Abstract

We develop a new proxy for the probability of informed trading based on contrarian trading and trade size. The proxy, called PCM, significantly captures the information asymmetry component of the bid-ask spread, becomes dramatically elevated around high-information events such as earnings announcements, and exhibits very similar patterns to Kyle’s lambda and price impact in time series. In double-sorted portfolios on illiquidity and PCM, excess returns increase from low to high PCM portfolios holding illiquidity constant, and in asset-pricing tests, the effect of PCM on returns remains significantly positive after controlling for illiquidity and other variables related to informed trading. A long-short portfolio strategy based on PCM generates a Fama-French three-factor alpha of 8.4 percent per year. Overall, PCM appears highly economically and statistically significant, thus lending support to the information asymmetry hypothesis that information risk is priced.

JEL Classification: G10, G14, G19

Keywords: Information asymmetry; Liquidity; Stock returns; Informed trading; Information risk; Contrarian trades; Kyle’s lambda; Asset pricing; Market microstructure; PIN; PCM
1 Introduction

The asset pricing literature focuses primarily on risk factors that affect an asset’s returns in the long run while abstracting from the trading mechanism of the market in which the asset is transacted. In contrast, the market microstructure literature focuses on how the trading processes itself affects the formation of transaction prices, while largely being silent on its asset pricing implication. Such a dichotomy is convenient for analysis but leaves open an important question: to what extent do microstructure features affect expected stock returns?

To be sure, there are at least two prominent hypotheses that have been put forward explaining the linkage between market microstructure and stock returns. Amihud and Mendelson (1986) propose the “liquidity” hypothesis that investors maximize expected returns net of transaction costs, and hence liquidity, which embodies such costs, should be priced. They present supporting evidence by showing that stock returns significantly increase in bid-ask spreads. In addition, Amihud (2002) constructs an illiquidity measure and shows that stock returns significantly increase in this measure. Easley and O’Hara (2004) propose the “information asymmetry” hypothesis that uninformed traders lose to informed traders and hence require compensation to hold stocks with greater information asymmetry. As a result, such information risk should be priced. Easley et al. (2002) develop the PIN measure of the probability of informed trading as a proxy for information asymmetry and find that stock returns significantly increase in PIN. It is worth noting that these two hypotheses are not competing, but are rather intimately related to each other in the literature.

Recently, the information asymmetry hypothesis has come under serious re-examination. In a well-known and widely cited study, Duarte and Young (2009) extend the structural model of Easley et al. (1996) and decompose PIN into an information asymmetry component and an illiquidity component that arises due to a symmetric order flow shock. They find that stock returns are significantly and positively related to the illiquidity component, but not to the information asymmetry component. Moreover, the effect of PIN disappears once illiquidity is included in the asset pricing tests. As such, they conclude that PIN is priced, not because it captures informed trading per se, but because it serves as proxy for the illiquidity effect that is unrelated to information. Similarly, using a sample of international stocks across 47 countries, Lai et al. (2014) find that their PIN estimates capture information asymmetry, but they are not priced in international markets. Thus, the question of whether and how information asymmetry affects expected stock returns still seems to be an open one. As Duarte and Young note (p. 136): "It is possible...that private information could indeed be related to expected returns. However, in this case, [PIN] would be [an] inappropriate [proxy] for information asymmetry." The goal of this paper is to address this question with a new alternative proxy for the probability of informed trading, which we develop and test below.

It should be noted, however, that tests of the liquidity hypothesis are not without their own complications either. Liquidity, like information asymmetry, is a slippery concept, mainly because it is not observed directly. One, therefore, must rely on noisy proxies such as the bid-ask spread or the Amihud (2002) illiquidity measure. If the noise in such measures were orthogonal to the signal, then we would have a reliable proxy. However, market microstructure theory posits that
information asymmetry and liquidity are intimately related, making it difficult to disentangle these
two signals from each other, even though they arise from two distinct economic forces (Biais et
al., 2005). On the one hand, when information asymmetry increases, the adverse selection problem
is heightened and this can lead to a fall in liquidity. On the other hand, there are also liquidity
concerns unrelated to information such as order-handling costs, inventory-holding costs, or market
power. In other words, one part of liquidity is information driven, while the other part is unrelated
to information. In addition, there is evidence that the effect of characteristic liquidity measures
such as Amihud illiquidity have declined dramatically over the years, especially in the last decade
(Ben-Rephael et al., 2015).

Many of the liquidity measures used in the literature suffer from such confounding effects. For
example, a large body of literature shows that bid-ask spreads consist of both an adverse-selection
component and other transaction-cost components unrelated to information (Glosten and Harris
1988; Hasbrouck 1991; George et al., 1991; Lin et al., 1995; Huang and Stoll, 1997; Madhavan et
his illiquidity measure is significantly and positively related to both Kyle (1985) lambda and the
fixed cost component in the bid-ask spread. Since Kyle lambda captures information asymmetry,
Amihud illiquidity also suffers from similar confounding effects.

Thus, it seems that the current measures of illiquidity and information asymmetry – bid-ask
spreads, Amihud (2002) illiquidity, and PIN – all suffer from confounding effects (from each other,
no less). One therefore must be cautious when using any of these proxies to test either of the two
above hypotheses. For example, when one sees that stock returns are significantly and positively
related to bid-ask spreads (Amihud and Mendelson, 1986; Easley et. al., 1996), or illiquidity (Ami-
hud, 2002; Duarte and Young, 2009), or PIN (Easley et al., 2002), the question that immediately
arises is whether this premium is due to information asymmetry or liquidity concerns unrelated to
information. To disentangle the two, what is needed is a cleaner measure of information asymmetry
that is not confounded by the illiquidity effect.

In light of the discussion above, the aim of this paper is twofold: (1) to develop a new proxy for
the probability of informed trading that is motivated by relevant theories and (2) to employ this
proxy to test the information asymmetry hypothesis while controlling for the important illiquidity
effect. To achieve the first aim, we rely on the literature on the behavior of informed traders to
motivate our proxy for the probability of informed trading. In particular, the literature suggests
that informed traders are more likely to engage in contrarian trades than herding (Campbell et
al., 1993; Avramov et al., 2006; Back et al., 2016), and they tend to submit medium-size trades.
We thereby develop our measure for the probability of informed trading based on contrarian trades
with medium size in two steps.

In the first step, as in Avramov et al. (2006) and Chang and Wang (2015), we adopt the
notion that buy (sell) trades made amid declining (rising) prices are contrarian in nature, while buy
(sell) trades initiated during rising (declining) prices suggest uninformed herding. In this spirit, we
define contrarian trades at the intraday, transaction-level frequency as buy trades in the presence
of negative unexpected returns and sell trades in the presence of positive unexpected returns. The daily probability of contrarian trading (PC) is then simply calculated as the proportion of the number of contrarian trades over the total number of trades during a day.

This PC measure captures the salient features of information asymmetry modeled in Back et al. (2016), which is a hybrid of PIN and Kyle’s (1985) model that allows for a probabilistic information event and an optimizing (possibly) informed trader. Their hybrid model implies that both price changes and order flows are needed to identify information asymmetry. Moreover, the optimizing informed trader in Back et al., as in Kyle’s model, is by definition a contrarian trader who buys when liquidity traders sell (as the price gets driven below the informed trader’s expected price based on his or her private information) and sells when liquidity traders buy (which pushes the price above the informed trader’s expected price). In the same spirit, as a baseline for identifying information asymmetry, our PC measure for the probability of contrarian trading by construction captures both notions of price changes and order flows.

Next, we further refine PC to obtain a finer measure of informed trading to account for the concern that some contrarian trades included in PC may come from uninformed traders who by chance happen to trade in a contrarian direction but who actually do not have private information. To do so, we rely on trade size to further differentiate those contrarian trades that are more likely to be submitted by informed traders from those by uninformed traders. The microstructure literature on stealth trading suggests that large informed traders prefer to split orders and submit medium size trades to mitigate the price impact of their trades (Barclay and Warner, 1993; Chakravarty, 2001). We thereby further condition PC on medium-size trades, which we define as trades consisting of 1,000-9,999 shares, resulting in the PCM measure. While Chakravarty uses 500 shares as the lower bound for medium size, Kyle and Obizhaeva (2016) posit in their market microstructure invariance hypotheses that the average bet size of long-term traders, and hence trade size (all else equal), must grow with overall trading activity in the market, which of course has increased dramatically over recent years. Thus, we raise the lower bound of medium-sized trades to 1,000 shares per trade to better capture this phenomenon.

It is worth noting that PCM, as a relatively model-free measure, is fundamentally different from the structural-model and statistical-based PIN of Easley et al. (1996, 2002). In particular, PCM is a standalone index based on the proportion of certain types of trades occurring during a trading day, and therefore can be computed directly from the data at daily frequency without the need for aggregating data over longer horizons and estimation techniques that require numerical optimization, which is an especially convenient feature when dealing with very large transaction-level databases of trades and quotes. Furthermore, PCM is a general measure of informed trading that does not depend on a specific group of market participants or a particular type of private information. That is, PCM may capture informed trading from investors, asset managers, limit order traders, or even market makers with superior information (e.g., Calcagno and Lovo, 2006) and may include private information about payoffs, endowments, orders, or market participants (e.g., Vayanos, 1999).
We calculate PCM at the daily frequency from January 1993 to December 2012 for stocks traded on the NYSE and AMEX markets. To be consistent with the asset pricing literature, we then derive a monthly PCM measure to be used for our empirical analysis. The market microstructure theory posits that stock bid-ask spreads are influenced by adverse selection due to informed trading and by other factors unrelated to information (Biais, et al., 2005). Thus, we gauge the effectiveness of PCM as a measure for informed trading by examining whether it is positively related to the bid-ask spread. We find that the effect of PCM on spreads is significant and positive at the 1% level. In light of the critique of Duarte and Young (2009), we include illiquidity, as well as firm size and other control variables so that we are able attribute the marginal effect of PCM on bid-ask spreads to private information trading, rather than liquidity factors unrelated to information. In a similar vein, Chang and Wang (2015) also find that spreads are significantly and positively related to their PC-based proxies after controlling for an array of trading and illiquidity effects. Taken together, the evidence renders consistent and strong support that PCM captures the adverse selection component of the bid-ask spread and thus serves as an effective proxy for informed trading.

There is also a large body of literature using earnings announcements as a proxy for information events. Kim and Verrecchia (1994, 1997) posit that earnings announcements can stimulate sophisticated traders to process public disclosure into private information, thus resulting in higher information asymmetry around earnings announcements. In light of this, we conduct study an event window around earnings announcement days to examine whether PCM is able to register such information-rich events. We find that PCM begins to rise slightly one day before the earnings announcement day, spikes dramatically upward on announcement days, and then subsides, remaining elevated above normal for roughly three days following an announcement. This pattern strongly suggests that PCM is able to accurately detect the presence of informed trading.

Back et al. (2016) note that Kyle’s (1985) lambda, which also reflects the degree of information asymmetry trading, is also the price impact of trades. They show that Kyle’s lambda and the price impact estimate (Holden and Jacobsen, 2014) exhibit similar patterns in time-series. As our PCM measure is consistent with the hybrid model of Back et al., we find that there is also a striking similarity in time-series patterns between PCM and both Kyle’s lambda and the price impact. Specifically, all three time series rise over the 1990s and drop sharply following the turn of the century, with a brief upward movement during the recent financial crisis of 2008. These results further support that PCM is an effective proxy for information asymmetry as it exhibits similar patterns to both Kyle’s lambda and the price impact.

We then turn our attention to the relationship between expected stock returns and informed trading. First, we sort stocks into monthly decile portfolios based on PCM and conduct a "10-1" long-short (self-funded) portfolio strategy to evaluate excess returns and portfolio alpha. We find that the long-short portfolio generates large excess returns and has a large positive and significant Fama-French three factor alpha of 0.702% (corresponding to an 8.4% abnormal return per year), which suggests that the effect of PCM on stock returns is economically significant as well. We then compute the excess returns on 25 double-sorted portfolios sorted monthly on PCM and illiquidity.
quintiles. Excess returns generally increase as we move from low to high PCM portfolios within a given illiquidity category.

We then conduct more formal tests of the information asymmetry hypothesis by conducting a series of monthly, firm-level Fama-MacBeth (1973) regressions of stock returns on PCM and other variables standard variables in the literature that are known to also influence the cross-section of stock returns. From these asset pricing tests, we find that PCM is positive and significant at the 1% level in explaining the cross-section of stock returns after controlling for illiquidity and the three Fama-French firm characteristics beta, size, and book-to-market ratio. Since the effect of PCM on returns is obtained after controlling for illiquidity, we are able to overcome the critique of Duarte and Young (2009) and thereby attribute the marginal effect of PCM on returns to informed trading, rather than liquidity concerns unrelated to information.

We then conduct a direct comparison between PCM and PIN. The estimation results show that the coefficient of PCM is always positive and significant, with or without illiquidity included in the regression, but the corresponding coefficient of PIN is invariably insignificant in our sample (with the inclusion of liquidity further reducing point estimates and test statistics). The direct comparison between PCM and PIN thus suggests that PCM is a measure of information asymmetry that is orthogonal to and more robust than PIN.

Recognizing recent technological innovations and regulatory changes that have affected the trading process since the advent of decimalization in January 2001, we also conduct our analysis using two subsamples to reflect periods prior to and after this important change, namely a subperiod from 1994-2000, and from 2001-2011. We find that the effect of PCM as a proxy for information asymmetry actually strengthens after the decimalization in 2001, suggesting that the recent technological innovations and regulatory changes do not necessarily imply a diminishing information risk premium. In fact, our evidence points to the opposite, i.e., that information risk remains substantial and significant.

In our subsample analysis, we also consider an alternative version of PCM using 500 shares as the lower bound for medium-sizes trades, as in the earlier stealth trading literature (Barclay and Warner, 1993; Chakravarty, 2001). We find that the size of the coefficient of PCM is more than twice that of the alternative PCM (denoted PCM500) in each sub-period. Furthermore, PCM is statistically significant in both sub-periods, whereas PCM500 is only significant in the latter sub-period. These results validate our increasing of the lower bound on medium-sized trades to construct PCM and are consistent with Kyle and Obizhaeva’s (2016) market microstructure invariance hypotheses that implies that the increased trading activity in recent years must lead, all else equal, to larger trade sizes for long-term traders who are motivated by information rather than liquidity.

For robustness, we then examine alternative explanations to information asymmetry by adding additional explanatory variables, including return standard deviation, turnover, coefficient of variation of turnover, and momentum. The estimation results show that the coefficient on PCM remains significant and positive after controlling for these additional variables, providing further evidence that PCM is a fundamental priced variable rather than a proxy for some omitted variables unrelated
to information. Moreover, since the effect of PCM remains significant and positive after accounting for these alternative explanations, there is robust and consistent evidence to support the information asymmetry hypothesis (Easley and O'Hara, 2004) that information risk is priced and informed trading is an important determinant of expected stock returns.

The paper proceeds as follows. Section 2 describes the data used in this study. Section 3 develops our PCM measure for the probability of informed trading. Section 4 evaluates the properties of PCM and other variables of interest used in this study. Section 5 investigates the link between PCM and informed trading. Section 6 examines the relationship between PCM and the cross-section of stock returns. Section 7 concludes.

2 Data

The intraday transaction data come from the Trades and Quotes (TAQ), database and information on other share characteristics (e.g., share code, exchange code, shares outstanding, etc.) are from the Center for Research in Security Prices (CRSP) database. The data span the period January 1993 to December 2012. We include NYSE and AMEX-listed domestic issues, excluding foreign companies, exchange traded funds, closed-end funds, and REITs (real estate investment trusts). Transactions occurring outside the normal opening and closing times of the exchange are omitted, along with transactions that have special conditions, corrections, or other indicators. To avoid complications associated with thinly traded, illiquid stocks, only shares for which there are at least 250 trades per month are included in the analysis.

As is standard in the empirical microstructure literature, we use the Lee and Ready (1991) algorithm to match trades and quotes and to determine whether a particular trade is buyer- or seller-initiated. For each firm in the sample, the total number of trades on a particular day is the sum of all buy and sell trades (as well as unsigned trades) occurring on that day. As described further below, our proxies for the probability of informed trading are then calculated using the number of certain types of buy and sell trades in proportion to the total number of trades on a given day.

For the adverse selection and asset pricing tests, additional data on firm and stock characteristics are required and obtained from the CRSP and COMPUSTAT databases. As is standard in the literature, the analysis is conducted using monthly cross-sections, so we aggregate our daily measure of the probability of informed trading to derive a monthly proxy, resulting in 244,900 firm-month observations (2,783 firm observations) upon merging the datasets. Considering only the subset of NYSE- and AMEX-listed firms for which all CRSP/COMPUSTAT data are available and our monthly proxies for the probability informed trading are calculable from the TAQ database results in cross-sectional samples ranging in size from 1,160 to 1,503 stocks across all months. This is comparable to Easley et al. (2002) and Duarte and Young (2009). With a sample period from 1993-2012, there are 238 months with which to conduct the analysis.
3 Constructing the PCM proxy for the probability of informed trading

To the extent that the identity of informed traders is not available to econometricians, we rely on the literature on the behavior of informed traders to motivate the development of our proxies for the probability of informed trading. In particular, we focus on two aspects of informed trading behavior to distinguish it from uninformed trading: (1) contrarian versus herding trades and (2) trade size choice. The literature suggests that informed traders are more likely to engage in contrarian trades than herding, and they tend to submit medium-size trades. In what follows, we detail the process of the development of our proxies for the probability of informed trading as guided by the relevant literature.

3.1 Information and trading behavior: Contrarian versus herding trades

Campbell et al. (1993) emphasize that changes in a stock’s price are caused by information that affects the valuation of the firm, or are due to the actions of liquidity or "non-informational" traders, who desire to buy or sell stock for exogenous reasons. In the former case, prices reflect new information and thus price reversals are less likely to be observed, if any. In the latter case, temporary demand and supply pressures are expected to be short-lived, and thus price reversals are more likely to be observed. Therefore, uninformed trading should be associated with negative serial correlation in individual stock returns, while no such dependence should be associated with informed trading.

Based on this intuition, Avramov et al. (2006) devise an empirical framework to delineate whether a particular trading day is dominated by broadly "contrarian" versus "herding" behavior on the part of investors. In particular, they define daily sell orders in face of unexpected positive returns as contrarian trades while the sell orders in face of unexpected negative returns as herding trades. Consistent with the model of Campbell et al. (1993), Avramov et al. (2006) show that unexpected returns associated with herding trades exhibit significant negative serial correlation, while the autocorrelation for contrarian trades is insignificant. Thus, it appears that contrarian (herding) trades are broadly akin to informed (uninformed) trades. While Avramov et al. (2006) use these criteria of informed trading only for daily sell orders to study its effect on volatility, Chang et al. (2014) apply this concept of contrarian trades further to both buy and sell orders at 15-minute intervals throughout the trading day and thereby calculate the probability of contrarian trades at each interval to construct a dynamic intraday probability of informed trading (DPIN). They find that DPIN is useful in explaining the relationship between idiosyncratic return variations and informed trading as conjectured by Roll (1988). Chang and Wang (2015) adopt a similar concept to construct a proxy for informed trading at the daily frequency based on the probability of contrarian trading and show that this proxy successfully captures the adverse selection component of bid-ask spreads and illiquidity due to information asymmetry.

In this paper, we build on Avramov et al. (2006) and Chang and Wang (2015) and first derive the
probability of contrarian trades at the daily frequency as the first foundation block of our measure of informed trading. Specifically, we isolate the unexpected component of returns as the residuals from the following regression at the daily frequency:

\[ R_{i,s} = \gamma_0 + \sum_{k=1}^{4} \gamma_{1i,k} D_{k}^{Day} + \sum_{k=1}^{12} \gamma_{2i,k} R_{i,s-k} + \varepsilon_{i,s}, \] (1)

where \( R_{i,s} \) is the return on stock \( i \) and day \( s \), and \( D_{k}^{Day} \) represents day-of-week dummy variables for Tuesday through Friday. Thus, the residual \( \varepsilon_{i,s} \) captures the variation in returns left over after average day-of-week effects and the effects of past returns have been accounted for and thereby serves as a proxy for unexpected returns.

Buy (sell) trades in the presence of negative (positive) unexpected returns are classified as broadly contrarian trades. On the other hand, buy (sell) trades in the presence of positive (negative) unexpected returns are broadly classified as herding trades. Formally, let \( NB_{i,s}, NS_{i,s}, \) and \( NT_{i,s} \) be the number of buy, sell, and total trades, respectively, for stock \( i \) on day \( s \). Then, the probability of contrarian trades (PC) can be expressed as follows:

\[ PC_{i,s} = \frac{NB_{i,s}}{NT_{i,s}}(\varepsilon_{i,s} < 0) + \frac{NS_{i,s}}{NT_{i,s}}(\varepsilon_{i,s} > 0), \] (2)

where \( (\varepsilon_{i,s} < 0) \) is an indicator variable that equals 1 when the unexpected return is negative and zero otherwise, and \( (\varepsilon_{i,s} > 0) \) takes on the value of 1 when unexpected returns are positive and zero otherwise.

PC captures the salient features of information asymmetry as modeled in Back et al. (2016). They note that the PIN model is driven by order flow imbalance, but order flow alone cannot identify information asymmetry if liquidity providers react to information asymmetry and informed traders react to liquidity. For example, an increase in information asymmetry can lead to a fall in liquidity, which in turn can lead to less informed trading and thereby may offset the increase in informed trading induced by higher information asymmetry in the first place. Therefore, the resulting order flow, which includes both informed trading and liquidity trading, need not reflect the change in information asymmetry.

Instead, Back et al. (2016) propose a hybrid of PIN and Kyle’s (1985) model that allows for a probabilistic information event and an optimizing (possibly) informed trader. This hybrid model implies that both price changes and order flows are two essential ingredients needed to identify information asymmetry. Moreover, the optimizing informed trader in both the Back et al. and Kyle models is by definition a contrarian trader who buys when liquidity traders sell (and drive the price down below its expected value based on the informed trader’s private information) and sells when liquidity traders buy (and push the price up beyond this value). In the same spirit, PC in Equation (2) is calculated based on both price changes and order flows, and it is defined precisely as the probability of contrarian trades.
3.2 Information and trade size

Some contrarian trades defined in Equation (2) may come from uninformed traders who happen to trade in a contrarian direction by chance but who actually do not have private information. To address this issue, we use trade size to further differentiate those contrarian trades that are more likely to be submitted by informed traders from those of uninformed traders. The relationship between information and trade size choice has been extensively studied in the microstructure literature, suggesting that large informed traders prefer to split orders and submit medium size trades to mitigate the price impact of their trades. Thus, by applying trade size as a filter to further separate informed from uninformed trades, the goal is to obtain a more precise proxy for informed trading.

Specifically, Kyle (1985) shows that it is optimal for a large insider to strategically smooth out his trading on private information. Admati and Pfeiderer (1988) show further that informed traders have incentive to disguise their trades by placing them among uninformed trades. Barclay and Warner (1993) propose the stealth-trading hypothesis, arguing that informed traders disguise their trades by submitting medium-sized trades, which are just small enough to camouflage their private information and just large enough to avoid high transaction costs. Barclay and Warner, Hasbrouck (1995), Chakravarty (2001), and Alexander and Peterson (2007) find evidence of stealth trading by institutional investors such that their medium-sized trades tend to have large cumulative price impact.

In light of this discussion, we refine our baseline PC measure based on a medium trade-size criterion to create the finer PCM measure, which serves as a closer proxy for informed trading than contrarian trading alone:

$$PCM_{i,s} = \frac{NB_{i,s}^{Med}}{NT_{i,s}}(\varepsilon_{i,s} < 0) + \frac{NS_{i,s}^{Med}}{NT_{i,s}}(\varepsilon_{i,s} > 0),$$

where $NB_{i,s}^{Med}$ and $NS_{i,s}^{Med}$ are the total number of medium sized buy and sell trades, respectively, in stock $i$ on day $s$, where an intraday trade is classified as medium sized if the number of shares traded is between 1,000 and 9,999 shares. We set our lower bound for the medium-size trade at 1,000 shares, which is greater than the 500 shares used in the earlier stealth trading literature (Barclay and Warner, 1993; Chakravarty, 2001). This adjustment is motivated by Kyle and Obizhaeva’s (2016) market microstructure invariance hypothesis that posits that the average bet size for informed traders, and by implication trade size (all else equal), necessarily grows as the overall trading activity in the market increases. Given the dramatic increase in market activity in recent years, especially after decimalization in 2001, this is a salient issue.\(^1\)

It is worth noting that PCM is a general measure of informed trading that does not depend on a specific group of market participants or a particular type of private information. That is, PCM may capture informed trading from investors, asset managers, limit order traders, or even market makers with superior information (e.g., Calcagno and Lovo, 2006) and may include private information

---

\(^1\)For robustness, we also consider an alternative PCM using 500 shares as the lower bound and find that PCM using 1000 shares as the lower bound indeed performs better as a proxy for information asymmetry. See Table 8.
about payoffs, endowments, orders, or market participants (e.g., Vayanos, 1999). Clearly, PCM as proxy for the probability of informed trading, is fundamentally different by design from PIN of Easley et al. (1996, 2002), which is based on a structural trading model that requires maximum likelihood estimation.

Another advantage of PCM stems from the fact that it is, in essence, a standalone index based on the proportion of certain types of trades occurring during a trading day, and therefore can be computed directly from the data at daily frequency without the need for aggregating data over longer horizons and estimation techniques that require numerical optimization. This becomes much more convenient when dealing with extremely large transaction-level databases such as TAQ. Furthermore, PCM is better suited to capture the dynamic nature of private information and its relationship with adverse selection. In this respect, PCM employs the standard Lee and Ready (1991) algorithm to classify trades rather than the volume-based classification scheme of Easley et al. (2011, 2012), which may be the source of the mechanical link between VPIN and volume effects, as noted in Andersen and Bondarenko (2014a,b).

4 Properties of PCM and other variables of interest

4.1 Distribution of PCM for the probability of informed trading

We calculate PCM at daily frequency from January 1993 to December 2012 for stocks traded on the NYSE and AMEX markets. Because the adverse selection and asset pricing tests are conducted at the (standard) monthly frequency, we average daily values to derive a corresponding monthly PCM, which is what we will refer to below and use in our subsequent analyses. Table 1 reports empirical properties of PCM. Panel (a) presents the distribution across firms and months and Panel (b) presents the distribution across firms. Both panels indicate substantial variation both across firms and across firms and time, suggesting at first glance that firms indeed differ along dimensions related to informed trading as measured by PCM.

4.2 Correlations between PCM and firm characteristics

Table 2 reports the correlations between firm-level PCM and other relevant firm characteristics with regard to adverse selection and asset pricing, namely firm size, trading volume, illiquidity, and the bid-ask spread. Several observations are in order. First, PCM is significantly and positively correlated with the bid-ask spread. Second, and more surprisingly, PCM is not correlated with illiquidity. In concert, these two findings suggest, at least preliminarily, that PCM is capturing the component of the bid-ask spread that is purely due to adverse selection and unrelated to illiquidity. This is a promising sign, at this stage, that PCM can disentangle the information asymmetry effect from the illiquidity effect, thus overcoming the Duarte and Young (2009) critique.

Third, PCM is significantly and negatively correlated with firm size. This result suggests that informed trades concentrate more on small firms. To the extent that small firms are subject to higher information asymmetry, it makes sense that informed traders have incentive to split orders
and submit medium-size trades to mitigate price impact. Lastly, both the bid-ask spread and illiquidity are significantly and negatively correlated with both firm size and trading volume, while spreads and illiquidity are both significantly and positively correlated with each other. These results are consistent with prior expectations.

### 4.3 Other variables used in adverse selection and asset pricing tests

For the purpose of identifying the effect of informed trading on the adverse-selection component of the bid-ask spread and the cross-section of stock returns, which is the focus of the rest of the paper, it is paramount to include an illiquidity measure in the asset pricing tests. As noted above, Duarte and Young (2009) find that the effect of PIN of Easley et al. (1996, 2002) on stock returns disappears once illiquidity is included in the regression, implying that PIN does not have any incremental explanatory power over illiquidity in explaining returns. Moreover, and related to this, our discussion above on the mutually confounding effects between information asymmetry and illiquidity also underscores the necessity of controlling for the latter in order to properly identify the former. Thus, in light of the both the Duarte and Young critique and the existence of such confounding effects, we include in all of our adverse selection and asset pricing tests the Amihud (2002) illiquidity measure, denoted ILLIQ, which for our purposes is calculated as the average ratio of the absolute daily return to daily dollar volume over the previous month. This is to make results comparable to PCM, which is also calculated by averaging daily values over a month.

Following Easley et al. (2002) and Duarte and Young (2009), the preranking portfolio beta for an individual stock in year $t$ is estimated using data from at least two years to, when possible, five years prior. Stock returns are regressed on the contemporaneous and lagged value-weighted CRSP NYSE/AMEX index. The preranking portfolio beta is then computed as the sum of the two coefficients. Based on these estimated betas, stocks are sorted into 40 portfolios, and monthly portfolio returns are calculated as equally-weighted returns of the stocks in each portfolio. For each of the 40 portfolios, the postranking portfolio beta is estimated from the full sample period by regressing portfolio returns on contemporaneous and lagged CRSP index returns. The corresponding portfolio beta, denoted $BETA$, is then the sum of these two coefficients. Individual stock betas are then replaced by the portfolio beta of the portfolio in which they belong. This sorting process is repeated at the beginning of every year, so that portfolio betas change across years. In addition to $BETA$, the other two standard Fama-French (1992) firm-characteristics are also included: $SIZE$ denotes the logarithm of the year-end market value of equity, and $BM$ represents the logarithm of book equity divided by market value of equity at the end of the previous year.

As further controls in the adverse selection and asset pricing test we add additional variables, including bid-ask spreads, return variability, and share turnover, that may proxy for liquidity or volume effects that may be correlated with our PC-based proxies and also affect the adverse selection component or returns. The bid-ask spread, $SPREAD$, is the average monthly bid-ask spread of stock. The stock return variability, $STD$, is the daily return standard deviation of stock over the previous year. Stock turnover, $TURN$, and the coefficient of variation of turnover, $CVTURN$, are
employed to control for volume effects: the former is the logarithm of the average monthly turnover divided by the number of shares outstanding over the previous three years, while the latter is defined as the logarithm of the coefficient of variation of the monthly turnover. Chordia et al. (2001) find that both turnover variables are negatively related to stock returns.

Table 3 reports the summary statistics for all additional variables used in the adverse selection and asset pricing tests across all stocks and months. There appears to be substantial and sufficient variation in each variable across firms during the 1993-2012 sample period for reliable estimation.

5 The link between PCM and informed trading

5.1 Bid-ask spreads, adverse selection, and the PCM measure

Market microstructure theory posits that stock bid-ask spreads are influenced by adverse selection due to informed trading and by other factors unrelated to information (Biais, et al., 2005). In light of this, one way to gauge the effectiveness of PCM as a proxy for informed trading is to test whether it is indeed positively related to the spreads as predicted. Since spreads may also be influenced by factors unrelated to information such as order-handing and inventory-control costs, we use other variables to control for such factors. Specifically, we conduct the following monthly Fama-MacBeth (1973) regressions:

\[
SPREAD_{i,t} = \alpha_t + \beta_1PCM_{i,t-1} + \beta_2ILLIQ_{i,t-1} + \beta_3SIZE_{i,t-1} + \gamma_iX_{i,t-1} + \epsilon_{i,t} \tag{4}
\]

for stock \( i \) in month \( t \), where the dependent variable \( SPREAD \) is the average bid-ask spread of a stock over the current month, \( PCM \) is the informed trading proxy over the previous month; \( ILLIQ \) is the Amihud (2002) illiquidity measure, which is calculated as the average ratio of the absolute daily return to daily dollar volume in the previous month; \( SIZE \) in each month is taken to be the logarithm of market value of equity at the end of the previous year. The vector of control variables \( X \) contains factors associated with the bid-ask spread, namely \( TURN \), which is the logarithm of the average monthly turnover divided by the number of shares outstanding from over the previous three years, and \( CVTURN \), the logarithm of the coefficient of variation of turnover; \( RET \), the percentage return in excess of the one-month T-bill rate in the previous month; \( STD \), the daily return standard deviation in the previous year; and a lag of \( SPREAD \), the average bid-ask spread in the previous month.

If PCM successfully captures the adverse selection component of the bid-ask spread that is due to informed trading, we would expect \( \beta_1 \) to be positive, even after controlling for other important factors that influence spreads and may be correlated with informed trading, notably size and illiquidity. In addition, to the extent that market makers take into account the illiquidity condition in the previous month, we would expect \( \beta_2 \) to be positive. To the extent that firm size is negatively related to the spreads, we would expect \( \beta_3 \) to be negative.
The results of the above regression are reported in Table 4. First, all coefficients are of the predicted sign. Most importantly, the estimate of $\beta_1$ on PCM is significant and positive at the 1% level. This result provides strong evidence that PCM is indeed capturing the adverse selection component of the bid-ask spreads that is due strictly to information asymmetry and orthogonal to other determinants of the bid-ask spread. In other words, PCM does not appear to be confounded by other factors such as illiquidity and firm size. At this stage, such results again suggest that PCM is a promising candidate for overcoming the Duarte and Young (2009) critique. Here, as far as spreads are concerned, the effect of PCM remains strong, even after controlling for illiquidity. This finding is similar to those in Chang and Wang (2015), who also find that bid-ask spreads are significantly and positively related to their PC-based informed trading proxies after controlling for an array of trading and illiquidity effects.

In addition, consistent with our expectations, the estimate of $\beta_2$ is significantly positive and the estimate of $\beta_3$ is significantly negative, both at the 1% level. These results confirm that that market makers do take into account the illiquidity condition in the previous month, ILLIQ-1, when setting their current spreads and firm size is negatively related to spread. Overall, these results lend strong support to the use of PCM as an effective proxy for informed trading.

5.2 PCM around earnings announcements

There is a large body of literature using earnings announcements as a proxy for information events. Kim and Verrecchia (1994, 1997), among others, posit that earnings announcements can stimulate sophisticated traders to process public disclosure into private information, thus resulting in higher information asymmetry around earnings announcements. Consistent with this prediction, Lee et al. (1993) and Yohn (1998) find an increase in bid-ask spreads in the three days around earnings announcements. Moreover, Krinsky and Lee (1996) find that only the adverse selection component of the spread increases significantly around earnings announcements.

Given the strong evidence that earnings announcements are information-asymmetry rich events, then if PCM is a useful measure of informed trading, we should detect an increase in PCM around earnings announcements. To test this prediction, we examine the behavior of PCM in an event window around earnings announcement days. Earnings announcement dates are collected from COMPUSTAT for all firms in the sample from 1993 to 2012. Denoting earning announcement days as Day 0, we analyze the behavior of PCM in event time using a [-10, +10] day window around earnings announcements days. We first compute the average PCM value for each stock and then calculate the ratio of each stock’s PCM on event days to each stock’s mean PCM. We then average these ratios across all stocks on each event day.

Figure 1 shows the value of this average ratio in event time along with 95% confidence bands. Most notably, there is a large and significant spike in PCM on earnings announcements days, with informed trading (as proxied by PCM) increasing by about 11% above normal on Day 0. Consistent with Kim and Verrecchia (1994, 1997), PCM begins rising prior to the announcement day and remains higher than normal but declining over subsequent days. In our case, PCM begins
rising the day before announcement day, peaks on the announcement day, and then declines but
remain significantly above normal for roughly three days following an announcement. These results
strongly suggest that PCM is directly measuring such information asymmetry. As a note of contrast,
using a similar event window approach, Back et al. (2015) find that PIN (or more precisely, relative
absolute order imbalance, which is used as a proxy for PIN (Aktas, 2007)) actually falls around
earnings announcements, and therefore performs poorly in capturing information asymmetry.

5.3 PCM, Kyle’s lambda, and price impact in time series

Kyle (1985) indicates that Kyle’s lambda reflects the degree of information asymmetry. Consistent
with this prediction, Cong et al. (2010) find that the estimated Kyle’s lambda is higher around
earnings announcements. Similarly, Back et al. (2016) develop a hybrid model of PIN and Kyle’s
(1985) model and calculate a Kyle-style lambda as a measure of information asymmetry. Back et
al. (2015) note that Kyle’s lambda is the price impact of trades and use the price impact estimate
suggested in Holden and Jacobsen (2014) as the empirical counterpart of Kyle’s lambda. They find
that the price impact estimate rises dramatically around earnings announcements, thus confirming
that price impact is a proxy for information asymmetry. Moreover, Back et al. (2016) find that
Kyle’s lambda and price impact both rise during the 1990s and then decline sharply following the
turn of the century, with a small upward movement during the recent financial crisis of 2008.

As noted above, PCM is consistent with the hybrid model of Back et al. (2016). To the extent
that both Kyle’s lambda and the price impact estimate proxy for information asymmetry, we should
then expect to see similar patterns over time between our PCM measure and both Kyle’s lambda
and the price impact estimate. To see whether this similarity indeed exists, we construct a daily
time series by calculating the average PCM across all stocks for each day in the sample. Figure 2
shows the daily time series for this average (i.e., market) PCM over the 1993-2012 period. There
is a striking similarity in time-series patterns between PCM and both Kyle’s lambda and the price
impact exhibited in Back et al. (2016) (see Figures 3f and 5, respectively). Specifically, all three
series rise during the 1990s and then decline sharply following the year 2000, with a small and brief
upward movement following the recent financial crisis in 2008. Moreover, there is a relatively high
degree of daily variability in the measure in the first half of the sample but then a severe decline in
variance after 2001. This pattern in variability closely mirrors the results in Back et al., who find
the same pattern in the inter-quartile range of Kyle’s lambda and price impact over the same period.
In addition, the evolution of the magnitude of our PCM proxy over time also matches closely that
of the price impact estimate. Both time series start around 15% in 1993, rise to 20% around 2000,
and subsequently drop below 5% following 2007. The close similarities in the time-series behavior
between PCM, Kyle’s lambda, and price impact provides further strong evidence to suggest that
PCM is an effective proxy for information asymmetry.

Overall, the collective results from the Fama-MacBeth regressions of bid-ask spreads on PCM,
the analysis of the behavior of PCM in event windows around earnings announcements, and the
time-series comparisons between PCM, Kyle’s lambda, and price impact establishes a consistent
and robust link between PCM and informed trading.

6 PCM and the cross-section of stock returns

Having established the link between PCM and informed trading, our next aim is to apply it to a larger question in the asset pricing literature: whether information risk is priced. The information asymmetry hypothesis predicts that expected stock returns should be positively related to informed trading because uninformed traders lose to informed traders and hence require compensation to hold stocks with greater information asymmetry (Easley and O'Hara, 2004).

6.1 PCM, illiquidity, firm size, and portfolio excess returns

6.1.1 Excess returns and three-factor alphas of portfolios sorted on PCM

As a first pass towards understanding the relationship between PCM and returns, we examine the behavior of portfolios sorted on PCM. Specifically, we sort stocks into 10 decile portfolios each month with the "1 (Low)" portfolio containing the 10% of stocks with the lowest PCM and the "10 (High)" portfolio containing the 10% of stocks with the highest PCM. We also form the "10-1" portfolio, which represents the long-short (self-funded, zero-investment) portfolio, with returns calculated by subtracting portfolio 1 returns from portfolio 10 returns. For each portfolio, we examine portfolio excess returns and evaluate the Fama-French three-factor (FF3) alpha to determine whether PCM risk contributes to returns in excess to the standard market risk (MKT), size (SMB), and growth (HML) factors.

Table 5 presents the sorting results. The mean (excess) return for each portfolio is reported in the first column, while the second column reports the FF3-factor alpha. Notably, the "10-1" long-short portfolio generates a large and statistically significant mean return 1.155% per month, and has a large and statistically significant alpha of .702% (corresponding to an 8.4% abnormal return per year). Thus, the results from these one-dimensional sorts suggest that stock returns do vary with information asymmetry as measured by the PCM proxy, and that PCM can be used to construct portfolios with abnormal risk-adjusted returns.

6.1.2 Excess returns and firm size from two-dimensional sorts on PCM and illiquidity

Given the Duarte and Young (2009) critique, a next step to further investigate the relationship between informed trading and returns must take into account the possible confounding effects of illiquidity. We thus follow the methodology in Fama and French (1992) and perform monthly, two-dimensional, sequential sorting of stocks into portfolios and calculate and compare their monthly excess returns. Specifically, for each month, stocks are sorted into five illiquidity groups from Low to High based on their Amihud (2002) illiquidity measure (ILLIQ) from the previous month, and then stocks in each of these ILLIQ groups are further sorted into five PCM groups from Low to High based on their average PCM in the previous month, resulting in 25 ILLIQ/PCM portfolios.
For each portfolio, average monthly excess return and firm size are calculated and these numbers are then averaged across the entire sample period.

The double-sorting results are reported in Table 6. Panel (a) shows that within each ILLIQ group, portfolio excess returns generally increase across the five Low to High PCM categories. However, for some of these groups, the increase is non-monotonic. Panel (b), which shows average firm size across portfolios, helps to explain why. For the lowest ILLIQ group, for which illiquidity is of the least concern (and firm size is the largest), high PCM is associated with larger firm size, which in turn is associated with lower returns. On the other hand, for the highest ILLIQ group, for which illiquidity is of the greatest concern (and firm size is the smallest), high PCM is associated with lower firm size, which in turn is associated with higher returns. For the three middle ILLIQ groups, we see that firm size fluctuates to a certain degree as we move across PCM categories.

Thus, although a positive relationship between PCM and returns is still observed after "controlling" for illiquidity in these two-dimensional sorts, we now see that firm size is also an important factor (as we should expect, of course). Thus, it appears that further regression analysis is warranted and necessary in order to control for possible confounding effects due to illiquidity, as well as firm size and other variables, when identifying the asset pricing implication of informed trading.

6.2 Asset pricing tests

6.2.1 Regression analysis methodology

The information asymmetry hypothesis predicts that expected stock returns should be positively related to the intensity of informed trading because uninformed traders lose to informed traders and hence require compensation to hold stocks with greater information asymmetry (Easley and O'Hara, 2004). We now test this hypothesis formally by conducting a series of monthly, firm-level Fama-MacBeth (1973) regressions to account for other variables that also influence stock returns. Our first baseline specification employs the three Fama-French (1992) firm characteristics – beta, firm size, and the book-to-market ratio – along with an illiquidity measure, as follows:

\[
\text{RET}_{i,t} = \alpha_t + \gamma_{1,t}PCM_{i,t-1} + \gamma_{2,t}ILLIQ_{i,t-1} + \gamma_{3,t}BETA_{i,t-1} + \gamma_{4,t}SIZE_{i,t-1} + \gamma_{5,t}BM_{i,t-1} + \varepsilon_{i,t} \]  

(5)

for stock \( i \) in month \( t \), where the dependent variable \( \text{RET} \) is the monthly percentage return of stock \( i \) in excess of the one-month T-bill rate; \( PCM \) is the measure of information asymmetry in the previous month; and \( ILLIQ \) is the monthly Amihud (2002) illiquidity measure, which is calculated as the average ratio of the absolute daily return to daily dollar volume over the previous month. \( BETA \) is the portfolio beta in the current year estimated from the full period using 40 post-ranking portfolios. For each month, \( SIZE \) is the logarithm of market value of equity at the end of the previous year, and \( BM \) is the logarithm of book equity divided by market value of equity at the end of the previous year.
Under the information asymmetry hypothesis, the coefficient of interest is the estimate of the parameter $\gamma_1$ in Equation (5). We should observe a positive and statistically significant coefficient if information asymmetry, as measured by PCM, indeed has a marginal effect on stock returns.

As noted above, for the purpose of identifying the effect of informed trading on the cross-section of stock returns, it is important to include an illiquidity measure in the asset pricing tests. Duarte and Young (2009) find that the effect of the PIN measure of Easley et al. (1996, 2002) on stock returns disappears once illiquidity is included in the regression. The fact that PIN is insignificant when controlling for illiquidity implies that PIN does not have any incremental explanatory power over illiquidity in explaining returns. As such, they conclude that PIN is priced, not because it captures information risk per se, but because it is a proxy for illiquidity effects that are unrelated to information. Moreover, and related to this, our discussion above on the mutually confounding effects between information asymmetry and illiquidity also underscores the necessity of controlling for the latter in order to properly identify the former. Thus, in light of both the Duarte and Young critique and the existence of such confounding effects, we include $\text{ILLIQ}$ in all of our asset-pricing tests.

6.2.2 Estimation results

The estimation results in Table 7 strongly confirm our preliminary findings from the portfolio sorts in Tables 5 and 6. That is, PCM is large and positive and highly significant in explaining the cross-section of stock returns after controlling for $\text{ILLIQ}$ and the three Fama-French firm characteristics, $\text{BETA}$, $\text{SIZE}$, and $\text{BM}$. In addition, $\text{ILLIQ}$ is positive and significant at the 1% level, confirming the strong effect of illiquidity on stock returns (Duarte and Young, 2009). Also consistent with previous findings in the literature, stock returns are significantly negatively related to $\text{SIZE}$ and positively related to $\text{BM}$, while $\text{BETA}$ is no different from zero, consistent with Fama and French (1992) and Easley et al. (2002). Most importantly, from Table 7 it is evident that since the result for PCM is obtained after controlling for $\text{ILLIQ}$ and other important factors such as $\text{SIZE}$, we are able to overcome the critique of Duarte and Young (2009) as well as other possible confounding effects implied in Table 6, and thereby can more confidently attribute the marginal effect of PCM on stock returns to informed trading, per se, rather than liquidity concerns or unobserved size effects unrelated to information.

6.3 PCM versus PIN

We next turn to a comparison between the two measures of information asymmetry, PCM and PIN. This allows us to assess the marginal explanatory power of each measure in explaining the cross section of returns, and whether there is any overlap in their information content (in the econometric sense). We run the comparison by adding PIN to Equation (5) and estimate the following Fama-MacBeth regression:
\[ RET_{i,t} = \alpha_t + \gamma_{1,t} PCM_{i,t-1} + \gamma_{2,t} PIN_{i,t-1} \\
+ \gamma_{3,t} ILLIQ_{i,t-1} + \gamma_{4,t} BETA_{i,t-1} + \gamma_{5,t} SIZE_{i,t-1} + \gamma_{6,t} BM_{i,t-1} + \varepsilon_{i,t} \]  

(6)

for stock \( i \) and month \( t \).\(^2\) The estimation results for various specifications based on the above regression are reported in Table 8. First, PIN is insignificant in our sample for Models (1) and (2) excluding PCM, with or without controlling for illiquidity. When PCM and PIN are included together in the regression in Models (3) and (4), the coefficient on PCM is positive and significant at the 1\% level irrespective of the illiquidity control, but PIN is again invariably insignificant. The magnitude of the estimates and \( t \)-statistics on PIN become drastically reduced in the presence of PCM when comparing across models. These results suggest that PCM has a stronger economically and statistically significant marginal effect in explaining the cross-section of stock returns than PIN, and that if any incremental information exists in PIN, it is largely subsumed in PCM. On the other hand, the information content in PCM is largely orthogonal to PIN. Thus, direct comparison between PCM and PIN suggests that the former is a more robust measure of information asymmetry.

### 6.4 Sub-sample analysis: before and after decimalization

So far, we have conducted our empirical analysis using data from January 1993 to December 2012. This time period encompasses tremendous technological innovation and regulatory changes in stock markets and trading processes. Chief among these, arguably, is the decimalization of stock prices that occurred in January 2001. Since then, new trading technologies and venues have advanced rapidly and trading activity has increased dramatically as a result. Some researchers argue that such technological innovation and regulatory changes may lead to better liquidity in the market (see, e.g., Angel et al., 2010; O’Hara, 2014). Indeed, Ben-Rephael et al. (2015) show that the effect of characteristic liquidity measures such as Amihud illiquidity have declined dramatically over the years, especially over the last decade. In light of these developments, we conduct a sub-sample analysis to examine whether and how the effect of information asymmetry, as measured by the PCM proxy, on stock returns may have been affected as a result.

Related to decimalization, the average trade size has also declined dramatically in recent years with many traders splitting their orders into a series of smaller sized trades. O’Hara (2014) reports that the average trade size of U.S. equity markets is now just over 200 shares. On the one hand, this might suggest that the effect of PCM, which uses a lower bound of 1,000 shares, may diminish or even disappear after decimalization. On the other hand, one may reasonably argue that much of the smaller sized trades are not motivated by information but rather by liquidity, such as those from high frequency trading (HFT). Moreover, Kyle and Obizhaeva (2016) posit that under market microstructure invariance the bet size of long-term traders must increase as the trading activity

\(^2\)The PIN data are from Brown and Hillegeist (2007), available for the period of 1993 to 2010 and downloaded from http://scholar.rhsmith.umd.edu/sbrown/pin-data?destination=node/998
accelerates. To the extent such long-term traders are motivated by information, then all else equal
the average informed trade size may plausibly grow with bet size for a similar order splitting strategy
(e.g., splitting a bet into 10 trades on average for each bet).

It is in the spirit of the latter consideration (i.e., increasing informed trade size) that we use 1000
shares as the lower bound for the medium-size of trades in our PCM measure, instead of the 500
share cutoff used in the earlier stealth trading literature (Barclay and Warner, 1993; Chakravarty,
2001). However, given the above discussion on changes in trading processes and technology, it
becomes necessary and interesting to not only to compare results between subperiods before and
after decimalization, but also to compare PCM as currently defined to an alternative version of
PCM defined using 500 shares as the lower bound for medium sized trades, which we denote by
PCM500.

We therefore re-perform the regression in Equation (6) using two subsamples to re‡ect the two
periods prior to and after decimalization, namely a subperiod from 1994-2000, and from 2001-2011.
In each sub-period, we run the regression using PCM and PCM500 separately as alternative proxies
for information asymmetry. We also include PIN in the subsample analysis to see whether its e‡ect
on returns may be time dependent, since we did not detect signi…cance in the full sample, and to
to continue the comparison between PCM in PIN within and across subperiods.

The results are reported in Table 9, with Panel (a) corresponding to the 1994-2000 subperiod, and
Panel (b) the 2001-2011 subperiod. Several observations are in order. First, and most importantly,
the e‡ect of PCM as a proxy for information asymmetry actually strengthens after the decimalization
in 2001. This is evident from the size and signi…cance of the coe¢ cient estimate on PCM (and
PCM500), which both increase dramatically in the later subperiod. Namely, the coe¢ cient on
PCM increases from 5.03 (p-value = .09) to 13.23 (p-value = .01). This result is in stark contrast
to the diminishing liquidity premium documented in Ben-Rephael et al. (2015). Thus, while recent
technological innovation (especially HFT) and regulatory changes may have lead to a diminishing
liquidity premium, there does not necessarily seem to be a diminishing information risk premium. If
anything, this …nding points in the opposite direction, i.e., that information risk remains substantial
and signi…cant.

Second, the size of the coe¢ cient of PCM is more than two times that of PCM500 in each sub-
period. For example, in the later sub-period, the coe¢ cient on PCM is 13.23 whereas the coe¢ cient
on PCM500 is 6.43. Furthermore, only PCM is statistically signi…cant in both subperiods, whereas
PCM500 is only signi…cant in the later subperiod. These results validate our use of 1000 shares,
instead of 500 shares, as the lower bound to de…ne medium-sized trades in constructing PCM.
They are also consistent with the larger bet size predicted by the market microstructure invariance
hypotheses of Kyle and Obizhaeva (2016).

Third, the effect of PIN is signi…cant at the 10% level in the …rst subperiod, and only with
PCM500 is included rather than PCM. This result is generally consistent with those in Duarte and
Young (2009) in that the effect of PIN appears marginally signi…cant in earlier years but seems to
have weakened in more recent data. Lastly, and quite interestingly, we find that the coe¢ cient on
ILLIQ is highly significant at the 1% level in the earlier subperiod, but becomes insignificant in the later subperiod. This result is consistent with the finding of a diminishing liquidity premium in Ben-Rephael et al. (2015). Overall, the above subsample analysis with alternative definitions of PCM indicates that our main findings from the full sample are robust and remain strong in the face of the important recent technological innovations and regulatory changes ushered in since the decimalization of stock prices in 2001.

6.5 Alternative explanations

Following Easley et al. (2002) we conduct a final robustness check on PCM by considering alternative variables that are known to be positively associated with returns and may have confounding effects on informed trading. Specifically, we expand the regression in Equation (5) to include the following variables: return variability, share turnover, which may proxy for liquidity or volume effects, as well as momentum. However, unlike Easley et al., we do not include the bid-ask spread in any of the regressions at this stage. As discussed above, market microstructure theory suggests that the spread consists of two main parts: one part is related to information asymmetry while the other is related to liquidity concerns unrelated to information (Biais, et al., 2005). Reflecting these two parts, in Section 5 we find that the spread is positively related to PCM, while others, e.g., Amihud and Mendelson (1986), find a positive relation between the spread and stock returns as evidence to support their liquidity hypothesis. Thus, having identified and measured both main components of the bid-ask spread, further adding spreads themselves into the regression again would introduce a severe multicollinearity problem. Here, our approach is similar to that in Duarte and Young (2009), who also do not include the spread when testing PIN in the presence of an illiquidity factor.

Specifically, stock return variability, STD, is the daily return standard deviation of stock i over the previous year. Stock turnover, TURN, and the coefficient of variation of turnover, CVTURN, are employed to control for volume effects: the former is calculated as the logarithm of the average monthly turnover divided by the number of shares outstanding over the previous three years, while the latter is the logarithm of the former’s coefficient of variation. Chordia et al. (2001) find that both turnover variables are negatively related to stock returns. Finally, while we have not explored possible linkages between informed trading and momentum effects, given that the latter is commonly used in the literature, we include an additional variable to control for momentum, RET27, which is the cumulative return of stock i over the previous five months.

Table 10 presents the results from incorporating these additional variables in our asset pricing tests. Notably, regardless of which of these additional variables is included, either in succession or altogether at once, the coefficients on PCM remain significant and positive at either the 1% or 5% levels. These corroborative results further strengthen the main findings above, indicating that PCM indeed captures a fundamental, priced variable in the form of information-asymmetry risk, rather than the effects of other omitted variables unrelated to information. Most importantly, since the marginal effect of PCM remains significant and positive after accounting for these alternative explanations, there is robust and consistent evidence to support the information asymmetry hypothesis.
(Easley and O’Hara, 2004) that information risk is priced and informed trading is an important determinant of expected stock returns.

7 Conclusion

The relationship between stock returns and information is a central issue in both the asset pricing and market microstructure literature. The information asymmetry hypothesis posits that information risk should be priced because uninformed traders lose to informed traders and hence require compensation to hold stocks with greater information asymmetry (Easley and O’Hara, 2004). The testing of this hypothesis, however, is often marred by some confounding effect inherent in the proxy used for informed trading. For example, Duarte and Young (2009) find that the PIN of Easley et al. (2002) is priced, not because it captures informed trading, but because it serves as a proxy for the illiquidity effect that is unrelated to information. In light of this, the aim of this paper is twofold: (1) to develop a new proxy for the probability of informed trading that is motivated by relevant theories and (2) to employ this proxy to test the information asymmetry hypothesis while controlling for the important illiquidity effect.

We develop a new proxy for informed trading, called PCM, based on the concept of contrarian trades (Campbell et al., 1993; Avramov et al., 2006; Back et al., 2016) and trade size choice, which is motivated by the stealth trading literature (Barclay and Warner, 1993; Chakravarty, 2001), along with the market microstructure invariance hypotheses (Kyle and Obizhaeva, 2016). PCM, then, in essence is conceived as the probability of observing medium-sized contrarian trades.

We establish the link between PCM and informed trading in several ways. First, we examine whether PCM is positively related the bid-ask spread, which is known to have an adverse selection component and an illiquidity component. We find that PCM is significant and positive in explaining the spreads even after controlling for Amihud (2002) illiquidity and other variables of interest. Thus, we can confidently attribute the marginal effect of PCM on the bid-ask spreads to private information trading (adverse selection component), rather than illiquidity unrelated to information (illiquidity component). In addition, we analyze the behavior of PCM in an event window surrounding earnings announcements, which are information rich events. Consistent with Kim and Verrecchia (1994, 1997), PCM begins rising the day before the earnings announcement day, peaks on the announcement day, and declines but remains significantly above normal for roughly three days following an announcement. This result strongly suggests that PCM directly measures information asymmetry during these periods. There is also a striking similarity in time-series patterns between PCM and both the Kyle’s lambda and the price impact estimated in Back, et al. (2016). All three time series rise over 1990s and drop sharply following the turn of the century, with a brief upward movement during the recent financial crisis. The collective evidence from the cross-sectional regression analysis of bid-ask spreads, the analysis of an event window around earnings announcements, and the time-series comparison with Kyle’s lambda and price impact establishes a consistent and robust link between PCM and information asymmetry.
As a preliminary check of the information asymmetry hypothesis, we first conduct a series of portfolio sorting exercises. Sorting stocks into decile portfolios based on PCM, the self-funded, long-short "10-1" portfolio exhibits large positive excess returns as well as a statistically and economically significant alpha of 0.702% (corresponding to an 8.4% abnormal return per year). Two-dimensional sorting on PCM and illiquidity indicates that excess returns generally increase as we move from low to high PCM portfolios.

We then test the information asymmetry hypothesis more formally by conducting a series of Fama-MacBeth (1973) regressions of stock returns on PCM and other variables that may also influence stock returns. We find that PCM is positive and significant in explaining stock returns after controlling for Amihud illiquidity and the three Fama-French firm characteristics, beta, firm size, and the book-to-market ratio. Since the effect of PCM on returns is obtained after controlling for illiquidity, we are able to overcome the critique of Duarte and Young (2009) and thereby attribute the marginal effect of PCM on returns to informed trading, rather than liquidity concerns unrelated to information. We also run a direct comparison between PCM and PIN. The estimation results show that the coefficient of PCM is always positive and significant, but the corresponding coefficient of PIN is invariably insignificant. This direct comparison suggests that PCM is a more robust and consistent measure of information asymmetry since PCM survives the critique of Duarte and Young (2009) in explaining stock returns, while PIN does not.

To account for the potential effect of decimalization in 2001, we split the sample into pre- and post-2011 sub-periods. The effect of PCM actually strengthens after the decimalization in 2001. This suggests that while recent technological innovations (especially HFT) and regulatory changes may have lead to a diminishing liquidity premiums, as documented in Ben-Rephael et al. (2015), they have not necessarily lead to a diminishing information risk premium. In fact, we find that information risk remains substantial and significant after decimalization.

Finally, we examine alternative explanations to information asymmetry by adding additional explanatory variables used in the literature. The estimation results show that the coefficient on PCM remains significant and positive after controlling for these additional variables. The collective evidence in this paper therefore renders robust and strong support to the information asymmetry hypothesis (Easley and O'Hara, 2004) that information risk is priced and informed trading is an important determinant of expected stock returns.
References


Table 1: Distribution of the PCM proxy for the probability of informed trading across firm-months and firms

(a) Monthly firm PCM

$NT_{Months} = 244,900$ firm-month obs.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>St. Dev</th>
<th>Min</th>
<th>25th pct.</th>
<th>75th pct.</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.074</td>
<td>0.046</td>
<td>0.072</td>
<td>0.000</td>
<td>0.009</td>
<td>0.131</td>
<td>0.553</td>
</tr>
</tbody>
</table>

(b) Firm PCM

$N = 2,783$ firm obs.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>St. Dev</th>
<th>Min</th>
<th>25th pct.</th>
<th>75th pct.</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.085</td>
<td>0.072</td>
<td>0.068</td>
<td>0.000</td>
<td>0.028</td>
<td>0.129</td>
<td>0.504</td>
</tr>
</tbody>
</table>

Notes: The table above contains means, medians, standard deviations, 25th percentiles and 75th percentiles, and minimums and maximums of the PCM proxy for the probability of informed trading computed (a) across all stocks and months (monthly firm PCM), and (b) across all firms (firm-level PCM). The PCM proxy for the probability of informed trading is defined in Eqs. (3).
Table 2: Correlation matrix for firm-level PCM proxy for the probability of informed trading and firm/stock characteristics

<table>
<thead>
<tr>
<th></th>
<th>Size</th>
<th>Volume</th>
<th>Illiquidity</th>
<th>Spread</th>
<th>Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCM</td>
<td>-0.192</td>
<td>-0.006</td>
<td>0.022</td>
<td>0.572</td>
<td>0.179</td>
</tr>
<tr>
<td></td>
<td>(&lt;.0001)</td>
<td>(0.7640)</td>
<td>(0.2471)</td>
<td>(&lt;.0001)</td>
<td>(&lt;.0001)</td>
</tr>
<tr>
<td>Size</td>
<td>0.426</td>
<td>-0.323</td>
<td>-0.754</td>
<td>-0.311</td>
<td>-0.311</td>
</tr>
<tr>
<td></td>
<td>(&lt;.0001)</td>
<td>(&lt;.0001)</td>
<td>(&lt;.0001)</td>
<td>(&lt;.0001)</td>
<td>(&lt;.0001)</td>
</tr>
<tr>
<td>Volume</td>
<td>-0.115</td>
<td>-0.223</td>
<td>-0.223</td>
<td>-0.64</td>
<td>-0.64</td>
</tr>
<tr>
<td></td>
<td>(&lt;.0001)</td>
<td>(&lt;.0001)</td>
<td>(&lt;.0001)</td>
<td>(0.0007)</td>
<td>(0.0007)</td>
</tr>
<tr>
<td>Illiquidity</td>
<td>0.315</td>
<td>0.208</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(&lt;.0001)</td>
<td></td>
<td>(&lt;.0001)</td>
<td></td>
<td>(&lt;.0001)</td>
</tr>
<tr>
<td>Spread</td>
<td></td>
<td></td>
<td></td>
<td>0.351</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(&lt;.0001)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table above reports the correlation coefficients between the firm-level PCM proxy for the probability of informed trading and other firm and stock-level characteristics related to adverse selection (p-values in parentheses). Firm size is the natural logarithm of market capitalization. Volume is the average daily turnover for the stock. Illiquidity is measured as in Amihud (2002) by calculating the average monthly ratio of absolute daily returns to the daily dollar volume of a stock. The spread is the average daily bid-ask spread expressed as a percentage of the bid quote. Rest as in Table 1.
Table 3: Summary statistics for firm and stock characteristics across all stocks and months used in adverse selection and asset pricing tests

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>RET</td>
<td>1.130</td>
<td>0.762</td>
<td>-98.30</td>
<td>640.4</td>
</tr>
<tr>
<td>BETA</td>
<td>1.051</td>
<td>1.016</td>
<td>0.221</td>
<td>2.678</td>
</tr>
<tr>
<td>BM</td>
<td>-0.801</td>
<td>-0.740</td>
<td>-7.784</td>
<td>6.380</td>
</tr>
<tr>
<td>ILLIQ</td>
<td>0.663</td>
<td>0.112</td>
<td>0.000</td>
<td>2246.9</td>
</tr>
<tr>
<td>SPREAD</td>
<td>1.241</td>
<td>0.834</td>
<td>0.020</td>
<td>9.531</td>
</tr>
<tr>
<td>STD</td>
<td>2.635</td>
<td>2.277</td>
<td>0.079</td>
<td>25.698</td>
</tr>
<tr>
<td>TURN</td>
<td>-0.142</td>
<td>-0.113</td>
<td>-5.659</td>
<td>4.338</td>
</tr>
<tr>
<td>CVTURN</td>
<td>3.806</td>
<td>3.763</td>
<td>2.525</td>
<td>5.942</td>
</tr>
</tbody>
</table>

Notes: This table reports means, medians, minimum value, and maximum value for the variables included in the asset pricing tests. All statistics are calculated across all stocks and months in the sample. RET is the monthly percentage return of stock \( i \) in excess of the one-month T-bill rate. BETA is the portfolio beta estimated from the full period using 40 post-ranking portfolios. SIZE is the logarithm of the year-end market value of equity. BM is the logarithm of book equity divided by market value of equity at the end of the previous year. ILLIQ is the Amihud (2002) illiquidity measure, calculated as the average ratio of the absolute daily return to daily dollar volume over the month. SPREAD is the average monthly bid-ask spread of stock \( i \). STD is the daily return standard deviation of stock \( i \) over a year. TURN is the logarithm of the average monthly turnover divided by the number of shares outstanding over the previous three years, and CVTURN, is the logarithm of the coefficient of variation of turnover.
Table 4: Analysis of the PCM proxy for the probability of informed trading and the adverse selection component of the bid-ask spread

<table>
<thead>
<tr>
<th>Dependent variable: $SPREAD$</th>
<th>PCM</th>
<th>$ILLIQ$</th>
<th>$SIZE$</th>
<th>$TURN$</th>
<th>$CVTURN$</th>
<th>$RET_{-1}$</th>
<th>$STD$</th>
<th>$SPREAD_{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.359</td>
<td>0.009</td>
<td>-0.026</td>
<td>-0.020</td>
<td>-0.005</td>
<td>-0.003</td>
<td>0.020</td>
<td>0.895</td>
</tr>
<tr>
<td></td>
<td>(5.42)</td>
<td>(3.58)</td>
<td>(-18.03)</td>
<td>(-11.68)</td>
<td>(-2.27)</td>
<td>(-27.08)</td>
<td>(9.08)</td>
<td>(140.74)</td>
</tr>
</tbody>
</table>

Notes: This table contains the results from monthly Fama-MacBeth regressions ($T = 238$ months) using the average PCM proxy for the probability of informed trading over the previous month, as explanatory variable. The dependent variable is $SPREAD$, which is the average bid-ask spread of stock $i$ in the current month. $ILLIQ$ is the Amihud (2002) illiquidity measure, which is calculated as the average ratio of the absolute daily return to daily dollar volume over the previous month. $SIZE$ is the logarithm of market value of equity at the end of the previous year. $TURN$ is the logarithm of the average monthly turnover divided by the number of shares outstanding from over the previous three years, and $CVTURN$, is the logarithm of the coefficient of variation of turnover. $RET_{-1}$ is the percentage return of stock $i$ in excess of the one-month T-bill rate over the previous month. $STD$ is the daily return standard deviation of stock $i$ over the previous year. $SPREAD_{-1}$ is the average bid-ask spread of stock $i$ in the previous month.
Table 5: Decile portfolios sorted on the PCM proxy for the probability of informed trading

<table>
<thead>
<tr>
<th></th>
<th>Mean return</th>
<th>FF3-Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Low)</td>
<td>0.776***</td>
<td>0.097</td>
</tr>
<tr>
<td></td>
<td>(2.60)</td>
<td>(0.79)</td>
</tr>
<tr>
<td>2</td>
<td>0.686**</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(2.25)</td>
<td>(-0.01)</td>
</tr>
<tr>
<td>3</td>
<td>0.829***</td>
<td>0.139</td>
</tr>
<tr>
<td></td>
<td>(2.63)</td>
<td>(1.11)</td>
</tr>
<tr>
<td>4</td>
<td>1.035***</td>
<td>0.314***</td>
</tr>
<tr>
<td></td>
<td>(3.20)</td>
<td>(2.80)</td>
</tr>
<tr>
<td>5</td>
<td>0.938***</td>
<td>0.184</td>
</tr>
<tr>
<td></td>
<td>(2.79)</td>
<td>(1.59)</td>
</tr>
<tr>
<td>6</td>
<td>0.865**</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>(2.45)</td>
<td>(0.64)</td>
</tr>
<tr>
<td>7</td>
<td>0.941**</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>(2.57)</td>
<td>(0.95)</td>
</tr>
<tr>
<td>8</td>
<td>1.075***</td>
<td>0.198</td>
</tr>
<tr>
<td></td>
<td>(2.75)</td>
<td>(1.33)</td>
</tr>
<tr>
<td>9</td>
<td>1.283***</td>
<td>0.251</td>
</tr>
<tr>
<td></td>
<td>(2.70)</td>
<td>(1.22)</td>
</tr>
<tr>
<td>10 (High)</td>
<td>1.931***</td>
<td>0.799**</td>
</tr>
<tr>
<td></td>
<td>(3.39)</td>
<td>(2.54)</td>
</tr>
<tr>
<td>10-1</td>
<td>1.155***</td>
<td>0.702**</td>
</tr>
<tr>
<td></td>
<td>(2.87)</td>
<td>(2.04)</td>
</tr>
</tbody>
</table>

Notes: This table reports the results from sorting stocks each month into 10 decile portfolios based on the PCM proxy for the probability of informed trading, with the "1 (Low)" portfolio containing the 10% of stocks with the lowest probabilities of informed trading and the "10 (High)" portfolio containing the 10% of stocks with the highest probabilities of informed trading according to PCM. There are a total of 238 months and each portfolio contains, on average, 90 to 94 stocks. Mean (excess) returns for each decile portfolio are reported in the first column, while the second column reports the Fama-French three-factor alpha (FF3-Alpha), which is the intercept coefficient from the regression of portfolio returns on the market risk-premium factor (MKT), value-growth factor (HML), and size factor (SMB). The portfolio "10-1" represents the long-short (self-funded, zero-investment) portfolio, with returns calculated by subtracting portfolio 1 returns from portfolio 10 returns. t-statistics reported in parentheses. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively.
Table 6: Average returns on portfolios sorted monthly on illiquidity and the PCM proxy for the probability of informed trading

(a) Excess returns

<table>
<thead>
<tr>
<th>ILLIQ/PCM</th>
<th>Low</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.694</td>
<td>0.865</td>
<td>0.843</td>
<td>0.941</td>
<td>0.871</td>
</tr>
<tr>
<td>2</td>
<td>0.795</td>
<td>1.087</td>
<td>0.777</td>
<td>0.840</td>
<td>0.933</td>
</tr>
<tr>
<td>3</td>
<td>0.568</td>
<td>0.885</td>
<td>0.713</td>
<td>1.039</td>
<td>1.333</td>
</tr>
<tr>
<td>4</td>
<td>0.815</td>
<td>0.716</td>
<td>0.856</td>
<td>0.957</td>
<td>1.896</td>
</tr>
<tr>
<td>High</td>
<td>0.838</td>
<td>0.816</td>
<td>1.058</td>
<td>1.806</td>
<td>2.993</td>
</tr>
</tbody>
</table>

(b) Firm size

<table>
<thead>
<tr>
<th>ILLIQ/PCM</th>
<th>Low</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>18,344.4</td>
<td>21,055.9</td>
<td>25,869.0</td>
<td>31,875.3</td>
<td>27,790.0</td>
</tr>
<tr>
<td>2</td>
<td>6,611.3</td>
<td>8,141.1</td>
<td>8,817.4</td>
<td>8,713.1</td>
<td>6,102.4</td>
</tr>
<tr>
<td>3</td>
<td>3,733.7</td>
<td>4,160.3</td>
<td>4,161.6</td>
<td>3,873.9</td>
<td>2,855.8</td>
</tr>
<tr>
<td>4</td>
<td>2,218.0</td>
<td>2,329.4</td>
<td>2,320.1</td>
<td>2,103.6</td>
<td>1,251.2</td>
</tr>
<tr>
<td>High</td>
<td>1,246.5</td>
<td>1,226.2</td>
<td>1,171.9</td>
<td>937.5</td>
<td>502.3</td>
</tr>
</tbody>
</table>

Notes: This table reports results on average excess returns and firm size from sorting stocks monthly into portfolios by the Amihud (2002) illiquidity measure, ILLIQ, which is calculated as the average ratio of the absolute daily return to daily dollar volume over the previous month, and then by the probability of informed trading as measured by PCM. For each month, stocks are sorted into five (quintile) illiquidity groups, and then each of these illiquidity groups is further sorted into five (quintile) informed-trading groups (consistent with the methodology in Fama and French (1992)), resulting in 25 portfolios for which average monthly excess return (in percent) and firm size (in $ millions) are calculated. For each portfolio, monthly characteristics are then averaged across the 1993-2008 sample period (238 months), with the results reported here. The average number of stocks in each portfolio ranges from 40 to 42 over the sample period.
Table 7: Asset pricing tests with PCM

<table>
<thead>
<tr>
<th>Model</th>
<th>PCM</th>
<th>ILLIQ</th>
<th>BETA</th>
<th>SIZE</th>
<th>BM</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.306</td>
<td>0.042</td>
<td>-0.215</td>
<td>0.188</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.46)</td>
<td>(1.25)</td>
<td>(-3.26)</td>
<td>(1.83)</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>6.456</td>
<td>0.369</td>
<td>0.259</td>
<td>-0.206</td>
<td>0.185</td>
</tr>
<tr>
<td></td>
<td>(2.39)</td>
<td>(4.24)</td>
<td>(0.79)</td>
<td>(-3.25)</td>
<td>(1.82)</td>
</tr>
</tbody>
</table>

Notes: This table reports time-series averages of the coefficients in cross-sectional asset pricing tests using the Fama-MacBeth (1973) procedure ($T = 238$ months), with $t$-statistics reported in parentheses. The dependent variable is $RET$, which is calculated as the monthly percentage return of stock $i$ in excess of the one-month T-bill rate. PCM is the proxy for the probability of informed trading defined in Eq. (3). $ILLIQ$ is the monthly Amihud (2002) illiquidity measure, which is calculated as the average ratio of the absolute daily return to daily dollar volume over the previous month. $BETA$ is the portfolio beta estimated from the full period using 40 post-ranking portfolios. $SIZE$ is the logarithm of market value of equity at the end of the previous year. $BM$ is the logarithm of book equity divided by market value of equity at the end of the previous year.
Table 8: Comparison of PCM and PIN in asset pricing tests

<table>
<thead>
<tr>
<th>Model</th>
<th>PCM</th>
<th>PIN</th>
<th>ILLIQ</th>
<th>BETA</th>
<th>SIZE</th>
<th>BM</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>2.407</td>
<td>0.453</td>
<td>-0.167</td>
<td>0.184</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.55)</td>
<td>(1.28)</td>
<td>(-2.16)</td>
<td>(1.74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>2.161</td>
<td>0.307</td>
<td>0.451</td>
<td>-0.144</td>
<td>0.194</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.39)</td>
<td>(3.09)</td>
<td>(1.29)</td>
<td>(-1.83)</td>
<td>(1.84)</td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>10.083</td>
<td>0.770</td>
<td>0.265</td>
<td>-0.166</td>
<td>0.176</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.04)</td>
<td>(0.48)</td>
<td>(0.80)</td>
<td>(-2.13)</td>
<td>(1.70)</td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>9.512</td>
<td>0.730</td>
<td>0.403</td>
<td>0.260</td>
<td>-0.170</td>
<td>0.173</td>
</tr>
<tr>
<td></td>
<td>(3.02)</td>
<td>(4.13)</td>
<td>(0.78)</td>
<td>(-2.16)</td>
<td>(1.66)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table reports time-series averages of the coefficients in cross-sectional asset pricing tests using the Fama-MacBeth (1973) procedure, with $t$-statistics reported in parentheses. The dependent variable is $RET$, which is calculated as the monthly percentage return of stock $i$ in excess of the one-month T-bill rate. PCM is the proxy for the probability of informed trading defined in Eq. (3). $ILLIQ$ is the monthly Amihud (2002) illiquidity measure, which is calculated as the average ratio of the absolute daily return to daily dollar volume over the previous month. $BETA$ is the portfolio beta estimated from the full period using 40 post-ranking portfolios. $SIZE$ is the logarithm of market value of equity at the end of the previous year. $BM$ is the logarithm of book equity divided by market value of equity at the end of the previous year. PIN is an alternative measure of the probability of information based trading in the previous year (PIN data are available from 1993 to 2010, resulting in 216 months used in the analysis).
Table 9: Comparison of PCM and PCM500 in split-sample asset pricing tests before and after decimalization

<table>
<thead>
<tr>
<th>Model</th>
<th>PCM</th>
<th>PCM500</th>
<th>PIN</th>
<th>ILLIQ</th>
<th>BETA</th>
<th>SIZE</th>
<th>BM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) Subperiod: 1994-2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>5.031</td>
<td>2.891</td>
<td>0.828</td>
<td>0.332</td>
<td>0.038</td>
<td>0.161</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.68)</td>
<td>(1.26)</td>
<td>(4.20)</td>
<td>(0.69)</td>
<td>(0.29)</td>
<td>(0.85)</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>2.510</td>
<td>2.582</td>
<td>0.130</td>
<td>0.451</td>
<td>-0.144</td>
<td>0.194</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.30)</td>
<td>(1.77)</td>
<td>(3.01)</td>
<td>(0.24)</td>
<td>(1.04)</td>
<td>(0.54)</td>
<td></td>
</tr>
<tr>
<td>(b) Subperiod: 2001-2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>13.231</td>
<td>-1.138</td>
<td>0.130</td>
<td>0.142</td>
<td>-0.256</td>
<td>0.139</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.55)</td>
<td>(-0.53)</td>
<td>(1.47)</td>
<td>(0.34)</td>
<td>(-2.75)</td>
<td>(1.21)</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>6.426</td>
<td>-1.638</td>
<td>-0.052</td>
<td>0.167</td>
<td>-0.204</td>
<td>0.174</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.09)</td>
<td>(-0.81)</td>
<td>(-1.18)</td>
<td>(0.41)</td>
<td>(-2.41)</td>
<td>(1.51)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table reports time-series averages of the coefficients in cross-sectional asset pricing tests using the Fama-MacBeth (1973) procedure on split samples based on subperiods before and after the introduction of decimalization in 2001, with \( t \)-statistics reported in parentheses. The first subperiod from 1994-2000 uses 84 months in the analysis, while the latter subperiod from 2001 to 2011 uses 130 months in the analysis. The dependent variable is \( RET \), which is calculated as the monthly percentage return of stock \( i \) in excess of the one-month T-bill rate. PCM is the proxy for the probability of informed trading defined in Eq. (3), while PCM500 modifies the original PCM proxy by defining medium trade size as 500 to 9,999 shares (as in Chakravarty, 2001). PIN is an alternative measure of the probability of information based trading in the previous year (PIN data are available from 1993 to 2010). The rest of the variables are as defined in Table 8.
### Table 10: Asset pricing tests using PCM with alternative explanatory variables

<table>
<thead>
<tr>
<th>Model</th>
<th>PCM</th>
<th>ILLIQ</th>
<th>BETA</th>
<th>SIZE</th>
<th>BM</th>
<th>STD</th>
<th>TURN</th>
<th>CVTURN</th>
<th>RET27</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.323</td>
<td>0.234</td>
<td>-0.178</td>
<td>0.183</td>
<td>0.119</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.72)</td>
<td>(0.86)</td>
<td>(-3.44)</td>
<td>(1.95)</td>
<td>(1.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>6.436</td>
<td>0.358</td>
<td>0.220</td>
<td>-0.214</td>
<td>0.174</td>
<td>-0.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.55)</td>
<td>(4.09)</td>
<td>(0.82)</td>
<td>(-4.02)</td>
<td>(1.88)</td>
<td>(-0.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>0.281</td>
<td>0.541</td>
<td>-0.250</td>
<td>0.167</td>
<td>-0.174</td>
<td>-0.260</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.30)</td>
<td>(1.73)</td>
<td>(-4.20)</td>
<td>(1.68)</td>
<td>(-1.96)</td>
<td>(-1.98)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>6.909</td>
<td>0.337</td>
<td>0.401</td>
<td>-0.253</td>
<td>0.158</td>
<td>-0.212</td>
<td>-0.326</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.63)</td>
<td>(3.93)</td>
<td>(-4.34)</td>
<td>(1.60)</td>
<td>(-2.52)</td>
<td>(-2.63)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td>0.142</td>
<td>0.398</td>
<td>-0.145</td>
<td>0.181</td>
<td>0.040</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.26)</td>
<td>(1.31)</td>
<td>(-2.21)</td>
<td>(1.91)</td>
<td>(0.10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6)</td>
<td>5.222</td>
<td>0.195</td>
<td>0.276</td>
<td>-0.142</td>
<td>0.173</td>
<td>0.038</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.97)</td>
<td>(1.77)</td>
<td>(0.95)</td>
<td>(-2.23)</td>
<td>(1.86)</td>
<td>(0.10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7)</td>
<td>0.116</td>
<td>0.325</td>
<td>-0.155</td>
<td>0.145</td>
<td>-0.167</td>
<td>-0.383</td>
<td>0.027</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.09)</td>
<td>(1.37)</td>
<td>(-3.01)</td>
<td>(1.66)</td>
<td>(-2.17)</td>
<td>(-3.33)</td>
<td>(0.07)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8)</td>
<td>5.155</td>
<td>0.153</td>
<td>0.323</td>
<td>-0.184</td>
<td>0.137</td>
<td>-0.152</td>
<td>-0.352</td>
<td>0.040</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.02)</td>
<td>(1.40)</td>
<td>(1.39)</td>
<td>(-3.45)</td>
<td>(1.59)</td>
<td>(-2.02)</td>
<td>(-3.12)</td>
<td>(0.11)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table reports time-series averages of the coefficients in cross-sectional asset pricing tests using the Fama-MacBeth (1973) procedure ($T = 238$ months), with t-statistics reported in parentheses. The dependent variable is $RET$, which is calculated as the monthly percentage return of stock $i$ in excess of the one-month T-bill rate. PCM is the proxy for the probability of informed trading defined in Eq. (3). $SPREAD$ is the average bid-ask spread of stock $i$ in the previous month. $STD$ is the daily return standard deviation of stock $i$ over the previous year. $TURN$ is the logarithm of the average monthly turnover divided by the number of shares outstanding over the previous three years, and $CVTURN$ is the logarithm of the coefficient of variation of turnover. $RET27$ is the gross cumulative compound return from seven to two months before the current month (total of 233 months when $RET27$ is included). The rest of the variables are as defined in Table 7.
Figure 1: PCM proxy for the probability of informed trading around earnings announcement days

Notes: This graph shows the average ratio of the PCM proxy for the probability of informed trading to its mean value across all stocks in the sample from 1993 to 2012. Values are plotted in event time in a [-10,+10] day window around earnings announcement days (Day 0). Dotted lines represent 95% confidence bands.
Figure 2: Time series plot of daily PCM, 1993-2012

Notes: This graph shows the average daily value of PCM across all stocks from 1993 to 2012.