading, $\mathrm{Var}\left(\mathrm{IA}_1\right)$, into three components by estimating the following regression for every stock individually:

$$IA_{1,t} = \alpha_0 + \sum_{i=1} \beta_i Commonality_{i,t} + \sum_{j=1} \gamma_j Trading_{j,t} + \sum_{k=1} \delta_k Structural_{k,t} + \varepsilon_t .$$

Thereafter, the following statistics are calculated:

$$\mathrm{REP}_{\mathrm{Commonality}} = \frac{\mathrm{Var}\left(\sum_{i=1} \hat{\beta}_i \mathrm{Commonality}_i\right)}{\mathrm{Var}\left(\mathrm{IA}_1\right)},\ \mathrm{REP}_{\mathrm{Trading}} = \frac{\mathrm{Var}\left(\sum_{j=1} \hat{\gamma}_j \mathrm{Trading}_j\right)}{\mathrm{Var}\left(\mathrm{IA}_1\right)},\ \mathrm{REP}_{\mathrm{Structural}} = \frac{\mathrm{Var}\left(\sum_{k=1} \hat{\delta}_k \mathrm{Structural}_k\right)}{\mathrm{Var}\left(\mathrm{IA}_1\right)},$$

where *REP* denotes the relative explanatory power related to *i* market-wide *Commonality* components, to *j* stock-level *Trading* characteristics and to *k Structural* characteristics. Numbers below are averages of the firm-level $REP_{Commonality}$, $REP_{Trading}$, and $REP_{Structural}$ calculated by regressing $IA_1$ estimated over 60 minutes for each stock individually on all explanatory variables and summing up the ratios of explained to total variance by commonality, trading, and structural components. *Unexplained Informed Trading* is the average of one minus the R-square from the stock-level ordinary least-square regressions. Results are shown by firm size decile, where *Size Decile* is calculated based on the average market capitalization a firm has over the entire sample period. Numbers in the table below are in percentages.

| | Average Relative Explanatory Power of (in %) | | | Unexplained Informed Trading (in %) |
|---|---|---|---|---|
| Size Decile | Market-wide Commonality | Stock-level Trading Characteristics | Firm-specific Structural Characteristics | |
| 1 | 27.7 | 64.4 | 7.9 | 78.1 |
| 2 | 32.8 | 59.6 | 7.6 | 75.7 |
| 3 | 40.8 | 52.3 | 6.9 | 73.7 |
| 4 | 43.7 | 49.3 | 7.0 | 73.8 |
| 5 | 47.1 | 45.3 | 7.6 | 73.1 |
| 6 | 47.6 | 44.8 | 7.6 | 71.8 |
| 7 | 53.4 | 39.7 | 6.9 | 72.1 |
| 8 | 57.7 | 36.1 | 6.2 | 70.9 |
| 9 | 62.4 | 31.0 | 6.6 | 69.3 |
| 10 | 60.9 | 33.6 | 5.4 | 68.0 |

## Table 8 – Cross-sectional Returns and Components of Informed Trading

This table shows the results of a Fama and MacBeth (1973)-type cross-sectional regression of monthly portfolio returns in excess of the one-month Treasury bill rate on a set of explanatory variables and the average level of informed trading during the previous month that is estimated within each month. The regression is estimated monthly between January 1995 and December 2005:

$$R_p^e = \kappa_0 + \kappa_1 BETA_{up,p} + \kappa_2 BETA_{down,p} + \kappa_3 Size_p + \kappa_4 BTM_{up} + \kappa_5 InfoTrade_{k,p} + \zeta_t,$$

where $BETA_{up}$ and $BETA_{down}$ are up-market and down-market betas, respectively, the subscript *p* denotes portfolio *p*, and $InfoTrade_k$ refers to the measure *k* of informed trading or the average level of daily effective spread during the previous month (see Table 2 for variable definitions). Panel A shows the results when the regressions are estimated using the monthly excess returns of 50 portfolios formed on the average level of $InfoTrade_k$ during the previous year, whereby $InfoTrade_k$ is captured by $PIN_1$, $PIN_2$, $PIN_3$, $PIN_{RAIN}$, or $PIN_{EXIT}$. Panel B shows the results when the regressions are estimated using the monthly excess returns of 50 portfolios formed on the average level of $InfoTrade_k$ during the previous month, whereby $InfoTrade_k$ is captured by the effective spread, $IA_1$, $IA_2$, $IA_3$, $IA_{RAIN}$, or $IA_{EXIT}$. $IA_{RAIN2}$ or $PIN_{RAIN2}$ is calculated by regressing all explanatory variables on $IA_1$. Portfolios are based on $IA_{RAIN}$ or $PIN_{RAIN}$ if $InfoTrade_k$ is not included in the regression. $IA_1$ is estimated over 60 minutes. *T*-values are based on the Litzenberger and Ramaswamy (1979)-adjustment. The coefficients, shown in columns *Coeff*, are in percentages with *p*-values based on a two-sided *t*-test of the coefficient being equal to zero in parentheses to the right of the respective coefficients. The column $R^2$ shows the average adjusted R-square from the individual cross-sectional regressions.

### Panel A – Cross-sectional Return-Association of $PIN_1$

| $PIN_1$ | | $PIN_2$ | | $PIN_3$ | | $PIN_{RAIN}$ | | $PIN_{EXIT}$ | | $PIN_{RAIN2}$ | | Book-to-Market | | Firm Size | | Beta Down-market | | Beta Up-market | | Intercept | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | $R^2$ |
| | | | | | | | | | | | | -0.69 | (0.00) | -0.37 | (0.00) | -6.73 | (0.00) | 3.42 | (0.00) | 3.27 | (0.01) | 23.2 |
| 2.38 | (0.06) | | | | | | | | | | | -0.78 | (0.00) | -0.23 | (0.08) | -6.99 | (0.00) | 2.59 | (0.00) | 2.59 | (0.10) | 29.9 |
| | | 2.23 | (0.04) | | | | | | | | | -0.76 | (0.00) | -0.25 | (0.06) | -7.00 | (0.00) | 2.50 | (0.00) | 2.81 | (0.07) | 29.9 |
| | | | | 1.92 | (0.04) | | | | | | | -0.56 | (0.00) | -0.16 | (0.24) | -6.78 | (0.00) | 3.33 | (0.00) | 1.61 | (0.33) | 29.5 |
| | | | | | | 2.52 | (0.03) | | | | | -0.72 | (0.00) | -0.26 | (0.05) | -5.96 | (0.00) | 3.21 | (0.00) | 2.00 | (0.18) | 26.5 |
| | | | | | | | | -6.11 | (0.34) | | | -0.55 | (0.00) | -0.82 | (0.00) | -5.81 | (0.00) | 3.05 | (0.00) | 6.76 | (0.00) | 27.5 |
| | | | | | | | | | | 2.08 | (0.03) | -0.56 | (0.00) | -0.25 | (0.04) | -7.19 | (0.00) | 3.02 | (0.00) | 2.64 | (0.06) | 26.2 |

*(continued)*

# Table 8 – Cross-sectional Returns and Components of Informed Trading (continued)

## Panel B – Cross-sectional Return-Association of $IA_1$

| Effective Spread | | $IA_1$ | | $IA_2$ | | $IA_3$ | | $IA_{RAIN}$ | | $IA_{EXIT}$ | | $IA_{RAIN2}$ | | Book-to-Market | | Firm Size | | Beta: Down-market | | Beta: Up-market | | Intercept | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | $R^2$ |
| -96.72 | (0.00) | | | | | | | | | | | | | -0.04 | (0.98) | 0.20 | (0.04) | -1.26 | (0.00) | 1.98 | (0.00) | -2.02 | (0.21) | 27.4 |
| | | | | | | | | | | | | | | -0.50 | (0.16) | 0.17 | (0.01) | -1.19 | (0.00) | 1.78 | (0.00) | -2.37 | (0.20) | 22.6 |
| | | 326.68 | (0.00) | | | | | | | | | | | 0.00 | (0.77) | -0.07 | (0.08) | -1.32 | (0.00) | 1.38 | (0.00) | 2.62 | (0.28) | 29.7 |
| | | | | 329.68 | (0.00) | | | | | | | | | -0.19 | (0.70) | -0.08 | (0.05) | -1.43 | (0.00) | 1.42 | (0.00) | 2.71 | (0.12) | 32.6 |
| | | | | | | 221.50 | (0.00) | | | | | | | 0.08 | (0.47) | -0.24 | (0.08) | -1.75 | (0.00) | 1.28 | (0.00) | 5.36 | (0.02) | 33.8 |
| | | | | | | | | 369.09 | (0.00) | | | | | -0.10 | (0.99) | -0.26 | (0.01) | -1.18 | (0.00) | 1.35 | (0.00) | -3.37 | (0.03) | 34.7 |
| | | | | | | | | | | -608.68 | (0.00) | | | -0.09 | (0.59) | 0.00 | (0.88) | -1.46 | (0.00) | 2.29 | (0.00) | 0.48 | (0.71) | 36.6 |
| | | | | | | | | | | | | 62.69 | (0.00) | -2.00 | (0.00) | -0.10 | (0.02) | -1.63 | (0.00) | 1.29 | (0.02) | -1.94 | (0.59) | 43.1 |

# Table 9 – Pooled Time-series Cross-sectional Association of Informed Trading and Returns

This table shows the results of fitting a pooled time-series cross-sectional generalized least square regression of monthly portfolio excess returns on monthly factor returns and the average level of informed trading during the previous month:

$$R^e_{s,j,t} = \alpha + \sum_{s=1}^{5}\sum_{j=1}^{5}\left(\psi_{s,j} R^m_{s,j,t} + \phi_{s,j} SMB_{s,j,t} + \lambda_{s,j} HML_{s,j,t}\right) + \theta Rank_{s,j,t} + \chi_{s,j} + \upsilon_{s,j,t},$$

$$R^e_{s,p,j,t} = \rho + \sum_{s=1}^{5}\sum_{p=1}^{5}\sum_{j=1}^{5}\left(\mu_{s,p,j} R^m_{s,p,j,t} + \vartheta_{s,p,j} SMB_{s,p,j,t} + \tau_{s,p,j} HML_{s,p,j,t}\right) + \varpi Rank_{s,p,j,t} + \sigma_{s,p,i} + o_{s,p,j,t}.$$

Subscript $s$ denotes the firm size group, subscript $p$ indicates the book-to-market group, subscript $j$ the rank of the level of informed trading, and $t$ is the time-index. Five monthly firm size groups and book-to-market groups are formed based on firm size and the book-to-market ratio at the end of the previous year. Similarly, the average level of informed trading of every stock during the previous month is used to calculate informed trading quintiles. The lowest level of informed trading is assigned informed trading quintile rank 1 and the highest level of informed trading is assigned quintile rank 5. Portfolio excess returns, $R^e$, are calculated as the equally-weighted cross-sectional mean of monthly stock returns in excess of the one-month Treasury bill rate of all stocks that are in the same firm size group and have the same informed trading quintile rank (to calculate $R^e_{s,j,t}$) or that are in the same firm size group, are in the same book-to-market group, and have the same informed trading quintile rank (to calculate $R^e_{s,p,j,t}$). $R^m_{s,j,t}$, $SMB_{s,j,t}$, and $HML_{s,j,t}$ are equal to the market excess returns, and the returns on the SMB and HML factor portfolios, if the respective portfolio belongs to firm size group $s$ and to informed trading quintile rank $j$ and zero otherwise. $R^m_{s,p,j,t}$, $SMB_{s,p,j,t}$, and $HML_{s,p,j,t}$ are equal to the market excess returns, and the returns on the SMB and HML factor portfolios if the respective portfolio belongs to firm size group $s$, to book-to-market group $p$, and to informed trading quintile rank $j$ and zero otherwise. Informed trading is captured by $IA_1$ estimated over 60 minutes. The estimation uses a generalized least square regression with a random intercept model to account for potential correlation within portfolios. The column *Portfolio Sort* shows whether the portfolios are sorted on size and informed trading (*Size and IA*) alone or whether the portfolios are formed based on size, book-to-market, and informed trading (*Size, BTM, and IA*). The portfolio-specific coefficients of $R^m$, *SMB*, and *HML* are suppressed for clarity of exposition. Coefficients (*Coeff*) are in percentages and the associated *p-value* is in parentheses to the right.

| | $IA_1$ | | $IA_2$ | | $IA_3$ | | $IA_{RAIN}$ | | $IA_{EXIT}$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| Portfolio Sort | Coefficient | *p*-value | Coefficient | *p*-value | Coefficient | *p*-value | Coefficient | *p*-value | Coefficient | *p*-value |
| Size and IA | 6.5 | (1.00) | –0.4 | (0.49) | 0.8 | (0.22) | 3.0 | (1.00) | –0.4 | (1.00) |
| Size, BTM, and IA | –1.1 | (0.68) | 42.2 | (1.00) | 0.8 | (0.80) | 3.0 | (0.01) | 1.3 | (0.56) |

# Table 10 – Robustness Checks

This table reports in Panel A the results of re-estimating regressions (2) to (6) in a different order, whereby the stock-level trading environment is regressed on $IA_1$ and used to construct $IA_2$, which is regressed on the variables capturing the market-wide trading environment. $IA_3$ is then constructed using the resulting regression coefficients and is then regressed on firm-level structural characteristics. The results reported below refer to $IA_1$a horizon of 60 minutes. Panel B reports the results of regressing daily averages of intra-day quote-returns on variables that capture market-wide commonality, stock-level trading characteristics, and firm-specific structural characteristics by firm size deciles using a stock-specific intercept according to:

$$QuoteReturn_{i,t} = \pi_{i,0} + \pi_1 MBA_t + \pi_2 MVOL_t + \pi_3 \mathrm{MVLA}_t + \pi_4 MOIB_t + \pi_5 VLA_{\mathrm{i,t}} + \pi_6 BA_{\mathrm{i,t}} + \pi_7 OIB_{\mathrm{i,t}} + \pi_8 TIC_{\mathrm{i,t}} + \pi_9 UED\mathrm{Spread}_{\mathrm{i,t}} + \pi_{10} VOL_{\mathrm{i,t}} +$$
$$\pi_{11}\mathrm{Insider}_{i,t} + \pi_{12}\mathrm{Outsider}_{i,t} + \pi_{13}\mathrm{Capex}_{i,t} + \pi_{14}\mathrm{R\&D}_{i,t} + \pi_{15} BTM_{i,t} + \pi_{16}\mathrm{Profit}_{i,t} + \pi_{17}\mathrm{Options}_{i,t} + \pi_{18} Size_{i,t} + v_{i,t},$$

where $QuoteReturn_{i,t}$ denotes the daily average of intra-day returns of the BBO quote mid-point and the BBO quote mid-point 15 minutes later of stock *i* on day *t*. The hypotheses associated with the various variables are shown in Table 1 the variable definitions are shown in Table 2. In the table below, the column *Decile* shows the size group with *Decile 10* referring to the largest size group. Regression coefficients shown in column *Coeff* are in basis points, *p*-values are to the right of the respective coefficients in parentheses and the $R^2$ is in percentages (based on the derivation by Nagelkerke (1991)). Panel C shows the results of a principal component analysis on the *RAIN* time-series. Thereby, only stocks that have observations for at least 97.5% of the sample period are used. The column *RAIN Based on* shows the horizon over which $IA_1$ is estimated that is used to construct *RAIN*, column *Number of Days* shows the number of days of the sample period included in the estimation, whereas *Number of Firms* shows how many individual firms are included. *Sample Variance Explained (%)* shows the percentage of the total *RAIN*-variance explained by the first, second, and third principal component (*PC*), whereby the second row shows the cumulative variance explained.

*(continued)*

# Table 10 – Robustness Checks (Continued)

## Panel A – Association of Information Asymmetry with Explanatory Variables in Different Order

| | | | Decile | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Regression | Variable | Statistic | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| First | BA | Coeff | 9.13 | 0.35 | -1.04 | -0.39 | -0.52 | -0.26 | -0.28 | -0.33 | -0.45 | -0.47 |
| Regression | | p-val | (0.00) | (0.01) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| | VOL | Coeff | 0.07 | 0.02 | 0.05 | 0.01 | 0.03 | 0.01 | 0.01 | 0.01 | 0.02 | 0.03 |
| | | p-val | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| | VOLA | Coeff | 1.18 | 1.20 | 1.55 | 5.56 | 8.57 | 9.31 | 12.71 | 13.99 | 17.06 | 20.81 |
| | | p-val | (0.03) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.01) | (0.00) | (0.00) | (0.00) |
| | OIB | Coeff | 0.07 | -0.16 | -0.05 | -0.13 | 0.04 | -0.01 | 0.01 | -0.04 | 0.00 | 0.06 |
| | | p-val | (0.61) | (0.02) | (0.33) | (0.01) | (0.31) | (0.77) | (0.84) | (0.16) | (0.99) | (0.00) |
| | TIC | Coeff | 1.21 | -1.36 | -0.23 | -0.25 | -0.17 | -0.07 | -0.10 | -0.20 | -0.32 | -0.21 |
| | | p-val | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.13) | (0.01) | (0.00) | (0.00) | (0.00) |
| | UED | Coeff | -5.01 | -0.71 | 0.29 | 0.13 | 0.23 | 0.13 | 0.22 | 0.22 | 0.13 | 0.27 |
| | Spread | p-val | (0.00) | (0.00) | (0.00) | (0.04) | (0.00) | (0.01) | (0.00) | (0.00) | (0.00) | (0.00) |
| | | $R^2$ | 36.43 | 38.84 | 38.84 | 38.10 | 37.45 | 36.74 | 35.94 | 34.59 | 33.46 | 32.71 |
| Second | MBA | Coeff | 3.09 | 0.08 | 0.34 | 0.17 | 0.13 | 0.12 | 0.03 | 0.09 | 0.18 | 0.40 |
| Regression | | p-val | (0.00) | (0.00) | (0.00) | (0.02) | (0.03) | (0.02) | (0.00) | (0.02) | (0.00) | (0.00) |
| | MVOL | Coeff | 2.11 | 1.05 | 1.03 | 0.18 | 0.03 | 0.07 | 0.10 | 0.24 | 0.55 | 0.70 |
| | | p-val | (0.00) | (0.00) | (0.00) | (0.01) | (0.00) | (0.00) | (0.03) | (0.00) | (0.00) | (0.00) |
| | VIX | Coeff | 0.05 | -0.67 | -0.40 | -0.30 | -0.08 | -0.11 | 0.32 | 0.18 | 0.17 | 0.18 |
| | | p-val | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.02) | (0.00) | (0.00) | (0.00) | (0.00) |
| | MOIB | Coeff | 0.07 | 0.77 | 0.38 | 0.21 | 0.17 | 0.09 | 0.05 | 0.03 | 0.15 | 0.42 |
| | | p-val | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.03) | (0.00) | (0.00) | (0.00) | (0.00) |
| | | $R^2$ | 35.15 | 39.25 | 38.92 | 38.09 | 37.48 | 36.70 | 35.84 | 34.72 | 33.59 | 32.75 |
| Third | Firm | Coeff | -4.52 | -3.71 | -3.11 | -2.74 | -2.54 | -2.24 | -1.59 | -0.94 | -0.87 | -0.06 |
| Regression | Size | p-val | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| | Profit | Coeff | -0.83 | -0.48 | -0.45 | -0.21 | -0.34 | -0.43 | -0.02 | 0.09 | 0.02 | -0.21 |
| | | p-val | (0.00) | (0.00) | (0.00) | (0.01) | (0.00) | (0.00) | (0.00) | (0.04) | (0.00) | (0.00) |
| | BTM | Coeff | 1.10 | -0.20 | 0.20 | 0.15 | 0.17 | -0.26 | 0.13 | 0.20 | -0.15 | -0.27 |
| | | p-val | (0.00) | (0.04) | (0.02) | (0.05) | (0.01) | (0.00) | (0.01) | (0.00) | (0.00) | (0.00) |
| | Outsider | Coeff | 1.23 | 0.12 | 0.34 | 0.58 | 0.16 | 0.32 | -0.10 | 0.23 | -0.18 | 0.10 |
| | | p-val | (0.00) | (0.32) | (0.00) | (0.00) | (0.01) | (0.00) | (0.05) | (0.00) | (0.00) | (0.00) |
| | Insider | Coeff | -4.10 | 1.00 | 0.13 | -0.75 | 0.20 | 0.05 | 0.32 | -0.07 | -0.14 | 0.10 |
| | | p-val | (0.00) | (0.00) | (0.23) | (0.00) | (0.01) | (0.41) | (0.00) | (0.18) | (0.00) | (0.00) |
| | R&D | Coeff | -1.79 | 0.07 | 0.53 | -0.30 | 0.54 | 0.29 | 0.55 | -0.21 | -0.09 | -0.23 |
| | | p-val | (0.00) | (0.66) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.04) | (0.00) |
| | Capex | Coeff | 1.78 | 0.53 | -0.56 | 0.26 | -0.58 | -0.04 | -0.24 | -0.06 | 0.14 | 0.28 |
| | | p-val | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.49) | (0.00) | (0.20) | (0.00) | (0.00) |
| | Options | Coeff | -5.47 | -0.06 | 0.86 | 1.24 | 0.15 | -0.47 | -0.69 | -1.57 | -0.52 | -0.35 |
| | | p-val | (0.00) | (0.80) | (0.00) | (0.00) | (0.32) | (0.00) | (0.00) | (0.00) | (0.00) | (0.01) |
| | | $R^2$ | 35.01 | 39.12 | 38.81 | 38.09 | 37.48 | 36.71 | 35.86 | 34.78 | 33.67 | 32.52 |

*(continued)*

# Table 10 – Robustness Checks (Continued)

## Panel B – Association of Quote-returns with Explanatory Variables

| | | Decile | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | | 7 | | 8 | | 9 | | 10 | |
| Variable | | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val |
| Market-wide commonality | Bid-ask spread | -1.8 | (0.00) | -1.4 | (0.00) | -1.2 | (0.00) | -0.7 | (0.00) | -0.9 | (0.00) | -0.8 | (0.00) | -0.6 | (0.00) | -0.5 | (0.00) | -0.5 | (0.00) | -0.3 | (0.00) |
| | Volume | -0.5 | (0.00) | -0.1 | (0.00) | -0.1 | (0.00) | -0.1 | (0.00) | 0.0 | (0.44) | 0.0 | (0.19) | 0.0 | (0.41) | 0.0 | (0.09) | 0.1 | (0.00) | 0.1 | (0.00) |
| | Volatility | -0.8 | (0.00) | -0.4 | (0.00) | -0.2 | (0.00) | 0.0 | (0.09) | 0.0 | (0.89) | 0.1 | (0.00) | 0.1 | (0.00) | 0.2 | (0.00) | 0.2 | (0.00) | 0.2 | (0.00) |
| | Order-imbalance | 1.0 | (0.00) | 1.1 | (0.00) | 1.1 | (0.00) | 1.2 | (0.00) | 1.1 | (0.00) | 1.1 | (0.00) | 1.1 | (0.00) | 1.1 | (0.00) | 1.1 | (0.00) | 1.2 | (0.00) |
| Stock-level trading characteristics | Volatility | 1.4 | (0.00) | 0.2 | (0.00) | 0.2 | (0.00) | 1.5 | (0.00) | 1.0 | (0.00) | 1.6 | (0.00) | 1.1 | (0.00) | 1.3 | (0.00) | 1.6 | (0.00) | 2.0 | (0.00) |
| | UED spread | -1.1 | (0.00) | -0.6 | (0.00) | -0.3 | (0.00) | 0.0 | (0.31) | -0.1 | (0.00) | -0.1 | (0.00) | 0.0 | (0.13) | 0.1 | (0.00) | 0.1 | (0.00) | 0.2 | (0.00) |
| | Bid-ask spread | 2.1 | (0.00) | 1.4 | (0.00) | 1.1 | (0.00) | 0.7 | (0.00) | 0.8 | (0.00) | 0.7 | (0.00) | 0.4 | (0.00) | 0.3 | (0.00) | 0.2 | (0.00) | -0.1 | (0.01) |
| | Volume | 0.3 | (0.00) | 0.4 | (0.00) | 0.4 | (0.00) | 0.4 | (0.00) | 0.4 | (0.00) | 0.4 | (0.00) | 0.3 | (0.00) | 0.3 | (0.00) | 0.2 | (0.00) | 0.1 | (0.00) |
| | Tick size | -1.7 | (0.00) | -0.8 | (0.00) | -0.5 | (0.00) | -0.6 | (0.00) | -0.5 | (0.00) | -0.4 | (0.00) | -0.3 | (0.00) | -0.3 | (0.00) | -0.3 | (0.00) | -0.1 | (0.00) |
| | Order-imbalance | 0.5 | (0.00) | 0.8 | (0.00) | 0.9 | (0.00) | 1.0 | (0.00) | 1.2 | (0.00) | 1.3 | (0.00) | 1.4 | (0.00) | 1.6 | (0.00) | 1.7 | (0.00) | 2.2 | (0.00) |
| Firm-specific structural characteristics | Firm size | 3.0 | (0.00) | 3.5 | (0.00) | 3.4 | (0.00) | 3.0 | (0.00) | 2.0 | (0.00) | 1.8 | (0.00) | 1.6 | (0.00) | 1.2 | (0.00) | 0.6 | (0.00) | 0.0 | (0.96) |
| | Profit | -0.4 | (0.02) | -0.3 | (0.00) | -0.2 | (0.00) | -0.4 | (0.00) | -0.3 | (0.00) | -0.2 | (0.00) | -0.4 | (0.00) | -0.4 | (0.00) | -0.2 | (0.00) | -0.1 | (0.02) |
| | BTM | 1.2 | (0.00) | 0.9 | (0.00) | 1.1 | (0.00) | 0.8 | (0.00) | 0.4 | (0.00) | 0.4 | (0.00) | 0.4 | (0.00) | 0.2 | (0.00) | 0.1 | (0.03) | -0.1 | (0.02) |
| | Outsider | -0.1 | (0.84) | 0.7 | (0.00) | 0.5 | (0.00) | 0.2 | (0.01) | -0.1 | (0.18) | 0.0 | (0.49) | 0.1 | (0.00) | 0.1 | (0.01) | 0.1 | (0.12) | -0.1 | (0.01) |
| | Insider | -0.7 | (0.19) | 0.0 | (0.93) | 0.2 | (0.04) | 0.2 | (0.10) | 0.0 | (0.81) | 0.2 | (0.01) | 0.2 | (0.00) | -0.1 | (0.26) | -0.2 | (0.00) | 0.0 | (0.58) |
| | R&D | 0.1 | (0.88) | 0.2 | (0.57) | -0.7 | (0.00) | -0.7 | (0.00) | 0.1 | (0.66) | 0.1 | (0.56) | -0.1 | (0.38) | -0.5 | (0.00) | -0.4 | (0.00) | 0.0 | (0.63) |
| | Capex | -0.6 | (0.00) | -0.2 | (0.01) | 0.0 | (0.64) | -0.1 | (0.04) | -0.1 | (0.36) | -0.1 | (0.02) | -0.3 | (0.00) | 0.0 | (0.72) | 0.0 | (0.59) | 0.0 | (0.93) |
| | Options | 0.0 | (0.00) | 0.0 | (0.00) | 0.0 | (0.00) | 0.0 | (0.00) | 0.0 | (0.00) | 0.0 | (0.00) | 0.0 | (0.00) | 0.0 | (0.00) | 0.0 | (0.00) | 0.0 | (0.00) |
| | $R^2$ | 1.5 | | 2.2 | | 2.3 | | 1.8 | | 1.3 | | 0.2 | | 1.5 | | 3.4 | | 6.2 | | 11.7 | |

## Panel C – Principal Component Analysis on *RAIN*

| | | | Sample Variance Explained (%) | | |
|---|---|---|---|---|---|
| RAIN Based on | Number of Days | Number of Stocks | $1^{st}$ PC | $2^{nd}$ PC | $3^{rd}$ PC |
| $IA_1$ over 15 minutes | 1,235.00 | 656.00 | 7.05 | 3.77 | 1.08 |
| | | | 7.05 | 10.82 | 11.90 |
| $IA_1$ over 30 minutes | 1,244.00 | 655.00 | 5.17 | 2.78 | 0.96 |
| | | | 5.17 | 7.95 | 8.92 |
| $IA_1$ over 60 minutes | 1,224.00 | 656.00 | 4.20 | 2.00 | 0.93 |
| | | | 4.20 | 6.20 | 7.13 |
| $IA_1$ over 1 day | 1,253.00 | 574.00 | 3.41 | 1.02 | 0.89 |
| | | | 3.41 | 4.43 | 5.32 |