Time-Varying Arrival Rates of Informed and Uninformed Trades

DAVID EASLEY, ROBERT F. ENGLE, MAUREEN O'HARA, AND LIUREN WU*

March 13, 2002; first draft: November, 2000

*David Easly is from Department of Economics, Cornell University, Ithaca, NY 14853; (607) 255-6283; dae3@cornell.edu. Robert Engle is from Stern School of Business, New York University, 44 West 4th Street, New York, NY 10012; (212) 998-0710; rengel@stern.nyu.edu. Maureen O'Hara is from Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853-6201; (607) 255-3645; mo19@cornell.edu. Liuren Wu is from Graduate School of Business, Fordham University, 113 West 60th Street, New York, NY 10023; (212) 636-6117; wu@fordham.edu. We thank Mark Ready, Schmuel Baruch and session participants at the 2002 AFA meetings for helpful comments. The latest version of the paper can be downloaded from http://www.bnet.fordham.edu/lwu.

ABSTRACT

In this paper we extend the model of Easley and O'Hara (1992) to allow the arrival rates of informed and uninformed trades to be time-varying and forecastable. We specify a generalized autoregressive bivariate process for the arrival rates of informed and uninformed trades and estimate the model on 16 actively traded stocks on the New York Stock Exchange over 15 years of transaction data. Our results show that uninformed trades are highly persistent. Uninformed order arrivals clump together, with high uninformed volume days likely to follow high uninformed volume days, and conversely. This behavior is consistent with the passive characterization of the uninformed found in the literature. But we do find an important difference in how the uninformed behave; they avoid trading when the informed are forecasted to be present. Informed trades also exhibit complex patterns, but these patterns are not consistent with the strategic behavior posited in the literature. The informed do not appear to hide in order flow, but instead they trade persistently.

We also investigate the correlation between the arrival rates of trades and trade composition on market volatility, liquidity and depth. We find that although volatility increases with the forecasted arrival rates of total trades, it is relatively independent of the forecasted composition of the trade. We use the opening bid-ask spread as a measure of market liquidity. We find that as the number of trades increases over time, the relative proportion of informed trades decreases and hence, spreads become narrower and the market becomes more liquid. Finally, we compute the price impact curve of consecutive buy orders and report the half life of the price impact as a measure of market depth. We find a positive correlation between the half life and total trades indicating that the market is deeper in presence of more trades.

Time-Varying Arrival Rates of Informed and Uninformed Trades

I. Introduction

A fundamental insight of the microstructure literature is that order flow is informative regarding subsequent price movements. This informational role arises because orders arrive from both informed and uninformed traders, and market observers can infer new information regarding the value of the asset from the composition and existence of trades. Thus, market parameters such as volume, volatility, market depth, and liquidity are all linked in the sense that each is influenced by the underlying order arrival processes. In this paper, we propose a dynamic microstructure model of trading, and we investigate how the dynamics of trades and trade composition interact with the evolution of market liquidity, depth and price volatility.

There are many reasons why understanding market liquidity, depth and price volatility are important. From a practical perspective, the cost of trading in a security is inextricably linked to these market variables, and market professionals devise trading strategies that explicitly incorporate these factors. Moreover, the volatility process is important not only for influencing the risk and return to an investor in the security, but also for understanding the behavior of derivative securities linked to the asset. From a more academic perspective, understanding the evolution of liquidity, depth and volatility provides insight into the price formation process as well as into concepts such as market efficiency. We argue in this paper that understanding these market parameters requires understanding a more basic market variable, the order arrival process.

To motivate our analysis, a useful construct is to view order arrivals as reflecting the behavior of sheep and wolves. In sequential trade models (for example, Glosten and Milgrom, 1985) and in Kyle (1985), the uninformed act as sheep, meekly heading to market where they will be preyed upon by the more informed wolves. The informed traders, the wolves, also head to the market, driven by the need to feast on the gains from their private information. In simple constructs, the sheep and wolves eschew strategic considerations and thus move competitively (and mechanistically) to market. Kyle (1985) introduced the concept of a strategic trader, or a smart wolf, who profits by timing his order arrivals so as to hide among the sheep. Whether packs of wolves can similarly profit by strategic trading is unclear as the equilibrium in Kyle's model breaks down when the number of informed traders becomes too large (see Back, Cao, and Willard, 2000). Despite scientific breakthroughs elsewhere, the sheep in microstructure models remain quite docile. Admati and Pfleiderer (1988) and Foster and Vishwanathan (1990) allowed for more introspective sheep who timed their trades to avoid the presence of wolves. In these models, sheep herd by sending in orders when other sheep are known to be present; the wolves respond by trading when the sheep trade. A problem in these models is either that multiple equilibria abound, or that no equilibrium exists at all.

How then do actual traders behave in securities markets? Do uninformed traders meekly head to market, or are they more sensitive to the dangers of certain trading environments? Are informed traders strategic in the sense of hiding amongst the uninformed, or do more carnal urges force them to act more competitively? These are empirical questions, and their answers have important implications for the resultant price processes in markets.

In this paper, we develop a theoretical and empirical framework for addressing these questions. Our model is a dynamic extension of the microstructure model of Easley and O'Hara (1992), in which a competitive market maker sets bid and ask prices based on her forecast of the composition of the traders (informed versus uninformed) and the probability of good or bad news. In Easley and O'Hara (1992) the arrival rates of traders are assumed to be constant and iid over time. Easley, Kiefer, and O'Hara (1997a) relaxed these stringent assumptions to allow for greater complexity in the arrival process for uninformed trades, and in particular allowed uninformed trades to be path dependent within trading days. Lei and Wu (2000) consider a model in which trades are independent within trading days, but have arrival rates that follow a Markov switching process with history dependent switching probabilities. In this research, we consider independent arrivals within trading days, but we allow the arrival rates of informed and uninformed traders to be time-varying and forecastable. In particular, we propose a generalized autoregressive bivariate vector process for (i) the arrival rates and (ii) the log of the arrival rates. We estimate both models on 16 actively traded stocks listed on the New York Stock Exchange over 15 years of transactions data. The performance of the two dynamic models is similar, suggesting that common features of the trade dynamics underlie all stocks investigated.

Our approach is a blending of model-based microstructure with the literature analyzing the econometric determinants of the joint dynamics between trades and prices. Examples of this research include

Hasbrouck (1991), Dufour and Engle (2000), Engle (2000), Engle and Russell (1998), Manganelli (2000), and Engle and Lange (2001). In common with that literature, we develop a half-life measure for market depth that is closely related to Engle and Lange's VNET measure (defined as the excess volume of buys and sells associated with a price movement). However, our approaches differ in that we derive and estimate the trade implications on prices and markets from a dynamic microstructure model, in contrast to their exogenous dynamic specification. Our work is also related to research linking order imbalances to market wide liquidity, see, for example, Chordia, Roll, and Subrahmanyam (2001). These authors relate overall trade imbalances to market returns and to market liquidity. Our analysis also involves order imbalances, but our model analyzes the richer order flow processes, rather than the static and exogenous total imbalance. Our model shows why particular components of order imbalances matter, thus providing an econometric structure for investigating order flow information.

We find a number of results on the arrival processes, the most important of which we highlight here. First, the arrival rates of both informed and uninformed trades are highly persistent. A heavy trading day is more likely to be followed by another heavy trading day. Furthermore, uninformed traders tend to follow their own type (herding), and they move to avoid informed traders. Intriguingly, uninformed traders refrain from entering the market after a day with many informed traders; in effect, the sheep remain in the barn when the trading climate is inclement. Informed traders, on the other hand, are not as responsive to the arrival of uninformed traders. These traders exhibit little strategic behavior, suggesting that information flow is well captured by models of competitive informed trading. This last result may be particularly important for empirical analyses, as it suggests that informed trade per se does not introduce complex patterns into either trades or the resultant prices.

Given the forecasted arrival rates, we then investigate the dynamic interactions between the arrival rates and market volatility, liquidity, and market depth. We find that forecasted arrival rates of both types of trades are positively correlated with intra-day volatility measures. Hence, potentially we could use forecasted arrival rates to enhance the forecasting of daily volatilities. We also find the expected result that market spreads are increasing in informed arrival rates, and the perhaps not so expected result that information events appear to be fully revealed by each day's end. We use Bayesian updating to calculate a measure of market depth we term the half-life. This measure is defined as the number of buys needed for the price impact to exceed some pre-specified maximum. Our analysis reveals a number of interesting properties of this market depth measure, with a particular finding being that it takes more trades to reveal information in a heavy trading day than in a light trading day (i.e. the market is deeper in the presence of heavier trading activities).

The paper is organized as follows. Section II describes the benchmark model of Easley and O'Hara (1992) and our dynamic extensions. Section III describes the data set and our estimation procedure. Section IV discusses the implications of our estimates for the arrival processes of informed and uninformed trades. Section V describes the implications of our estimates for market volatility, liquidity and depth. Section VI explores potential applications and future research.

II. Model Formulation

A. The Static Model

We follow Easley and O'Hara (1992) and Easley, Kiefer and O'Hara (1996, 1997a, 1997b) in modeling a market in which a competitive market maker trades a risky asset with uninformed and informed traders. Trade occurs over T discrete trading days and, within each trading day, trade occurs in continuous time. Information events occur between trading days with probability α. When these events occur they are either bad news, with probability δ, or good news with probability, $1 - \delta$. Traders informed of bad news sell and those informed of good news buy. We assume that orders from these informed traders follow a Poisson process with daily arrival rate μ. Uninformed traders trade for liquidity reasons. We assume that buy and sell orders from uninformed traders each arrive at the market according to a Poisson process with daily arrival rate ε. A more extensive discussion of this structure can be found in Easley, Kiefer and O'Hara (1996, 1997a, 1997b).

On day t, conditional on the parameter vector of the model, $\Theta \equiv [\mu, \varepsilon, \alpha, \delta]^\top$, the probability of observing B buys and S sells is given by

$$
\begin{aligned}
\Pr[\mathbf{y}_t = (B,S)|\Theta] \quad = \quad & \alpha(1-\delta)e^{-(\mu+2\varepsilon)}\frac{(\mu+\varepsilon)^B(\varepsilon)^S}{B!S!} \\
& +\alpha\delta e^{-(\mu+2\varepsilon)}\frac{(\mu+\varepsilon)^S(\varepsilon)^B}{B!S!} \\
& +(1-\alpha)e^{-2\varepsilon}\frac{(\varepsilon)^{B+S}}{B!S!}, \qquad (1)
\end{aligned}
$$

where $\mathbf{y}_t$ denotes the observation vector (number of buys and sells) for day t. The probability can be regarded as a mixture of three Poisson probabilities, weighted by the probability of having a "good news day" $\alpha(1-\delta)$, a "bad news day" $\alpha\delta$, and a "no news day" $(1-\alpha)$.

The model is static in the sense that each day the arrivals of an information event and trades, conditional on information events, are drawn from identical and independent distributions. The likelihood function can hence be written as a simple product of the above probability density over days. The log likelihood function, after dropping a constant term and rearranging, can be written as

$$
\begin{aligned}
\mathcal{L}(\{\mathbf{y}_t\}_{t=1}^T|\Theta) \quad = \quad & \sum_{t=1}^T[-2\varepsilon + M\ln x + (B+S)\ln(\mu+\varepsilon)] \qquad (2) \\
& +\sum_{t=1}^T \ln\left[\alpha(1-\delta)e^{-\mu}x^{S-M} + \alpha\delta e^{-\mu}x^{B-M} + (1-\alpha)x^{B+S-M}\right],
\end{aligned}
$$

where $M \equiv \min(B,S) + \max(B,S)/2$, and $x \equiv \frac{\varepsilon}{\mu+\varepsilon} \in [0,1]$. Given an information event, x captures the ratio of the arrival rates of "wrong trades" to that of "right trades." A trade is "wrong" when it is a buy in the presence of a bad signal or a sell in the presence of a good signal; a trade is "right" when it conforms with the information signal. The factoring of x^M is done to increase the computing efficiency and reduce truncation error, especially when the numbers of buys and sells are large.

B. Information Content of Trades

According to this model, data on daily arrivals of buys and sells contain information about the underlying parameters of the model. Let $TT = S + B$ denote the total number of trades per day. Then $\mathbb{E}[TT]$ is equal to the sum of the Poisson arrival rates of informed and uninformed trades:

$$\mathbb{E}[TT] = \alpha(1-\delta)(\varepsilon+\mu+\varepsilon)+\alpha\delta(\mu+\varepsilon+\varepsilon)+(1-\alpha)(\varepsilon+\varepsilon) = \alpha\mu+2\varepsilon.$$

Furthermore, the expected value of the trade imbalance $K = S - B$ is given by:

$$\mathbb{E}[K] = \alpha\mu(2\delta-1).$$

Hence, when the probability of bad news δ is not exactly one half, the mean of trade imbalance provides information on the arrival of informed trades. A more informative quantity is the absolute value of the trade imbalance. The expectation on absolute differences of Poisson variables takes rather complicated forms(Katti 1960), but the following approximate relation holds when μ is large:

$$\mathbb{E}[|K|] \doteq \alpha\mu.$$

These relationships provide a basis for estimating and interpreting the arrival rates of trades. The absolute trade imbalance $|K|$ contains information on the arrival of informed trades, $\alpha\mu$. In contrast, the variable $(TT-|K|)$ is essentially the balanced trade in the market and contains information on the arrival of uninformed trades. We incorporate these two trade quantities into our arrival rates forecasting specifications.

C. Time-Varying Arrival Rates of Trades

The constant arrival rates in the above model imply that both the number of balanced trades $TT-|K|$ and the trade imbalance $|K|$ should be iid over time. Yet, such a structure seems both rigid and unrealistic. A more complete structure would allow for autocorrelation for both series and cross-correlation between them, in addition to a time trend for each series. To allow for these effects, we specify a dynamic vector process for the arrival rates of informed and unformed traders. The arrival rate of informed trades is $\alpha\mu$ and the arrival rate of the uninformed trades is 2ε. Let $\psi = [\alpha\mu, 2\varepsilon]^\top$ denote the vector of the two arrival rates. To remove any deterministic trend on arrival rate of trades, we model the detrended arrival rates $\widetilde{\psi}_{it} = \psi_{it} e^{-g_i t}, i = 1, 2$, as a vector stationary process, where the vector $g \equiv [g_1, g_2]^\top$ captures the growth rates of the two arrival rates.

C.1. A generalized autoregressive specification on arrival rates of trades

We propose a bivariate generalized autoregressive forecasting relation on the detrended arrival rates of informed and uninformed trades:

$$\widetilde{\psi}_t = \omega + \sum_{k=1}^{p} \Phi_k \widetilde{\psi}_{t-k} + \sum_{j=0}^{q-1} \Gamma_j \widetilde{\mathbf{Z}}_{t-j}, \tag{3}$$

where $\widetilde{\psi}_t$ denotes the detrended time t forecast of the arrival rate vector at time $t+1$, $\mathbf{Z}_t \equiv [|K_t|, TT_t - |K_t|]^\top$ denote the time-t observables (number of trade imbalance and balanced trades), and $\widetilde{\mathbf{Z}}_{it} = \mathbf{Z}_{it} e^{-g_i t}, i = 1, 2$. As a first order approximation, $\mathbb{E}_{t-1}\left[\widetilde{\mathbf{Z}}_t\right] \doteq \widetilde{\psi}_{t-1}$, and the specification in (3) is analogous to the GARCH specification of Bollerslev (1986) on conditional volatilities. As in GARCH models, the above forecasting relation can be rewritten as an $ARMA(\max[p,q], q)$ process:

$$\widetilde{\psi}_t \doteq \omega + \sum_{k=1}^{\max[p,q]} \widehat{\Phi}_k \widetilde{\psi}_{t-k} + \sum_{j=0}^{q} \Gamma_j \xi_{t-j}, \tag{4}$$

where

$$\widehat{\Phi}_k = \begin{cases} \Phi_k + \Gamma_{k-1} & \text{if} \quad k \leq q \\ \Phi_k & \text{if} \quad k > q \end{cases},$$

and $\xi_t \equiv \widetilde{\mathbf{Z}}_t - \mathbb{E}_{t-1}\left[\widetilde{\mathbf{Z}}_t\right] \doteq \widetilde{\mathbf{Z}}_t - \widetilde{\psi}_{t-1}$ denotes the forecasting error. The stationarity of the process requires that the eigenvalues of $\widehat{\Phi}_k$ be less than one. We set $p = q = 1$ for the model calibration. Adding back the time trend, we can rewrite the forecasting relation as

$$\psi_t = \omega e^{gt} + \Phi \psi_{t-1} e^{g} + \Gamma \mathbf{Z}_t, \tag{5}$$

where the products ωe^{gt} and $\psi_{t-1} e^{g}$ are short hands for element by element operations.

Given the forecasting relation in (5), the log likelihood function has the same form as in (2), with the constant arrival rates replaced by their respective conditional forecasts. However, equation (5) forecasts the product of the two parameters $\alpha\mu$ while in the likelihood function we need values of the two parameters separately. To separate them, we assume that α, the probability of an information event, is constant over time.[1]

A backward equation analogous to (3) also holds for the same time series:

$$\widetilde{\psi}_t = \omega + \sum_{k=1}^{p} \Phi_k \widetilde{\psi}_{t+k} + \sum_{j=0}^{q-1} \Gamma_j \widetilde{\mathbf{Z}}_{t+j}.$$

We use this backward equation to forecast the starting value of ψ. In particular,

$$\psi_t = \omega e^{gt} + \Phi \psi_{t+1} e^{-g} + \Gamma \mathbf{Z}_t. \tag{6}$$

Refer to Box and Jenkins (1976), Chapter 6.3, for details. Again, the products ωe^{gt} and $\psi_{t+1}e^{-g}$ both denote element by element operations.

C.2. A generalized autoregressive specification on the logarithm of arrival rates

Since arrival rates only exist in the positive hyperplane, $\psi \in \mathbb{R}^{2+}$, we propose an alternative forecasting relation on the logarithm of the arrival rates:

$$\ln \widetilde{\psi}_t = \omega + \sum_{k=1}^{p} \Phi_k \ln \widetilde{\psi}_{t-k} + \sum_{k=0}^{q-1} \Gamma_j \mathbf{M}_{t-j}, \tag{7}$$

where $\mathbf{M}_t$ is an approximate martingale difference vector formulated from $\mathbf{Z}_t$:

$$\mathbf{M}_{it} = \frac{\mathbf{Z}_{it}}{\psi_{i(t-1)}} - 1, \quad i = 1, 2, \tag{8}$$

[1]The growth of trades suggests that both uninformed and informed trades are time varying. Informed trades could vary because of variation in μ or α, or both. We find it more plausible that the arrival rate of informed traders is time varying than that information events are time varying. However, it is also possible that information events follow a stochastic process that we miss-identify as variation in informed trades with this assumption.

where the subscript i denotes the i-th element of the vector. The approximate martingale property follows readily from the fact that $\mathbb{E}_{t-1}[\mathbf{Z}_t] \doteq \psi_{t-1}$. The specification is analogous to the EGARCH model of Nelson (1991) for conditional volatilities.

We again set $p = q = 1$ for the calibration. Plug in the time trend, we have:

$$\ln \psi_t = \omega + \Phi g + (I - \Phi) gt + \Phi \ln \psi_{t-1} + \Gamma \mathbf{M}_t. \tag{9}$$

Again, a backward equation is used to determine the initial value:

$$\ln \psi_t = \omega - \Phi g + (I - \Phi) gt + \Phi \ln \psi_{t+1} + \Gamma \mathbf{E}_t,$$

where $\mathbf{E}_t$ is a backward analogue to $\mathbf{M}_t$: $\mathbf{E}_{it} \equiv \mathbf{Z}_{it} / \psi_{i(t+1)} - 1$.

III. Data and Estimation

We estimate our two models using data from 16 stocks: Ashland (ASH), Exxon Mobil (XOM), Duke Energy (DUK), Enron (ENE), AOL Time Warner (AOL), Philip Morris (MO), ATT (T), Pfizer (PFE), Southwest Air (LUV), AMR (AMR), Dow Chemical (DOW), CitiGroup (C), JP Morgan Chase (JPM), Wal Mart (WMT), Home Depot (HD), and General Electric (GE). We chose representative stocks from a variety of industries which each had high trading volume and were listed on the NYSE. The latter criterion is intended to avoid differences introduced by different trading platforms. Trade data for these stocks are taken from the TAQ transactions database for the period January 3rd, 1983, to December 24th, 1998 (3891 business days). A minimum level of trading activity is necessary to extract the information changes from each day, so we exclude days when there are either no buys or no sells. The least active stock is Enron, from which we drop 69 inactive days, then Wal Mart (19 days), Exxon Mobil (18 days), Southwest Air (7 days), Pfizer (4 days), ATT (4 days), Philip Morris (3 days), JP Morgan Chase (2 days), Exxon Mobil (1 day), and Ashland (1 day). Furthermore, the data for AOL Time Warner, CitiGroup, and Home Depot start late. The starting dates are, respectively, September 16th, 1996, October 29th, 1996, and April 19th, 1984.

The TAQ data provide a complete listing of quotes, depths, trades, and volume at each point in time for each traded security. For our analysis, we require the number of buys and sells for each day, but the TAQ data record only transactions, not who initiated the trade. This classification problem has been dealt with in a number of ways in the literature, with most methods using some variant on the uptick or downtick property of buys and sells. In this article, we use a technique developed by Lee and Ready (1991). Those authors propose defining trades above the midpoint of the bid-ask spread to be buys and trades below the midpoint of the spread to be sells. Trades at the midpoint are classified depending upon the price movement of the previous trade. Thus, a midpoint trade will be a sell if the midpoint moves down from the previous trade (a downtick) and will be a buy if the midpoint moves up. If there is no price movement then we move back to the prior price movement and use that as our benchmark. We apply this algorithm to each transaction in our sample to determine the daily numbers of buys and sells.[2]

To investigate the interactions between trades and prices, we also download the daily open (O), high (H), low (L), and close (C) prices from Bloomberg corresponding to the same stocks and time periods. In particular, we proxy the intra-day volatility by the absolute returns on the open-close ($|\ln C/O|$) and high-low ($\ln H/L$).

We begin by analyzing the properties of the trade variables. Table 1 reports the summary statistics of the trade quantities $\mathbf{Z} = [|K|, TT - |K|]$, or the imbalanced and balanced trade variables. We observe the following features:

1. *Trades are increasing.* The daily number of balanced trades $TT - |K|$ in general grows faster than the trade imbalance K. The estimated annual growth rate for the balanced trade ranges from 2.4% for DOW to 94% for AOL. The growth rate for the trade imbalance ranges from negative for XOM (-3.66%) and DOW (-1.51%) to 133% for AOL.

2. *The number of balanced trades is more volatile than trade imbalance.* For all stocks investigated, the standard deviation of the balanced trades is much larger than the standard deviation of the trade imbalance. Standard deviations are measured on the detrended residuals. Furthermore, the intercept of the detrending regression is also larger for the number of balanced trades $TT - |K|$

[2] The first trade at each day is excluded from our sample as it is determined by a different mechanism.

than for the trade imbalance $|K|$, implying that the number of balanced trades dominates the total trades.

3. *Trades are highly persistent.* The number of balanced trades are more persistent than the trade imbalance. The first order autocorrelation for balanced trade ranges from 0.697 to 0.953 while that for the trade imbalance ranges from 0.145 and 0.772. Autocorrelations are measured on the detrended residuals. This suggests a complexity to the order arrival process that is not well captured by static models. It also suggests that informed and uninformed trade behavior may exhibit interesting complex dynamics, an issue we address in the next section.

4. *Balanced trades and trade imbalances are cross-correlated.* The two quantities are generally positively correlated. The cross-correlation coefficient between the balanced trade $TT - |K|$ and the trade imbalance $|K|$ ranges from -0.004 for XOM to 0.802 for Citigroup.

We next estimate the parameters of the two dynamic order arrival models in (5) and (9) by maximizing the log likelihood function in (2). We refer to the GARCH analogue as Model A and the EGARCH as Model B. The results are summarized in Table II for Model A and in Table III for Model B. The log likelihoods from the two models are very close to each other, neither consistently dominates the other across the 16 stocks. Estimates for the two models also imply similar properties for the arrival rates.

IV. The Dynamics of the Arrival Rates

We now turn to analyzing the behavior of the order arrival processes. The structural models derived in Section II provide a framework for analyzing the potentially complex processes characterizing order flow. As reported above, the models can be estimated with reasonable precision, allowing us to test the importance and significance of alternative behavioral hypotheses. Our focus here is on the dynamics of informed and uninformed order flow, and in particular on the factors that influence the correlation structures of trades.

A. Are order flows correlated over time?

The basic structure analyzed in standard microstructure model is of orders arriving in a probabilistic fashion from informed and uninformed traders. A simple construct of this behavior is to characterize orders by Poisson arrival processes, where the arrival rates are constant, but may differ across informed and uninformed traders. The models we derive here allow for much greater complexity in behavior, and in particular we allow trades to be both auto-correlated and cross-correlated.

The persistence of the arrival rate for uninformed traders is captured by $\widehat{\Phi}_{22} = \Phi_{22} + \Gamma_{22}$ for Model A and Φ_{22} for Model B. For informed traders, this persistence is captured by $\widehat{\Phi}_{11} = \Phi_{11} + \Gamma_{11}$ and Φ_{11}, respectively for Models A and B. The large, and positive estimates for all of these variables shows that orders overall are highly auto-correlated, with a high arrival day more likely to be followed by another high arrival date. This behavior is not unexpected given that many studies have shown volume to be significantly, and positively auto-correlated. But this result is at variance with the predictions of microstructure models in which trades are viewed as iid. Perhaps more importantly, the result suggests that trade patterns are predictable across trading days.

We now turn to analyzing the specific behaviors exhibited by the informed and uninformed traders. For most stocks, the arrival rate of uninformed trades is much more persistent that that of informed trades. This suggests that uninformed trade is more likely to exhibit serial patterns across trading days than is informed trade. Thus, the characterization of uninformed traders as sheep herding together is consistent with these results. The uninformed do tend to move together, either trading or not trading, but doing so persistently. Informed trade is also persistent, but the persistence is much lower than it is for uninformed trade. This is consistent with new information being largely incorporated into security prices by the end of a trading day.

A natural concern in interpreting these estimates is the stationarity of the underlying processes. As noted earlier, theoretical models suggest both a wide range of possible equilibrium trading strategies, as well as the possibility that no equilibrium exists at all. This issue can be addressed by examining the eigenvalues for Models A and B. For the two processes to be stationary, the two eigenvalues of $\widehat{\Phi}$ for Model A and the two eigenvalues of Φ for Model B need to be less than one. Table IV reports these eigenvalues for the 16 stocks in our sample. The second eigenvalue is very close to one for

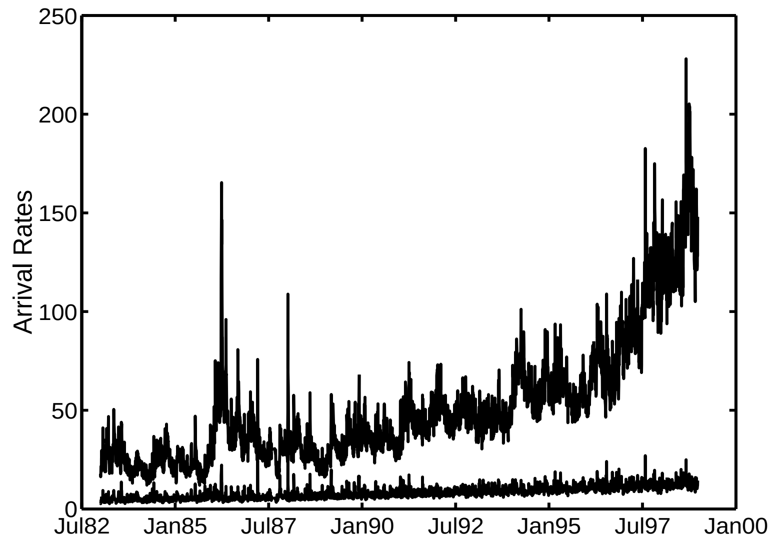

Figure 1. Arrival Rates of Trades
The arrival rates of uninformed trades (top) and informed trades (bottom) are forecasted over time based on Model B for Ashland, with parameter estimates in Table III.

most stocks. For three stocks under Model A and two under Model B, the second eigenvalue becomes slightly greater than one, implying non-stationarity. These results further confirm that the arrival rates are highly persistent.

Figure 1 shows a typical time series of the forecasted order arrival rates for one stock in our sample, Ashland Oil. Here the distinctions between the two series are apparent. Uninformed trade dominates the total order flow, a finding consistent with the general functioning of liquid markets. The arrival rates of the uninformed trades not only grow faster over time, but they fluctuate more as well. This greater volatility suggests that uninformed trade may exhibit complex dependencies, an issue we now consider.

B. The sheep and the wolf: Do uninformed traders try to avoid informed traders?

From above we know that a high arrival day is likely to be followed by another high arrival day. But what if the high total trade arrivals reflect greater informed trade? Will the uninformed continue to arrive en mass to the market, or will they be dissuaded from trading by the expected presence of informed traders? The impact of previous informed arrival rates on the current uninformed arrival rates is captured by $\widehat{\Phi}_{21}$ for Model A and by Φ_{21} for Model B. These variables measure the cross correlation effects of lagged informed trade on uninformed trade. Thus, a direct test of this avoidance hypothesis is to examine the sign and significance of these variables.

The estimates of $\widehat{\Phi}_{21}$ for Model A and of Φ_{21} for Model B are remarkably negative for all stocks. This is evidence of the uninformed systematically avoiding trading when the informed are expected to be present. This behavior is not predicted by microstructure models, which view the only determinant of uninformed trading as the presence of other uninformed traders.[3] But the strategy seems sensible nonetheless. Why venture into the trading arena when it is more likely to be populated by wolves? A better strategy is to simply stay away, and that appears to be what the uninformed tend to do.

The arrival of uninformed traders is also influenced by the most recent realizations on the number of balanced trades and trade imbalances. In particular, the impact of the trade imbalance $|K|$ on the uninformed arrival is captured by $\Gamma_{21} - \Gamma_{22}$ in Model A. The results here are mixed. The estimates for the difference are negative for seven firms and positive for nine. Hence, the impacts of the trade imbalance are ambiguous. Estimates from Model B are harder to interpret as the impact of $|K|$ would be captured by

$$\frac{\Gamma_{21}}{\alpha\mu_t} - \frac{\Gamma_{22}}{2\varepsilon_t}$$

and hence is actually time-varying by model design. The time series averages of this impact are very small and have mixed signs across stocks.

Alternatively, the impact of the total number of trades on the arrival rate of uninformed traders is captured by Γ_{22} under both models. Here the estimates are unambiguously positive for all stocks, implying that an increase in the total number of trades today forecasts an increase in the arrival of uninformed traders tomorrow.

[3] An exception to this characterization is Lei and Wu (2000) who allow uninformed trade to be affected by factors such as momentum and loss aversion.

C. The wolf revisited: Do informed traders stalk uninformed traders?

If informed traders are strategic, then their trades should depend upon the order arrivals of the uninformed. In particular, Kyle's model dictates that the informed choose their orders to blend in with the uninformed, suggesting that there would be strong cross-correlation effects of uninformed trade on informed trade. Alternatively, if the informed act competitively then their trading strategy is more mechanistic: trade until the price reaches the new true value, and then stop. This trading strategy suggests little cross-effects, but would be characterized by strong autocorrelation effects.

We can test for just such behaviors in our model. The impact of previous day's uninformed order arrival on today's informed arrivals is captured by $\widehat{\Phi}_{12}$ for Model A and by Φ_{12} for Model B. If the informed act strategically, we would expect these variables to be large and positive. The estimates reveal a different story. The estimates in both Models are small, and they are not consistently positive or negative across stocks. This is consistent with the simpler, competitive model of informed trade. Informed traders act on information, and hence they do not respond to the activity of uninformed traders.

Informed trades do tend to be affected by overall volume. The impact of the total number of trades on the arrival rate of informed traders is captured by Γ_{12} under both models. The estimates are positive for all stocks and under both models. Thus, an increase in the total number of trades today forecasts an increase in the arrival rates of both informed and uninformed trades tomorrow. The impact of the trade imbalance variable is more problematic. This effect is captured by $\Gamma_{11} - \Gamma_{12}$ under model A. The results here are predominantly positive. Under Model B, the impact is captured by

$$\frac{\Gamma_{11}}{\alpha\mu_t} - \frac{\Gamma_{12}}{2\varepsilon_t},$$

which again is time varying and harder to interpret. The time series averages have mixed signs across stocks.

In summary, we have found that the order arrival processes exhibit a wide range of complex behaviors. Uninformed trades tend to be highly persistent, and volatile. Uninformed order arrivals clump together, with high volume days more likely to follow high volume days, and conversely. This behavior is consistent with the sheep characterization found in the literature, but with a significant difference: the uninformed here are smart enough to try to avoid the wolf. The informed also exhibit complex patterns, but these patterns are not consistent with the strategic behavior posited in the literature. The informed do not appear to hide in the order flow, but instead trade persistently. There is a smaller autocorrelation effect of informed order arrivals across days, consistent with information being revealed during the trading day.

V. Interactions Between Trades and Market Volatility, Liquidity, and Depth

We now turn to analyzing the relation between trades and market parameters such as volatility, liquidity, and depth. As noted in the Introduction, each of these parameters is linked to trades because the order arrival process greatly influences subsequent market behavior. Our model provides a way to characterize this inter-dependence, as well as a framework for testing specific hypotheses.

A. Intra-day volatility

To investigate the interaction between trades and price volatility, we construct two intra-day volatility measures: (1) the absolute return on daily open-close $|\ln O/C|$, and (2) daily high-low, $\ln H/L$. Our model provides detrended order arrival forecasts, so we use these predicted estimates to characterize how the market's beliefs regarding order arrivals affect price volatility.

A natural starting point is to consider what relation we would expect to find between trades and volatility. A market with high order flow is generally viewed as a deep market, or one in which orders can be accommodated without large impacts on price. In the same vein, it could be argued that greater order flow brings greater potential for buyers and sellers to cross, and so again price effects would be small. These arguments suggest that forecasted high order arrival rates would result in low market volatility. Alternatively, there are two arguments for suggesting the opposite relation. First, large order flow may expose market makers to significant inventory imbalances. To mitigate this exposure, market makers may widen spreads or otherwise move prices, which would increase price volatility. Greater order flow may also signal the presence of new information, and prices would naturally gravitate toward new equilibrium levels. Such information-linked effects would also suggest a positive relation between order arrivals and price volatility.

Table V reports the correlation between the detrended arrival rates forecasts and the two (realized) intra-day volatility measures. One result is immediately apparent: there is a positive relation between order arrivals and volatility. This positive relation holds for 15 of the 16 stocks in our sample, and is robust across both model specifications. Numerous researchers have shown that there is a positive empirical relation between volume and volatility. But our results suggest that this relation is actually deeper, that volatility is positively correlated with the predicted order arrival processes.

If information effects dominate inventory effects, then we would expect this link to be more positive for information-based order arrivals than it is for uninformed order arrivals. The evidence in Table V does not support this. While this effect holds true for some stocks, it does not for others. A more direct test of the link between informed arrivals and volatility is to look at the correlation between the intra-day volatility measures and the composition of trades. In particular, we define β_t to the the forecasted proportion of informed trades, where β_t is defined as:

$$\beta_t \equiv \frac{\alpha\mu_t}{\alpha\mu_t + 2\varepsilon_t}.$$

Table VI reports these correlations. The signs of the correlations vary across stocks, yielding no clear prediction. However, the estimated correlations are small in any case, suggesting little link between trade composition and intra-day price volatility. These results indicate that while volatility increases with the forecasted arrival rates of total trades, it is relatively independent of the forecasted trade composition.

In the same table, we also report the correlation of trade composition forecasts β with the total number of realized trades TT, as well as the realized proportion of trade imbalance, $|K|/TT$. The correlation with TT is negative, implying that the relative proportion of informed trades decreases with increasing total trades. The correlations with $|K|/TT$ are mostly positive as expected since β_t can be regarded as an approximate forecast of $|K|/TT$.

B. Market liquidity and bid-ask spread

Market liquidity is often measured by the bid-ask spread of the security prices with markets in which the bid-ask spread is small being interpreted as liquid markets. We use our model to derive the bid-ask spread as a function of the trade sequence and arrival rate forecasts. We then use forecasted arrival rates to determine how the components of order arrival affect market liquidity.

An application of Bayes rule shows that the probabilities of a good (g) and a bad (b) information event conditional on a sell order at time t, $sell_t$, are given by, respectively,

$$\Pr(g|sell_t) = \frac{\Pr_t(g)\,\varepsilon}{\Pr_t(b)\,\mu+\varepsilon}, \quad \Pr(b|sell_t) = \frac{\Pr_t(b)\,(\varepsilon+\mu)}{\Pr_t(b)\,\mu+\varepsilon}, \tag{10}$$

where $\Pr_t(g)$ and $\Pr_t(b)$, denote, respectively, the prior probabilities at time t of a good and a bad information event. Let $\overline{V}$ be the expected asset value conditional on good news and $\underline{V}$ be the expected value conditional on bad news. Let $V^* \equiv (1-\delta)\overline{V} + \delta\underline{V}$ denote the unconditional expected value of the asset.

In a competitive market, the bid price must provide the market maker a zero expected profit conditional on a trade at the bid; that is, the arrival of a sell order. The bid price is thus the expected value of the asset conditional on history and on the arrival of a sell order:

$$\begin{aligned} bid_t &= \Pr(g|sell_t)\,\overline{V} + \Pr(b|sell_t)\,\underline{V} + \Pr(n|sell_t)\,V^* \\ &= V^* + \left(\overline{V} - \underline{V}\right)\frac{\delta\Pr_t(g)\,\varepsilon - (1-\delta)\Pr_t(g)\,(\varepsilon+\mu)}{\Pr_t(b)\,\mu+\varepsilon}, \end{aligned} \tag{11}$$

where $\Pr(n|sell_t) = 1 - \Pr(g|sell_t) - \Pr(b|sell_t)$ is the probability of no information event.

Analogously, the probabilities of a good and a bad information event conditional on a buy order, buy_t, are given by

$$\Pr(g|buy_t) = \frac{\Pr_t(g)\,(\varepsilon+\mu)}{\Pr_t(g)\,\mu+\varepsilon}; \quad \Pr(b|buy_t) = \frac{\Pr_t(b)\,\varepsilon}{\Pr_t(g)\,\mu+\varepsilon}. \tag{12}$$

The ask price is therefore,

$$ask_t \quad = \quad V^* + \left(\overline{V} - \underline{V}\right) \frac{\delta \Pr_t(g)(\varepsilon + \mu) - (1-\delta)\Pr_t(b)\varepsilon}{\Pr_t(g)\mu + \varepsilon}. \tag{13}$$

The bid-ask spread can now be computed from equations (11) and (13). In particular, at the opening, the unconditional probabilities of good and bad information events are, respectively,

$$\Pr(g) = (1-\delta)\alpha, \quad \Pr(b) = \delta\alpha.$$

The opening bid-ask spread (OS) is therefore given by

$$OS = \left(\overline{V} - \underline{V}\right)\delta(1-\delta)\alpha\mu\left[\frac{(\alpha\mu + 2\varepsilon)}{((1-\delta)\alpha\mu + \varepsilon)(\delta\alpha\mu + \varepsilon)}\right].$$

If we further assume that $\delta = 1/2$, i.e. bad and good news have equal probabilities, the opening bid-ask spread simplifies to

$$OS = \left(\overline{V} - \underline{V}\right)\frac{\alpha\mu}{\alpha\mu + 2\varepsilon} = \left(\overline{V} - \underline{V}\right)\beta. \tag{14}$$

Hence, the opening bid-ask spread is proportional to the significance of the information event $\left(\overline{V} - \underline{V}\right)$ and the forecasted proportion of informed arrivals, β. As Table VI illustrates that β is negatively correlated with the total number of trades for most stocks, assuming relative time stability on the significance of the information event $(\overline{V} - \underline{V})$, market liquidity increases with increasing trades. When $\delta \neq 1/2$, the impacts are not exactly captured by β, but similar observations apply. Generally, more active markets are more liquid markets.

C. Market depth and price impacts of consecutive trade orders

When a trader tries to load or unload a large position by putting in consecutive buy or sell orders to the market, the price change could be significant. The price impact of a sequence of trade orders can be computed by repeated application of equations (10)-(13).

We take consecutive buy orders as an example. Let $\Pr_t^{N-1}(g)$ and $\Pr_t^{N-1}(b)$ denote the probabilities of a good and a bad information event conditional on $N-1$ consecutive buy orders. From (13), we have that the price impact of N consecutive buys is

$$\begin{aligned} ask_t^N &= V^* + (\overline{V} - \underline{V})\gamma_t^N, \\ \gamma_t^N &= \frac{\delta \Pr_t^{N-1}(g)(\varepsilon+\mu) - (1-\delta)\Pr_t^{N-1}(b)\,\varepsilon}{\Pr_t^{N-1}(g)\,\mu+\varepsilon}. \end{aligned}$$

The probabilities $\Pr_t^{N-1}(g)$ and $\Pr_t^{N-1}(b)$ can be readily updated via Bayes rule as in (12), starting with the unconditional priors at the opening. As the number of consecutive buy orders increases, the probability of a good information event increases and approaches unity while the probability of a bad information event approaches zero. The price impact γ_t^N converges to δ and the price converges to the expected upper bound of the asset $\overline{V}$. The speed of convergence governs the depth of the market and is determined by the arrival rate forecasts (μ,ε) on that day.

In Figure 2, we plot a typical price impact curve, γ_t^N, as a function of the number of consecutive buy orders, N, based on arrival rates forecasts on Ashland on December 24th, 1998 (the last day of observation). On that day, the forecasted arrival rate of informed trades is 14.33 and that of uninformed trades is 148.00. As shown in Figure 2, the price of the asset converges to the high value $\overline{V}$ after fewer than 20 consecutive buys.

Similar curves can be computed for N consecutive sells and for any sequence of buys and sells. Knowledge of the price impact curve is obviously very important for institutional portfolio managers in designing strategies of loading or unloading large positions.

Engle and Lange (2001) define a market depth measure VNET, which is intended to capture the net order flow associated with a fixed price movement. On each day, given the arrival rate forecasts, we construct an analogous measure of market depth: the half life $(\tau_{1/2})$ of the price impacts for consecutive buys. Our measure is defined as the number of buys N needed for the price impact γ_t^N to exceed half of its maximum (δ). Nevertheless, our half life measure and VNET differ in at least two important aspects. First, VNET is defined on the excess trading volume while we are only concerned with the number of trades. Trade size does not play a role in our analysis. A second difference is that VNET implicitly assumes that the sequence of trades does not matter, only the net trade imbalance affects prices. In

Figure 2. Price Impacts of Consecutive Buys

The line depicts the price impact of consecutive buys (γ_t^N) for Ashland, based on the arrival rate forecasts from Model B on the last day of the data (December 24th, 1998). The parameter estimates of the model are reported in Table III.

our model, however, the exact sequence of trading history also plays a role in the price movement. We therefore specifically define the half life as a function of consecutive number of buys, not on net order flows.

Figure 3 depicts two typical times series of our market depth forecasts for Enron on the left and Pfizer on the right, implied by Model B estimates. For both stocks, the market depth measured by half life has increased in the 90s.

Tables VII reports the mean half life for each stock as well as its correlations with trades and price volatilities. The half lives implied from the two models differ from each other, but they exhibit similar overall trends: Stocks such as Pfizer and AOL Time Warner have a deeper market than stocks such as Ashland. Furthermore, the half lives for most stocks are positively correlated with the total number of trades TT. It takes more trades to reveal information on a heavy trading day, i.e. the market is deeper in presence of heavier trade activities. On the other hand, the correlation between the half lives and the ratio of trade imbalance to total trades are mostly negative, implying that the market is deeper when we

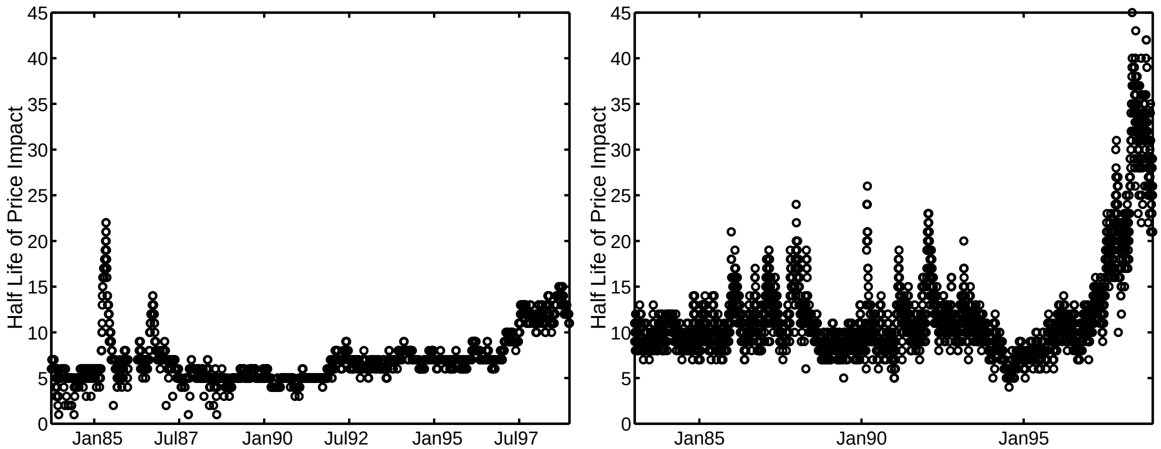

Figure 3. Time Varying Forecasts of Market Depth
Market depth is measured as the half life ($\tau_{1/2}$) of the price impact of consecutive buy orders, defined as the number of consecutive buys needed for the impact to exceed half of its maximum. The half life is computed for Enron on the left panel and Pfizer on the right panel, based on estimates of Model B, reported in Table III.

have fewer informed trades. In summary, an increase in total trades increases the market depth while an increase in informed trades reduces it.

VI. Residual Analysis

One way to investigate the robustness of our specification is to check for structure in the residuals of forecasted order flows. If our specification captures the data well, we should find minimal structure from the following forecasting residuals on the absolute trade imbalance $|K|$ and balanced trades $TT - |K|$:

$$e_{it} = \frac{\mathbf{Z}_{it} - \mathbb{E}_{t-1}\left[\mathbf{Z}_{it}\right]}{\mathbb{E}_{t-1}\left[\mathbf{Z}_{it}\right]}, \quad i = 1, 2, \quad \text{with} \quad \mathbf{Z}_t = \left[|K_t|, TT_t - |K_t|\right].$$

The expected value on total trades is known analytically: $\mathbb{E}_{t-1}\left[TT_t\right] = \alpha\mu_{t-1} + 2\varepsilon_{t-1}$. We determine the expected value of $|K|$ by simulation. The residuals are represented as a percentage of their respective forecasted value. Table VIII reports the summary statistics for both residuals under Model A and under Model B. For each stock, the first row reports the properties of e_{1t} and the second row reports the properties of e_{2t}.

Compared to the summary statistics of **Z** in Table I, the forecasting residuals exhibit much less structure. In particular, the serial dependence (the first order autocorrelation) is significantly smaller, and in many cases not significantly different from zero. The cross correlations between the two residuals are also smaller than those between the two elements of the raw data **Z**.

Nevertheless, one can still discern some remaining structure in the residuals. In particular, both models seem to induce similar biases on the mean of the forecasting residuals. Both models under-forecast the absolute trade imbalance but slightly over-forecast the balanced trades.

VII. Conclusion

The dynamic models in this paper identify important forecasting relations in the arrival rates of trades. While the models have implications for the trading behavior and price impacts of trades, for future research, more information can be revealed and additional interesting applications can be found by combining our dynamic quantity models of trades with dynamic models of prices.

As an example, recall that the opening bid-ask spread (OS) each day can be written as a product of the significance of the information event and the composition of arrival rates.

$$OS_t = (\overline{V} - \underline{V})\beta_t,$$

where β_t captures the relative proportion of informed arrivals. Thus, given the forecasts of arrival rates, together with information on the high and low expected values, we can compute the opening bid-ask spread. Or upon the observation of the opening bid-ask spread, one can infer the high low difference, which is also a measure of the significance of the information event. More significant news in either direction would generate a higher high-low spread. We hence obtain a forecast of the significance of the information event.

In an actively traded market, given the presence of an information event, the significance of the information event will ultimately be revealed to the market at the close. In case of a positive information event, the transaction price converges to $\overline{V}$ and in case of a negative information event, it converges to $\underline{V}$. Therefore, the forecasted arrival rates and the opening bid-ask spread together reveal important information about the open-close spread. Let $V^* = (1-\delta)\overline{V} + \delta\underline{V}$ be the opening price, then the open-close spread is $(1-\delta)(\overline{V} - \underline{V})$ in case of a bad news, $\delta(\overline{V} - \underline{V})$ in case of a good news, and zero in case of no news.

As the open-close spread can also be regarded as a daily volatility forecast, it is interesting to compare a forecast from trade quantities and opening bid-ask spread with forecasts from the price process, such as volatility forecasts via GARCH type models. One can further investigate whether incorporating the information from the trade process increases the forecasting efficiency of GARCH type models. Furthermore, if the derivatives market only prices in GARCH type forecasts, information revealed from the trade process can potentially be used to design profitable trading strategies. For example, when the forecasted open-close spread is higher than already priced in the market, one can long a daily put, a call, or a straddle and delta hedge. All these positions profit from increasing volatilities but are more or less immune from directional bets.

References

Abramowitz, M., and I. A. Stegum, 1974, *Handbook of Mathematicall Functions*. (Dover New York).

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

Back, Kerry, C. Henry Cao, and Gregory A. Willard, 2000, Imperfect competition among informed traders, *Journal of Finance* 55, 2117–2156.

Bollerslev, Tim, 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307–327.

Box, George E. P., and Gwilym M. Jenkins, 1976, *Time Series Analysis: Forecasting and Control*. (Holden-Day San Francisco) revised edn.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2001, Trading activity and expected stock returns, *Journal of Financial Economics* 59, 3–28.

Dufour, Alfonso, and Robert F. Engle, 2000, Time and the price impact of a trade, *Journal of Finance* 55, 2467–2598.

Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, 1996, Cream-skimming or profit-sharing? the curious role of purchased order flow, *Journal of Finance* 51, 811–833.

Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, 1997a, The information content of the trading process, *Journal of Empirical Finance* 4, 159–185.

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

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

Engle, Robert F., 2000, The econometrics of ultra high frequency data, *Econometrica* 68, 1–22.

Engle, Robert F., and Joe Lange, 2001, Predicting VNET: A model of the dynamics of market depth, manuscript, New York University and Federal Reserve Board.

Engle, Robert F., and Jeffrey R. Russell, 1998, Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica* 66, 1127–1162.

Foster, D., and S. Vishwanathan, 1990, A theory of the intraday variations in volume, variance, and trading costs in securities markets, *Review of Financial Studies* 3, 593–624.

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

Hasbrouck, Joel, 1991, Measuring the information content of stock trades, *Journal of Finance* 46, 179–208.

Karlis, Dimitris, and Ioannis Ntzoufras, 2000, Discrete distributions with applications in sports, manuscript, Athens University of Economics and Business Greece.

Katti, S. K., 1960, Moments of absolute difference and sbsolute deviation of discrete distributions, *Annals of Mathematical Statistics* 31, 78–85.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1336.

Lee, Charles M. C., and Mark A. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733–746.

Lei, Qin, and Guojun Wu, 2000, The behavior of uninformed investors and time-varying informed trading activities, manuscript, University of Michigan.

Manganelli, Simone, 2000, Time, volume and price impact of trades, manuscript, European Central Bank Germany.

Nelson, Daniel B., 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347–370.

Table I
Summary Statistics of Trading Activities

Entries report the summary statistics of the trade quantities $\mathbf{Z} = [|K|, V - |K|]$, where $|K| = |S - B|$ is the trade imbalance (difference between number of sells and buys) and $TT = S + B$ is the total number of trades (sells plus buys) at each day. Under each ticker, the first row reports the properties of trade imbalance $|K|$ while the second row reports the properties of the number of balanced trades $TT - |K|$. The second column (g) reports the growth rates, estimated from the following regression:

$$\ln \mathbf{Z}_{it} = a + g_i t + e_t, \quad i = 1, 2.$$

The third column (a) reports the regression intercept estimate. The fourth column (St Dev) reports the standard deviation of the regression residual e_t. The fifth column (Auto) reports the first order autocorrelation of the residual. The last column (ρ) reports the cross-correlation between the trade imbalance $|K|$ and the number of balanced trades $TT - |K|$, measured on the detrended residuals.

Ticker	g,%	a	St Dev	Auto	ρ
ASH	5.073	0.921	10.190	0.145	0.206
	11.495	2.721	37.044	0.809	—
XOM	-3.662	3.685	47.322	0.326	-0.004
	6.447	5.149	197.227	0.885	—
DUK	3.743	1.551	15.216	0.224	0.183
	10.419	3.200	57.442	0.882	—
ENE	11.557	0.870	16.761	0.291	0.326
	16.285	2.812	82.516	0.908	—
AOL	133.194	2.896	131.974	0.571	0.683
	93.718	5.408	688.675	0.906	—
MO	14.643	2.323	83.095	0.579	0.455
	15.132	4.655	340.383	0.899	—
T	6.033	3.369	78.816	0.433	0.132
	4.495	5.808	235.872	0.815	—
PFE	13.650	2.170	76.184	0.683	0.625
	13.944	4.431	375.726	0.953	—
LUV	17.934	0.360	21.802	0.452	0.416
	18.387	2.476	88.850	0.873	—
AMR	5.503	2.071	27.079	0.267	0.369
	7.186	4.388	128.836	0.836	—
DOW	-1.513	2.928	31.871	0.419	0.125
	2.394	5.121	88.271	0.697	—
C	22.445	1.482	76.227	0.772	0.802
	24.244	3.341	314.672	0.951	—
JPM	12.619	1.609	33.315	0.473	0.554
	13.800	3.778	151.941	0.898	—
WMT	11.009	2.490	58.606	0.514	0.210
	15.338	4.057	207.550	0.907	—
HD	21.105	1.387	57.029	0.658	0.533
	22.693	3.206	179.999	0.887	—
GE	10.925	2.557	57.672	0.398	0.328
	12.771	5.057	452.945	0.947	—

Table II
Maximum Likelihood Estimates for Model A

Entries are maximum likelihood estimates of Model A:

$$\psi_t = \omega e^{gt} + \Phi\psi_{t-1}e^{g} + \Gamma\mathbf{Z}_t,$$

where $\psi_t \equiv [\alpha\mu_t, 2\varepsilon_t]^\top$ denotes time t forecasts of the arrival rates of informed and uninformed trades at time $t+1$ and $\mathbf{Z} \equiv [|K|, TT - |K|]^\top$ denotes the realized trade imbalance and number of balanced trades at time t. In the parentheses are standard errors. The last row reports the log likelihood value.

Θ	ASH	XOM	DUK	ENE	AOL	MO	ATT	PFE
δ	0.5511	0.7743	0.5349	0.4816	0.5371	0.3834	0.5951	0.4482
	(0.0142)	(0.0092)	(0.0127)	(0.0136)	(0.0000)	(0.0132)	(0.0111)	(0.0145)
α	0.4092	0.5266	0.4867	0.4481	0.5203	0.4922	0.4908	0.4074
	(0.0103)	(0.0090)	(0.0099)	(0.0098)	(0.0000)	(0.0093)	(0.0087)	(0.0098)
g_1	0.0072	0.0001	0.0471	0.0523	0.0154	0.1445	0.0078	0.1389
	(0.0044)	(0.0043)	(0.0031)	(0.0041)	(0.0000)	(0.0009)	(0.0078)	(0.0014)
g_2	0.0093	0.0027	0.0491	0.0537	0.1593	0.1424	0.0321	0.1388
	(0.0042)	(0.0040)	(0.0030)	(0.0041)	(0.0000)	(0.0007)	(0.0033)	(0.0013)
ω_1	2.1190	2.4286	2.3074	1.9913	3.0877	2.8442	0.8761	2.1160
	(0.0957)	(0.1300)	(0.0956)	(0.0861)	(0.0000)	(0.0688)	(0.1187)	(0.0640)
ω_2	7.8509	8.1612	7.8323	8.8338	10.1759	9.4953	5.5258	12.4808
	(0.5016)	(0.4496)	(0.4637)	(0.5569)	(0.0000)	(0.1034)	(0.4442)	(0.2546)
$\widehat{\Phi}_{11}$	0.5204	0.6117	0.5046	0.5378	0.4863	0.6387	0.5042	0.5081
	(0.0179)	(0.0040)	(0.0156)	(0.0152)	(0.0002)	(0.0033)	(0.0032)	(0.0048)
$\widehat{\Phi}_{12}$	0.0348	0.0413	0.0371	0.0329	0.0666	0.0260	0.0595	0.0314
	(0.0028)	(0.0009)	(0.0025)	(0.0021)	(0.0000)	(0.0006)	(0.0013)	(0.0009)
$\widehat{\Phi}_{21}$	-1.7298	-1.2705	-1.6347	-2.0162	-1.9612	-0.9257	-1.8897	-2.8179
	(0.1279)	(0.0339)	(0.1008)	(0.1351)	(0.0000)	(0.0262)	(0.0425)	(0.0909)
$\widehat{\Phi}_{22}$	1.1219	1.1360	1.1193	1.1417	1.2552	1.0549	1.2227	1.1769
	(0.0123)	(0.0022)	(0.0101)	(0.0116)	(0.0001)	(0.0011)	(0.0028)	(0.0039)
Γ_{11}	0.0768	0.1302	0.0913	0.0719	0.1120	0.1305	0.0926	0.0575
	(0.0033)	(0.0024)	(0.0033)	(0.0028)	(0.0000)	(0.0025)	(0.0017)	(0.0015)
Γ_{12}	0.0720	0.0826	0.0718	0.0646	0.0815	0.0997	0.0877	0.0482
	(0.0028)	(0.0015)	(0.0024)	(0.0024)	(0.0000)	(0.0019)	(0.0016)	(0.0012)
Γ_{21}	0.3022	0.4449	0.3335	0.3431	0.4376	0.3948	0.3671	0.3698
	(0.0067)	(0.0023)	(0.0057)	(0.0052)	(0.0000)	(0.0013)	(0.0013)	(0.0017)
Γ_{22}	0.3316	0.3590	0.3308	0.3574	0.2938	0.4627	0.4253	0.3471
	(0.0035)	(0.0012)	(0.0036)	(0.0029)	(0.0001)	(0.0006)	(0.0007)	(0.0009)
$\mathcal{L}(\times 10^5)$	5.9201	64.0319	9.9586	12.5957	31.3832	98.6538	112.6279	74.8664

Table II (continued)
Maximum Likelihood Estimates for Model A

Entries are maximum likelihood estimates of Model A:

$$\psi_t = \omega e^{gt} + \Phi \psi_{t-1} e^{g} + \Gamma \mathbf{Z}_t,$$

where $\psi_t \equiv [\alpha\mu_t, 2\varepsilon_t]^\top$ denotes time t forecasts of the arrival rates of informed and uninformed trades at time $t+1$ and $\mathbf{Z} \equiv [|K|, TT-|K|]^\top$ denotes the realized trade imbalance and number of balanced trades at time t. In the parentheses are standard errors. The last row reports the log likelihood value.

Θ	LUV	AMR	DOW	C	JPM	WMT	HD	GE
δ	0.2998	0.3827	0.5529	0.4275	0.5375	0.5864	0.5600	0.5008
	(0.0138)	(0.0153)	(0.0129)	(0.0143)	(0.0131)	(0.0106)	(0.0156)	(0.0130)
α	0.4276	0.4707	0.4161	0.4960	0.5191	0.5814	0.3397	0.4342
	(0.0096)	(0.0101)	(0.0096)	(0.0104)	(0.0102)	(0.0090)	(0.0104)	(0.0097)
g_1	0.0682	0.0996	0.0486	0.0809	0.0908	0.0614	0.0759	0.1229
	(0.0035)	(0.0039)	(0.0027)	(0.0042)	(0.0028)	(0.0013)	(0.0018)	(0.0022)
g_2	0.0701	0.1010	0.0452	0.0848	0.0918	0.0714	0.0790	0.1233
	(0.0034)	(0.0038)	(0.0023)	(0.0041)	(0.0027)	(0.0010)	(0.0018)	(0.0020)
ω_1	2.0010	2.7846	2.2215	2.6584	2.9133	2.8769	2.7226	2.0085
	(0.0718)	(0.1719)	(0.1242)	(0.0897)	(0.0957)	(0.0617)	(0.0929)	(0.0734)
ω_2	6.7676	9.4418	10.2262	9.7331	9.4784	6.4879	10.6474	8.9251
	(0.2306)	(0.5416)	(0.4643)	(0.3617)	(0.3080)	(0.0899)	(0.1842)	(0.2661)
$\widetilde{\Phi}_{11}$	0.5514	-0.3745	0.5461	0.5143	0.3432	0.7717	0.4794	0.5210
	(0.0085)	(0.0267)	(0.0071)	(0.0078)	(0.0086)	(0.0029)	(0.0048)	(0.0040)
$\widetilde{\Phi}_{12}$	0.0444	0.2131	0.0366	0.0577	0.0697	0.0210	0.0364	0.0301
	(0.0020)	(0.0069)	(0.0014)	(0.0016)	(0.0022)	(0.0005)	(0.0013)	(0.0008)
$\widetilde{\Phi}_{21}$	-1.4905	-4.5369	-1.7943	-1.7880	-2.1052	-0.4211	-2.0071	-1.9364
	(0.0594)	(0.1841)	(0.0642)	(0.0694)	(0.0708)	(0.0118)	(0.0778)	(0.0589)
$\widetilde{\Phi}_{22}$	1.1461	1.7012	1.1393	1.2133	1.2219	1.0334	1.1371	1.1186
	(0.0068)	(0.0259)	(0.0054)	(0.0070)	(0.0074)	(0.0007)	(0.0033)	(0.0020)
Γ_{11}	0.0960	0.1202	0.0900	0.0630	0.1106	0.1370	0.0801	0.0973
	(0.0026)	(0.0029)	(0.0022)	(0.0016)	(0.0026)	(0.0022)	(0.0025)	(0.0023)
Γ_{12}	0.0863	0.0980	0.0716	0.0721	0.0982	0.1045	0.0727	0.0680
	(0.0022)	(0.0023)	(0.0018)	(0.0017)	(0.0022)	(0.0017)	(0.0023)	(0.0016)
Γ_{21}	0.3294	0.4634	0.3817	0.2637	0.3840	0.2878	0.3282	0.4300
	(0.0033)	(0.0025)	(0.0031)	(0.0022)	(0.0030)	(0.0016)	(0.0018)	(0.0020)
Γ_{22}	0.3478	0.4002	0.3806	0.3358	0.3677	0.3886	0.3248	0.3994
	(0.0018)	(0.0012)	(0.0014)	(0.0013)	(0.0016)	(0.0010)	(0.0010)	(0.0008)
$\mathcal{L}(\times 10^5)$	12.9931	29.3080	38.1600	35.0912	30.5118	53.0571	38.1229	115.8519

Table III
Maximum Likelihood Estimates for Model B

Entries are maximum likelihood estimates of Model B:

$$\ln\psi_t = \omega + \Phi g + (I - \Phi)gt + \Phi \ln\psi_{t-1} + \Gamma \mathbf{M}_t,$$

where $\psi_t \equiv [\alpha\mu_t, 2\varepsilon_t]^\top$ denotes the time t forecasts of the arrival rates of informed and uninformed trades at time $t+1$ and $\mathbf{M}_{it} \equiv \mathbf{Z}_{it}/\psi_{i(t-1)} - 1$ is a martingale difference formulated from $\mathbf{Z} \equiv [|K|, TT - |K|]^\top$. In the parentheses are standard errors. The last row reports the log likelihood value.

Θ	ASH	XOM	DUK	ENE	AOL	MO	ATT	PFE
δ	0.5086	0.7037	0.4727	0.7731	0.3955	0.4553	0.5209	0.8315
	(0.0147)	(0.0107)	(0.0127)	(0.0116)	(0.0635)	(0.0149)	(0.0115)	(0.0096)
α	0.3896	0.6158	0.4880	0.2248	0.2409	0.4919	0.5303	0.1382
	(0.0101)	(0.0087)	(0.0099)	(0.0086)	(0.0267)	(0.0097)	(0.0086)	(0.0070)
g_1	0.0705	0.0831	0.0637	0.1138	0.0943	0.1425	0.0391	0.0821
	(0.0016)	(0.0012)	(0.0025)	(0.0011)	(0.0072)	(0.0006)	(0.0004)	(0.0010)
g_2	0.1020	0.0829	0.0999	0.1651	0.0637	0.1814	0.0457	0.2244
	(0.0016)	(0.0019)	(0.0041)	(0.0009)	(0.0082)	(0.0017)	(0.0006)	(0.0025)
ω_1	1.0086	0.4368	0.1463	0.7891	1.6345	0.6154	0.1450	1.3101
	(0.0278)	(0.0131)	(0.0281)	(0.0263)	(0.0649)	(0.0097)	(0.0134)	(0.0327)
ω_2	0.5553	0.1691	0.1125	0.7238	0.3766	0.2941	0.1297	0.3011
	(0.0136)	(0.0039)	(0.0181)	(0.0184)	(0.0127)	(0.0049)	(0.0071)	(0.0078)
Φ_{11}	0.6718	0.7882	0.5642	0.7637	0.5218	0.6761	0.6031	0.4319
	(0.0106)	(0.0024)	(0.0174)	(0.0072)	(0.0035)	(0.0012)	(0.0024)	(0.0027)
Φ_{12}	-0.1725	0.0471	0.2095	-0.1931	-0.0468	0.0710	0.2406	-0.1359
	(0.0091)	(0.0029)	(0.0160)	(0.0076)	(0.0047)	(0.0013)	(0.0033)	(0.0026)
Φ_{21}	-0.1386	-0.0583	-0.2693	-0.1523	-0.0897	-0.1594	-0.2104	-0.1358
	(0.0061)	(0.0009)	(0.0183)	(0.0094)	(0.0014)	(0.0008)	(0.0020)	(0.0014)
Φ_{22}	0.8862	1.0040	1.1238	0.8079	0.9797	1.0390	1.1190	0.9730
	(0.0041)	(0.0009)	(0.0127)	(0.0038)	(0.0010)	(0.0007)	(0.0020)	(0.0006)
Γ_{11}	0.0784	0.1820	0.0825	0.0227	0.0749	0.0942	0.0812	0.0288
	(0.0030)	(0.0028)	(0.0030)	(0.0015)	(0.0084)	(0.0019)	(0.0013)	(0.0015)
Γ_{12}	0.4527	0.6226	0.3468	0.2495	0.5278	0.7779	0.6126	0.4768
	(0.0091)	(0.0042)	(0.0096)	(0.0122)	(0.0032)	(0.0018)	(0.0025)	(0.0029)
Γ_{21}	0.0451	0.0585	0.0566	0.0254	0.0194	0.0511	0.0451	0.0092
	(0.0015)	(0.0009)	(0.0015)	(0.0010)	(0.0021)	(0.0010)	(0.0008)	(0.0005)
Γ_{22}	0.3325	0.3230	0.2880	0.3503	0.1939	0.4986	0.4308	0.3343
	(0.0032)	(0.0011)	(0.0032)	(0.0026)	(0.0007)	(0.0004)	(0.0007)	(0.0007)
$\mathcal{L}(\times 10^5)$	5.9190	64.0295	9.9590	12.5788	31.3359	98.6070	112.6333	74.7764

Table III (continued)
Maximum Likelihood Estimates for Model B

Entries are maximum likelihood estimates of Model B:

$$\ln\psi_t = \omega + \Phi g + (I-\Phi)gt + \Phi\ln\psi_{t-1} + \Gamma\mathbf{M}_t,$$

where $\psi_t \equiv [\alpha\mu_t, 2\varepsilon_t]^\top$ denotes the time t forecasts of the arrival rates of informed and uninformed trades at time $t+1$ and $\mathbf{M}_{it} \equiv \mathbf{Z}_{it}/\psi_{i(t-1)} - 1$ is a martingale difference formulated from $\mathbf{Z} \equiv [|K|, TT - |K|]^\top$. In the parentheses are standard errors. The last row reports the log likelihood value.

Θ	LUV	AMR	DOW	C	JPM	WMT	HD	GE
δ	0.7215	0.2314	0.3422	0.5545	0.3508	0.3596	0.5635	0.3970
	(0.0156)	(0.0129)	(0.0121)	(0.0140)	(0.0136)	(0.0109)	(0.0165)	(0.0132)
α	0.1618	0.2811	0.4220	0.5225	0.5034	0.5807	0.3002	0.4369
	(0.0079)	(0.0081)	(0.0093)	(0.0074)	(0.0098)	(0.0085)	(0.0101)	(0.0093)
g_1	0.1521	0.0294	0.0208	0.0012	0.1399	0.0707	0.1092	0.1014
	(0.0015)	(0.0015)	(0.0010)	(0.0000)	(0.0015)	(0.0005)	(0.0007)	(0.0006)
g_2	0.1679	0.0130	0.0199	0.1277	0.1829	0.0743	0.1536	0.1379
	(0.0009)	(0.0027)	(0.0014)	(0.0016)	(0.0029)	(0.0015)	(0.0007)	(0.0008)
ω_1	0.7282	0.0826	0.0673	0.3374	0.1494	0.8912	0.4953	1.0596
	(0.0124)	(0.0258)	(0.0312)	(0.0050)	(0.0128)	(0.0091)	(0.0069)	(0.0100)
ω_2	0.3802	0.0689	0.0794	0.2884	0.0966	0.6926	0.3240	0.5735
	(0.0048)	(0.0139)	(0.0158)	(0.0042)	(0.0077)	(0.0074)	(0.0036)	(0.0064)
Φ_{11}	0.9184	0.3790	0.5753	1.0698	0.7296	0.6643	0.8860	0.7703
	(0.0029)	(0.0055)	(0.0062)	(0.0029)	(0.0039)	(0.0023)	(0.0017)	(0.0017)
Φ_{12}	−0.2519	0.3072	0.2339	−0.1306	0.1159	0.0425	−0.0654	−0.0747
	(0.0038)	(0.0058)	(0.0085)	(0.0032)	(0.0048)	(0.0019)	(0.0017)	(0.0015)
Φ_{21}	−0.0186	−0.3331	−0.2120	0.0576	−0.1614	−0.2391	−0.0539	−0.1131
	(0.0013)	(0.0056)	(0.0049)	(0.0025)	(0.0038)	(0.0023)	(0.0010)	(0.0014)
Φ_{22}	0.8697	1.1600	1.1084	0.8905	1.0673	1.0196	0.9468	0.9533
	(0.0016)	(0.0042)	(0.0050)	(0.0028)	(0.0033)	(0.0014)	(0.0009)	(0.0008)
Γ_{11}	0.0525	0.0685	0.0958	0.0758	0.0742	0.1135	0.0556	0.0263
	(0.0026)	(0.0021)	(0.0023)	(0.0012)	(0.0017)	(0.0017)	(0.0019)	(0.0006)
Γ_{12}	0.5637	0.4782	0.5315	0.4578	0.5558	0.4699	0.4362	0.5197
	(0.0053)	(0.0033)	(0.0052)	(0.0024)	(0.0057)	(0.0021)	(0.0021)	(0.0054)
Γ_{21}	0.0253	0.0419	0.0517	0.0623	0.0475	0.0845	0.0328	0.0160
	(0.0013)	(0.0013)	(0.0012)	(0.0010)	(0.0010)	(0.0013)	(0.0011)	(0.0004)
Γ_{22}	0.3235	0.3543	0.3676	0.4123	0.3730	0.4761	0.3145	0.3737
	(0.0012)	(0.0010)	(0.0013)	(0.0009)	(0.0011)	(0.0009)	(0.0006)	(0.0008)
$\mathcal{L}(\times 10^5)$	12.9704	29.3073	38.1629	35.0781	30.5117	53.0567	38.1154	115.8491

Table IV
Stationarity of the Dynamic Processes in Models A and B

Entries are the eigenvalues of the autocorrelation matrix $\widehat{\Phi} = \Phi + \Gamma$ in Model A and Φ in Model B. The eigenvalues should be less than one for the processes to be stationary.

Ticker	Model A		Model B	
ASH	0.6473	0.9950	0.591	0.802
XOM	0.7464	1.0013	0.696	0.613
DUK	0.6281	0.9958	0.513	0.711
ENE	0.6821	0.9974	0.735	0.400
AOL	0.7401	1.0014	0.967	0.333
MO	0.7080	0.9855	0.688	0.595
T	0.7347	0.9921	0.799	0.696
PFE	0.6901	0.9950	0.850	0.732
LUV	0.6996	0.9979	0.967	0.990
AMR	0.3310	0.9957	0.992	0.959
DOW	0.6936	0.9918	0.989	1.004
C	0.7259	1.0018	0.987	1.005
JPM	0.5672	0.9979	0.821	0.971
WMT	0.8115	0.9936	0.986	0.976
HD	0.6209	0.9956	0.997	0.988
GE	0.6437	0.9959	0.983	0.991

Table V
Correlations Between Arrival Rate Forecasts and Price Volatility

Entries are the correlations between the arrival rates (detrended) of informed and uninformed trades ($\widetilde{\psi}$) and absolute returns on daily open-close ($|\ln(O/C)|$) and high-low ($\ln(H/L)$).

Ticker	Model A				Model B			
	$\|\ln(O/C)\|$		$\ln(H/L)$		$\|\ln(O/C)\|$		$\ln(H/L)$	
	$\alpha\widetilde{\mu}$	$2\widetilde{\varepsilon}$	$\alpha\widetilde{\mu}$	$2\widetilde{\varepsilon}$	$\alpha\widetilde{\mu}$	$2\widetilde{\varepsilon}$	$\alpha\widetilde{\mu}$	$2\widetilde{\varepsilon}$
ASH	0.0619	0.0225	0.1285	0.0535	0.1244	0.0826	0.2142	0.1586
XOM	0.1604	0.1471	0.2537	0.2329	0.0652	0.1268	0.1423	0.2757
DUK	0.1526	0.1279	0.2439	0.1961	0.1444	0.1284	0.2707	0.2507
ENE	0.0537	0.0355	0.1867	0.1614	0.0377	0.1123	-0.0038	0.2507
AOL	0.0712	0.0811	0.1657	0.1725	0.0649	0.0738	0.1568	0.1477
MO	0.1407	0.1544	0.2577	0.2795	0.1249	0.1137	0.1985	0.2181
T	0.1324	0.1345	0.1970	0.2513	0.1176	0.1252	0.2383	0.2720
PFE	0.1831	0.1998	0.3408	0.3479	0.0433	0.0799	0.0987	0.1893
LUV	0.1133	0.1081	0.2424	0.2220	0.0883	0.0662	0.1608	0.1847
AMR	0.1549	0.1629	0.3067	0.3161	0.1193	0.1596	0.2441	0.3109
DOW	0.1472	0.1678	0.2668	0.2865	0.1238	0.1556	0.2228	0.2560
C	0.1314	0.1255	0.2384	0.2273	0.0943	0.1240	0.1563	0.2336
JPM	0.1490	0.1436	0.2369	0.2268	0.0966	0.0727	0.1376	0.0886
WMT	0.1575	0.1569	0.2712	0.2768	0.1446	0.1594	0.2549	0.2819
HD	-0.0003	-0.0368	0.0556	-0.0061	0.0565	0.0109	0.1228	0.0592
GE	0.1849	0.2237	0.3647	0.4132	0.0334	0.2189	0.0989	0.4106

Table VI
Correlations Between Trade Composition Forecasts and Total Trades, Trade Imbalances, and Price Volatilities

Entries are the correlation between the forecasted fraction of informed trades β:

$$\beta_t \equiv \frac{\alpha\mu_t}{\alpha\mu_t + 2\varepsilon_t},$$

and the total number of trades (TT), the ratio of trade imbalance to total trades ($|K|/TT$), and absolute returns on daily open-close ($|\ln O/C|$) and high-low ($\ln H/L$).

Ticker	Model A				Model B			
	TT	$\frac{\|K\|}{TT}$	$\|\ln\frac{O}{C}\|$	$\ln\frac{H}{L}$	TT	$\frac{\|K\|}{TT}$	$\|\ln\frac{O}{C}\|$	$\ln\frac{H}{L}$
ASH	-0.6555	0.3063	0.0226	0.0685	-0.6225	0.2560	0.0270	0.0535
XOM	-0.3327	0.5292	-0.0159	-0.0253	-0.3657	0.4794	-0.0389	-0.0763
DUK	-0.7017	0.2982	-0.0663	-0.0823	-0.7541	0.3002	-0.0607	-0.0628
ENE	-0.7511	0.3611	-0.0297	-0.1325	-0.7034	0.2938	-0.0599	-0.2027
AOL	0.0527	-0.0410	-0.0159	0.0608	-0.5577	-0.0401	0.0073	0.0136
MO	-0.1555	0.2476	-0.0414	-0.0541	-0.3559	0.0929	-0.0723	-0.1430
T	0.1787	0.3190	0.0349	0.0619	-0.2218	0.3811	-0.0375	-0.0718
PFE	-0.3987	-0.0006	-0.1537	-0.2317	-0.5366	0.0125	-0.1792	-0.2736
LUV	-0.7212	0.2766	-0.0668	-0.1420	-0.4089	0.1771	0.0230	-0.0084
AMR	-0.1129	0.0057	-0.0664	-0.1039	-0.4014	0.0326	-0.1382	-0.2597
DOW	0.0663	0.1936	-0.0255	-0.0127	-0.2244	0.2434	-0.0405	-0.0246
C	-0.6319	0.2207	-0.0347	-0.0726	-0.6620	0.2104	-0.0634	-0.1293
JPM	-0.4126	0.1484	-0.0134	-0.0145	-0.5045	0.1537	-0.0277	-0.0430
WMT	-0.5137	0.4569	-0.0172	-0.0405	-0.6896	0.3921	-0.0621	-0.1172
HD	-0.6102	0.2517	0.1141	0.1449	-0.6456	0.2604	0.1124	0.1473
GE	-0.2605	0.2532	-0.0597	-0.0693	-0.6094	0.1910	-0.1821	-0.2971

Table VII
Properties of the Half Life of Price Impacts

Entries report the mean half life ($\tau_{1/2}$) of price impacts and its correlations with the number of trades (TT), the ratio of trade imbalance to total trades ($|K|/TT$) and absolute returns on daily open-close ($|\ln O/C|$) and high-low ($\ln H/L$). The price impact of N consecutive buy orders is defined as

$$\gamma_t^N = \frac{\delta\,\Pr_t^{N-1}(g)\,(\varepsilon+\mu) - (1-\delta)\,\Pr_t^{N-1}(b)\,\varepsilon}{\Pr_t^{N-1}(g)\,\mu+\varepsilon},$$

where $\Pr_t^{N-1}(g)$ and $\Pr_t^{N-1}(b)$ are, respectively, the probabilities of the good news and bad news at date t, conditional on $N-1$ consecutive buys, which is updated via Bayes rule. The half life $\tau_{1/2}$ is then defined as the minimum number of consecutive buys N such that $\gamma_t^N > \delta/2$.

Ticker	Model A					Model B				
	Mean	TT	$\frac{\lvert K\rvert}{TT}$	$\lvert\ln\frac{O}{C}\rvert$	$\ln\frac{H}{L}$	Mean	TT	$\frac{\lvert K\rvert}{TT}$	$\lvert\ln\frac{O}{C}\rvert$	$\ln\frac{H}{L}$
ASH	3.4242	0.6477	-0.3010	-0.0162	-0.0437	3.3910	0.6720	-0.2472	-0.0020	0.0060
XOM	5.9628	0.3770	-0.4816	0.0139	0.0232	5.1263	0.4831	-0.4263	0.0574	0.1034
DUK	3.4436	0.6562	-0.2650	0.0625	0.0905	3.2431	0.7598	-0.2376	0.0498	0.0745
ENE	3.5372	0.7787	-0.3524	0.0198	0.1197	6.9759	0.7473	-0.3038	0.0560	0.2031
AOL	4.2668	-0.1496	0.0529	0.0129	-0.0779	9.0777	0.7422	-0.0056	0.0666	0.0975
MO	3.7621	0.1476	-0.1987	0.0470	0.0578	4.4108	0.3434	-0.0660	0.0647	0.1302
T	4.8598	-0.1331	-0.2742	-0.0161	-0.0311	4.2812	0.2519	-0.2713	0.0308	0.0771
PFE	4.1724	0.4464	-0.0437	0.1308	0.1909	11.8016	0.7905	-0.0791	0.1979	0.2934
LUV	2.4784	0.7001	-0.2560	0.0503	0.1075	6.0131	0.5481	-0.2741	0.0203	0.0823
AMR	3.0041	0.0147	0.0001	0.0060	-0.0066	6.3631	0.4722	-0.0423	0.1489	0.2724
DOW	4.6692	-0.0154	-0.1406	0.0318	0.0224	3.8990	0.1728	-0.1847	0.0337	0.0197
C	3.1464	0.6713	-0.1675	0.0574	0.1041	4.2188	0.9017	-0.1309	0.1144	0.2111
JPM	3.8781	0.3217	-0.1282	0.0023	-0.0063	3.2632	0.4919	-0.1037	0.0384	0.0694
WMT	3.7766	0.5143	-0.4067	0.0292	0.0512	2.9543	0.7133	-0.3772	0.0769	0.1361
HD	3.8954	0.5473	-0.2489	-0.0909	-0.1262	3.9940	0.6434	-0.2800	-0.0882	-0.1053
GE	5.4387	0.2552	-0.1917	0.0451	0.0598	4.9054	0.7272	-0.2074	0.1716	0.2915

Table VIII
Residual Analysis

Entries report the sample estimates of mean (Mean), standard deviation (St Dev), first order autocorrelation (Auto), and cross-correlation coefficient (ρ) of the percentage forecasting residuals of the absolute trade imbalance $|K|$ and the balanced trade $TT - |K|$:

$$e_{it} = \frac{\mathbf{Z}_{it} - \mathbb{E}_{t-1}[\mathbf{Z}_{it}]}{\mathbb{E}_{t-1}[\mathbf{Z}_{it}]}, \quad i = 1,2, \quad \text{with} \quad \mathbf{Z}_t = [|K_t|, TT_t - |K_t|].$$

The expected value on total trades is: $\mathbb{E}_{t-1}[TT_t] = \alpha\mu_{t-1} + 2\varepsilon_{t-1}$. The expected value of $|K|$ is determined by simulation. For each stock, the first row reports the properties of the first element and the second row reports that of the second element of the residual **e**. The arrival rates are forecasted based on parameters reported in Tables II and III.

Ticker	Model A				Model B			
	Mean	St Dev	Auto	ρ	Mean	St Dev	Auto	ρ
ASH	0.373	1.159	-0.000	-0.118	0.455	1.253	0.036	-0.113
	-0.076	0.371	0.028	—	-0.081	0.369	0.010	—
XOM	0.318	1.055	0.189	-0.400	0.130	0.898	0.138	-0.394
	-0.031	0.229	0.074	—	-0.031	0.229	0.073	—
DUK	0.353	1.165	0.079	-0.296	0.341	1.152	0.080	-0.296
	-0.065	0.310	-0.010	—	-0.064	0.310	0.014	—
ENE	0.368	1.132	0.043	-0.204	1.986	2.651	0.116	-0.166
	-0.063	0.331	0.028	—	-0.143	0.306	0.046	—
AOL	0.303	0.971	-0.008	0.278	3.775	3.557	0.067	0.161
	-0.015	0.340	0.108	—	-0.124	0.321	0.253	—
MO	0.165	0.970	0.134	-0.039	0.247	1.076	0.233	-0.105
	-0.029	0.263	0.043	—	-0.094	0.251	0.020	—
T	0.342	1.104	0.274	-0.214	0.216	0.984	0.272	-0.211
	-0.036	0.228	0.090	—	-0.034	0.227	0.055	—
PFE	0.492	1.200	0.177	-0.046	4.519	4.809	0.225	-0.107
	-0.038	0.268	0.058	—	-0.093	0.261	0.100	—
LUV	0.130	0.948	0.078	-0.053	2.491	3.135	0.032	-0.041
	-0.076	0.428	0.069	—	-0.184	0.381	0.104	—
AMR	-0.022	0.879	0.026	0.053	3.390	3.966	0.067	0.040
	-0.030	0.339	0.018	—	-0.123	0.307	0.022	—
DOW	0.299	1.092	0.138	-0.113	0.279	1.063	0.120	-0.113
	-0.031	0.251	0.051	—	-0.027	0.251	0.048	—
C	0.260	0.973	0.098	-0.143	0.752	1.751	0.360	-0.103
	-0.053	0.332	0.074	—	-0.137	0.311	0.016	—
JPM	0.179	0.947	0.076	-0.115	0.229	0.988	0.116	-0.102
	-0.036	0.277	0.065	—	-0.030	0.280	0.028	—
WMT	0.191	0.915	0.170	-0.221	0.118	0.881	0.243	-0.203
	-0.043	0.292	0.067	—	-0.050	0.288	-0.010	—
HD	0.717	1.419	0.262	-0.065	0.844	1.545	0.306	-0.019
	-0.064	0.378	0.107	—	-0.072	0.373	0.094	—
GE	0.363	1.148	0.105	-0.156	0.284	1.086	0.236	-0.145
	-0.021	0.210	0.068	—	-0.024	0.210	0.063	—