Journal of Empirical Finance 29 (2014) 80–94

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

A dynamic intraday measure of the probability of informed trading and firm-specific return variation

Sanders S. Chang [a,*], Lenisa V. Chang [b], F. Albert Wang [a]

[a] Department of Economics and Finance, University of Dayton, USA
[b] Department of Economics, University of Cincinnati, USA

ARTICLE INFO

Article history:
Received 1 August 2013
Received in revised form 19 February 2014
Accepted 21 February 2014
Available online 28 February 2014

JEL classification:
G10
G14
G19

Keywords:
Informed trading
Private information
Price discovery
High-frequency
Firm-specific return variation
Price non-synchronicity

ABSTRACT

A central question in financial economics is how private information is incorporated into asset prices. A common method of measuring private information is the PIN measure, which uses statistical estimation of a sequential trade model of the trading process to estimate the probability of informed trading. A notable limiting feature of PIN is that one must aggregate very fine intraday data over very long macro horizons in order to estimate it. In this paper, our aim is to develop and implement a dynamic intraday measure of the probability of informed trading that circumvents this aggregation issue and allows for the measurement of information based trading activity at much higher frequencies. We then apply our dynamic intraday measure of the probability of informed trading to examine the relationship between private information and firm-specific return variation.

1. Introduction

A central question in financial economics is related to the role of information in markets and the process by which information is incorporated into asset prices. The market microstructure literature identifies two broad investor types: informed traders, who buy and sell assets based on information they possess regarding the asset's true future intrinsic value, and liquidity (or noise) traders, who trade for purposes unrelated to information such as meeting liquidity needs. Empirically, however, while detailed transaction level data are available at the intraday level (i.e., trade-by-trade basis), these data do not indicate whether a particular trade is initiated by an individual who is informed or not. To make this determination, researchers must commonly infer from the data whether trades are more likely information or liquidity based.

One of the most common and widely accepted methods of doing this is the PIN measure of Easley et al. (1997a,b) and Easley et al. (2002), which estimates the probability of informed trading based on a sequential trade model drawn from Glosten and Milgrom (1985) and Easley and O'Hara (1987). With a measure of information based trading in hand, researches have extensively applied PIN to study the effect of informed trading in a broad range of areas in finance, including stock price informativeness, corporate governance and investment decisions, stock market volatility, and insider trading, just to name a few.

However, the traditional PIN measure has some well known limiting features. Most notably, in order to estimate it one must aggregate very fine intraday data, which occur at approximately five-minute intervals within the trading day, across multiple days

* Corresponding author at: Department of Economics and Finance, University of Dayton, 300 College Park, Dayton, OH, 45469-2251, USA.
E-mail addresses: schang1@udayton.edu (S.S. Chang), lenisa.chang@uc.edu (L.V. Chang), fwang1@udayton.edu (F.A. Wang).

http://dx.doi.org/10.1016/j.jempfin.2014.02.003
0927-5398/

(Easley et al., 1997a,b). The resulting estimate measures informed trading over a very long macro horizon — typically from one month to one year. Arguably, the variation and information content of intraday trades is diluted, or possibly even lost, when combining over such long time periods, especially in modern financial markets where information is short-lived and traders act with increasing alacrity. Indeed, with regard to the necessity of using many days in their maximum likelihood procedure, Easley et al. (1997a) concede the tradeoff between estimation accuracy and economic reasonableness: "[w]hat is also apparent, however, is that while it may be sensible to use large sample methods to estimate [certain parameters in the PIN model], it is less so for [other parameters in the model]. The presumed stationarity of information is unlikely to be true over a long sample period, dictating a natural limit to the number of days we can sensibly employ." As such, horizons of multiple months, or up to one year in the case of Easley et al. (2002), might seem to push the envelope of such sensibility.

In addition, over such long macro horizons it is likely that the actual impact of short-lived private information may become diluted or masked by other factors. For example, Duarte and Young (2009) argue and find that PIN can be decomposed into two further components: one that represents private information, as originally postulated, but another component that is a proxy for market illiquidity (i.e., disruptions in the supply and demand conditions in stock markets) that is unrelated to information. Since market illiquidity is certainly a more long-lived phenomenon than private information, this is suggestive that Duarte and Young's finding may be an artifact of the aggregation process.

The first aim of our paper is to develop and implement a dynamic intraday version of PIN, which we refer to as DPIN, that circumvents the aggregation issue described above and allows us to estimate the probability of informed trading at much finer frequencies — specifically, at 15-minute intervals throughout the trading day. Since such frequencies are more in line with the speed at which traders react to and digest information in modern financial markets, our dynamic DPIN measure may be better suited to more accurately capture information based trading activity at higher frequencies, even within the trading day.

Our method for constructing DPIN comes from an extension of the approach of Avramov et al. (2006), which is based on the trading model of Campbell et al. (1993) and used to study the effect of informed *selling* on daily stock price volatility. The contribution of our paper in this respect is to recognize that this approach can be further employed to derive a dynamic intraday measure of the probability of informed trading, essentially by calculating the proportion of trades that are classified as information based over a given time interval. The DPIN measure has the advantage of being dynamic and flexible — it can be aggregated over many intervals to make comparisons with existing macro-horizon models, yet it is also capable of capturing cross-sectional and time-series variation in the probability of informed trading at much higher daily and intraday frequencies. Another advantage is that the DPIN measure does not require any form of numerical optimization for its estimation and thus is relatively simple and quick to compute from the transactions data, thus providing a more straightforward and less time consuming alternative method for dealing with tremendously large datasets involving intraday transactions.

Upon specifying several versions of the DPIN measure, we find numerical estimates that are generally consistent with existing measures for the probability of informed trading. When aggregated to similarly long horizons as the PIN in Easley et al. (2002), several of our DPIN measures are remarkably close in terms of location, spread, and skewness when combining across firms and years. Another version of our DPIN measure turns out to be quite close numerically to that of Nyholm (2002), who also attempts to measure the probability of informed trading at the transaction level. We take these results as evidence that our proposed DPIN measures are not too far off the mark and conform to the range of previous estimates found in the literature. However, a distinguishing feature of our analysis is that we are also able to disaggregate the time horizon, allowing an examination of the intraday pattern of informed trading, as well as how this pattern has evolved over time. As we will discuss, such results provide new empirical evidence for existing microstructure theories on the intraday behavior of information based trading activity.

With a new dynamic intraday measure of the probability of informed trading in hand, the second aim of the paper is to apply this measure to study an open question in finance regarding the role of private information and a particular aspect of asset returns, namely firm-specific return variation, or price non-synchronicity. Roll (1988) finds that firm-specific stock price movements are generally not associated with identifiable news releases and thus surmised that private information might play a crucial role in explaining such movements. Ever since Roll's conjecture, researchers have taken (for granted) the notion that firm specific-return variation might be caused by private information. For example, Morck et al. (2000), Durnev et al. (2003, 2004), and Chen et al. (2007) use firm-specific return variation as a proxy for stock price informativeness to study an array of issues at the intersection of private information, price formation, corporate earnings forecasts, capital budgeting, and investment policy. Tellingly, each of these papers acknowledges that their respective analyses rest upon the validity of the notion that price non-synchronicity is indeed caused by private information, and ultimately only provide indirect and circumstantial evidence of their possible link. Indeed, Durnev et al. (2004) state further the caveat that the "conceptual arguments and empirical studies [cited above] constitute a *subtle case* [emphasis added] for accepting firm-specific return variation as a proxy for stock price informativeness…(p.66)."

While many of these indirect arguments are conceptually appealing and point convincingly to a relationship between private information and price non-synchronicity, very few studies have directly investigated the relationship between the two variables, especially at higher frequencies. Thus, as an application of our dynamic intraday measure of informed trading, we examine the empirical link between private information and firm-specific return variation to provide more in-depth and direct evidence on the validity of Roll's (1988) conjecture.

The rest of the paper is organized as follows. Section 2 describes the data used in the study. In Section 3, we construct and compute the various DPIN measures and compare them to previous measures of informed trading in the literature. We also examine their intraday properties and implications for the theoretical microstructure literature in this regard. Section 4 uses the DPIN measures to study the relationship between price non-synchronicity and private information, providing direct evidence for Roll's (1988) conjecture. Section 5 concludes.

2. Data

The intraday transaction data for this paper 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 2008. We restrict our attention to NYSE-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. Lastly, 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. Since we are interested in intraday horizons, each trading day is divided into 26 fifteen-minute intervals, with each buy or sell trade being assigned to one of these intervals depending on when the trade occurred during the trading day. For each firm in the sample, the total number of trades in any 15-minute interval is the sum of all buy and sell trades (as well as unsigned trades) occurring within the corresponding time span. Fifteen-minute returns are obtained by log-differencing the last recorded midpoint prices of consecutive intervals. Proceeding in the above fashion yields 14,405,663 firm-interval observations with which to conduct our analysis.

3. A dynamic intraday measure of the probability of informed trading

3.1. Construction of DPIN measures

Our approach for constructing a dynamic, intraday measure of the probability of informed trading extends previous theoretical work by Campbell et al. (1993) and empirical work by Avramov et al. (2006). The basic intuition behind the Campbell et al. model is 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 "noninformational" 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. Thus, 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 for aggregating intraday data to delineate whether a particular trading day, overall, is dominated by broadly "contrarian" versus "herding" behavior on the part of investors. Consistent with the Campbell et al. (1993) model, Avramov et al. show that unexpected returns associated with herding days exhibit significant negative serial correlation, while the autocorrelation for contrarian days is insignificant. Thus, it appears that contrarian trades are closely akin to informed trades and herding trades are a good representation of uninformed trades. Avramov et al. also refine this broad delineation of informed trading by considering additional dimensions known in the literature to be associated with the behavior of investors, such as the disposition effect and the size of trades. They then apply these methods to determine whether informed or uninformed trading can explain asymmetric volatility effects in daily returns.[1]

Below, we adopt the Avramov et al. (2006) approach but adapt it for use at a higher frequency. First, unlike Avramov et al., who focus solely on the effects of sell trades on volatility, our aim is to delineate both informed *buy and* sell transactions from their uninformed counterparts, thus allowing for a standalone measure of the probability of informed trading. Second, since our measure is constructed at a much higher (15-minute) frequency, it will allow us to study the intraday pattern of informed trading, providing empirical evidence and allowing comparisons for previous theoretical work in this particular area. Third, given the dynamic, high-frequency nature of our measure, it will be better suited to capture the short-lived nature of private information and its effect on firm-specific return variation (as we will discuss in Section 4).

3.1.1. Baseline DPIN measure

In order to arrive at our first broad delineation of informed (contrarian) versus uninformed (herding) trades, we first isolate the unexpected component of returns as the residuals from the following regression:

$$R_{i,j} = \gamma_0 + \sum_{k=1}^{4} \gamma_{1i,k} D_k^{Day} + \sum_{k=1}^{26} \gamma_{2i,k} D_k^{Int} + \sum_{k=1}^{12} \gamma_{3i,k} R_{i,j-k} + \varepsilon_{i,j}, \tag{1}$$

where $R_{i,j}$ is the return on stock i at intraday interval j ($j = 1,...,26$), D_k^{Day} represents day-of-week dummy variables for Tuesday through Friday, and D_k^{Int} represents dummy variables corresponding to the particular 15-minute interval at which returns are

[1] To be clear, this approach does not imply that the simple contrarian reaction to past returns is by definition informed. As pointed out by an anonymous referee, a useful interpretation is to assume that there is an unobserved high frequency dynamic of a variable that induces informed traders to trade. Our goal, then, is to construct a proxy for this unobserved variable based on what is observed in the data ex post. To the extent that contrarian trading, along with accounting for disposition and size effects, is associated with informed trading (which has been established in the literature), periods where the data show signs of such effects therefore have a higher likelihood of informed trading.

measured on a given day t. Thus, the residual $\varepsilon_{i,j}$ captures the variation in returns leftover after average day-of-week effects, average time-of-day effects, and the effects of past returns have been accounted for and therefore serves as a proxy for unexpected returns.

Our baseline measure of the dynamic probability of informed trading ($DPIN_{BASE}$) is then constructed as follows. Extending Avramov et al. (2006), buy (sell) trades in the presence of negative (positive) unexpected returns are classified as informed trades. On the other hand, buy (sell) trades in the presence of positive (negative) unexpected returns are classified as uninformed trades. Formally, let $NB_{i,j}$, $NS_{i,j}$, and $NT_{i,j}$ be the number of buy, sell, and total trades, respectively, for stock i at interval j. Then, our baseline DPIN measure is constructed as follows:

$$DPIN_{BASE_{i,j}} = \frac{NB_{i,j}}{NT_{i,j}}\left(\varepsilon_{i,j} < 0\right) + \frac{NS_{i,j}}{NT_{i,j}}\left(\varepsilon_{i,j} > 0\right), \tag{2}$$

where $(\varepsilon_{i,j} < 0)$ is an indicator variable that equals 1 when the unexpected return is negative and zero otherwise, and $(\varepsilon_{i,j} > 0)$ takes on the value of unity when unexpected returns are positive and zero otherwise. The rationale behind Eq. (2) is that buy (sell) trades made amid declining (rising) prices are contrarian in nature, and thus indicative of informed trading as shown by Avramov et al. On the other hand, buy (sell) trades initiated during rising (declining) markets suggest uninformed herding. Thus, a straightforward measure of the probability of informed trading during any given 15-minute interval can be simply obtained by calculating the proportion of contrarian trades taking place during that interval, as in our baseline DPIN measure, $DPIN_{BASE}$, above.

3.1.2. DPIN with disposition effect and trend chasing

The baseline DPIN measure in Eq. (2) is somewhat broad in the sense that while buying (selling) amid falling (rising) prices is necessary for a trade to be considered contrarian/informed, it is not sufficient to make such a conclusion. To better pinpoint informed trading activity, we consider several refinements to our baseline DPIN measure. The first of these refinements is accounting for the disposition effect in the selling of shares and trend chasing in share purchases.

For the selling of shares, Avramov et al. (2006) note that the behavioral finance literature documents several well known cognitive biases among unsophisticated investors, chief among these being loss aversion, in which investors are reluctant to realize losses. Thus, the resulting disposition effect suggests that uninformed investors will be less willing to sell shares following price declines, and more likely to sell after price increases. Thus sells taking place when unexpected returns are negative and past cumulative returns are positive reflect both herding and the disposition effect and are likely initiated by uninformed investors.

On the buying side, various behavioral explanations of trend chasing exist, such as anchoring, herding and feedback trading, confirmation bias, and overreaction. In any case, to the extent that investors perceive past price increases as a positive signal and are more likely to buy additional shares, such trades are more likely to be initiated by uninformed investors. Thus, buys taking place when unexpected returns are positive and past cumulative returns are positive reflect both herding and trend chasing and are likely to be initiated by uninformed investors.

Accounting for the disposition effect and trend chasing allows for a finer partition of informed and uninformed trades, thus yielding a refined measure of the probability of informed trading, $DPIN_{DISP}$, which is constructed as follows:

$$DPIN_{DISP_{i,j}} = \left[\frac{NB_{i,j}}{NT_{i,j}}\left(\varepsilon_{i,j} < 0\right) + \frac{NS_{i,j}}{NT_{i,j}}\left(\varepsilon_{i,j} > 0\right)\right]\left(R_{i,j-10:j-1} < 0\right), \tag{3}$$

where $(R_{i,j-10;j-1} < 0)$ is an indicator variable that takes on the value of unity if the cumulative return over the last ten intervals is negative and zero otherwise. Eq. (3) builds on the baseline approach in Eq. (2) for broadly classifying informed trades but imposes the additional condition that trades made when past cumulative returns are negative are "even more" informed. Specifically, buying (selling) that takes place amid declining (rising) prices and when past cumulative returns are negative are more likely to reflect informed trades rather than trend chasing (the disposition effect) on the part of uninformed investors.

3.1.3. DPIN with trading size effects

Another type of partition that considers trade size can be made to also obtain finer classification of informed trades. Using the finding in Easley and O'Hara (1987) that informed traders are more likely to submit larger orders, we can construct one more additional measure of the probability of informed trading by imposing an additional condition on Eq. (2) that accounts for trade size:

$$DPIN_{SIZE_{i,j}} = \left[\frac{NB_{i,j}}{NT_{i,j}}\left(\varepsilon_{i,j} < 0\right) + \frac{NS_{i,j}}{NT_{i,j}}\left(\varepsilon_{i,j} > 0\right)\right]\left(LT_{i,j}\right), \tag{4}$$

where $(LT_{i,j})$ is a "large trades" indicator variable that equals 1 if the total trade size for stock i over interval j is larger than the stock's median interval trade size throughout the same trading day, and zero otherwise. Again, Eq. (4) builds on the broad approach in Eq. (2), but adds the nuance that large contrarian buys and sells are more likely to be initiated by informed investors.[2,3]

---

[2] Avramov et al. (2006) provide further tests on the feasibility of the trading size refinement by conducting daily autocorrelation tests. They find that large contrarian trades lead to zero serial correlation in unexpected returns, while herding trades, whether large or small, are associated with negative autocorrelation. The rationale is that large contrarian trades reflect information and therefore, unlike uninformed trades, tend not to be followed by price reversals.

[3] We use the median rather than average trade size to delineate large and small trades as the later will be skewed in the presence of a few extremely large trades or when there are many consecutive periods with very few trades.

3.2. The DPIN measure as a proxy for informed trading

Private information arrival by its nature is a short-lived random shock across stocks and across time. In this spirit, we have constructed our DPIN measure to capture such unobservable high frequency dynamics of private information arrival. To verify that the DPIN measure is a reasonable proxy for informed trading, we conduct several empirical examinations and find that to a large extent the DPIN measure is a reasonable proxy for the unobservable information trading as follows. First, the aggregate DPIN measure is consistent with the prior literature of informed trading, i.e., the PIN measure of Easley et al. (2002). Second, the unconditional (average) DPIN measure is associated with firm characteristics in terms of the degree of opaqueness. For example, high DPIN stocks are likely to be associated with high opaqueness such as small size, low volume, and high illiquidity. Third, the DPIN measure (conditional on size) is able to capture the widely known U-shaped intraday pattern of information trading. These collaborating results lend a strong support to our use of the DPIN measure to capture the unobservable high frequency dynamic of the underlying private information arrival. We next relate our findings to the previous literature and discuss the relationship between our DPIN measure and firm characteristics. Intraday properties of the DPIN measures are examined in the next subsection (Section 3.3).

3.2.1. Results and comparison of DPIN with previous models

The formulations in the preceding section lead to various measures of the probability of informed trading. These measures have the advantage of being dynamic and flexible — they can be aggregated over many intervals to make comparisons with existing macro-horizon models, yet they are also capable of capturing cross-sectional and time-series variation in the probability of informed trading at the much higher daily and intraday frequencies (in our case, 15-minute intervals). Another advantage is that the various DPIN measures presented above do not require any form of numerical optimization for their estimation and are thus relatively simple and quick to compute from the transactions data, thus providing a more straightforward and less time consuming alternative method for dealing with tremendously large datasets such as TAQ.

To facilitate comparisons with other existing estimates of the probability of informed trading in the literature, Table 1(a) presents summary statistics for yearly cross-sectional DPIN measures averaged across years 1993 to 2008. Not surprisingly, given the broad nature of the baseline DPIN measure, $DPIN_{BASE}$ yields the largest mean and median estimates of the probability of informed trading. The most noteworthy feature in Table 1(a) is that the means and medians of two of the refined DPIN measures, namely $DPIN_{DISP}$ and $DPIN_{SIZE}$, are quite close to the PIN estimates of Easley et al. (2002, p. 2208), who find a parameter mean and median of 0.191 and 0.185, respectively. Restricting the sample to the 1993–98 period (which is the overlap between our sample and that of Easley et al.), yields even closer results: we find 0.185 and 0.206, respectively. Interestingly, although constructed at a much higher, intraday frequency the refined DPIN measures come surprisingly close to approximating the properties of the macro-horizon PIN when aggregated to similarly long time horizons, although the cross-sectional dispersion of the DPIN measures appear lower than that for PIN (standard deviations of 0.027 and 0.057, respectively). Also worth mentioning is that our DPIN measures are also in proximity to the probability of informed trading found in Nyholm (2002, p.495), who computes a mean value of 0.1106 for high volume stocks and 0.1380 for low volume stocks.[4]

Continuing with a comparison of the properties of DPIN with existing measures in the literature, Fig. 1 shows the yearly cross-sectional average DPIN measures over the years 1993 to 2008. Again, the results are quite similar to those of Easley et al. (2002), who find very little year-to-year variation in their PIN estimates (see their Fig. 3 Panel A on p. 2204). Thus, it again appears that when aggregated over comparably long horizons as the PIN estimate of Easley et al., our DPIN measures exhibit a similar stability across time. Across stocks, Fig. 2 plots the distribution of average DPIN measures for all stocks in the sample. From the plotted histograms, it appears that each of the various DPIN measures yields adequate cross-sectional variation; importantly, stocks appear to differ very noticeably along the dimensions associated with the various DPIN measures. Again with respect to Easley et al. (2002), the most striking similarity occurs with the distribution of our $DPIN_{SIZE}$ measure. Comparing our Fig. 2(c) with their Fig. 4 Panel C (p. 2207), which shows the cross-sectional distribution of PIN, not only are the two histograms centered in the same approximate location but there is a similar left skew in both plots, with both also exhibiting a relatively long right tail. However, the overall dispersion of PIN appears higher, consistent with the fact that it has a higher standard deviation.

3.2.2. DPIN measures and firm characteristics

We next split the sample into two groups by taking High (Low) DPIN stocks to be stocks whose average intraday DPIN measure are above (below) the sample average for all stocks. Table 1(b) reports mean firm characteristics for high and low DPIN stocks. It is clear that High DPIN stocks are associated with firm characteristics consistent with higher levels of opaqueness. Stocks with higher DPIN measures are associated with much smaller firm size (by an order of magnitude in most cases), low volume, and much higher illiquidity (as measured by Amihud (2002) as the ratio of absolute daily returns to the daily dollar volume of a

[4] Nyholm (2002) is similarly interested in a dynamic estimate of the probability of informed trading at the transaction level. Placing the trade-indicator model of Huang and Stoll (1997) in a Markov-switching framework, he estimates the probability of informed trading using the evolution of the smoothed conditional state probabilities through time. Nyholm uses 108 stocks and one-month of TAQ data in the analysis.

Table 1
Summary statistics for DPIN measures and firm characteristics.

(a) Yearly cross-sectional DPIN across all years.

Measure	Mean	Median	Std. dev.	Min	Max
$DPIN_{BASE}$	0.458	0.455	0.036	0.274	0.733
$DPIN_{DISP}$	0.215	0.215	0.029	0.073	0.401
$DPIN_{SIZE}$	0.231	0.227	0.027	0.133	0.500

(b) DPIN and firm characteristics

Measure	High/low	No. firms	Size	Illiquidity	Volume
$DPIN_{BASE}$	High	1899	819,299	7.051	103,740
	Low	2306	5,448,227	0.760	591,708
$DPIN_{DISP}$	High	2046	1,265,812	6.094	162,587
	Low	2159	5,340,255	1.241	569,050
$DPIN_{SIZE}$	High	1721	534,357	7.715	67,367
	Low	2484	5,313,942	0.751	581,898

Notes: The variable descriptions are as follows: $DPIN_{BASE}$ is the baseline measure of the dynamic probability of informed trading; $DPIN_{DISP}$ is a refinement to the baseline measure that accounts for the disposition effect (trend chasing) in sell (buy) transactions; $DPIN_{SIZE}$ is a refinement to the baseline measure that accounts for trade size that uses the median interval trade size to delineate large trades from small trades. Panel (a) above reports time-series averages across years 1993 to 2008 of cross-sectional means, medians, standard deviations, minimums, and maximums for the various DPIN measures. Panel (b) reports mean firm characteristics for firms whose average intraday DPIN measure is above (High) and below (Low) the sample average for all firms. The measure of illiquidity is from Amihud (2002), which is calculated as the ratio of absolute daily returns to the daily dollar volume of a stock.

stock). These results bolster the case that the various DPIN measures reasonably capture the unobserved dynamics of informed trading and thus serve as a suitable proxy for it.

3.3. Intraday properties of the DPIN measures

An advantage of the DPIN measures constructed above is that they provide estimates of the probability of informed trading at much finer frequencies. Thus, we can examine the intraday patterns of informed trading and provide new evidence with regard to the theoretical microstructure literature in this area.

Table 2 reports summary statistics for the (disaggregated) intraday DPIN measures across all 15-minute intervals and stocks. Of course, the means of each of the measures are very similar to those in Table 1(a), while the standard deviations are much higher given the higher variability that is expected at the intraday level across intervals and stocks. Notably, the medians for the two refined DPIN measures are zero, indicating that the majority of 15-minute intervals do not exhibit any signs of informed trading activity whatsoever. Typically, for any given stock on any given trading day, the pattern of the probability of informed trading exhibits a cycle in which there are several consecutive 15-minute intervals where the probability of informed trading is nonzero, and then reverts to zero for several more intervals, etc. Stocks with generally higher average intraday DPIN measures tend to display longer runs of consecutive "nonzero" intervals, and shorter runs of consecutive "zero" intervals. This finding is in contrast to the PIN model, which postulates a constant Poisson arrival rate of informed trading throughout the trading day. However, our finding is more consistent with the notion of strategic timing on the part of informed investors of when to take

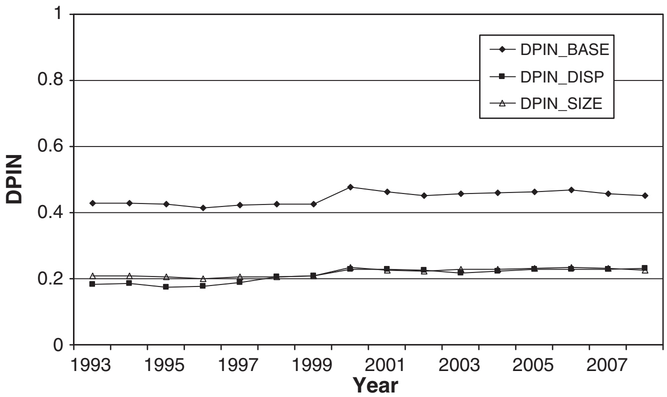

Fig. 1. Yearly cross-sectional average DPIN measures over time. Notes: The figure above shows yearly cross-sectional average DPIN measures for each year from 1993 to 2008. Variable definitions are in Table 1.

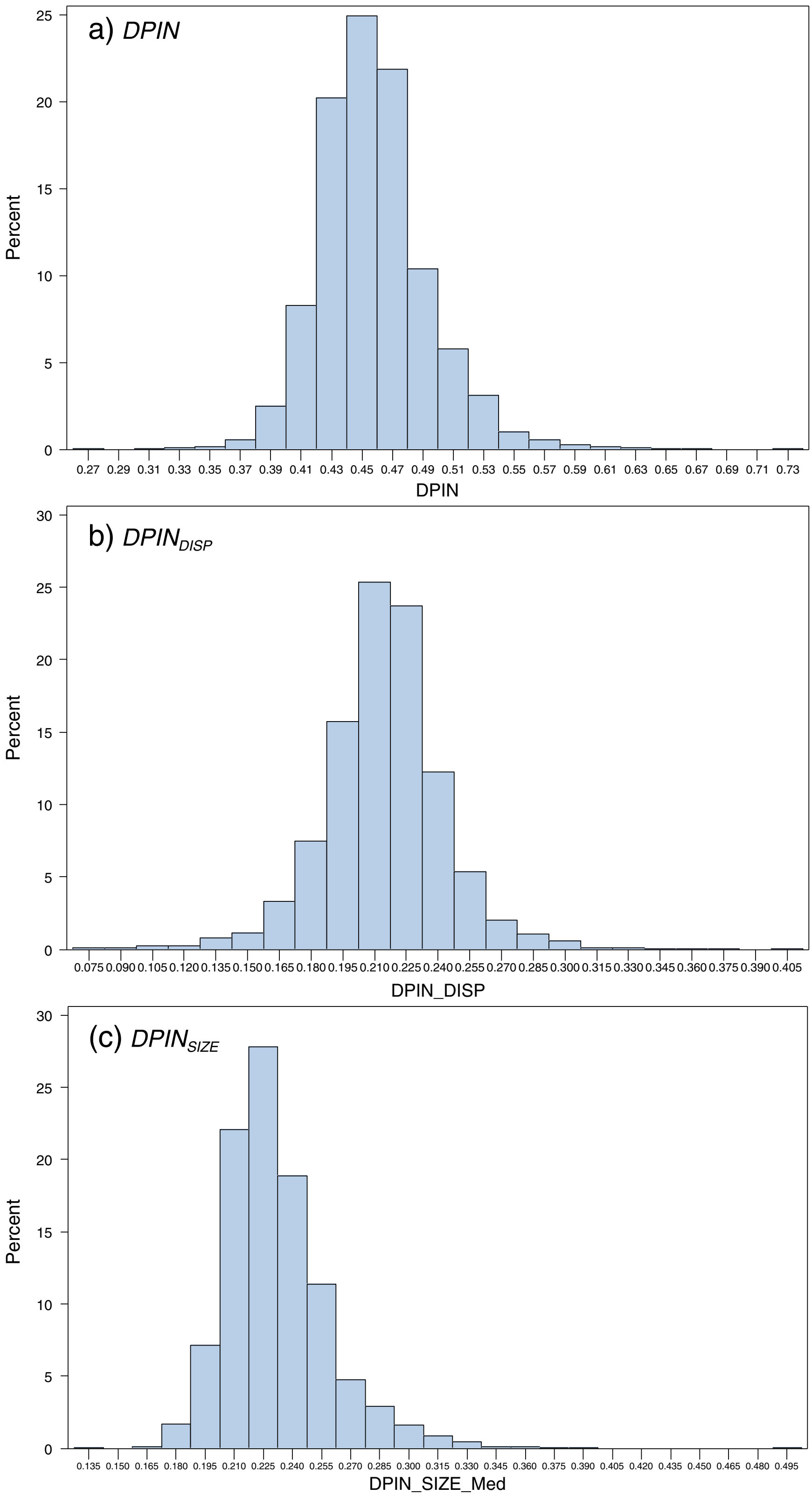

Fig. 2. Distribution of intraday DPIN measures across stocks. Notes: The figures above show the distributions of the average intraday DPIN measures across all stocks (i.e., the histogram frequency represents proportion of stocks with a given average DPIN measure). Rest as in Table 1.

Table 2
Intraday DPIN measures across all stocks and intervals.

Measure	Mean	Median	St. dev	25th percentile	75th percentile
$DPIN_{BASE}$	0.447	0.431	0.297	0.250	0.600
$DPIN_{DISP}$	0.212	0.000	0.301	0.000	0.417
$DPIN_{SIZE}$	0.222	0.000	0.289	0.000	0.429

Notes: The table above contains means, medians, standard deviations, 25th percentiles, and 75th percentiles of the intraday DPIN measures computed across all stocks and intervals (i.e., across all 15-minute intervals, across every trading day, across all years, and across every stock). Rest as in Table 1.

advantage of the camouflage provided by uninformed investors to exploit their asymmetric information advantage (Kyle, 1985; Back, 1992; Lei and Wu, 2005).[5]

Further results on the intraday pattern of the probability of informed trading are reported in Fig. 3, which shows the various DPIN measures at each 15-minute interval throughout a trading day aggregated across all stocks and time periods. Such an analysis gives a glimpse of the market-wide dynamics of informed trading throughout a typical trading day. Similar to their macro-horizon counterparts, $DPIN_{BASE}$ and $DPIN_{DISP}$ exhibit a high degree of stability throughout the trading day. However, $DPIN_{SIZE}$ exhibits an interesting U-shaped pattern, with a relatively high degree of informed trading (conditional on large trades) taking place closer to the opening (9:00 am to 11:30 am) and closing (2:45 pm to 4:00 pm) hours of the trading day, and relatively less informed activity (conditional on large trades) occurring during the middle portion of the day in between (especially during the lunch hours from 12:00 pm to 2:00 pm).

Examining how this U-shaped intraday pattern of informed trading has evolved over the years, Fig. 4 considers four four-year sub-periods of $DPIN_{SIZE}$. It is clearly evident from both panels that the intraday U-shaped pattern of the probability of informed trading has become much more pronounced over time. Over each of the sub-periods, there is a marked increase in informed trading activity (conditional on large trades) during the opening and closing hours of the trading day. On the other hand, the probability of informed trading (conditional on large trades) appears to have declined considerably over time during the middle of the day.

It is important to note that the increasing U-shaped pattern in the DPIN measure occurs when conditioning on large trade size, while no such pattern is exhibited in $DPIN_{BASE}$ or $DPIN_{DISP}$, both of which are unrelated to size. This seemingly perplexing result is in fact consistent with the predictions of the prior literature on market microstructure theory. In his seminal paper, Kyle (1985) posits that the informed investor will trade strategically to camouflage his/her trading among those of the uninformed investors. Admati and Pfleiderer (1988) argue further that uninformed investors have an incentive to concentrate their trades in order to minimize their trading costs. To the extent that inventory-motivated trades are likely to occur in anticipation of and following non-trading periods, uninformed investors naturally concentrate their trades toward the beginning and the end of the trading day. Consequently, the informed investor can take advantage of the thick market during both ends of the trading day by submitting large sizes of trade without revealing much private information. This also means that in the middle of the trading day when uninformed trading is sparse, the informed investor will strategically avoid large trades to minimize information revelation. Such strategic behavior on the part of informed traders is consistent with the U-shaped intraday pattern in $DPIN_{SIZE}$.

Once we understand the strategic behavior of informed investors behind the U-shaped intraday trading pattern conditional on large trades, we naturally ask why no such pattern is exhibited in the DPIN measures without conditioning on size, i.e., in $DPIN_{BASE}$ and $DPIN_{DISP}$. If informed investors only concentrated their trading at both ends of the trading day, then we would have observed the same intraday U-shaped pattern of informed trading regardless of trade size. The fact that we do not observe such a pattern in $DPIN_{BASE}$ and $DPIN_{DISP}$ might seem to refute such a simple scenario of informed trading. Moreover, if information arrival is random and short-lived during the trading day, then informed investors have an incentive to trade whenever information arises since delayed trading can mean a reduced information advantage, or none at all. Knowing this, however, informed investors also recognize the impact of their trading on prices, especially when the market is thin during the middle of the trading day. To resolve this dilemma, informed traders can strategically break up intended large orders into a series of small trades to minimize information revelation. Such strategic behavior of the informed traders can actually lead to an inverse U-shaped intraday trading pattern when conditioned on small trade sizes. That is, conditional on small trades, the probability of information trading is high in the middle of the trading day and low at both ends. To verify this prediction in our DPIN framework, we define a corresponding measure for small trades, $DPIN_{SMALL}$, such that

$$DPIN_{SMALLi,j} = \left[\frac{NB_{i,j}}{NT_{i,j}}\left(\varepsilon_{i,j} < 0\right) + \frac{NS_{i,j}}{NT_{i,j}}\left(\varepsilon_{i,j} > 0\right)\right]\left(ST_{i,j}\right), \tag{5}$$

where $(ST_{i,j})$ is a "small trades" indicator variable that equals 1 if the total trade size for stock i over interval j is smaller than the

[5] Lei and Wu (2005) develop a theoretical model in which informed investors monitor market movements and respond rationally to any change in the arrival of uninformed traders. They point out that because theoretical frameworks in Glosten and Milgrom (1985) and Easley et al. (1996) assume that traders are chosen probabilistically to submit orders, informed traders cannot respond to camouflage provided by the uninformed traders in such models.

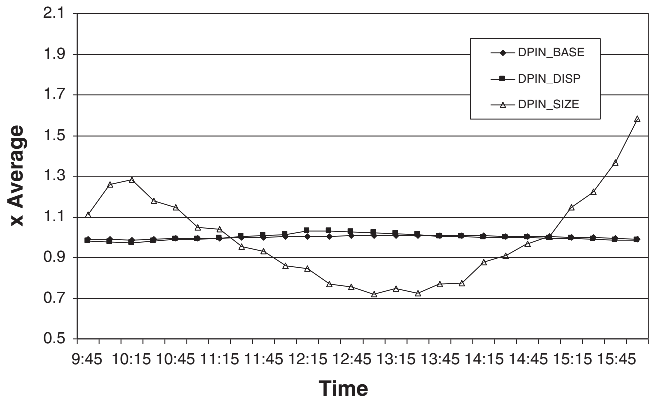

Fig. 3. Intraday average DPIN measures across all stocks and trading days. Notes: The figure above shows the intraday pattern of the various DPIN measures across 26 15-minute intervals throughout the trading day. Each point corresponds to an average DPIN measure across all stocks and trading days for the given 15-minute interval. Each point is reported in proportion to the average DPIN measure calculated across the 26 15-minute intervals. Rest is as in Table 1.

stock's median interval trade size throughout the same trading day, and zero otherwise. We then plot the results of the intraday pattern of this measure for small trades in Fig. 5(a) for the whole sample and in Fig. 5(b) for sub-periods. Indeed, the figures clearly show an inverse U-shaped pattern in both measures. Furthermore, Fig. 5(b) indicates that the inverse U-shaped intraday pattern of the probability of informed trading has become much more pronounced over time. Over each of the sub-periods, there is a marked increase in informed trading activity during the middle hours of the trading day conditional on small trades.

The strategic behavior of informed traders who break up large orders into a series of small trades to hide their information is first posited in the stealth trading hypothesis of Barclay and Warner (1993) and well documented in later empirical studies by Chakravarty (2001), Alexander and Peterson (2007), Hansch and Choe (2007), and Blau et al. (2009). Among these studies, our sample period (1993 to 2008) overlaps best with that of Hansch and Choe, who examine the period from 1993 to 2003. Interestingly, they find that the distribution of stealth informed trading shifts from medium-sized trades to small trades around the year 2000, partly due to increased access to information after the millennium during the Internet era. This phenomenon may explain what we observe in Fig. 5(b) with the inverse U-shaped pattern of information based trading becoming more pronounced over time. Since large trades are the opposite of small trades, as small informed trades become more pronounced during the middle of the trading day over time, large informed trades at the same time are likely to become more pronounced towards both ends of the trading day — a result which is indeed confirmed in Figs. 3 and 4.

Once we realize that the U-shaped intraday pattern of large informed trades and the inverse U-shaped pattern of small informed trades are two sides of the same coin, or two distinct pieces of the whole puzzle, we obtain a more complete picture of informed trading. That is, informed investors do not solely trade at both ends, nor do they trade only in the middle of the trading day. They will trade anytime during the day whenever short-lived information arises for fear of losing their information advantage. During both ends of the trading day when the market is thick, the informed traders can afford to submit large trades without revealing much information. However, during the middle of the trading day when the market is thin, the informed

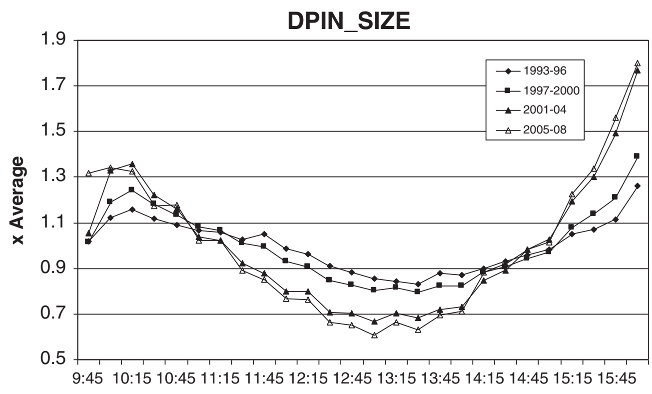

Fig. 4. Intraday average DPIN measures conditioned on large trades for various subperiods. Notes: The figure above shows the evolution of the intraday pattern of the $DPIN_{SIZE}$ measure over four four-year subperiods: 1993–96, 1997–2000, 2001–04, and 2005–08. Rest as in Fig. 3.

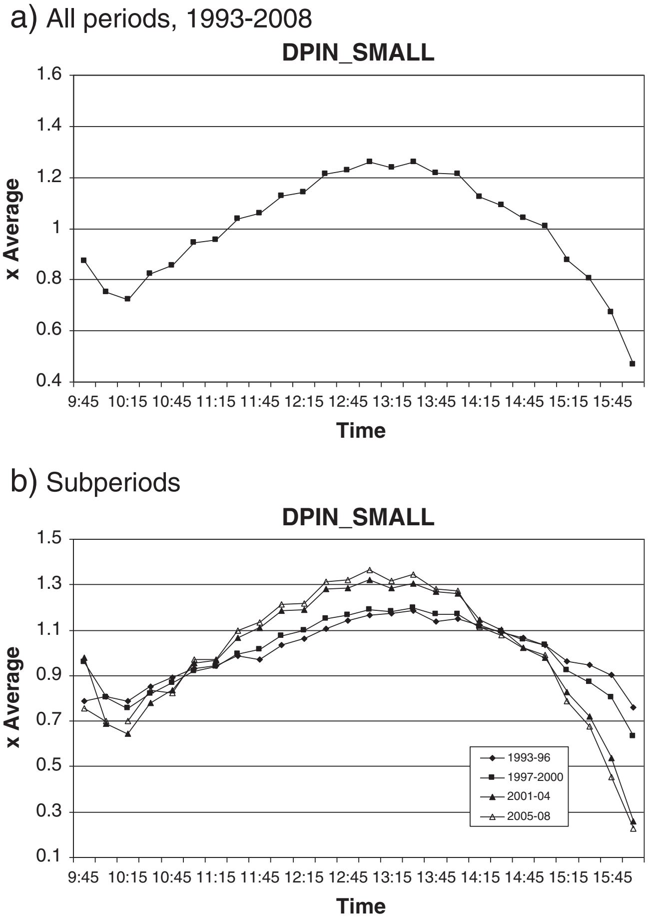

Fig. 5. Intraday average DPIN measures across all stocks and trading days, conditioned on small trade size. Notes: The figure in panel (a) above shows the intraday pattern of $DPIN_{SMALL}$, which is conditioned on small trade size as in Eq. (5), across 26 15-minute intervals throughout the trading day. Each point corresponds to the average across all stocks and trading days for the given 15-minute interval. Each point is reported in proportion to the average calculated across the 26 15-minute intervals. The figure in panel (b) shows the evolution of the intraday pattern of $DPIN_{SMALL}$ over four four-year subperiods: 1993–96, 1997–2000, 2001–04, and 2005–08.

traders cannot afford to submit large trades and hence have to strategically break up their large orders into a series of small trades. Combining these two opposite forces or U-shaped patterns result in a flat probability of informed trading throughout the trading day, dictated mainly by the random arrival of information. Thus, we obtain a flat intraday pattern of informed trading in both of the unconditional DPIN measures, $DPIN_{BASE}$ or $DPIN_{DISP}$, as shown in Fig. 3. In this sense, then, such an unconditional, flat intraday pattern of informed trading may actually lend support to the theoretical models of Glosten and Milgrom (1985) and Easley et al. (1996) that traders are chosen probabilistically to submit orders.

Lastly, it is worth noting that while the U-shaped intraday pattern has been widely documented in earlier literature concerning the role of informed trading, for example, Wood et al. (1985), Harris (1986), Jain and Joh (1988), Gerety and Mulherin (1992), Chan et al. (1995), and Wang (1998), few have uncovered a corresponding, inverse U-shaped intraday pattern in informed trading for small trades. One notable exception is Blau et al. (2009) who use price changes as a proxy for informed trading and find a similar U-shaped intraday pattern in large informed trades and an inverse U-shaped intraday pattern in small informed trades. To our knowledge, we are the first study that uses a direct measure of the probability of informed trading to document such complementary intraday patterns of informed trading in its totality.

4. DPIN and daily firm-specific return variation

As discussed above, ever since Roll (1988) surmised that firm specific-return variation might be caused by private information, researchers have taken (for granted) the former as a proxy for the latter to study an array of issues in finance. While many appealing conceptual and indirect arguments have been made for such an approach, very few studies (to our knowledge) directly investigate the relationship between the two variables, especially at the higher frequency, intraday level. In this section, we examine the empirical link between informed trading and firm-specific return variation to provide more in-depth and direct evidence on the validity of Roll's conjecture.

4.1. Methodology

To construct our measure of daily firm-specific return variation, we use the standard approach in the literature based on the R-squared statistic from a market model regression (e.g., Roll (1988), Durnev et al. (2003, 2004), and Chen et al. (2007)). Namely, for each firm on day t we perform the following regression:

$$R_{i,j} = \gamma_{i,0} + \gamma_{i,1} R_{m,j} + \varepsilon_{i,j}, \tag{6}$$

where $R_{i,j}$ is the return of firm i at intraday interval j (within day t) and $R_{m,j}$ is the market return at interval j using both an equally weighted and value weighted portfolio. Daily firm-specific return variation for firm i on day t is then defined as $FSRV_{i,t} = \log[(1 - R^2_{i,t}) / R^2_{i,t}]$, where $R^2_{i,t}$ is the daily R-squared statistic for firm i on day t from the regression in Eq. (6). Thus, $FSRV$ captures the unexplained variation in a firm's returns that remains after market returns have been accounted for. [6] As a result, for each day in the sample, we have a measure of the firm-specific return variation for each stock.

The daily probability of informed trading for a given stock is taken to be the average DPIN across all intervals over a particular trading day. Thus, stocks with higher interval DPINs within a trading day will tend to have higher daily DPINs. We use the three DPIN measures in Eqs. (2), (3), and (4), and report the results using each below.

To directly examine the empirical relationship between daily firm-specific return variation and informed trading, we use the Fama and MacBeth (1973) regression framework, which involves performing cross-sectional regressions to obtain parameter estimates for each day, and then taking the time-series average across all days to arrive at parameter estimates and their sampling distributions for the entire sample. There are a total of $N = 4{,}191$ stocks and $T = 3{,}994$ days in the regressions. The Fama–MacBeth regressions are specified as follows:

$$FSRV_{i,t} = \alpha + \beta_{1,t} DPIN_{i,t} + \beta_{2,t} DPIN_{i,t-1} + \gamma_t' X_{i,t} + \varepsilon_{i,t}, \tag{7}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where i denotes firm and t denotes day, $DPIN$ corresponds to one of the four DPIN measures constructed above (with one lag of $DPIN$ included to account for lagged effects), and X is a vector of control variables. The control variables are: total daily volume (VOL); the Amihud (2002) illiquidity measure (ILL), which is calculated as the ratio of absolute daily returns to the daily dollar volume of a stock; firm size ($SIZE$); stock price (PRC); and returns (RET). To account for short-term dynamics, we also include in the vector of controls one lag each of $FSRV$, VOL, ILL, and RET.

The Fama–MacBeth approach is widely used in the empirical asset pricing literature, especially in the "large T, large N" panel data context (i.e., a large number of time-series observations and cross-sectional observations (which characterizes our dataset)), as opposed to fixed or random effects estimation (which appears to be more widely used in the "large N, small T" settings commonly arising in corporate finance or empirical microeconomics). The properties of the Fama–MacBeth approach are explored further in Skoulakis (2008) and Petersen (2009). Of concern is the computation of unbiased standard errors in the presence of cross-sectional and serial correlations; ignoring such correlations can result in underestimation of the standard errors and thus inflated t-statistics and invalid inferences.

In particular, Petersen (2009) shows that the Fama–MacBeth standard errors are robust to cross-sectional correlation across firms (or stocks) in a given time period (as this is what Fama–MacBeth was originally designed to handle), but are biased downwards in the presence of an unobserved firm fixed effect that induces serial correlation for a given stock across periods, even after Newey and West (1987) adjustments are made for serial correlation and heteroskedasticity of unknown form (i.e., HAC estimation). As such, Petersen notes that Fama–MacBeth regressions containing persistent data (which may arise from such firm fixed effects) are most likely to suffer from biased standard errors.

To address this potential problem, we conduct two robustness checks. First, we perform the Fama–MacBeth regression on first-differenced data, as this helps to remove any firm fixed effects and reduces persistence in the data, thereby possibly leading to potentially more reliable standard errors. In addition, first-differencing has the advantage of mitigating potential microstructure effects (e.g., the bid–ask bounce), while allowing an examination of the dynamics of how day-to-day changes in informed trading affect day-to-day changes in firm-specific return variation across firms. Specifically, we perform Fama–MacBeth regressions based on the following model:

$$\Delta FSRV_{i,t} = \alpha + \beta_{1,t} \Delta DPIN_{i,t} + \beta_{2,t} \Delta DPIN_{i,t-1} + \gamma_t' Z_{i,t} + \varepsilon_{i,t}, \tag{8}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where, $\Delta DPIN$ corresponds to the daily first-difference of the corresponding DPIN measure (with one lag of $\Delta DPIN$ included to account for the effects of lagged daily changes), and Z is a vector of control variables that includes the

[6] There are a maximum of 26 return observations for each firm-day. Some stocks do not have transactions, and thus returns, in all intervals, result in missing data for those intervals and insufficient observations with which to perform the regression in Eq. (6). To avoid problems associated with having too small of a sample, we restrict our attention to firm-days with greater than 10 interval return observations per day.

Table 3
Results from Fama-MacBeth regressions of firm-specific return variation on various DPIN measures.

	(a) Equally weighted market model			(b) Value-weighted market model		
	(1)	(2)	(3)	(1)	(2)	(3)
$DPIN_{BASE}$	0.558***			0.096***		
	(28.55)			(4.69)		
$DPIN_{BASE,\ t-1}$	0.467***			0.100***		
	(24.29)			(5.08)		
$DPIN_{DISP}$		0.270***			0.111***	
		(16.91)			(7.45)	
$DPIN_{DISP,t-1}$		0.093***			0.001	
		(7.64)			(0.13)	
$DPIN_{SIZE}$			0.640***			−0.013
			(21.63)			(−0.42)
$DPIN_{SIZE,t-1}$			0.569***			0.033
			(21.90)			(1.14)
$FSRV_{t-1}$	0.018***	0.019***	0.019***	0.012***	0.012***	0.012***
	(24.79)	(25.41)	(25.06)	(20.60)	(20.81)	(20.53)
VOL	−0.064***	−0.067***	−0.065***	−0.006	−0.006	−0.008**
	(−17.41)	(−18.31)	(−17.77)	(−1.57)	(−1.43)	(−2.07)
VOL_{t-1}	−0.045***	−0.048***	−0.045***	0.0003	0.0007	0.0004
	(−10.92)	(−11.63)	(−10.98)	(0.08)	(0.16)	(0.10)
ILL	0.383***	0.625***	0.469***	−1.489***	−1.335***	−1.410***
	(3.00)	(4.86)	(3.66)	(−10.17)	(−10.40)	(−9.63)
ILL_{t-1}	−0.147	0.112	−0.076	−1.346***	−0.013***	−1.266***
	(−1.20)	(0.91)	(−0.62)	(−10.81)	(−10.64)	(−10.27)
$SIZE$	−0.026***	−0.020	−0.023*	−0.021	−0.025	−0.020
	(−2.03)	(−1.56)	(−1.82)	(−1.24)	(−1.37)	(−1.17)
PRC	−0.002***	−0.002***	−0.002***	−0.0003***	−0.0003***	−0.0004***
	(−20.73)	(−21.71)	(−21.70)	(−5.01)	(−1.37)	(−5.73)
RET	−0.914***	−0.545***	−0.918***	−0.402***	−0.260***	−0.412***
	(−13.57)	(−8.12)	(−13.62)	(−6.48)	(−4.53)	(−6.63)
RET_{t-1}	−0.306***	−0.216***	−0.302***	−0.131***	−0.146***	−0.137***
	(−7.43)	(−4.81)	(−7.34)	(−3.61)	(−3.72)	(−3.77)
$Wald$	1411.5***	343.24***	947.39***	22.31***	55.80***	1.48

Notes: The table above contains the results from the Fama–MacBeth regressions specified in Equation (7). Reported coefficients are time-series averages of daily cross-sectional regression coefficients, with corresponding Newey–West t-statistics reported in parentheses. The dependent variable is firm-specific return variation, which is defined as $FSRV = \log[(1 - R^2)/R^2]$, where R^2 is the daily R-squared statistic from the intraday market model regression using an (a) equally weighted and (b) value-weighted market portfolio in Equation (6), respectively. Each column reports the results from using one of the three DPIN measures in the paper (definitions of the DPIN measures in Table 1). The variable VOL is daily volume (divided by 10^6); ILL is the Amihud (2002) illiquidity measure, which is defined as the absolute daily return divided by the daily dollar volume (multiplied by 10^4); $SIZE$ is firm size (divided by 10^8); PRC is share price; and RET is daily returns. $Wald$ denotes the robust Wald statistic for testing the hypothesis that the contemporaneous and lagged DPIN measures are jointly significant. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

change in total daily volume (ΔVOL) and the daily change in the Amihud (2002) illiquidity measure (ΔILL), with one lag of each included. Contemporaneous and lagged returns (RET) are also included, but not further differenced. Firm size ($SIZE$) and stock price (PRC) are not included as these are insignificant in the regression, while their first differences are omitted due to possible multi-collinearity issues, as these are highly correlated with returns.

Second, we use the intermediate demeaning method developed in Skoulakis (2008), who performs the Fama–MacBeth regressions on time-demeaned data. In the presence of an unobserved fixed firm effect, Skoulakis shows that such an approach (by effectively eliminating the fixed effect) also yields accurate asymptotic approximations. In addition, evidence from simulations indicates that standard econometric tools such as Newey–West HAC estimators produce reliable t-statistics. The specification is as follows:

$$FSRV_{i,t}^{\sim} = \alpha + \beta_{1,t} DPIN_{i,t}^{\sim} + \beta_{2,t} DPIN_{i,t-1}^{\sim} + \gamma_t^{'} V_{i,t}^{\sim} + \varepsilon_{i,t}, \tag{9}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where for each variable for a stock on day t, we subtract the corresponding time-series mean over all days from the variable. The demeaned variable is represented by a "tilde" superscript. The vector of controls $V^{\sim}$ contains demeaned volume, illiquidity, and size, as well as one lag each of demeaned volume and demeaned illiquidity. As in Eq. (8), returns are included but are not demeaned since they are mean-reverting with very little persistence.

4.2. Estimation results

The results from the Fama–MacBeth estimation of Eq. (7) are contained in Table 3; panel (a) uses an equally weighted market portfolio in the market model regression in Eq. (6), while panel (b) uses a value-weighted market portfolio. Newey and West

(1987) t-statistics are reported in parentheses below the corresponding parameter estimate. It is evident from columns (1), (2), and (3) in Table 3(a) that all three DPIN measures – $DPIN_{BASE}$, $DPIN_{DISP}$, and $DPIN_{SIZE}$ – and their lags are positive and statistically significant at the 1% level, consistent with Roll's (1988) conjecture. The last row in Table 3(a) reports the robust Wald statistic for testing the null hypothesis that the coefficients on the DPIN measures and their lags are jointly zero (it is Chi-square distributed with two degrees of freedom). The reported test statistics indicate that each DPIN measure and its lag are jointly significant in the regression at the 1% level. All control variables in each column are statistically significant, except for lagged illiquidity, and of the expected sign. Lagged firm-specific return variation has a positive coefficient, indicating persistence in the *FSRV* variable. High volume stocks appear to have lower firm-specific return variation. Stocks with higher Amihud (2002) illiquidity tend to have higher firm-specific return variation. Finally, firm size, price, and current and lagged returns have a negative effect.

From Table 3(b), we see that when using a value-weighted market portfolio two of the DPIN measures – $DPIN_{BASE}$ and its lag and $DPIN_{DISP}$ – are still positive and statistically significant at the 1% level. Wald tests indicate that $DPIN_{BASE}$, $DPIN_{DISP}$, and their respective lags are jointly significant at the 1% level. On the other hand, $DPIN_{SIZE}$ and its lag are now both individually and jointly insignificant. In general, the control variables are all significant, except now for *SIZE*. Also, lagged *VOL* now has a positive coefficient while contemporaneous illiquidity has a negative coefficient.

Overall, from Table 3 it appears that the majority of DPIN measures are positive and highly significant after controlling for other factors that are associated with firm-specific return variation and potentially correlated with our DPIN measures. Thus, these results provide direct evidence at the daily level that informed trading is indeed an important cause of firm-specific return variation, as originally suggested by Roll (1988).

The results from the Fama–MacBeth estimation of Eq. (8) are contained in Table 4. In Table 4(a), using an equally weighted market portfolio, columns (1), (2), and (3) indicate that $\Delta DPIN_{BASE}$, $\Delta DPIN_{DISP}$, and $\Delta DPIN_{SIZE}$, and each of their respective lags, are positive and significant at the 1% level for explaining firm-specific return variation (except for the lag of $\Delta DPIN_{DISP}$, which is significant at the 10% level). Wald tests indicate that all three DPIN measures and their respective lags are jointly significant at the 1% level. The first-differenced control variable ΔVOL and its lag, as well as lagged ΔILL, are all significant, indicating that firms that experience a contemporaneous and lagged increase in daily volume have lower daily firm-specific return variation, while shares that experience an increase in lagged illiquidity have higher firm-specific return variation, suggesting mean reversion in the response of *FSRV* to the illiquidity measure. Finally, firms with high contemporaneous (lagged) returns tend to experience declines (increases) in firm-specific return variation. In Table 4(b), we use a value-weighted market portfolio and repeat the analysis as in

Table 4
Results from Fama–MacBeth regressions of first-differenced firm-specific return variation on changes in first-differenced DPIN measures.

	(a) Equally weighted market model			(b) Value-weighted market model		
	(1)	(2)	(3)	(1)	(2)	(3)
$\Delta DPIN_{BASE}$	0.150***			−0.010		
	(8.31)			(−0.57)		
$\Delta DPIN_{BASE,\ t-1}$	0.054***			−0.010		
	(2.99)			(−0.56)		
$\Delta DPIN_{DISP}$		0.194***			0.118***	
		(11.07)			(7.18)	
$\Delta DPIN_{DISP,t-1}$		0.033*			0.027	
		(1.77)***			(1.51)	
$\Delta DPIN_{SIZE}$			0.145***			−0.052**
			(5.32)			(−1.98)
$\Delta DPIN_{SIZE,t-1}$			0.072***			−0.003
			(2.72)			(−0.12)
ΔVOL	−0.080***	−0.080***	−0.0008***	−0.036***	−0.035***	−0.038***
	(−18.79)	(−18.66)	(−19.01)	(−8.17)	(−7.90)	(−8.36)
ΔVOL_{t-1}	−0.058***	−0.058***	−0.058***	−0.023***	−0.023***	−0.022***
	(−11.88)	(−11.94)	(−11.85)	(−5.30)	(−5.29)	(−5.21)
ΔILL	−0.181	−0.175	−0.144	−0.450***	−0.488***	−0.470***
	(−1.05)	(−1.02)	(−0.83)	(−2.66)	(−2.89)	(−2.62)
ΔILL_{t-1}	−0.344**	−0.374**	−0.332*	−0.283*	−0.311*	−0.291*
	(−1.97)	(−2.12)	(−1.91)	(−1.77)	(−1.94)	(−1.76)
RET	−1.365***	−1.176***	−1.364***	−0.658***	−0.528***	−0.660***
	(−16.79)	(−14.86)	(−16.76)	(−8.65)	(−7.42)	(−8.66)
RET_{t-1}	0.613***	0.332***	0.618***	0.285***	0.104	0.289***
	(8.24)	(4.69)	(8.30)	(4.07)	(1.59)	(4.13)
Wald	77.87***	125.79***	34.61***	0.644	53.90***	3.93

Notes: The table above contains the results from the Fama–MacBeth regressions specified in Equation (8). Reported coefficients are time-series averages of daily cross-sectional regression coefficients, with corresponding Newey–West t-statistics reported in parentheses. The dependent variable is the first-difference of firm-specific return variation, $\Delta FSRV$, where *FSRV* is calculated from daily R-squared statistics from the intraday market model regression using an (a) equally weighted and (b) value-weighted market portfolio in Equation (6), respectively. Each column reports the results from using first-differences of one of the DPIN measures in the paper (see Table 1 for definitions of the DPIN measures). The variable ΔVOL is the daily change in volume and ΔILL is the daily change in the Amihud (2002) illiquidity measure. Rest as in Table 3. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 5
Results from Fama–MacBeth regressions of demeaned firm-specific return variation on demeaned DPIN measures.

	(a) Equally weighted market model			(b) Value-weighted market model		
	(1)	(2)	(3)	(1)	(2)	(3)
$\widetilde{DPIN}_{BASE}$	0.460***			0.025		
	(28.32)			(1.53)		
$\widetilde{DPIN}_{BASE,t-1}$	0.369***			0.026*		
	(23.76)			(1.66)		
$\widetilde{DPIN}_{DISP}$		0.236***			0.084***	
		(15.62)			(6.13)	
$\widetilde{DPIN}_{DISP,t-1}$		0.420***			0.027*	
		(26.89)			(1.71)	
$\widetilde{DPIN}_{SIZE}$			0.497***			−0.103***
			(20.68)			(−4.41)
$\widetilde{DPIN}_{SIZE,t-1}$			0.435***			−0.061***
			(20.03)			(−2.65)
$\widetilde{VOL}$	−0.042***	−0.042***	−0.042***	−0.037***	−0.037***	−0.038***
	(−11.67)	(−11.98)	(−11.93)	(−10.18)	(−10.03)	(−10.38)
$\widetilde{VOL}_{t-1}$	−0.024***	−0.023***	−0.024***	−0.023***	−0.024***	−0.023***
	(−7.33)	(−7.15)	(−7.22)	(−7.28)	(−7.43)	(−7.19)
$\widetilde{ILL}$	0.648***	0.763***	0.701***	−0.344***	−0.366***	−0.323***
	(5.96)	(6.95)	(6.45)	(−3.54)	(−3.77)	(−3.32)
$\widetilde{ILL}_{t-1}$	−0.025	−0.045	0.004	−0.203**	−0.184*	−0.171
	(−0.25)	(−0.45)	(0.04)	(−2.07)	(−1.87)	(−1.75)
$\widetilde{SIZE}$	−0.066***	−0.061***	−0.061***	−0.212***	−0.213***	−0.210***
	(−3.91)	(−3.70)	(−3.56)	(−11.43)	(−11.79)	(−11.34)
$\widetilde{PRC}$	−0.002***	−0.002***	−0.002***	−0.001***	−0.001***	−0.001***
	(−22.14)	(−22.47)	(−22.69)	(−7.00)	(−6.82)	(−7.28)
RET	−0.901***	−0.610***	−0.907***	−0.375***	−0.280***	−0.385***
	(−13.65)	(−9.38)	(−13.71)	(−6.11)	(−4.90)	(−6.26)
RET_{t-1}	−0.328***	−0.344***	−0.328***	−0.121***	−0.124***	−0.128***
	(−8.16)	(−8.60)	(−8.14)	(−3.43)	(−3.50)	(−3.60)
Wald	1367.8***	967.30***	830.45***	5.07*	40.53***	26.50***

Notes: The table above contains the results from the Fama–MacBeth regressions specified in Equation (9). Reported coefficients are time-series averages of daily cross-sectional regression coefficients, with corresponding Newey–West t-statistics reported in parentheses. The dependent variable is the time-demeaned firm-specific return variation, $\widetilde{FSRV}$, where $FSRV$ is calculated from daily R-squared statistics from the intraday market model regression using an (a) equally weighted and (b) value-weighted market portfolio in Equation (6), respectively. The variable $\widetilde{VOL}$ is demeaned volume, $\widetilde{ILL}$ is the demeaned Amihud (2002) illiquidity measure, and $\widetilde{SIZE}$ is demeaned firm size and $\widetilde{PRC}$ is demeaned price. Rest as in Table 3. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4(a). Now, only $\Delta DPIN_{DISP}$ is significant and of the correct sign, although it and its lag are highly jointly significant. The signs, magnitudes, and statistical significance of the control variables are similar to those reported in Table 4(a) and discussed above.

The results from the Fama–MacBeth estimation of Eq. (9) are contained in Table 5. Using an equally weighted market portfolio, Table 5(a) indicates that all three demeaned DPIN measures and their lags are highly statistically significant, both individually and jointly. All other controls are also significant (except lagged illiquidity) and of similar sign to previous specifications. Using a value-weighted portfolio, Table 5(b) indicates that the demeaned $DPIN_{DISP}$ and its lag, as well as the lag of demeaned $DPIN_{BASE}$ remain statistically significant at conventional levels and of the correct sign, and both variables and their lags are jointly significant at conventional levels.

Overall, and on balance, the results in Tables 4 and 5 corroborate the evidence from the original analysis in Table 3. Namely, a clear majority of the measures of informed trading activity remain positive and highly significant after controlling for other factors that are associated with changes in firm-specific return variation and potentially correlated with our DPIN measures. Again, these results provide further direct evidence at the daily level that informed trading is indeed an important determinant of firm-specific return variation.

Finally, since our hypothesis is that higher informed trading causes higher firm-specific return variation vis-à-vis Roll (1988), and not the other way around, we next examine the issue of reverse causality. For this purpose, we use first-differenced data and for each of the three DPIN measures, we regress changes in DPIN on lagged changes in firm-specific return variation and the control variables ΔVOL and ΔILL (and one lag of each). The results of these regressions are reported in Table 6. The majority of coefficients on firm-specific return variation are negative, while those that are positive are either insignificant or less significant. Overall, it thus appears that an increase in firm-specific return variation causes a *decline*, rather than an increase, in informed trading. Thus, reverse causality does not appear to be a problem in our study.

5. Conclusion

The aim of our paper has been twofold. First, we constructed a dynamic intraday measure of the probability of informed trading that is relatively straightforward to implement and circumvents the aggregation problem of other existing measures. As

Table 6
Results from Fama–MacBeth regressions for testing for reverse causality.

	$\Delta DPIN_{BASE}$	$\Delta DPIN_{DISP}$	$\Delta DPIN_{SIZE}$
Equally weighted:			
$\Delta FSRV_{t-1}$	−0.085***	−0.168***	−0.023*
	(−5.00)	(−5.74)	(−1.95)
Value-weighted:			
$\Delta FSRV_{t-1}$	0.014	−0.107***	0.022*
	(0.83)	(−3.70)	(1.88)

Notes: The table above contains the results from the Fama—MacBeth regressions with the DPIN measures as dependent variables and lagged firm-specific return variation, Δ*FSRV*, as an independent variable. Reported coefficients are time-series averages of daily cross-sectional regression coefficients, with corresponding Newey–West *t*-statistics reported in parentheses. The vector of controls includes Δ*VOL* and Δ*ILL*, along with one lag of each (results of these are not reported but available upon request). Rest as in Tables 3 and 4. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

such, our DPIN measure gives a finer and perhaps more accurate view of the dynamics of private information in modern financial markets, especially at higher frequencies. Given the importance of private information in the theoretical and empirical finance literature, such a measure is potentially useful in understanding how markets incorporate information into prices and has broad applicability to a large range of topics in market microstructure, asset pricing, and corporate finance. In this vein, we employed our dynamic intraday measure of informed trading to examine the empirical link between private information and firm-specific return variation. Unlike previous studies cited above that provide indirect and circumstantial support for a relationship between the two, our results using DPIN provide more in-depth and direct evidence on the validity of Roll's (1988) conjecture.

Acknowledgments

The authors would like to thank the editor (R.T. Baillie) and two anonymous referees for helpful comments and suggestions. Jung-Yu Liu provided excellent research assistance. Any mistakes, of course, are our own.

References

Admati, A., Pfleiderer, P., 1988. A theory of intraday patterns: Volume and price volatility. Rev. Financ. Stud. 1, 3–40.
Alexander, G., Peterson, M., 2007. An analysis of trade-size clustering and its relation to stealth trading. J. Financ. Econ. 84, 435–471.
Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. J. Financ. Mark. 5, 31–56.
Avramov, D., Chordia, T., Goyal, A., 2006. The impact of trades on daily volatility. Rev. Financ. Stud. 19, 1241–1277.
Back, K., 1992. Insider trading in continuous time. Rev. Financ. Stud. 5, 387–409.
Barclay, M., Warner, J., 1993. Stealth trading and volatility: which trades move prices? J. Financ. Econ. 34, 281–306.
Blau, B., Van Ness, B., Van Ness, R., 2009. Intraday stealth trading: which trades move prices during periods of high volume? J. Financ. Res. 32, 1–21.
Campbell, J., Grossman, S., Wang, J., 1993. Trading volume and serial correlation in stock returns. Q. J. Econ. 31, 281–318.
Chakravarty, S., 2001. Stealth trading: which trader's trades move prices? J. Financ. Econ. 61, 289–307.
Chan, K., Christie, W., Schultz, P., 1995. Market structure and the intraday pattern of bid–ask spreads for NASDAQ securities. J. Bus. 68, 35–60.
Chen, Q., Goldstein, I., Jiang, W., 2007. Price informativeness and investment sensitivity to stock prices. Rev. Financ. Stud. 20, 619–650.
Duarte, J., Young, L., 2009. Why is PIN priced? J. Financ. Econ. 91, 119–138.
Durnev, A., Morck, R., Yeung, B., Zarowin, P., 2003. Does greater firm-specific return variation mean more or less informed stock pricing. J. Account. Res. 41, 797–836.
Durnev, A., Morck, R., Yeung, B., 2004. Value enhancing capital budgeting and firm-specific stock return variation. J. Financ. 59, 65–105.
Easley, D., O'Hara, M., 1987. Price, trade size, and information in securities markets. J. Financ. Econ. 19, 69–90.
Easley, D., Kiefer, N., O'Hara, M., Paperman, J., 1996. Liquidity, information and infrequently traded stocks. J. Financ. 51, 1405–1436.
Easley, D., Kiefer, N., O'Hara, M., 1997a. One day in the life of a very common stock. Rev. Financ. Stud. 10, 805–835.
Easley, D., Kiefer, N., O'Hara, M., 1997b. The information content of the trading process. J. Empir. Financ. 4, 159–186.
Easley, D., Hvidkjaer, S., O'Hara, M., 2002. Is information risk a determinant of asset returns? J. Financ. 57, 2185–2221.
Fama, E., MacBeth, J., 1973. Risk, return, and equilibrium: empirical tests. J. Polit. Econ. 81, 607–636.
Gerety, M., Mulherin, H., 1992. Trading halts and market activity: an analysis of volume at the open and the close. J. Financ. 47, 1765–1784.
Glosten, L., Milgrom, P., 1985. Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders. J. Financ. Econ. 14, 71–100.
Hansch, O., Choe, H., 2007. Which trades move stock prices? Stealth trading revisited. Working Paper. Pennsylvania State University.
Harris, L., 1986. A transaction data study of weekly and intraday patterns of stock returns. J. Financ. Econ. 16, 99–118.
Huang, R., Stoll, H., 1997. The components of the bid-ask spread: A general approach. Rev. Financ. Stud. 10, 995–1034.
Jain, P., Joh, G., 1988. The dependence between hourly prices and trading volume. J. Financ. Quant. Anal. 23, 269–283.
Kyle, A., 1985. Continuous auctions and insider trading. Econometrica 53, 1315–1335.
Lee, C., Ready, M., 1991. Inferring trade direction from intraday data. J. Financ. 46, 733–746.
Lei, Q., Wu, G., 2005. Time-varying informed and uninformed trading activities. J. Financ. Mark. 8, 153–181.
Morck, R., Yeung, B., Yu, W., 2000. The information content of stock markets: why do emerging markets have synchronous stock price movements? J. Financ. Econ. 59, 215–260.
Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Nyholm, K., 2002. Estimating the probability of i\nformed trading. J. Financ. Res. 25, 485–505.
Petersen, M., 2009. Estimating standard errors in finance panel data sets: Comparing approaches. Rev. Financ. Stud. 22, 435–480.
Roll, R., 1988. R^2. J. Financ. 43, 541–566.
Skoulakis, G., 2008. Panel data inference in finance: least-squares vs. Fama–MacBeth. Working paper. University of Maryland.
Wang, F., 1998. Strategic trading, asymmetric information and heterogeneous prior beliefs. J. Financ. Mark. 1, 321–352.
Wood, R., McInish, T., Ord, J., 1985. An investigation of transaction data for NYSE stocks. J. Financ. 40, 723–740.