Journal of Financial Markets 8 (2005) 153–181

Journal of FINANCIAL MARKETS

www.elsevier.com/locate/econbase

Time-varying informed and uninformed trading activities ☆

Qin Lei*, Guojun Wu

Department of Finance, Ross School of Business at the University of Michigan, 701 Tappan Street, Ann Arbor, MI 48109, USA

Available online 5 November 2004

Abstract

We develop a framework to investigate time-varying interactions between informed and uninformed trading activities. By estimating the model for 40 NYSE stocks, we demonstrate that the buy and sell arrival rates of the uninformed traders are different and time-varying. Informed traders strategically match the level of the uninformed arrival rate with a certain probability. Uninformed traders tend to adopt contrarian strategy in reaction to high prior stock returns, but employ momentum strategy in reaction to high prior market returns. The estimated time-varying probability of informed trading is a good predictor for various measures of bid–ask spreads, and is a better measure of information asymmetry than several existing measures.

JEL classification: D8; G14

Keywords: Information asymmetry; Arrival rates; Trading activities; PIN

☆We would like to thank Bruce N. Lehmann (the editor) and an anonymous referee for their insightful comments and suggestions. We also would like to thank Kerry Back, Sugato Bhattacharyya, David Hirshleifer, Gautam Kaul, Nicholas M. Kiefer, M. P. Narayanan, Venkatesh Panchapagesan, Paolo Pasquariello, Nejat Seyhun, Tyler Shumway, Anjan V. Thakor, James Weston and seminar participants at the Asia Pacific Finance Association meeting in Tokyo, the European Financial Management meeting in London, the Financial Econometrics Conference in Waterloo, Canada, University of Michigan, and Washington University in St. Louis for useful discussions. All errors are our own.

*Corresponding author. Tel.: +1 734 647 9597.

E-mail address: leiq@umich.edu (Q. Lei).

1386-4181/$ - see front matter
doi:10.1016/j.finmar.2004.09.002

1. Introduction

A crucial topic in market microstructure is the relationship between informed and uninformed trading activities. From the early observations of Bagehot (1971) to the theoretical work of Kyle (1985) and Easley and O'Hara (1987), researchers generally agree that informed traders exploit their informational advantage and trade optimally to profit from uninformed investors.

The interaction between asymmetrically informed traders, however, has been mostly investigated in theoretical frameworks.[1] The relatively few empirical studies mostly focus on cross-sectional analysis and use very short samples. Using ten years of transaction data, this paper provides empirical perspectives on the time-varying trading behavior of informed and uninformed traders. Moreover, this paper estimates a measure of time-varying probability of information-based trading, which could be a very useful tool in empirical market microstructure analysis. Specifically, we address the following questions: Does uninformed trading change over time? How does it change? Does the probability of information-based trading change over time? If yes, has this probability more explanatory power than competing measures of information asymmetry? Not surprisingly, we find that both informed and uninformed trading change dramatically over a ten-year period. The uninformed buy and sell arrival rates are different and time-varying. The estimated probabilities of information-based trading are closely related to contemporaneous bid–ask spreads and can predict spreads for the next trading day.

In this paper, we extend the seminal work of Easley et al. (1996), who inquire why there is a larger spread observed for less-frequently traded stocks than for active ones. In the absence of market power, market makers charge a bid–ask spread in stock transactions to recover losses to informed traders and inventory costs.[2] In Easley et al. (1996), informed traders submit buy orders upon receiving a positive information signal and submit sell orders upon a negative information signal. Uninformed traders, on the other hand, submit both buy and sell orders regardless of whether information arrives or what type of information arrives. By assuming that orders submitted by informed traders, uninformed buyers and uninformed sellers follow three independent Poisson processes with exogenous, fixed arrival rates, Easley et al. (1996) derive a very useful framework to examine the information

[1] Wang (1993) shows that information asymmetry among investors can increase price volatility. Less informed traders may rationally behave like price chasers, which may in turn increase market volatility. See also Diamond and Verrecchia (1981), Glosten and Milgrom (1985), Kyle (1985), Admati (1991), Easley and O'Hara (1992), and Easley et al. (1997).

[2] Bid–ask spread arises naturally due to both inventory (see Smidt, 1971; Garman, 1976; Zabel, 1981; Mendelson, 1982; O'Hara and Oldfield, 1986; Madhavan and Smidt, 1993) and asymmetric information concerns (Bagehot, 1971; Glosten and Milgrom, 1985; Easley and O'Hara, 1987). Glosten and Harris (1988) decompose the bid–ask spread into two parts: one part due to informational asymmetries, and the remainder attributable to inventory carry costs, market maker risk aversion, and monopoly rents. Using a maximum likelihood technique, they find that the adverse selection component of the bid–ask spread is not economically significant for small trades, but increases with trade size. See also Hasbrouck (1991a,b), Barclay et al. (1990), Madhavan and Smidt (1991, 1993), Jones et al. (1994), and Madhavan and Sofianos (1998).

content of stock trading.[3] They find that the risk of information-based trading is lower for active stocks than it is forinfrequently traded securities. Although high-volume stocks tend to have higher probabilities of information events and higher arrival rates of informed traders, they are more than offset by the higher arrival rates of uninformed traders. From the perspective of the market maker, less active stocks are riskier since there is a higher probability that any trade comes from an informed trader.

It seems restrictive, however, to assume a constant arrival rate for uninformed traders since the amount of uninformed trading could change dramatically under different market conditions. For instance, with the advent of internet trading, hundreds of thousands of small investors traded technology stocks in the late 1990s when the NASDAQ Composite Index reached over 5,000. The subsequent stock market correction brought the index down to 1200 in mid 2002. Many small investors reduced their trading activities, and the market meltdown in technology stocks forced many former day-traders to quit trading altogether. The time-invariant arrival rates also restricted Easley et al. (1996) to study a short sample covering less than three months of daily trading data, since a longer sample period would make it implausible to assume constant arrival rates.

The level of uninformed buy and sell arrival rates may be affected by momentum (e.g., Jegadeesh and Titman, 1993, 1995, 2001; Chan et al., 1996; Hong and Stein, 1999; Rouwenhorst, 1999; Caginalp et al., 2000) and contrarian strategies (e.g., Lo and MacKinlay, 1990; Lakonishok et al., 1994). It may also be affected by factors such as investor sentiment (e.g., Siegel, 1992, Barberis et al., 1998), overconfidence (e.g., Daniel et al., 1998), loss aversion, and mental accounting (e.g., Barberis and Huang, 2001; Barberis et al., 2001). We assume that the arrival rate of uninformed buy orders switches between two levels in a Markov process, with endogenous time-varying transition probabilities. We model the difference between buy and sell arrival rates for uninformed traders to be time-varying and dependent on market variables such as lagged cumulative returns. Our framework can be generalized to more than two states, but we find that the parsimonious assumption of two states is sufficient to capture the main effects of time-varying uninformed trading.

We argue that informed traders closely monitor market movements and can respond rationally to any change in the arrival rate of uninformed traders. The theoretical frameworks in Glosten and Milgrom (1985) and Easley et al. (1996) assume that traders are chosen probabilistically to submit an order of one share at each session, so informed traders cannot respond to camouflage provided by the uninformed traders. Our framework allows informed traders to match the arrival rate of uninformed investors by assuming that the arrival rate of informed traders also can switch between two levels, with the transition probabilities reflecting informed matching activities (to be defined shortly).

Kyle (1985) and Back (1992) demonstrate elegantly the strategic trading behavior of informed traders in a risk-neutral environment, and the setup of this paper

[3]The Easley et al. (1996) framework was extended by Weston (2001) to include a class of discretionary liquidity traders.

amounts to an empirical test of their theoretical predictions. We can examine how well informed traders use the camouflage provided by uninformed investors. With the endogenously determined arrival rates for both types of traders, we are able to evaluate the evolution of information content in daily stock trading. Easley et al. (1996) perform a cross-sectional study on the relationship between the probability of information content and the daily opening spread. Since we use a much longer sample and allow time-varying probability of information-based trading, we are able to conduct the analysis in both the time series and the cross-sectional dimensions. Moreover, we can study the predictability of the estimated probabilities of information-based trading for various measures of stock spreads.

Our empirical estimation of 40 NYSE stocks shows that the uninformed traders tend to adopt contrarian strategy in reaction to high prior own stock returns, but employ momentum strategy in reaction to high prior market returns. Informed traders seem to take good advantage of the camouflage since the estimated probability of informed matching response ranges from 0.72 to 0.98. The estimated probability of informed trading has predictive power over mean bid–ask spreads in the next trading day, and dominates competing measures of information asymmetry in terms of explanatory power. Moreover, we use the estimated time-varying probability of informed trading series as a proxy for informational asymmetry to analyze its impact on the serial correlation of daily stock returns. Consistent with existing research on the issue with other measures of informational asymmetry such as firm size and the bid–ask spread, we find that higher probability of informed trading is associated with higher return autocorrelation.

The rest of the paper is organized as follows. Section 2 extends the framework in Easley et al. (1996) by allowing the uninformed arrival rates to be time-varying. Informed traders may match the level of the uninformed arrival rate with certain probability so as to make better use of the camouflage provided by uninformed traders. In Section 3, we discuss the sample selection and other data-related issues. Section 4 analyzes the maximum-likelihood estimation results and the cross-sectional variation of matching response by informed traders. In Section 5, we examine the predictability of the estimated probabilities of informed trading for various measures of spreads, and the effect of information asymmetry on return autocorrelation. Section 6 concludes the paper. Some technical details are provided in the appendix.

2. The model

There is one risky asset and one risk-free asset (as the *numeraire*) in the market. The risk-free rate is set to be zero for simplicity. At the beginning of each trading day $t \in [1, T]$, in which time is continuously indexed by $0 < i < 1$, nature decides whether to release a news event concerning the value of the risky asset to informed traders. The probability of a news event occurring is α, and when the news event arrives it may be bad news with probability δ or good news with probability $1 - \delta$. Hence the prior probabilities of having a negative, positive, and no news event are $p(-) \equiv \alpha\delta$, $p(+) \equiv \alpha(1 - \delta)$ and $p(\times) \equiv 1 - \alpha$, respectively. The buy and sell trades from the uninformed traders as well as the one-sided trades (buy at good news and sell at bad news) from the informed traders are assumed to follow three mutually independent Poisson processes.

2.1. Two-state Markov switching for the uninformed arrival

On each trading day, the uninformed arrival rate is assumed to follow a two-state Markov switching process, with the time-varying transition probabilities governed by factors such as prior stock return and prior market return. The uninformed arrival rate can be at the high level $\varepsilon_t^{\mathrm{h}}$ or at the low level $\varepsilon_t^{\mathrm{l}}$ on trading day t. Note that the high and low levels of the arrival rate are constant for the entire sample, despite the time subscript they carry for notational clarity.

Define the time-varying transition probabilities for the uninformed traders as

$$\pi_t \equiv \begin{bmatrix} \pi_t^{**} & 1-\pi_t^{**} \\ 1-\pi_t^{*} & \pi_t^{*} \end{bmatrix},$$

where

$$\pi_t^{*} \equiv \Pr(\varepsilon_t^{\mathrm{l}} \mid \varepsilon_{t-1}^{\mathrm{l}}) = f(\beta_{\mathrm{l}}'\mathbf{z}_t) \quad \text{and} \quad \pi_t^{**} \equiv \Pr(\varepsilon_t^{\mathrm{h}} \mid \varepsilon_{t-1}^{\mathrm{h}}) = f(\beta_{\mathrm{h}}'\mathbf{z}_t). \tag{1}$$

In the definitions above, π_t^{*} is the probability of the uninformed arrival rate transiting from the low level at period $t-1$ to the low level at period t, and π_t^{**} is the probability transiting from the high level at period $t-1$ to the high level at period t. $\mathbf{z}_t$ is a vector of instruments that are observable at the end of period $t-1$. If we use a vector of ones as the only instrument, then the model yields constant transition probabilities. Non-constant instruments, such as cumulative asset return and cumulative market return (value-weighted NYSE/AMEX/NASDAQ returns) in the previous 20-trading-day period, allow the modeling of time-varying transition probabilities. We use the logistic transformation $f(\cdot)$ to ensure the transition probabilities have appropriate values. β_{h} and β_{l} are the parameters associated with the instruments.

Uninformed investors may base their buying and selling activities on publicly observable information, such as past stock and market returns. We allow the uninformed buy arrival rate to be different from the uninformed sell arrival rate. However, the modeling of two separate uninformed arrival rates, each with its own Markov switching process, forces us to track the interaction between them. As a consequence, it significantly reduces the tractability of the empirical model and renders the estimated results very difficult to interpret. We take a modeling compromise, and allow only the uninformed buy arrival rate ($\varepsilon_t^{\mathrm{h},B}$ for the high state and $\varepsilon_t^{\mathrm{l},B}$ for the low state) to follow the aforementioned two-state Markov switching process. The uninformed sell arrival rate ($\varepsilon_t^{\mathrm{h},S}$ for the high state and $\varepsilon_t^{\mathrm{l},S}$ for the low state) switches into the same state as the uninformed buy arrival rates, but the level of the two arrival rates may differ. In particular, we assume that

$$\varepsilon_t^{\mathrm{h},S} = \varepsilon_t^{\mathrm{h},B}\exp(\gamma_{\mathrm{h}}'\mathbf{z}_t) \quad \text{and} \quad \varepsilon_t^{\mathrm{l},S} = \varepsilon_t^{\mathrm{l},B}\exp(\gamma_{\mathrm{l}}'\mathbf{z}_t). \tag{2}$$

Note that we are using the same set of instruments $\mathbf{z}_t$ here as in the context of modeling time-varying transition probabilities. The assumption that the uninformed buy and sell arrival rates will switch to the same state simultaneously, either high or low, is not as restrictive as it seems. Researchers on the trading behavior of small investors find that retail trading activities increase after up-markets (Odean, 1999; Grinblatt and Keloharju, 2001). They are more likely to sell stocks to realize gains and buy stocks because of overconfidence. During down-markets, retail investors are reluctant to sell to realize losses due to loss aversion, and they buy less due to a lack of interest or attention (Odean, 1998; Barber and Odean, 2003). Our simple structure also brings forth multiple benefits: achieving empirical tractability and ease of interpretation, and allowing us to focus on the extent of (percentage) difference between the uninformed buy and sell arrival rates in response to changes in market fundamentals. We call this difference "uninformed arrival rate differential" henceforth.

Since the amount of uninformed trading can vary dramatically from time to time, our switching model should improve the Easley et al. (1996) model that has constant arrival rates. A recent study by Easley et al. (2002a) also relaxes the assumption of fixed arrival rates in Easley et al. (1996) by using two GARCH specifications. We use a Markov-switching model instead to investigate how the behavior of the uninformed investors affects the strategic trading behavior of the informed traders. Our use of the Markov switching model is motivated by the observation that uninformed traders become more or less interested in trading depending on past performance of the stock and the overall stock market. Our focus is the interaction between informed and uninformed trading, and a Markov switching process makes it easy to estimate and interpret the interaction. In the next sub-section, we extend the model further to examine how well informed traders make use of the camouflage from uninformed traders by allowing informed traders to engage in level-matching activities. The Markov-switching model makes it a straightforward and easy task to interpret the matching probabilities. Our framework also can easily accommodate the use of instruments in the modeling of transition probabilities, and it provides a very intuitive interpretation for the estimated relationship.

2.2. Level-matching activities by informed traders

Informed traders may adjust their trading activities according to the amount of camouflage provided by uninformed traders (e.g., Kyle, 1985). In the market microstructure model of Glosten and Milgrom (1985) traders cannot trade more than one unit of the asset per period. Without this assumption, informed traders will buy (sell) an infinite amount upon the arrival of good (bad) news. So we preserve the assumption of unit trade in this paper and attempt to model the strategic behavior of informed traders by studying the fraction of time when both types of traders make synchronized moves in terms of the state level of arrival rates. That is, we allow the informed arrival rate on day t to take on a high level μ_t^{h} (or a low level μ_t^{l}), corresponding to the high level $\varepsilon_t^{\mathrm{h}}$ (or low level $\varepsilon_t^{\mathrm{l}}$) of the uninformed arrival rates, a situation we call "level-matching activities" in this paper. The empirically estimated probability of the matching response should shed some light on how well the informed traders are making use of the camouflage.

There are four possible combinations of the uninformed and the informed arrival rates in our model, denoted as $s_t^a \equiv (\varepsilon_t^{\mathrm{h}}, \mu_t^{\mathrm{h}})$, $s_t^b \equiv (\varepsilon_t^{\mathrm{h}}, \mu_t^{\mathrm{l}})$, $s_t^c \equiv (\varepsilon_t^{\mathrm{l}}, \mu_t^{\mathrm{h}})$, and $s_t^d \equiv (\varepsilon_t^{\mathrm{l}}, \mu_t^{\mathrm{l}})$.[4] Clearly, the two states s_t^a and s_t^d are the matching states and the rest are non-matching states. Now we have a modeling choice in terms of tracking the interaction between the activities of the informed traders and those of the uninformed traders. Though it is not impossible to impose a Markov-switching structure on the four composite states, there is no easy way to extract intuition regarding the informed–uninformed interaction from the complex four-state transition matrix. This is a problem similar to the one we face when modeling the relationship between the uninformed buy arrival rate and the uninformed sell arrival rate. Yet we must not resort to the same solution as before because an assumption that the informed traders move in a fully synchronized fashion with the state level of the uninformed traders completely destroys our goal of uncovering the extent of strategic movement by the informed traders. Given that a hard-wired complete synchronization between the informed and the uninformed traders is out of the question, we decide to impose some structural restrictions on the four-state transition matrix such that the informed traders match the level of their arrival rate with the level of the uninformed arrival rate only a fraction of the time. We hope to study the estimated value of this fraction for matching activities, a parameter we call "probability of matching response", to draw useful inferences on the informed traders' strategic behavior, while allowing the uninformed arrival rate to switch freely according to the two-level Markov process described earlier.

Provided that the informed traders know the level of the uninformed arrival rate at the beginning of each trading day, they can make a decision on how to exploit the information advantage they have. In particular, we define the probability of matching response as

$$\Pr(\mu_t^{\mathrm{h}} \mid \varepsilon_t^{\mathrm{h}}) = \Pr(\mu_t^{\mathrm{l}} \mid \varepsilon_t^{\mathrm{l}}) \equiv \rho. \tag{3}$$

Moreover, the time-varying probability of transiting from state s_{t-1}^a to s_t^a is equal to

$$\begin{aligned}\Pr(s_t^a \mid s_{t-1}^a) &= \Pr(\varepsilon_t^{\mathrm{h}}, \mu_t^{\mathrm{h}} \mid \varepsilon_{t-1}^{\mathrm{h}}, \mu_{t-1}^{\mathrm{h}}) \\ &= \Pr(\varepsilon_t^{\mathrm{h}} \mid \varepsilon_{t-1}^{\mathrm{h}}) \cdot \Pr(\mu_t^{\mathrm{h}} \mid \varepsilon_t^{\mathrm{h}}, \varepsilon_{t-1}^{\mathrm{h}}, \mu_{t-1}^{\mathrm{h}}) = \pi_t^{**}\rho.\end{aligned} \tag{4}$$

The first equal sign follows the definition of composite states. The second equal sign applies because the uninformed traders don't have knowledge about the current level of the informed arrival rate and we assume that the uninformed arrival rates switch freely according to the two-state Markov process. Implicitly, it is assumed that the trading behavior of the uninformed investors is governed by whatever information variables we use for the uninformed transition probabilities and the uninformed arrival rate differential. The third equal sign follows from the definition of matching

[4]Note that we use the abbreviated notation for the uninformed arrival rates, without distinguishing the uninformed buy and sell arrival rates, due to the assumption of adaptive movement in Eq. (2). We use the abbreviated notation whenever it is not likely to cause confusion.

response probability as well as the assumption that the informed traders pay no attention to arrival rates in the previous trading day.

To get a better grasp of the parameter ρ, we may analyze its inherent limitations. First, its existence hinges upon the independence assumption about the uninformed investors. If the uninformed investors have knowledge of the informed arrival rate and use this knowledge to decide their choice of arrival rate and buy–sell differential, then the second equal sign in the equation above does not go through and the so-called matching response probability cannot stand. One way of getting around this problem is to re-classify the trader types. That is, any uninformed traders who are sophisticated enough to infer the informed arrival rate and actively use that inference for trading decisions should probably be classified as informed traders.[5]

Second, given that the parameter ρ is a statistical construct, arising from the decomposition of the transition probability of composite states, we are still not very sure of its path-independent nature and its constancy. The path-independent assumption that the matching response has no reliance on previous arrival rates can be partially motivated by the information advantage the informed traders possess because they submit orders based upon private signals that may arrive on each trading day. So it is reasonable to assume that the informed investors do not care about their own arrival rate in the previous trading day. But given that the informed investors care about the existing level of uninformed arrival rates, which is related to the uninformed arrival rate in the previous period, why would the informed investors ignore the uninformed arrival rate in the previous trading day? Moreover, what makes us believe that the probability of matching response as defined is constant throughout the sample period? To address these two questions, we have estimated more complex versions of the model on a subset of our sample stocks. In particular, we allow the matching response to be dependent upon the uninformed arrival rate in the previous period, such as $\Pr(\mu_t^{\mathrm{h}} \mid \varepsilon_t^{\mathrm{h}}, \varepsilon_{t-1}^{\mathrm{h}}) \neq \Pr(\mu_t^{\mathrm{h}} \mid \varepsilon_t^{\mathrm{h}}, \varepsilon_{t-1}^{\mathrm{l}})$, and even make the probability of matching response vary over time. We find that the path-dependent matching response probabilities are not significantly different from each other, and the sample standard deviation of the time-series of matching response is very small. So we conclude these two assumptions related to the construction of ρ are reasonable.

Admittedly, the probability of matching response as proposed is an imperfect way of describing the interaction between the informed traders and the uninformed. Nevertheless we want to examine whether the estimated probability of matching response by the informed traders is statistically significant and whether its magnitude is as large as suggested by theory. If ρ is close to 1, then we have evidence that the informed traders are using the camouflage of the uninformed traders. Admati and Pfleiderer (1988) present a theory in which concentrated trading patterns arise

[5]The argument for reclassifying trader types is stated from a theoretical perspective, and we do not attempt to empirically classify each trade into the informed or the uninformed categories. The distinction between the arrival rates for these two types of traders is achieved via the sample likelihood function (to be discussed later), utilizing the fact that the informed traders submit trades on only one side based upon the type of private signal they receive.

endogenously as a result of the strategic behavior of (uninformed) noise traders and the informed traders. Although their theory is used mainly to explain intra-day trading patterns, our model presents an empirical test of their results for inter-day patterns (see also Back and Pedersen, 1998).

Finally, the time-varying nature of the uninformed transition probabilities and buy–sell arrival rate differential allows us to deliver a time series measuring the probability of information-based trading. That is, the probability that each trade on day t is information-based can be expressed as

$$\mathrm{TPIN}_t = \frac{\alpha\mu_t^{\mathrm{e}}}{\varepsilon_t^{\mathrm{e},B} + \varepsilon_t^{\mathrm{e},S} + \alpha\mu_t^{\mathrm{e}}}. \quad (5)$$

In this definition, α is the probability of news arrival; μ_t^{e} is the expected arrival rate for the informed traders on trading day t; $\varepsilon_t^{\mathrm{e},B}$ is the expected buy arrival rate for the uninformed traders on trading day t; and $\varepsilon_t^{\mathrm{e},S}$ is the expected sell arrival rate for the uninformed traders on trading day t. This definition is identical in spirit to the constant version of PIN defined in Easley et al. (1996), and it has an intuitive interpretation as the fraction of informed trades among all trades. The short-hand TPIN is used in this paper to signify its time-varying nature, as opposed to the constant PIN. The technical details about the composite transition matrix as well as the definition of the expected arrival rates are provided in the appendix.

2.3. Trading process and estimation methodology

The basic structure of the trading process on a typical trading day is depicted in Fig. 1. Nodes above the dotted line occur only once per trading day, and nodes on the dotted line are repeated many times within each trading day, following independent Poisson processes by assumption. Depending upon the previous level $\varepsilon_{t-1}^{?}$ of the uninformed arrival, the uninformed arrival may transit into the high level with probability $p(\varepsilon_t^{\mathrm{h}} \mid \varepsilon_{t-1}^{?})$ or the low level with probability $p(\varepsilon_t^{\mathrm{l}} \mid \varepsilon_{t-1}^{?})$. The informed traders engage in level-matching activities with a constant probability ρ. There may be no news arrival with probability $p(\times)$, good news arrival with probability $p(+)$, or bad news with probability $p(-)$. The last row of numbers is the sum of the expected arrival rate of the informed and the uninformed traders under each scenario, with notational abbreviation for the uninformed arrival rates.

Fig. 1 forms the basis of our empirical estimation. In the appendix, we construct the likelihood function in a computation-efficient way. Given the likelihood function, all the parameters can be estimated using an optimization procedure such as Maxlik in GAUSS, the statistical package we use for this study. Due to the one-sided nature of the informed trades, the sample log-likelihood function helps us distinguish the informed arrival rates from the uninformed arrival rates. Although it is not possible to empirically identify the particular composite state for each trading day, we can nevertheless make a probabilistic statement about the informed and the uninformed arrival rates for each trading day for the entire sample.

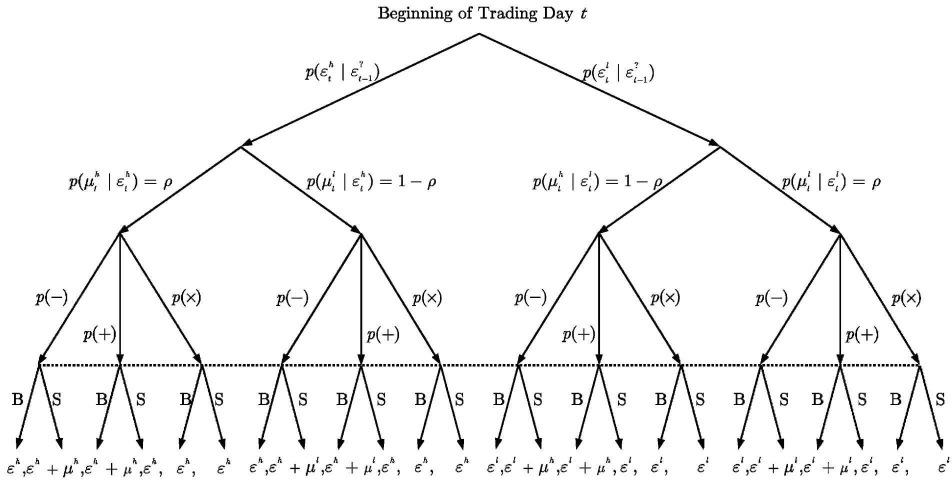

Fig. 1. Structure of trading process on a typical day.

In the figure, ε^{h} and ε^{l} are the high and low levels of the uninformed arrival rate, μ^{h} and μ^{l} are the high and low levels of the informed arrival rate. Depending upon the arrival rate level $\varepsilon_{t-1}^{?}$ in the previous day, the uninformed arrival may transit into the high state with probability $(\varepsilon_t^{\mathrm{h}} \mid \varepsilon_{t-1}^{?})$ or the low state with probability $p(\varepsilon_t^{\mathrm{l}} \mid \varepsilon_{t-1}^{?})$. The informed traders engage in level-matching activities with probability ρ. There may be no news arrival with probability $p(\times)$, good news arrival with probability $p(+)$, or bad news with probability $p(-)$. Nodes above the dotted line occur only once per trading day and nodes on the dotted line are repeated many times each day. The buy and sell orders are indicated by B and S, respectively. The last row indicates the aggregate, expected arrival rate of traders under each scenario, with the buy and sell nature for ε^{h} and ε^{l} abbreviated.

3. Data construction

Our sample of firms was selected from 611 common shares listed on the NYSE (Share Code 10 or 11) with complete monthly returns data from the Center for Research in Securities Prices (CRSP) database and trading data from the New York Stock Exchange Trade and Quote (NYSE TAQ) database. These stocks have no change in their respective ticker symbols during the ten-year period between 1993 and 2002. We sort these stocks into 10 deciles according to the monthly average turnover ratio (the number of shares traded to the shares outstanding) and randomly pick 10 stocks from each of the 2nd, 4th, 6th and 8th turnover deciles. We use turnover as a sample selection criterion because Easley et al. (1996) demonstrate that low volume is indicative of high information asymmetry and turnover has been used as a measure of volume in the literature (see, for example, Blume et al., 1994; Lee and Swaminathan, 2000; Lo and Wang, 2000).

Before matching trades with quotes, we collapse the closely adjacent (within five seconds) trades that were executed at the same price without intervening revision of quotes. This is done to mitigate the problem of misclassifying a large trade on one side (buy or sell), which involved multiple participants, as separate trades. We also allow for a systematic delay of five seconds before each submitted quote was time-stamped. We treat as relevant only the most recent quote that was at least five seconds old relative to the recorded trade time.

We follow the methodology proposed by Lee and Ready (1991) in determining if a trade is buyer-initiated or seller-initiated. Trades with transaction prices above the midpoint of the relevant bid and ask are called buy trades, and trades with transaction prices below the midpoint are called sell trades. Trades with a price at the midpoint of bid and ask are classified with a "tick test", i.e., trades with a price higher (or lower) than the most recent trade with a different price are called buy (or sell) trades.

Table 1 reports some summary statistics about the stocks in our sample. The summary statistics show that there are a lot of variations in firm characteristics, both within each turnover decile and across the turnover deciles. Stocks with higher daily mean turnover tend to have higher stock prices and more transactions (both buy trades and sell trades). The mean stock price ranges from \$4.20 to \$63.88, so our sample does not include penny stocks. The variation in the sample standard deviation for the variables presented in Table 1 has a pattern similar to the variation in the means of these variables.

4. Estimation results

In this section, we discuss properties of the estimation for each stock in our sample as well as the implied time series related to the information content of stock trading. We defer to the next section some applications of the estimated probability of informed trading.

4.1. Maximum likelihood estimates

Table 2 presents the maximum likelihood estimates for all the constant parameters in our model including the probability of news arrival, the probability of arrived news being negative, different levels of the informed and the uninformed arrival rates, and the probability of matching response. These constant parameters are estimated using the entire sample for each stock, and they are highly statistically significant at the 1% level.[6] There is a large variation for the estimated parameters across stocks. For example, the probability of news arrival ranges between 0.28 for stock ticker AGL and 0.52 for stock ticker LUB. The high level of the uninformed buy arrival rate can be as high as 101 trades per day for stock ticker OII, or as low as 9 trades for stock ticker SL. The high level of informed arrival rate varies between 92 trades for stock ticker KMT and 16 trades for stock ticker SL. We do not report the levels of the uninformed sell arrival rates, because we model the sell arrival rates at the same state level as the uninformed buy arrival rates, while allowing the difference

[6]We omitted reporting the standard errors for these parameter estimates in order to conserve space, and the complete estimation results are available upon request.

Table 1
Summary statistics

In this table, turnover is the ratio of daily share volume to shares outstanding (in basis points); is the average daily dollar price; nbuy is the average number of daily buy orders; and n sell is the average number of daily sell orders. The sample period is from January 1, 1993 to December 31, 2002, and nobs is the number of daily observations available for each stock in the respective turnover decile.

Decile	Ticker	nobs	Mean				Standard deviation			
			turnover	prc	nbuy	nsell	turnover	prc	nbuy	nsell
2	EBF	2494	18.60	11.16	9.85	9.72	24.04	2.51	13.10	10.77
	LDL	2489	18.68	18.88	9.20	9.73	27.95	8.32	9.79	9.31
	OCQ	2496	15.50	17.91	9.28	9.91	20.53	5.83	9.85	8.12
	OXM	2489	17.55	23.31	7.55	7.64	21.00	5.50	7.14	6.69
	RUS	2495	16.02	21.28	12.92	13.51	20.94	6.29	13.58	11.90
	SL	2376	15.51	8.50	3.32	3.73	22.55	3.48	4.44	4.13
	SMP	2488	16.21	17.00	8.23	8.10	20.04	4.63	7.75	6.55
	STL	2453	15.42	17.18	7.87	7.81	18.00	7.92	10.56	9.22
	TYL	2469	16.16	4.20	9.35	8.81	28.24	1.94	14.66	10.59
	VCD	2496	15.23	10.29	11.29	10.21	21.00	4.48	10.88	8.46
4	AGL	2495	24.87	18.69	11.62	10.73	37.26	6.08	15.34	12.29
	AIZ	2494	25.83	16.38	7.60	7.74	29.70	6.40	6.83	6.36
	BW	2496	24.24	16.98	14.21	14.16	40.16	4.27	15.74	13.76
	CPY	2494	25.23	19.84	9.50	9.88	34.02	4.46	8.92	7.22
	DVI	2491	24.96	13.98	10.61	10.80	40.22	4.03	13.74	11.42
	IMC	2495	25.86	20.98	19.41	18.70	27.23	4.24	20.56	17.90
	LUB	2496	25.94	16.17	19.46	19.90	26.32	6.69	12.12	11.70
	PWN	2496	27.13	9.06	17.12	17.14	35.50	2.70	17.73	13.93
	RML	2496	24.74	24.28	31.05	29.35	33.46	6.39	24.51	18.34
	UFI	2496	25.09	21.44	32.43	30.13	29.67	10.65	27.70	22.87
6	AIR	2496	32.21	18.04	22.51	20.83	37.62	7.97	21.51	16.62
	BDG	2496	32.80	43.80	23.60	20.28	31.38	11.72	21.42	15.76
	ESL	2495	33.84	18.91	21.36	19.63	40.54	8.05	25.38	21.70
	FRC	2472	34.70	20.86	11.85	11.13	55.95	7.23	16.34	14.48
	HUF	2496	32.86	12.22	9.65	11.07	46.53	4.13	11.06	8.66
	HXL	2490	36.55	10.51	15.38	15.01	60.30	6.93	15.47	13.75
	LZ	2496	33.87	31.12	58.03	54.65	28.98	4.86	48.11	42.31
	NC	2495	34.06	63.88	24.46	22.09	45.05	24.21	21.30	18.98
	OMI	2496	31.52	15.04	27.66	23.24	37.32	3.85	30.80	21.66
	RTI	2468	35.04	12.25	19.99	19.09	56.77	7.31	20.64	17.41
8	ALN	2496	45.04	17.95	29.63	26.37	53.92	9.73	28.07	18.89
	BGG	2496	44.36	49.17	54.83	47.05	41.68	14.16	43.68	34.41
	KMT	2496	44.72	34.69	48.75	43.27	53.17	8.83	46.30	39.14
	NEV	2493	49.91	23.20	24.98	23.40	66.11	10.32	23.83	19.58
	OII	2496	44.46	16.55	35.07	31.91	40.52	4.73	44.18	38.21
	OS	2496	41.40	13.76	22.83	22.95	50.81	7.18	19.45	15.60
	POP	2496	39.86	16.19	18.22	16.17	48.25	4.65	19.35	14.01
	SRR	2496	46.47	10.61	39.27	37.71	46.99	3.57	28.94	20.45
	WLMN	2496	41.04	19.09	37.25	34.68	36.84	5.12	30.13	22.16
	WWW	2496	46.01	20.71	40.31	35.15	49.53	7.82	39.19	31.70

Table 2
Maximum likelihood estimation results

In this table, α is the probability of news arrival; δ is the probability of the arrived news being negative; $\varepsilon^{h,B}(\varepsilon^{l,B})$ is the uninformed arrival rates for buyers at the high (low) state; $\mu^h(\mu^l)$ is the informed arrival rate at the high (low) state; and ρ is the probability of informed matching response. The estimates for these variables are significant at the 1% level. TPIN is the sample mean expected probability of information-based trading, and λ_1 and λ_2 are likelihood ratio statistics testing against our model specification (see the text) that are asymptotically distributed as $\chi^2(8)$ and $\chi^2(4)$, respectively.

Decile	Ticker	α	δ	$\varepsilon^{h,B}$	$\varepsilon^{l,B}$	μ^h	μ^l	ρ	TPIN	λ_1	λ_2
2	EBF	0.36	0.34	25.31	4.61	35.62	7.40	0.98	0.21	835.39	236.58
	LDL	0.43	0.47	18.37	4.23	24.12	7.29	0.93	0.25	242.13	54.61
	OCQ	0.37	0.28	13.58	2.99	21.79	6.93	0.85	0.23	443.84	251.00
	OXM	0.36	0.45	14.53	3.81	18.22	6.60	0.93	0.22	93.04	22.65
	RUS	0.39	0.31	25.33	5.31	33.08	9.84	0.94	0.22	748.55	144.16
	SL	0.31	0.46	8.90	1.67	15.60	4.05	0.97	0.24	173.50	66.24
	SMP	0.37	0.35	14.17	3.82	19.59	6.87	0.95	0.22	39.91	28.56
	STL	0.32	0.36	18.65	2.57	24.95	6.68	0.95	0.22	244.42	105.48
	TYL	0.31	0.31	19.96	2.32	41.37	8.09	0.93	0.28	286.70	106.32
	VCD	0.43	0.31	19.38	4.62	24.18	8.06	0.87	0.25	315.79	156.45
4	AGL	0.28	0.40	39.68	6.79	55.96	10.34	0.98	0.18	692.26	596.80
	AIZ	0.45	0.28	11.70	3.50	18.33	6.29	0.89	0.26	77.91	72.18
	BW	0.37	0.30	29.66	5.93	40.67	10.44	0.95	0.22	814.49	57.26
	CPY	0.39	0.46	19.40	5.49	28.07	7.93	0.95	0.22	206.90	125.60
	DVI	0.35	0.38	25.53	4.39	40.92	9.63	0.94	0.25	576.52	389.84
	IMC	0.38	0.31	41.86	8.58	46.73	13.10	0.92	0.20	172.80	121.96
	LUB	0.52	0.25	26.65	11.11	30.79	10.74	0.87	0.21	243.46	205.67
	PWN	0.42	0.33	31.40	7.81	44.85	11.63	0.90	0.24	602.88	85.66
	RML	0.51	0.23	46.52	13.73	50.50	17.22	0.86	0.23	426.97	360.37
	UFI	0.49	0.25	49.42	13.81	62.12	19.09	0.83	0.25	620.56	577.18
6	AIR	0.42	0.18	31.11	6.26	45.46	13.75	0.81	0.25	469.25	113.62
	BDG	0.43	0.17	40.94	10.14	48.01	13.83	0.90	0.22	519.22	264.08
	ESL	0.46	0.16	44.05	5.33	47.75	12.99	0.90	0.27	1325.74	1277.89
	FRC	0.35	0.32	29.26	3.50	36.46	9.47	0.94	0.24	658.54	501.07
	HUF	0.42	0.33	22.28	4.88	42.20	8.98	0.96	0.25	204.52	35.34
	HXL	0.45	0.27	23.61	4.87	32.05	11.07	0.83	0.27	313.74	182.57
	LZ	0.48	0.15	96.23	22.14	72.44	26.99	0.82	0.19	1548.07	196.63
	NC	0.40	0.31	40.43	9.21	43.35	14.39	0.90	0.20	188.02	183.38
	OMI	0.40	0.24	61.43	10.31	52.90	14.60	0.90	0.20	1358.22	139.73
	RTI	0.37	0.34	24.11	3.39	57.65	14.78	0.59	0.30	430.72	430.47
8	ALN	0.40	0.31	46.37	13.10	85.33	18.75	0.83	0.27	375.70	366.71
	BGG	0.40	0.27	83.67	25.56	87.92	27.05	0.79	0.19	447.83	446.98
	KMT	0.45	0.21	83.23	17.35	91.53	26.75	0.87	0.23	1596.04	507.39
	NEV	0.42	0.23	31.37	5.25	54.20	14.86	0.72	0.27	189.57	182.79
	OII	0.48	0.14	100.55	10.18	65.84	21.00	0.85	0.23	370.01	311.40
	OS	0.39	0.42	34.51	11.39	51.33	15.47	0.83	0.23	290.04	282.50
	POP	0.40	0.24	35.00	8.16	48.63	12.67	0.93	0.24	739.86	44.98
	SRR	0.49	0.28	51.86	20.55	69.89	20.53	0.83	0.22	107.43	106.18
	WLMN	0.43	0.29	60.62	20.30	77.18	21.29	0.89	0.20	267.92	85.06
	WWW	0.43	0.38	70.56	14.01	64.51	21.73	0.82	0.21	409.94	395.67

between the uninformed buy and sell arrival rates to be time-varying and responsive to market fundamentals. The sample mean TPIN ranges from 0.18 for stock ticker AGL to 0.30 for stock ticker RTI, whereas the constant probability of matching response by the informed traders varies between 0.72 for stock ticker NEV and 0.98 for stock tickers EBF and AGL. The matching response is very close to 1, a level indicative of highly synchronized movement between the uninformed and the informed traders. Therefore, informed traders do make good use of the camouflage provided by uninformed trades, as suggested by Kyle (1985), Admati and Pfleiderer (1988), and Back and Pedersen (1998).

We note with interest that the estimated probability of matching response is negatively correlated with the estimated probability of informed trading among our sample of 40 stocks, with a correlation coefficient −0.39. Consistent with the finding in Easley et al. (1996) that less frequently traded stocks have more asymmetric information, we obtain a cross-sectional correlation of −0.31 between the estimated probability of informed trading and the dollar volume. To understand better the relationship between the informed matching activities and firm characteristics, we estimate a cross-sectional regression as follows,

$$\rho_i = \underset{\underset{(65.21)}{0.9442}}{\beta_0} + \underset{\underset{(-4.23)}{-6.93E-04}}{\beta_1} \mathrm{INVOL}_i + \underset{\underset{(3.35)}{1.43E-04}}{\beta_2} \mathrm{UNVOL}_i + \varepsilon_i$$

$$\text{where } i = 1, \ldots, 40. \tag{6}$$

In this equation, the independent variable ρ is the estimated probability of matching response, and the informed dollar volume INVOL is the product of sample mean TPIN and average dollar volume (in thousands of dollars). The uninformed dollar volume UNVOL is the product of implied probability of uninformed trading, i.e., sample mean of 1 – TPIN, and the average dollar volume. The estimates are listed directly below the coefficients with t-stats in parentheses. The adjusted R^2 for this regression is 0.44 and the F-stat for three coefficients being zero jointly is 16.03.

The highly significant estimated coefficients for this regression design suggest that the extent of matching activities by the informed traders is higher among stocks with large uninformed dollar volume, while the increased presence of the informed traders undercuts the efforts of matching. In the absence of a rigorous model stipulating contributors to the matching response, it seems a plausible conjecture that the higher competition among informed traders makes it harder for all informed traders as a group to match with the uninformed trading activities in a synchronized fashion. On the other hand, an increased level of uninformed trading provides more camouflage for informed traders to undertake matching response activities.

4.2. *Influence of market fundamentals*

Details about the estimates related to information variables, including cumulative prior-20-trading-day stock return (CDR) and cumulative prior-20-trading-day market return (CMR), are presented in Fig. 2. A constant is also included as part of the instrument matrix for the uninformed transition probabilities (UT) and the

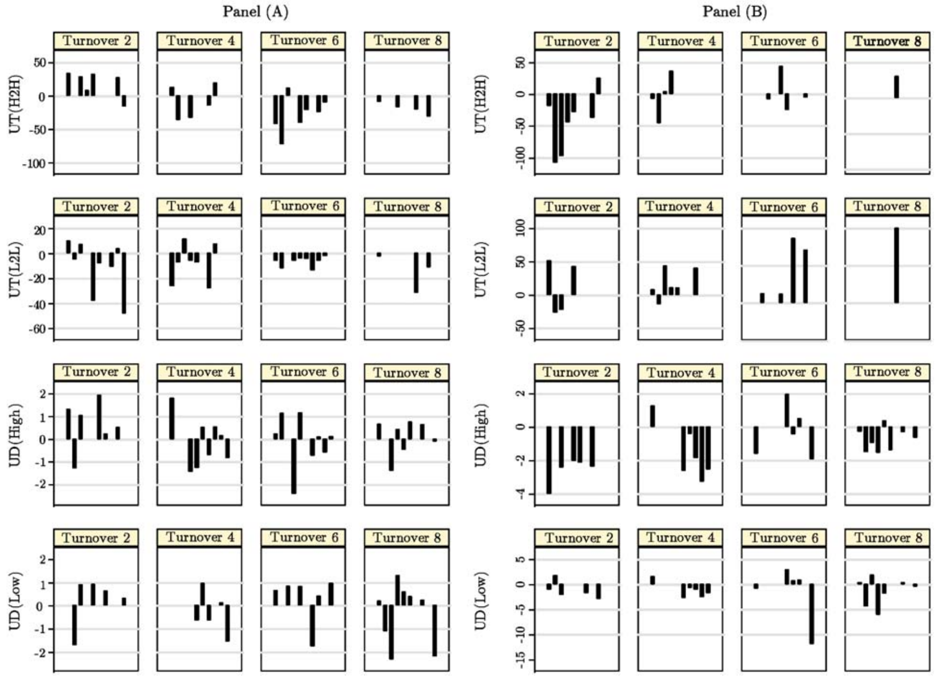

Fig. 2. Significant estimates for prior returns as instruments.

In the figure CDR is the cumulative prior-20-trading-day stock return net three-month T-bill rate; CMR is the cumulative prior-20-trading-day market return net of three-month T-bill rate; UT(H2H) and UT(L2L) are the estimated coefficients associated with CDR in Panel (A) or CMR in Panel (B) on the uninformed transition probability remaining at the high and the low state, respectively; and UD(High) and UD(Low) are the estimated coefficients associated with CDR in Panel (A) or CMR in Panel (B) on the uninformed arrival rate differential at the high and the low state, respectively. We assign zero value to the estimated coefficients that are not statistically significantly different from zero at the 5% level. The columns correspond to each of the 2nd, 4th, 6th, and 8th turnover deciles.

uninformed arrival rate differential (UD), but we do not report the estimated constants for brevity. We believe it is more effective to present graphically the estimates than to use tables of results.[7] In order to deliver a clear pattern, we do not show estimated coefficients that are not significantly different from 0 at the 5% level, and they are replaced with zero values.

The first two rows of Panel (A) in Fig. 2 depict the influence of prior stock returns on the uninformed transition probabilities, and the last two rows of Panel (A) present the influence of prior stock returns on the uninformed arrival rate differential. Each row of the panel contains four individual plots with a common scale, one for each turnover decile, and stocks within each turnover decile are plotted

[7]The full estimation results are available upon request.

by ticker symbol in alphabetic order. The first impression from Panel (A) of Fig. 2 is that we see a lot more bars (standing for estimates significant at the 5% level) than empty slots (standing for insignificant estimates). Therefore, the uninformed traders are quite responsive to prior stock returns.

The coefficient for the probability of staying at the high state arrival rate is denoted by label UT(H2H) in the figure, and the coefficient for the probability of staying at the low state arrival rate is denoted by label UT(L2L). The pattern of predominantly negative signs for both UT(H2H) and UT(L2L) suggests that the uninformed traders react differently to high prior stock returns, taking a path-dependent response. Specifically, they are more likely to switch into a different state instead of staying at the current one, as long as high prior stock returns are observed.

In the last two rows of Panel (A), the coefficient on the uninformed arrival rate differential at the high level is denoted by label UD(High), and the coefficient at the low level of uninformed arrival is denoted by UD(Low). We observe that in reaction to high prior stock returns, the uninformed traders on 19 stocks will sell significantly more than buy when the existing uninformed arrival rate is high, as indicated by a significantly positive coefficient on UD(High). The uninformed traders on 11 stocks significantly buy more than sell as indicated by a significantly negative coefficient on UD(High). That is, we see more use of the contrarian strategy than the momentum strategy by the uninformed traders when the amount of uninformed trading is high. When the existing uninformed arrival rate is low, we see that the uninformed traders on 16 stocks significantly sell more than buy, as indicated by a significantly positive coefficient on UD(Low). The uninformed traders on only eight stocks buy more than sell, with a significantly negative coefficient on UD(Low). Again, we see more use of the contrarian strategy than the momentum strategy when the amount of uninformed trading is low. Therefore, the evidence from our sample of stocks suggests that uninformed traders tend to adopt a contrarian strategy in reaction to high past stock returns. It is interesting to note that Lo and MacKinlay (1990) find that a contrarian strategy is profitable in very short horizon such as one month.

In Panel (B) of Fig. 2, we report the results on the response of the uninformed traders to past market returns. It seems that the uninformed traders prefer switching into or staying at the low level of arrival rate. Given that many stocks have an insignificant coefficient associated with prior market returns, we should avoid over-interpreting the impact of prior market returns on the uninformed transition probabilities.

The more interesting results lie in the third and fourth rows of Panel (B) in Fig. 2. We see 20 stocks with significantly negative UD(High) and four stocks with significantly positive UD(High), indicating that the uninformed traders are predominantly using momentum strategy in reaction to high prior market returns if the uninformed trading activities were high in the previous period. In the period following low uninformed trading activities, the uninformed traders still employ the momentum strategy more often since 15 stocks have significantly negative UD(Low) whereas eight stocks have significantly positive UD(Low). Comparing the magnitude of impact on UD(High) to that on UD(Low), we find a much stronger momentum effect when the existing uninformed traders are active. Comparing the magnitude of impact on stocks across different turnover deciles, we find a stronger momentum effect on the lower turnover deciles. The broad message from Panel (B) of Fig. 2 is that the uninformed traders tend to adopt the momentum strategy in reaction to high prior market returns, and submit more buy orders than sell orders.

Various studies have shown that different investors have very different trading styles. Choe et al. (1999) find strong daily evidence that Korean and foreign institutional investors use "positive feedback and herding" trading strategies. Moreover, in a study on the daily and intra-day trading of NASDAQ 100 stocks, Griffin et al. (2003) find the presence of trend-chasing behavior by institutional investors as well as evidence supporting contrarian strategy adopted by individual investors. Therefore, it is not surprising to find different trading behavior across different stocks, since they could potentially have very different institutional ownership structure. Grinblatt and Keloharju (2001) find small investors more willing to buy after stocks reach their monthly lows and more willing to sell after monthly highs (see Panel D of their Table 2). This is consistent with our finding that prior 20-trading-day own stock return induces more contrarian trading activities by the uninformed. Researchers such as Jegadeesh (1990) and Lehmann (1990) find that short-term own-stock contrarian strategies yield abnormal returns. This short-term reversal may reflect corrections for "over-reactions" in stock prices or inefficiency in the market for liquidity around large price changes. Some uninformed investors may earn profit for providing the liquidity service. We would like to caution not reading too much into the pattern since we only study 40 stocks using our structural model. The literature on momentum and contrarian strategies routinely studies thousands of stocks by forming portfolios without a parametric framework. In the following subsection we conduct maximum likelihood tests on the specification of the model. We will show that for each stock fundamentals such as cumulative stock and market returns are significantly related to uninformed trading activities.

4.3. Model specification test

Consider model A with constant transition probabilities (UT corresponding to Eq. (1)) and constant buy–sell differentials (UD corresponding to Eq. (2), where we use a vector of ones as the instruments $\mathbf{z}_t$ for both UT and UD. In model B, we allow for time-varying transition probabilities and constant buy–sell differentials by feeding a vector of ones as $\mathbf{z}_t$ for UD, but feeding a matrix of instruments $\mathbf{z}_t$ for UT, which consists of a vector of ones as well as prior stock returns CDR and prior market returns CMR. In the main model of our paper (model C), we use a matrix of instruments $\mathbf{z}_t$, consisting of a vector of ones and prior returns CDR and CMR, for both UT and UD so as to achieve time-varying transition probabilities and time-varying buy–sell differential. These three versions of the model are nested, and we can construct the likelihood ratio statistics to test our model C against the two alternative specifications.

The likelihood ratio statistic λ_1, constructed as twice the difference between the sample log-likelihood for model B and model C, is asymptotically distributed as $\chi^2(4)$ with critical value of 13.28 at the 1% level. Because the only difference between model B and model C is whether we use the prior stock returns and market returns for UD, the rejection of model B relative to model C at the 1% level for each stock (see the column second to the last in Table 2) provides strong statistical support for the time-varying nature of UD. It may be the case that past stock returns CDR alone do not make UD time-varying for some stocks, nor do past market returns CMR alone. However, the fact that model C is preferred to model B for every stock manifests the joint significance of past stock returns (CDR) and past market returns (CMR) in making UD time-varying. The equivalent graphical interpretation of this test result is that there is no single stock that has no bars (i.e., insignificant estimates) at all for the last two rows across both panels in Fig. 2.

Similarly, we construct the likelihood ratio statistic λ_2 as twice the difference between the sample log-likelihood for model A and model C, which is asymptotically distributed as $\chi^2(8)$ with critical value of 20.09 at the 1% level. The rejection of model A relative to model C at the 1% level for every stock (see the last column of Table 2) lends additional support for the time-varying nature of both UT and UD. Again, the equivalent graphical interpretation of this test result is that there is no single stock that has no bars (i.e., insignificant estimates) at all for all four rows across both panels in Fig. 2.

5. Applications

In this section, we focus on a few direct applications of the model. We first explore the contribution of information asymmetry to various measures of bid–ask spreads by providing evidence from both the cross-sectional and the time-series perspectives. We then examine the predictive power of the estimated probability of informed trading (TPINs) for mean spreads in the next period, alongside competing measures of information asymmetry. Finally, we use the estimated TPINs as a direct measure of information asymmetry and analyze its impact on the persistence of stock returns.

5.1. *The explanatory power of PINs*

To examine the extent to which information asymmetry in stock trading activities may affect various measures of stock spreads, we run a fixed-effect panel regression that is consistent with the spirit of Easley et al. (1996),

$$\begin{aligned}\Sigma_{i,t} = {} & \beta_0 + \beta_1\,\mathrm{VTPIN}_{i,t} + \beta_2\,\mathrm{VPIN}_{i,t} + \beta_3\mathrm{VOL}_{i,t} + \beta_4\delta_2(t) + \beta_5\delta_3(t) \\ & + \beta_6\delta_2(t)\,\mathrm{VTPIN}_{i,t} + \beta_7\delta_2(t)\,\mathrm{VPIN}_{i,t} + \beta_8\delta_3(t)\,\mathrm{VTPIN}_{i,t} \\ & + \beta_9\delta_3(t)\,\mathrm{VPIN}_{i,t} + \sum_{i=2}^{40}\gamma_i\delta(i) + \eta_{i,t}, \quad \text{where } i = 1,\ldots,40. \end{aligned} \tag{7}$$

The dependent variable Σ is the stock spread of different types extracted from the NYSE TAQ database. The independent variable VPIN is the product of stock price and probability of informed trading PIN estimated from the Easley et al. (1996) framework. Note that a constant PIN is estimated for each quarter of trading data and assigned to each trading day within that particular quarter. VTPIN is the product of stock price and probability of informed trading TPIN estimated from our time-varying model. VOL is the dollar volume defined as the product of stock price and share volume (in thousand shares). $\delta_2(t)$ is an indicator for the period with tick size being one-sixteenth on the New York Stock Exchange (i.e., between June 24, 1997 and January 28, 2001) and $\delta_3(t)$ is an indicator for the period of post-decimalization (i.e., on and after January 29, 2001). Empirical studies such as Angel (1997) show that the minimum tick size has significant impact on the size of the bid–ask spreads. $\delta(i)$ is the dummy variable corresponding to each of the 39 stocks.

According to Easley et al. (1996), we expect to see a positive slope on the measure of information asymmetry and a negative slope on the dollar volume VOL. The rationale is that market makers should quote higher spread to offset higher losses to informed trades, and frequently traded stocks command lower spread due to a lesser extent of information asymmetry. As competing measures of information asymmetry, VPIN and VTPIN are expected to have positive coefficients. If one of the two measures completely subsumes the other in explaining spread, then we expect to see a significant positive coefficient for the dominant measure and an insignificant one for the other.

The addition of tick size dummy variables δ_2 and δ_3 and their respective interaction terms with the two measures of information asymmetry is used to control for the regime changes in tick size. We expect that the reduction of tick size leads to lower spreads and that the post-decimalization period should exhibit an even stronger reduction in spreads than the period with one-sixteenth tick size.

It is plausible to infer that with tick size reduction liquidity providers face more competition with each other, and thus in the regime with smaller tick size market makers face more constraints in raising the spread in order to recoup the loss to informed traders. Market makers may still increase the spread when they face more informed trading, yet they are not able to increase it by as much as in the smaller tick size regimes. We can make two empirical predictions. First, the coefficient associated with the interaction term between the two tick size dummies and the proxy for information asymmetry should be negative. Second, the absolute magnitude of this coefficient should be larger for the post-decimalization period than for the one-sixteenth tick size period.

Panel (A) of Table 3 presents the fixed-effect regression results using all available daily observations between January 1993 and December 2002, and corrected for heteroskedasticity and autocorrelation of first order. VOL has the predicted negative sign, and is significant for all three measures of spread. Whereas VPIN is significant and positive in explaining the daily closing and daily mean spread, its explanatory power for the opening spread is zero. In contrast, VTPIN is highly significant and positive in all three measures of spread, and the size of the coefficients is much larger than that for VPIN. Because the sample mean of VTPIN is larger than that of VPIN (4.5 versus 3.5) the overall explanatory power of VTPIN dominates that of VPIN.

The coefficients for both tick size dummies are highly significant and negative, with a larger magnitude in the post-decimalization period. The interaction terms between VTPIN and the two tick size dummies are significantly negative for closing

Table 3
Information content and stock spreads

In this table, we present results for a fixed-effect panel regression with modification to the original Easley et al. (1996) design, $\Sigma_{i,t} = \beta_0 + \beta_1 \mathrm{VTPIN}_{i,t} + \beta_2 \mathrm{VPIN}_{i,t} + \beta_3 \mathrm{VOL}_{i,t} + \beta_4 \delta_2(t) + \beta_5 \delta_3(t) + \beta_6\, \delta_2(t)\, \mathrm{VTPIN}_{i,t} + \beta_7 \delta_2(t)\, \mathrm{VPIN}_{i,t} + \beta_8 \delta_3(t)\, \mathrm{VTPIN}_{i,t} + \beta_9 \delta_3(t)\, \mathrm{VPIN}_{i,t} + \sum_{i=2}^{40} \gamma_i \delta(i) + \eta_{i,t}$. Σ is the stock spread of different types extracted from the NYSE TAQ database. VPIN is the product of stock price and probability of informed trading PIN estimated from the Easley et al. (1996) framework. VTPIN is the product of stock price and estimated probability of informed trading TPIN from our extended model. VOL is the dollar volume defined as the product of stock price and share volume (in thousand shares). $\delta_2(t)$ is a dummy variable that takes the value 1 for the period of tick size being $\frac{1}{16}$ on the NYSE (i.e., between June 24, 1997 and January 28, 2001). $\delta_3(t)$ is a dummy variable that takes the value 1 for the period of post-decimalization on the NYSE (i.e., on and after January 29, 2001). $\delta(i)$ is the dummy variable for each of the 39 stocks. Panel (A) uses 99,379 daily observations between January 1993 and December 2002, and Panel (B) uses 1597 quarterly observations for the same period. In all the panel regressions, we control for heteroskedasticity and autocorrelations of order one. The t-statistics are reported in parentheses.

Regressors	Opening spread		Closing spread		Mean spread	
Panel (A): contemporaneous relationship (daily)						
VTPIN	0.011	(5.49)	0.006	(15.27)	0.006	(19.19)
VPIN	0.002	(0.80)	0.005	(10.79)	0.005	(13.25)
VOL	−1.83E-06	(−2.86)	−2.35E-06	(−19.18)	−4.42E-07	(−8.17)
δ_2	−0.044	(−3.89)	−0.040	(−24.31)	−0.045	(−33.18)
δ_3	−0.173	(−12.61)	−0.095	(−47.88)	−0.106	(−64.07)
δ_2 VTPIN	−0.004	(−1.09)	−0.006	(−10.83)	−0.005	(−10.75)
δ_2 VPIN	0.008	(1.78)	0.006	(8.56)	0.006	(10.20)
δ_3 VTPIN	−0.004	(−0.88)	−0.010	(−11.48)	−0.010	(−16.10)
δ_3 VPIN	0.036	(5.03)	0.009	(8.01)	0.010	(11.69)
Panel (B): contemporaneous relationship (quarterly)						
VTPIN	0.014 (5.84)	0.010	(12.18)	0.010	(12.76)	
VPIN	−0.002	(−0.83)	0.000	(0.53)	0.000	(−0.10)
VOL	−8.63E-06	(−4.67)	−8.13E-06	(−12.37)	−5.95E-06	(−10.59)
δ_2	−0.053	(−3.57)	−0.051	(−10.58)	−0.059	(−12.62)
δ_3	−0.177	(−10.02)	−0.090	(−15.48)	−0.092	(−16.01)
δ_2 VTPIN	0.002	(0.58)	0.000	(−0.02)	0.001	(1.02)
δ_2 VPIN	0.003	(0.77)	0.000	(−0.37)	0.000	(0.33)
δ_3 VTPIN	0.005	(1.08)	−0.003	(−1.64)	−0.002	(−1.26)
δ_3 VPIN	0.031	(5.12)	0.001	(0.57)	0.000	(0.25)

and mean spreads, with larger magnitude during the post-decimalization period. Hence the results on VTPIN are strong and are consistent with empirical predictions. In contrast, VPIN has significant and positive coefficients for its interaction with tick size dummies. This is counter-intuitive since it implies that VPIN predicts higher influence of information asymmetry for the regime with smaller tick size.

In computing constant PIN over each calendar quarter we use all data available for that quarter. The repeated estimation of the constant PIN may help reveal structural shifts not entirely captured by our model. One may attribute the weak results of VPIN to our practice of assigning a constant PIN for each trading day within a quarter as it effectively reduces the time series variation of VPIN. To address this concern, we average the time series of VTPIN estimated from our model to form quarterly series that is comparable with the quarterly VPIN series. We do not feel the constant PIN is unduly disadvantaged by this procedure since the sampling interval is one quarter. The results for the quarterly fixed-effect panel regressions are presented in Panel (B) of Table 3. VTPIN has highly significant and positive slopes in explaining all three measures of spread, and VPIN has no residual explanatory power for any measure of spread. The tick dummies are still highly significant and negative, with stronger impact during the post decimalization period. The interaction terms are no longer significant for all but one case.

In sum, there is evidence suggesting that the time-varying probability of informed trading TPIN is a more powerful measure of information asymmetry than the constant PIN.

5.2. Predictive power of competing measures of information asymmetry

Given that our model is computationally involved, we want to examine if the time-varying probability of informed trading is more informative when compared to other measures of information asymmetry. We run the following fixed-effect panel regression as a horse race among competing measures of information asymmetry for predicting the mean spread of the next trading day.

$$\begin{aligned}\Sigma_{i,t+1} = {} & \beta_{0,i} + \beta_1 \mathrm{VTPIN}_{i,t} + \beta_2 \mathrm{VPIN}_{i,t} + \beta_3 \mathrm{VOL}_{i,t} + \beta_4\, \mathrm{AMS}_{i,t} \\ & + \beta_5\, \mathrm{AVOL}_{i,t} + \beta_6\, \mathrm{OIMB}_{i,t} + \beta_7\, \mathrm{ME}_{i,t} + \beta_8\, \mathrm{RVOL}_{i,t} + \beta_9 \delta_2(t) \\ & + \beta_{10}\delta_3(t) + \beta_{11}\delta_2(t)\,\mathrm{VTPIN}_{i,t} + \beta_{12}\delta_2(t)\,\mathrm{VPIN}_{i,t} + \beta_{13}\delta_3(t) \\ & \times \mathrm{VTPIN}_{i,t} + \beta_{14}\delta_3(t)\,\mathrm{VPIN}_{i,t} + \sum_{i=2}^{40} \gamma_i \delta(i) + \eta_{i,t+1}, \end{aligned} \tag{8}$$

where $i = 1, \ldots, 40$. The dependent variable Σ is the mean bid–ask spread computed from the NYSE TAQ database. In addition to VTPIN, VPIN, VOL, $\delta_2(t)$, $\delta_3(t)$ and $\delta(i)$ as defined earlier, we also include the following explanatory variables. AMS is the abnormal mean spread computed as the deviation of current mean bid–ask spread from the moving average of past 20-trading-day mean bid–ask spread; AVOL is the abnormal volume computed as the deviation of current dollar volume from the moving average of past 20-trading-day mean dollar volume; OIMB is the order imbalance or absolute net order flow in number of trades; ME is the market value of equity; and RVOL is the volatility of returns in the past 20-trading-day period.

The dollar volume VOL and tick size dummies δ_2 and δ_3 are expected to have negative signs for their respective coefficients. Each of the five competing measures of information asymmetry, VTPIN, VPIN, AMS, AVOL and OIMB, should have a positive sign when predicting the future spread. Chordia

Table 4
Competing measures of information asymmetry

In this table, we report results for predicting the mean bid–ask spread on the next trading period using a set of competing measures of information asymmetry in a fixed-effect panel regression framework. The regression design is $\Sigma_{i,t+1} = \beta_0 + \beta_1 \mathrm{VTPIN}_{i,t} + \beta_2 \mathrm{VPIN}_{i,t} + \beta_3 \mathrm{VOL}_{i,t} + \beta_4 \mathrm{AMS}_{i,t} + \beta_5 \mathrm{AVOL}_{i,t} + \beta_6 \mathrm{OIMB}_{i,t} + \beta_7 \mathrm{ME}_{i,t} + \beta_8 \mathrm{RVOL}_{i,t} + \beta_9 \delta_2(t) + \beta_{10} \delta_3(t) + \beta_{11} \delta_2(t) \mathrm{VTPIN}_{i,t} + \beta_{12} \delta_2(t) \mathrm{VTPIN}_{i,t} + \beta_{13} \delta_3(t) \mathrm{VTPIN}_{i,t} + \beta_{14} \delta_3(t) \mathrm{VPIN}_{i,t} + \sum_{i=2}^{40} \gamma_i \delta(i) + \eta_{i,t+1}$. Σ is the mean bid–ask spread computed from the NYSE TAQ database. VPIN is the product of stock price and probability of informed trading PIN estimated from the Easley et al. (1996) framework. VTPIN is the product of stock price and estimated probability of informed trading TPIN from our extended model. VOL is the dollar volume defined as the product of stock price and share volume (in thousand shares). AMS is the abnormal mean spread computed as the deviation of current mean bid–ask spread from the moving average of past 20-trading-day mean bid–ask spread. AVOL is the abnormal volume computed as the deviation of current dollar volume from the moving average of past 20-trading-day mean dollar volume. OIMB is the order imbalance or absolute net order flow in number of trades. ME is the market value of equity. RVOL is the volatility of returns in the past 20-trading-day period. $\delta_2(t)$ is a dummy variable that takes the value 1 for the period of tick size being $\frac{1}{16}$ on the NYSE (i.e., between June 24, 1997 and January 28, 2001). $\delta_3(t)$ is a dummy variable that takes the value 1 for the period of post-decimalization on the NYSE (i.e., on and after January 29, 2001). $\delta(i)$ is the dummy variable for each of the 30 stocks. Panel (A) uses 98,638 daily observations between January 1993 and December 2002, and Panel (B) uses 1,557 quarterly observations for the same period. In all the panel regressions, we control for heteroskedasticity and autocorrelations of order one. The t-statistics are reported in parentheses.

Regressors	Panel (A)		Panel (B)	
VTPIN	0.011	(42.00)	0.009	(10.43)
VPIN	0.003	(8.49)	0.001	(2.35)
VOL	−5.48E-06	(−27.77)	1.12E-06	(1.34)
AMS	−0.160	(−40.45)	0.164	(1.75)
AVOL	5.14E-06	(26.29)	−8.45E-07	(−0.45)
OIMB	−1.48E-04	(−12.94)	−1.54E-03	(−8.19)
ME	−2.65E-08	(−28.73)	−3.75E-08	(−8.11)
RVOL	0.215	(11.31)	−0.167	(−2.20)
δ_2	−0.040	(−36.13)	−0.034	(−6.90)
δ_3	−0.103	(−75.33)	−0.077	(−12.18)
δ_2 VTPIN	−0.005	(−12.97)	−0.001	(−0.86)
δ_2 VPIN	0.005	(9.71)	0.002	(1.43)
δ_3 VTPIN	−0.008	(−15.01)	0.000	(0.05)
δ_3 VPIN	0.007	(10.12)	−0.001	(−0.42)

et al. (2002) argue that order imbalances reduce liquidity, so the predicted sign for absolute order imbalance is positive. Large stocks tend to be more liquid so it is reasonable to conjecture a negative coefficient associated with market equity ME. Inventory theory suggests that a risk-averse market maker will set a higher spread for stocks with higher past return volatility, so the expected sign for RVOL is positive.

The panel regression results in Table 4 indicate that all the explanatory variables are highly significant and have the expected sign. The only exceptions are the abnormal mean spread and the absolute order imbalance, both of which have significantly negative signs. Note that Chordia et al. (2002) also find a negative, albeit insignificant, coefficient for the order imbalance.[8] The interaction terms between VPIN and the tick dummies are significantly negative, with an intensified effect in the post-decimalization period. The interaction terms between VPIN and tick dummies are again significantly positive, contrary to the empirical predictions outlined earlier.

We conclude that the time-varying probability of informed trading TPIN is a better and more robust measure in predicting future mean spread, even after controlling for other competing measures of information asymmetry.

5.3. Volume, information asymmetry and autocorrelation of returns

Trading volume is often watched carefully by traders and academics alike. Not surprisingly, there is a large literature in finance devoted to volume (see, for example, Blume et al., 1994; Lee and Swaminathan, 2000; Lo and Wang, 2000). Campbell et al. (1993) investigate the relationship between stock market trading volume and serial correlation of daily stock returns. They find that the first-order daily return autocorrelation tends to decline with volume. They propose a model with risk-averse "market makers" who charge a premium for accommodating "liquidity" or "non-informational" traders. The resulting changing expected returns reward market makers for playing this role. Therefore, the stock price decline on a high-volume day is more likely than the stock price decline on a low-volume day to be associated with an increase in the expected stock return.

More recently, Llorente et al. (2002) analyze the dynamic relation between daily volume and first-order autocorrelation for individual stock returns. They present a model in which returns generated by non-informational trades tend to reverse themselves, while returns generated by informational trades tend to continue themselves. This relationship is intuitive in that the stock prices reflect new information via informed trades only in a gradual fashion. In the days immediately after the informed trades, stock prices tend to continue their decline (if the informed trades revealed bad news) or rise (if the informed trades revealed good news). On the other hand, large volumes of uninformed trades cause only short-lived price pressure so it is more likely to see a price reversal after uninformed trades. Their empirical tests show that the cross-sectional variation in the relation between volume and return autocorrelation is positively related to the extent of informed trading, where they use volume and spread as indirect measures of information asymmetry in a two-stage regression.

Since VTPIN is a direct measure of information asymmetry according to the evidence presented earlier, we run the following fixed-effect panel regression to

[8]Chordia et al. (2002) study a group of NYSE listed S&P500 component stocks. See their Table 3 for the negative coefficient for the post-transformed order imbalance when it is used to predict the next day percentage changes in value-weighted quoted spreads.

capture the essence of Llorente et al. (2002). After correcting for heteroskedasticity and autocorrelation of first order, we have the following relation:

$$r_{i,t+1} = \underset{\substack{0.0004\\(4.53)}}{\gamma_0} + r_{i,t}\Big(\underset{\substack{-0.0517\\(-8.61)}}{\gamma_1} + \underset{\substack{0.0102\\(8.05)}}{\gamma_2}\,\mathrm{VTPIN}_{i,t}\Big) + \sum_{i=2}^{40} \beta_i \delta(i) + \varepsilon_{i,t+1}$$

$$\text{where } i = 1, \ldots, 40. \tag{9}$$

The dependent variable r is the daily holding period return extracted from CRSP, and $\delta(i)$ is the stock-specific dummy variable. A selected set of estimates is listed directly below the coefficients with their t-statistics in parentheses. The results show a significantly negative auto-correlation γ_1 for daily stock returns. The impact of information asymmetry γ_2 is highly significant and positive, consistent with the intuition that higher level of information asymmetry induces higher return autocorrelation.

The impact of volume on the return auto-correlation can be studied indirectly by comparing the γ_2 coefficient across different turnover deciles. It is plausible to conjecture that stocks in the high turnover decile are likely to have a lower fraction of informed trading, so we predict that the γ_2 coefficient should decline as we move to higher turnover deciles. To test this specific hypothesis, we augment the empirical design with three dummies τ_4, τ_6 and τ_8, one for each of the 4th, 6th, and 8th turnover deciles, respectively, so that

$$r_{i,t+1} = \underset{\substack{0.0004\\(4.52)}}{\gamma_0} + r_{i,t}\Big[\underset{\substack{-0.0549\\(-7.92)}}{\gamma_1} + \big(\underset{\substack{0.0167\\(6.63)}}{\gamma_{2,2}} + \underset{\substack{-0.0052\\(-1.95)}}{\gamma_{2,4}}\,\tau_4 + \underset{\substack{-0.0066\\(-2.77)}}{\gamma_{2,6}}\,\tau_6 + \underset{\substack{-0.0080\\(-3.25)}}{\gamma_{2,8}}\,\tau_8\big)\mathrm{VTPIN}_{i,t}\Big] + \sum_{i=2}^{40} \beta_i \delta(i) + \varepsilon_{i,t+1}. \tag{10}$$

Some selected estimates are listed directly below the coefficients with t-statistics in parentheses. Again, we see a significantly negative auto-correlation γ_1 for daily stock returns. We also observe the monotonically declining impact of information asymmetry on return autocorrelation as the turnover decile becomes higher, as the γ_2 coefficients are 0.0167, 0.0115, 0.0049 and -0.0031 for the 2nd, 4th, 6th and 8th turnover decile, respectively. In sum, the time series of VTPIN enables us to conduct a sharper test, providing strong support for the argument of Llorente et al. (2002).

The time series of probability of informed trading also can be used in studies investigating if the information asymmetry risk is priced and if it affects asset returns. For example, Easley et al. (2002b) repeatedly estimate the probability of informed trading for each year in their sample period using the same framework as Easley et al. (1996) and reach the conclusion that a 10% difference in this probability between two stocks leads to a 2.5% difference in annual returns. Their analysis could be extended with the time-varying probability of informed trading. We leave this for future research.

6. Concluding remarks

Building upon the seminal work of Easley et al. (1996), we develop a framework to investigate time-varying informed and uninformed trading activities and the relationship between them. We allow the buy and sell arrival rates for the uninformed traders to follow a Markov switching process. Both the uninformed buy–sell arrival rate differential and the uninformed transition probabilities depend on past performance of the stock and the overall market. The informed traders may match the level of the uninformed arrival rate with certain probability so as to make better use of the camouflage provided by the uninformed traders.

Our empirical estimation of 40 NYSE stocks shows that the buy and sell arrival rates of the uninformed traders are different and time-varying. The uninformed traders tend to adopt contrarian strategy in reaction to high prior own stock returns, but employ momentum strategy in reaction to high prior market returns. Informed traders seem to take good advantage of the camouflage since the estimated probability of informed matching response ranges from 0.72 to 0.98. We find that the estimated time-varying probability of informed trading is a good predictor for various measures of bid–ask spreads, and is a better and more powerful measure of information asymmetry than the constant probability of informed trading. The estimated time-varying probability of informed trading has predictive power for mean bid–ask spreads in the next trading day, even after controlling for competing measures of information asymmetry. Finally, we use the estimated time-varying probability of informed trading as a measure of informational asymmetry to analyze its impact on the serial correlation of daily stock returns. The development of TPIN enables us to conduct a sharper and more robust test on this important issue.

Appendix A

A.1. Probability of informed trading

Using similar arguments in deriving Eq. (4), we can obtain other elements of the transition matrix for the four-composite-state model. The matrix v_t of transition probabilities can be written as follows:

$$v_t \equiv \begin{bmatrix} \pi_t^{**}\rho & \pi_t^{**}(1-\rho) & (1-\pi_t^{**})(1-\rho) & (1-\pi_t^{**})\rho \\ \pi_t^{**}\rho & \pi_t^{**}(1-\rho) & (1-\pi_t^{**})(1-\rho) & (1-\pi_t^{**})\rho \\ (1-\pi_t^{*})\rho & (1-\pi_t^{*})(1-\rho) & \pi_t^{*}(1-\rho) & \pi_t^{*}\rho \\ (1-\pi_t^{*})\rho & (1-\pi_t^{*})(1-\rho) & \pi_t^{*}(1-\rho) & \pi_t^{*}\rho \end{bmatrix}, \tag{11}$$

where each row stands for a composite state at trading day $t-1$ and each column stands for a composite state at trading day t. Note that the first and second rows in this transition matrix are the same, as are the third and fourth rows. This result comes from the assumption that the probability of matching response is path-independent. In particular, the informed traders form their response ρ not based upon the state level of the uninformed arrival rate in the previous trading day, and the uninformed transition probabilities are the same across these rows.

Denote as $\mathbf{p}_t^s$ the vector of probabilities for each composite state at period t

$$\mathbf{p}_t^s \equiv [p(s_t^a)\ p(s_t^b)\ p(s_t^c)\ p(s_t^d)].$$

The evolution of these state probabilities can be written as

$$\mathbf{p}_t^s = \mathbf{p}_{t-1}^s v_t = \mathbf{p}_0^s \prod_{m=1}^{t} v_m, \quad \forall t \geqslant 1, \tag{12}$$

where we assume $\mathbf{p}_0^s = [0.25\ 0.25\ 0.25\ 0.25]$, without loss of generality.

By combining the elements of $\mathbf{p}_t^s$, we can obtain the probabilities $p(\varepsilon_t^{\mathrm{h}})$, $p(\varepsilon_t^{\mathrm{l}}), p(\mu_t^{\mathrm{h}})$ and $p(\mu_t^{\mathrm{l}})$. We define the expectation of the uninformed and the informed arrival rates as

$$\varepsilon_t^{\mathrm{e},B} \equiv \varepsilon_t^{\mathrm{h},B} p(\varepsilon_t^{\mathrm{h}}) + \varepsilon_t^{\mathrm{l},B} p(\varepsilon_t^{\mathrm{l}}), \quad \varepsilon_t^{\mathrm{e},S} \equiv \varepsilon_t^{\mathrm{h},S} p(\varepsilon_t^{\mathrm{h}}) + \varepsilon_t^{\mathrm{l},S} p(\varepsilon_t^{\mathrm{l}})$$
$$\text{and} \quad \mu_t^{\mathrm{e}} \equiv \mu_t^{\mathrm{h}} p(\mu_t^{\mathrm{h}}) + \mu_t^{\mathrm{l}} p(\mu_t^{\mathrm{l}}).$$

Finally, the probability that each trade on day t is information-based can be expressed as $\mathrm{TPIN}_t = \alpha\mu_t^{\mathrm{e}}/(\varepsilon_t^{\mathrm{e},B} + \varepsilon_t^{\mathrm{e},S} + \alpha\mu_t^{\mathrm{e}})$.

A.2. Sample likelihood

Conditional on the state of a trading day t, say s_t^a, the likelihood of observing B_t buy trades and S_t sell trades is

$$\begin{aligned} g(s_t^a) = {} & (1-\alpha)\exp(-\varepsilon_t^{\mathrm{h},B} - \varepsilon_t^{\mathrm{h},S}) \frac{(\varepsilon_t^{\mathrm{h},B})^{B_t}}{B_t!} \frac{(\varepsilon_t^{\mathrm{h},S})^{S_t}}{S_t!} \\ & + \alpha\delta \exp(-\varepsilon_t^{\mathrm{h},B} - \varepsilon_t^{\mathrm{h},S} - \mu_t^{\mathrm{h}}) \frac{(\varepsilon_t^{\mathrm{h},B})^{B_t}}{B_t!} \frac{(\varepsilon_t^{\mathrm{h},S} + \mu_t^{\mathrm{h}})^{S_t}}{S_t!} \\ & + \alpha(1-\delta)\exp(-\varepsilon_t^{\mathrm{h},B} - \varepsilon_t^{\mathrm{h},S} - \mu_t^{\mathrm{h}}) \frac{(\varepsilon_t^{\mathrm{h},B} + \mu_t^{\mathrm{h}})^{B_t}}{B_t!} \frac{(\varepsilon_t^{\mathrm{h},S})^{S_t}}{S_t!}. \end{aligned} \tag{13}$$

As the number of trades gets larger, this conditional likelihood becomes harder to compute due to the factorial, the exponential and the power functions. To improve computational efficiency, we rewrite it as

$$g(s_t^a) = c_t \cdot m(s_t^a) \cdot h(s_t^a)/(B_t! S_t!), \tag{14}$$

where the common factor c_t makes $\ln(c_t)$ easy to compute, and the state-dependent factors $m(s_t^a)$ and $h(s_t^a)$ are constructed to moderate the size of the exponential functions and the power functions. The definitions of these factors are

$$
\begin{aligned}
c_t &\equiv \exp(-\varepsilon_t^{*,B} - \varepsilon_t^{*,S} - \mu_t^*)(\varepsilon_t^{*,B})^{B_t}(\varepsilon_t^{*,S})^{S_t},\\
m(s_t^a) &\equiv \exp\left[-(\varepsilon_t^{\mathrm{h},B} - \varepsilon_t^{*,B}) - (\varepsilon_t^{\mathrm{h},S} - \varepsilon_t^{*,S}) - (\mu_t^{\mathrm{h}} - \mu_t^*)\right]\left(\frac{\varepsilon_t^{\mathrm{h},B}}{\varepsilon_t^{*,B}}\right)^{B_t}\left(\frac{\varepsilon_t^{\mathrm{h},S}}{\varepsilon_t^{*,S}}\right)^{S_t},\\
h(s_t^a) &\equiv \left[(1-\alpha)\exp(\mu_t^{\mathrm{h}}) + \alpha\delta\left(1 + \frac{\mu_t^{\mathrm{h}}}{\varepsilon_t^{\mathrm{h},S}}\right)^{S_t} + \alpha(1-\delta)\left(1 + \frac{\mu_t^{\mathrm{h}}}{\varepsilon_t^{\mathrm{h},B}}\right)^{B_t}\right],\\
\varepsilon^{*,B} &\equiv \tfrac{1}{2}(\varepsilon_t^{\mathrm{h},B} + \varepsilon_t^{\mathrm{l},B}),\\
\varepsilon^{*,S} &\equiv \tfrac{1}{2}(\varepsilon_t^{\mathrm{h},S} + \varepsilon_t^{\mathrm{l},S}),\\
\mu^* &\equiv \tfrac{1}{2}(\mu_t^{\mathrm{h}} + \mu_t^{\mathrm{l}}).
\end{aligned}
$$

Using a similar method, we also can rewrite and calculate the conditional likelihood for other state types, namely $g(s_t^b)$, $g(s_t^c)$ and $g(s_t^d)$. Taking into account the probabilities for each state type, the unconditional likelihood for day t is

$$
L\{(B_t, S_t) \mid \Theta\} = \sum_{k \in \{a,b,c,d\}} g(s_t^k) \cdot p(s_t^k), \tag{15}
$$

where Θ is the vector of parameters to be estimated. By purging some constants that do not contain parameters to be estimated, we can write the sample log-likelihood function over a sample of T trading days in the following computation-friendly form,

$$
\begin{aligned}
\mathscr{L}(\Theta) &= \ln\left[\prod_{t=1}^{T} L\{\Theta \mid B_t, S_t\}\right]\\
&= \sum_{t=1}^{T}\left[-\varepsilon_t^{*,B} - \varepsilon_t^{*,S} - \mu_t^* + B_t \ln(\varepsilon_t^{*,B}) + S_t \ln(\varepsilon_t^{*,S})\right]\\
&\quad + \sum_{t=1}^{T} \ln\left[\sum_{k \in \{a,b,c,d\}} m(s_t^k) h(s_t^k) p(s_t^k)\right].
\end{aligned} \tag{16}
$$

References

Admati, A.R., 1991. The informational role of prices. Journal of Monetary Economics 28, 347–361.

Admati, A.R., Pfleiderer, P., 1988. A theory of intraday patterns: volume and price variability. Review of Financial Studies 1, 3–40.

Angel, J.J., 1997. Tick size, share prices, and stock splits. Journal of Finance 52, 655–681.

Back, K., 1992. Insider trading in continuous time. Review of Financial Studies 5, 387–409.

Back, K., Pedersen, H., 1998. Long-lived information and intraday patterns. Journal of Financial Markets 1, 385–402.

Bagehot, W., 1971. The only game in town. Financial Analysts Journal 27, 12–14, 22.

Barber, B.M., Odean, T., 2003. All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. University of California at Davis working paper.

Barberis, N., Huang, M., 2001. Mental accounting, loss aversion, and individual stock returns. Journal of Finance 56, 1247–1292.

Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. Journal of Financial Economics 49, 307–343.

Barberis, N., Huang, M., Santos, T., 2001. Prospect theory and asset prices. Quarterly Journal of Economics 116, 1–53.

Barclay, M.J., Litzenberger, R.H., Warner, J.B., 1990. Private information, trading volume, and stock-return variances. Review of Financial Studies 3, 233–253.

Blume, L., Easley, D., O'Hara, M., 1994. Market statistics and technical analysis: the role of volume. Journal of Finance 49, 153–181.

Caginalp, G., Porter, D., Smith, V., 2000. Momentum and overreaction in experimental asset markets. International Journal of Industrial Organization 18, 187–204.

Campbell, J.Y., Grossman, S.J., Wang, J., 1993. Trading volume and serial correlation in stock returns. Quarterly Journal of Economics 108, 905–939.

Chan, L.K.C., Jegadeesh, N., Lakonishok, J., 1996. Momentum strategies. Journal of Finance 51, 1681–1731.

Choe, H., Kho, B.C., Stulz, R.M., 1999. Do foreign investors destabilize stock markets? The Korean experience in 1997. Journal of Financial Economics 54, 227–264.

Chordia, T., Roll, R., Subrahmanyam, A., 2002. Order imbalance, liquidity, and market returns. Journal of Financial Economics 65, 111–130.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. Journal of Finance 53, 1839–1885.

Diamond, D.W., Verrecchia, R.E., 1981. Information aggregation in a noisy rational expectations economy. Journal of Financial Economics 9, 221–235.

Easley, D., O'Hara, M., 1987. Price, trade size, and information in securities markets. Journal of Financial Economics 19, 69–90.

Easley, D., O'Hara, M., 1992. Time and the process of security price adjustment. Journal of Finance 47, 577–605.

Easley, D., Kiefer, N.M., O'Hara, M., Paperman, J.B., 1996. Liquidity, information and infrequently traded stocks. Journal of Finance 51, 1405–1436.

Easley, D., Kiefer, N.M., O'Hara, M., 1997. One day in the life of a very common stock. Review of Financial Studies 10, 805–835.

Easley, D., Engle, R.F., O'Hara, M., Wu, L., 2002a. Time-varying arrival rates of informed and uninformed trades. Cornell University working paper.

Easley, D., Hvidkjaer, S., O'Hara, M., 2002b. Is information risk a determinant of asset returns? Journal of Finance 57, 2185–2221.

Garman, M.B., 1976. Market microstructure. Journal of Financial Economics 3, 257–275.

Glosten, L.R., Harris, L.E., 1988. Estimating the components of the bid/ask spread. Journal of Financial Economics 21, 123–142.

Glosten, L.R., Milgrom, P.R., 1985. Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71–100.

Griffin, J.M., Harris, J.H., Topaloglu, S., 2003. The dynamics of institutional and individual trading. Journal of Finance 58, 2285–2320.

Grinblatt, M., Keloharju, M., 2001. What makes investors trade? Journal of Finance 56, 589–616.

Hasbrouck, J., 1991a. Measuring the information content of stock trades. Journal of Finance 46, 179–207.

Hasbrouck, J., 1991b. The summary informativeness of stock trades: an econometric analysis. Review of Financial Studies 4, 571–595.

Hong, H., Stein, J.C., 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. Journal of Finance 54, 2143–2184.

Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. Journal of Finance 45, 881–898.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 65–91.

Jegadeesh, N., Titman, S., 1995. Overreaction, delayed reaction, and contrarian profits. Review of Financial Studies 8, 973–993.

Jegadeesh, N., Titman, S., 2001. Profitability of momentum strategies: an evaluation of alternative explanations. Journal of Finance 56, 699–720.

Jones, C.M., Kaul, G., Lipson, M.L., 1994. Information, trading, and volatility. Journal of Financial Economics 36, 127–154.

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315–1335.

Lakonishok, J., Shleifer, A., Vishny, R.W., 1994. Contrarian investment, extrapolation, and risk. Journal of Finance 49, 1541–1578.

Lee, C.M.C., Ready, M.J., 1991. Inferring trade direction from intraday data. Journal of Finance 46, 733–746.

Lee, C.M.C., Swaminathan, B., 2000. Price momentum and trading volume. Journal of Finance 55, 2017–2069.

Lehmann, B.N., 1990. Fads, martingales, and market efficiency. Quarterly Journal of Economics 105, 1–28.

Llorente, G., Michaely, R., Saar, G., Wang, J., 2002. Dynamic volume-return relation of individual stocks. Review of Financial Studies 15, 1005–1047.

Lo, A.W., MacKinlay, A.C., 1990. When are contrarian profits due to stock market overreaction? Review of Financial Studies 3, 175–205.

Lo, A.W., Wang, J., 2000. Trading volume: definitions, data analysis, and implications of portfolio theory. Review of Financial Studies 13, 257–300.

Madhavan, A., Smidt, S., 1991. A Bayesian model of intraday specialist pricing. Journal of Financial Economics 30, 99–134.

Madhavan, A., Smidt, S., 1993. An analysis of changes in specialist inventories and quotations. Journal of Finance 48, 1595–1628.

Madhavan, A., Sofianos, G., 1998. An empirical analysis of NYSE specialist trading. Journal of Financial Economics 48, 189–210.

Mendelson, H., 1982. Market behavior in a clearing house. Econometrica 50, 1505–1524.

Odean, T., 1998. Are investors reluctant to realize their losses? Journal of Finance 53, 1775–1798.

Odean, T., 1999. Do investors trade too much? American Economic Review 89, 1279–1298.

O'Hara, M., Oldfield, G.S., 1986. The microeconomics of market making. Journal of Financial and Quantitative Analysis 21, 361–376.

Rouwenhorst, K.G., 1999. International momentum strategies. Journal of Finance 53, 267–284.

Siegel, J.J., 1992. Equity risk premia, corporate profit forecasts, and investor sentiment around the stock crash of October 1987. Journal of Business 65, 557–570.

Smidt, S., 1971. Which road to an efficient stock market: free competition or regulated monopoly? Financial Analysts Journal 27, 18–20, 64–69.

Wang, J., 1993. A model of intertemporal asset prices under asymmetric information. Review of Economic Studies 60, 249–282.

Weston, J.P., 2001. Information, liquidity, and noise. Rice University working paper.

Zabel, E., 1981. Competitive price adjustment without market clearing. Econometrica 49, 1201–1221.