

Time-Varying Arrival Rates of Informed and Uninformed Trades

DAVID EASLEY

Cornell University

ROBERT F. ENGLE

New York University

MAUREEN O'HARA

Cornell University

LIUREN WU

Baruch College, CUNY

ABSTRACT

We propose a dynamic econometric microstructure model of trading, and we investigate how the dynamics of trades and trade composition interact with the evolution of market liquidity, market depth, and order flow. We estimate a bivariate generalized autoregressive intensity process for the arrival rates of informed and uninformed trades for 16 actively traded stocks over 15 years of transaction data. Our results show that both informed and uninformed trades are highly persistent, but that the uninformed arrival forecasts respond negatively to past forecasts of the informed intensity. Our estimation generates daily conditional arrival rates of informed and uninformed trades, which we use to construct forecasts of the probability of information-based trade (PIN). These forecasts are used in turn to forecast market liquidity as measured by bid-ask spreads and the price impact of orders. We observe that PINs vary across assets and over time, and most importantly that they are correlated across assets. Our analysis shows that one principal component explains much of the daily variation in PINs and that this systemic liquidity factor may be

We thank Mark Ready, Schmuël Baruch, and seminar participants at New York University and the 2002 AFA meetings for helpful comments. Address correspondence to Robert F. Engle, Stern School of Business, New York University, 44 West 4th Street, Suite 9-62, NY 10012-1126, or e-mail: rengle@stern.nyu.edu.

doi: 10.1093/jfinec/nbn003

Advance Access publication February 26, 2008

© The Author 2008. Published by Oxford University Press. All rights reserved. The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oupjournals.org

important for asset pricing. We also find that PINs tend to rise before earnings announcement days and decline afterwards. (*JEL*: C51, C53, G10, G12, G14)

KEYWORDS: Arrival rates, informed trades, uninformed trades, autoregressive process, market depth, liquidity

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 econometric microstructure model of trading, and we investigate how the dynamics of trades and trade composition interact with the evolution of market liquidity, market depth, and order flows.

There are many reasons why understanding market liquidity and depth 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. From a more academic perspective, understanding the evolution of liquidity and its interaction with information flow provides insight into the price formation process as well as into more fundamental asset pricing issues as formulated by Easley, Hvidkjaer, and O'Hara (2002), O'Hara (2003), and Acharya and Pedersen (2005). We argue in this paper that understanding market parameters such as liquidity requires understanding a more basic market variable, the order arrival process.

Our dynamic microstructure model follows Easley and O'Hara (1992) by letting the arrival of informed and uninformed traders dictate the order flow and the price formulation. Different from them, however, our model explicitly allows the arrival rates of informed and uninformed trades to be time-varying and predictable. We propose a forecasting relation for the bivariate arrival rate process which is analogous to the GARCH (Bollerslev 1986) specifications on volatilities. We estimate the parameters that govern the forecasting dynamics using a maximum likelihood method. The likelihood function is determined by the probability of having a given set of buy and sell orders each day, as a function of the arrival rate forecasts. Thus, our model specification allows us to forecast the arrival rates of informed and uninformed orders, and then to forecast the resultant measures of liquidity based on these order arrival processes.

Our modeling approach is a blending of model-based microstructure (see, for example, Easley and O'Hara 1992) with the literature analyzing the econometric determinants of the joint dynamics between trades and prices. Examples of the latter include Hasbrouck (1991), Dufour and Engle (2000), Engle (2000), Engle and Russell (1998), Manganello (2000), Engle and Lange (2001), Chordia, Roll, and Subrahmanyam (2000, 2001a, 2001b, 2002, 2005), Chordia and Subrahmanyam (2004), Hasbrouck and Seppi (2001), and Korajczyk and Sadka (2006). In common with

this econometric literature, our model generates direct forecasts on market liquidity and depth. Different from them, however, we do not rely on exogenous dynamic specifications of trade and price linkages. Instead, our inclusion of a GARCH-style specification into a microstructure model allows us to show why particular components of order imbalance matter, thus providing an econometric structure for investigating order flow information and its resultant effects on market liquidity and depth.

To illustrate the potential of our methodology, we estimate the dynamic model for 16 actively traded stocks using daily numbers of buys and sells over 15 years from January 1983 to December 1998. We find that both the informed and uninformed order flows are highly persistent. More trade today generates more trade tomorrow by both kinds of traders. However, the uninformed arrival forecasts respond negatively to past forecasts on the informed arrival. Informed trade arrival responds more to past order imbalance than it does to overall trade volumes, with the impulse responses to both variables positive and the decay exponential. Uninformed trade responds more to past uninformed trade than it does to past informed trade. The impulse responses suggest a slower decay to the uninformed trading behavior.

We use the estimated model to generate forecasts on the arrival rates of informed and uninformed traders. Based on the arrival rate forecasts, we compute forecasts of the probability of information-based trading (PIN), which has been shown to have explanatory power for both spreads and returns. We also use the arrival rate forecast to predict trading-cost relevant measures such as bid-ask spreads and price impacts. For example, our microstructure model directly links the arrival rates of informed and uninformed traders to the bid-ask spread, and so our arrival rate forecasts can be used to predict bid-ask spreads. We illustrate the power of this approach by predicting opening spreads for a sample of stocks, and we find significantly positive results for most stocks. Similarly, given the arrival rate forecasts, we can use Bayesian updating to calculate the price impact of any given sequence of order flows. As an illustration, we define a measure of market depth we term the half-life. This measure is defined as the number of consecutive buys needed for the price impact to exceed half of the exogenously specified maximum impact. The half-life estimates provide a compact forecast of the market depth based on the forecasts of arrival rates of informed and uninformed traders.

We also illustrate the value of our dynamic model of trading by showing how our estimated PINs vary around earnings announcement days. One might expect PINs to be high before earnings announcements, and low afterwards as earnings announcements turn private information about earnings into public information. In a recent working paper, Benos and Jochec (2007) ask whether constant PINs estimated from the static model over time periods of at least 28 trading days before and after earnings announcement have this property. They find that their PIN estimates do not have the expected property. Our belief is that this occurs because the variation in trade based on private information occurs in short periods before and after announcements and using long periods to estimate PINs obscures this effect. Using our dynamic model, we find significant variation in PIN, in the predicted direction, in the week or so before and after earnings announcement

days. This result suggests that with our dynamic specification PIN can be used in event studies.

We believe that our results will have an impact in three areas of finance. First, institutional investors need to predict trading costs in order to evaluate the efficiency of alternative trading strategies. In order to do this, it is necessary to predict the price impact of hypothetical trades. Our approach allows us to do a better job of making these predictions than standard microstructure models. We provide an illustrative example in Section 4. Second, the liquidity of assets is important for risk management as one of the risks associated with an asset position is the cost of reversing the position. We can predict the PIN, which in turn allows us to forecast liquidity. Third, our more sophisticated model of PIN shows that PINs are both autocorrelated and cross-correlated. Since PIN can be viewed as a simple measure of liquidity, our results show that liquidity covaries across assets. Acharya and Pedersen (2005) argue that liquidity risk matters for asset pricing and our PIN analysis shows that there is a systemic liquidity factor. Further, our new PINs should allow us to improve on the asset pricing results of Easley, Hvidkjaer, and O'Hara (2002).

The paper is organized as follows. We begin in Section 1 by setting out our dynamic microstructure models. Section 2 describes the data set and our estimation procedure. Section 3 provides our estimation results on the order arrival processes, and we examine the impulse response functions to shocks to trade imbalances and overall volume levels. Section 4 investigates the application of the arrival rate forecasts to the prediction of bid-ask spreads and price impacts. This section also illustrates how to use our dynamic model of PINs in an event study. Section 5 provides some diagnostic analysis of the forecasting results. Section 6 concludes.

1 MODEL FORMULATION

In this section, we propose a dynamic microstructure model of trading. We use this model as a vehicle to investigate how the dynamics of trades and trade composition interact with the evolution of market liquidity and depth. From a practical perspective, portfolio managers observe the order flow of buys and sells on an asset, but not information on what type of player is behind each order and why that player sends a particular order. The idea of building the dynamic microstructure model is to provide a theoretical base according to which portfolio managers can infer the unobservable arrival rates of different types of players from the publicly observable streams of buys and sells. From an academic perspective, the microstructure framework enables us to separate information risk and liquidity risk, and their different impacts on asset pricing.

To build our dynamic model, we use the model of Easley and O'Hara (1992) as our benchmark, but allow the arrival rates of different types of trades to follow autoregressive processes. Every day agents update their parameter estimates based on past information before embarking on their trading day. We can use the microstructure model in a conditional form to construct the likelihood function of the observed order flows. By maximizing the likelihood function, we identify

the parameters that govern the dynamic processes of the arrival rates. Using the estimated model, we can generate forecasts on the arrival rates, information flow, market liquidity, and depth.

1.1 The Static Model Benchmark

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 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).

Under this model, the probability of observing B number of buys and S number of sells at a given date t is given by

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

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.

1.2 Time-Varying Arrival Rates of Trades

The benchmark model assumes constant arrival rates for both informed and uninformed traders. In reality, agents continually gain information about the trading environment and consequently update their estimates of these arrival rates. To capture this effect econometrically, we specify how the arrival rates evolve and what the key information sources are about the arrival rates. With the dynamics specification, the arrival rates in Equation (1) become conditional arrival rate forecasts, and the probabilities of buys and sells vary over time with the conditional arrival rate forecasts.

1.2.1 The information content of trades. According to the benchmark microstructure model, data on daily numbers of buys and sells contain important information about the underlying arrival rates of informed and uninformed traders.

Let $TT = S + B$ denote the total number of trades per day. The expected value of the total trades, $\mathbb{E}[TT]$, 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 on rather complicated forms (see Katti 1960), but the first-order term of this expectation relates directly to the arrival of the informed trades: $\mathbb{E}[|K|] \doteq \alpha\mu$.

These relations provide the key information sources that agents would use to update their arrival rate estimates. In this paper, we model the arrival rate dynamics with a forecasting specification that uses past values of balanced and imbalanced trade as well as past arrival forecasts to forecast informed and uninformed arrival rates. It seems reasonable to allow arrival rates to depend on these variables as traders can observe them and can thus condition their trading choices on this data.

1.2.2 A generalized autoregressive specification on arrival rates of trades.

The arrival rate of informed trades is $\alpha\mu$ and the arrival rate of the uninformed trades is 2ε . We use $\psi = [\alpha\mu, 2\varepsilon]^\top$ to denote the vector of the two arrival rates. To remove any deterministic trend in arrival rates, we model the detrended arrival rates $\tilde{\psi}_{it} = \psi_{it}e^{-g_i t}$, $i = 1, 2$, as a vector stationary process, where the vector $g \equiv [g_1, g_2]^\top$ captures the growth rates of the two intensities.

In order to allow our arrival rate forecasts to depend on past observables, we specify that the detrended arrival rate forecasts follow bivariate vector autoregressive process with predetermined forcing variables,

$$\tilde{\psi}_t = \omega + \sum_{k=1}^p \Phi_k \tilde{\psi}_{t-k} + \sum_{j=0}^{q-1} \Gamma_j \tilde{\mathbf{Z}}_{t-j}, \quad (2)$$

where $\tilde{\psi}_t$ denotes the detrended time- t forecast of the arrival rate vector at time $t + 1$, $\mathbf{Z}_t \equiv [|K_t|, TT_t - |K_t|]^\top$ denotes the time- t observed absolute trade imbalance and balanced trades, and $\tilde{\mathbf{Z}}_{it} = \mathbf{Z}_{it}e^{-g_i t}$, $i = 1, 2$, denotes the detrended trade quantities. This equation is directly analogous to a GARCH equation (Bollerslev 1986), where unobservable quantities (arrival rates) are modeled as a function of observables (imbalanced and balanced trades). In principle, as in GARCH-type specifications, we can incorporate any predetermined observables into the forecasting equation as long as they are informative about the informed and uninformed trade arrivals.

To compute multistep forecasts of the arrival rates, it is necessary to forecast future values of $\tilde{\mathbf{Z}}$ based on the model. As a first-order approximation, $\mathbb{E}_{t-1}[\tilde{\mathbf{Z}}_t] \doteq \tilde{\psi}_{t-1}$. Then, as in GARCH models, the above forecasting relation can be rewritten as an $ARMA(\max[p, q], q)$ process:

$$\tilde{\psi}_t \doteq \omega + \sum_{k=1}^{\max[p, q]} \Phi_k^* \tilde{\psi}_{t-k} + \sum_{j=0}^q \Gamma_j \xi_{t-j}, \quad (3)$$

where

$$\Phi_k^* = \begin{cases} \Phi_k + \Gamma_{k-1} & \text{if } k \leq q \\ \Phi_k & \text{if } k > q \end{cases}$$

and $\xi_t \equiv \tilde{\mathbf{Z}}_t - \mathbb{E}_{t-1}[\tilde{\mathbf{Z}}_t] \doteq \tilde{\mathbf{Z}}_t - \tilde{\psi}_{t-1}$ denotes the forecasting error. The stationarity of the process requires that the eigenvalues of Φ_k^* be less than one.

For model estimation, we set $p = q = 1$. Adding back the time trend, we can rewrite the forecasting relation as

$$\psi_t = \omega \odot e^{g_t} + \Phi[\psi_{t-1} \odot e^{g_t}] + \Gamma \mathbf{Z}_t, \quad (4)$$

where \odot is the Hadamard product.

Equation (4) forecasts the product of the parameter α and the arrival rate of informed traders μ . However, the likelihood function needs separate inputs for the two quantities. To separate them, we assume that α , the probability of an information event, is constant over time. In reality, informed trades could vary because of variations in either the arrival rate of informed traders μ or the probability of an information event α , or both. We find it more plausible that the arrival rate of informed traders is time varying than that the probability of an information event is time varying. Some information events are more important than others. We use the time-varying arrival rate of informed traders to capture the variation in the importance of the information events. More important information events attract more informed traders. Nevertheless, it is possible that the probability of having an information event also follows a stochastic process that we miss-identify as variation in informed traders with this assumption.

1.3 Maximum Likelihood Estimation

With daily observations on the number of buys and sells, we use a maximum likelihood method to estimate the parameters that govern the dynamics of the arrival rates of informed and uninformed trades $[\omega, g, \Phi, \Gamma]$, the probability of an information event α , and the probability of bad news δ . First, given initial guesses on the model parameters, we use Equation (4) to forecast the informed and uninformed trade arrival rates at each time t based on information at time $(t-1)$ to obtain $[\alpha\mu_{t-1}, 2\epsilon_{t-1}]$. Second, conditional on the time- $(t-1)$ forecasts of the time- t arrival rates, we compute the time- $(t-1)$ conditional probability of having B_t buys and S_t sells at time t according to the benchmark microstructure

model,

$$\begin{aligned} \Pr[\mathbf{y}_t = (B_t, S_t) | \mathcal{F}_{t-1}] &= \alpha(1 - \delta)e^{-(\mu_{t-1} + 2\varepsilon_{t-1})} \frac{(\mu_{t-1} + \varepsilon_{t-1})^{B_t} (\varepsilon_{t-1})^{S_t}}{B_t! S_t!} \\ &\quad + \alpha\delta e^{-(\mu_{t-1} + 2\varepsilon_{t-1})} \frac{(\mu_{t-1} + \varepsilon_{t-1})^{S_t} (\varepsilon_{t-1})^{B_t}}{B_t! S_t!} \\ &\quad + (1 - \alpha)e^{-2\varepsilon_{t-1}} \frac{(\varepsilon_{t-1})^{B_t + S_t}}{B_t! S_t!}, \end{aligned} \quad (5)$$

where \mathcal{F}_{t-1} denotes the time- $(t - 1)$ filtration. Equation (5) represents a direct extension of Equation (1), where the constant arrival rates of informed and uninformed traders are replaced by their conditional forecasts.

We construct the aggregate log likelihood function on the time series of buys and sells as a summation of the logarithm of the daily conditional probabilities given in (5):

$$\mathcal{L}(\{\mathbf{y}_t\}_{t=1}^T | \Theta) = \sum_{t=1}^T \ln \Pr[\mathbf{y}_t = (B_t, S_t) | \mathcal{F}_{t-1}], \quad (6)$$

where T denotes the number of daily observations and Θ denotes the vector of model parameters, $\Theta \equiv [\delta, \alpha, g, \omega, \Phi, \Gamma]$. We obtain the parameter estimates by maximizing this aggregate likelihood function on the number of buys and sells.

Although the estimation procedure is straightforward, we often encounter numerical problems when performing the estimation in practice. The three components of the conditional probability in Equation (5) all have the factorials of buys and sells in the denominator and have the arrival rates raised to the power of buys and sells in the numerator. As the number of buys and sells become very large numbers for some heavily traded stocks, the computation generates overflow errors for both the numerator and the denominator. Furthermore, the exponential operation on the negative of the arrival rates can also generate underflow errors when the arrival rates are large.

To circumvent the numerical difficulty, we factor out a common term from the three components of the conditional probability, $e^{-2\varepsilon_{t-1}} (\mu_{t-1} + \varepsilon_{t-1})^{B_t + S_t} / (B_t! S_t!)$, and rewrite the log likelihood function as,

$$\begin{aligned} \mathcal{L}(\{\mathbf{y}_t\}_{t=1}^T | \Theta) &= \sum_{t=1}^T [-2\varepsilon_{t-1} + (B_t + S_t) \ln(\mu_{t-1} + \varepsilon_{t-1})] \\ &\quad + \sum_{t=1}^T \ln [\alpha(1 - \delta)e^{-\mu_{t-1}} x_{t-1}^{S_t} + \alpha\delta e^{-\mu_{t-1}} x_{t-1}^{B_t} + (1 - \alpha)x_{t-1}^{B_t + S_t}] \\ &\quad - \ln[B_t! S_t!], \end{aligned} \quad (7)$$

with $x_t \equiv \varepsilon_t / (\mu_t + \varepsilon_t) \in [0, 1]$. For model estimation, we also drop the last term $-\ln[B_t! S_t!]$ as it does not vary with the choice of model parameters.

Our model formulation combines the strength of GARCH-type specifications in forecasting arrival rate dynamics with a microstructure setting to generate a likelihood function that is tightly linked to the interactions between informed and uninformed traders. The GARCH specification in Equation (4) makes a static microstructure model dynamic and enables a highly stylized microstructure story to capture observed order flow behaviors. On the other hand, the microstructure backdrop provides guidance on the forecasting dynamics specifications and informative observable choices. It also generates structural interpretations on the estimated model parameters.

2 DATA AND ESTIMATION

We select 16 actively traded stocks to illustrate our approach to estimating the arrival rates dynamics and forecasting trading costs.¹ These stocks are 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 choose representative stocks from a variety of industries that 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 over 15 years 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 244 inactive days, then JP Morgan Chase (244 days), Ashland (65 days), Duke Energy (61 days), Wal Mart (19 days), Exxon Mobil (18 days), Southwest Air (7 days), Pfizer (4 days), ATT (4 days), and Philip Morris (3 days). Furthermore, the data for AOL Time Warner, CitiGroup, and Home Depot start late. The starting dates are, respectively, September 16, 1996; October 29, 1986; and April 19, 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. The 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, we move back to the prior

¹Fifteen of these stocks were randomly selected from the most active stocks on the NYSE. Ashland was included for comparison with the results of Easley, Kiefer, and O'Hara (1997a).

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

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

- *Trades are increasing.* The daily number of balanced trades $TT - |K|$ 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.
- *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 $TT - |K|$ than for the trade imbalance $|K|$, implying that the number of balanced trades dominates the total trades.
- *Trades are highly persistent.* 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.
- *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.

The above observations suggest a level of complexity to the order arrival process that is not well captured by static models. The observations also suggest that informed and uninformed trade behaviors exhibit complex dynamic interactions, which are the key motivations for our dynamic specifications of the arrival rates. The observation that balanced and imbalanced trades show both serial and cross-sectional dependence indicates that the arrival rates of informed and uninformed trades are not constant over time, but instead follow some correlated, autoregressive dynamics. The observation that the trades are increasing over time prompts us to also incorporate a deterministic time trend in the arrival rate dynamics specification.

Using the time series of balanced and imbalanced trades on each of the 16 stocks, we maximize the log likelihood defined in Equation (7) to estimate the parameters that govern the dynamics of the arrival rates of informed and uninformed trades. These estimated parameters indicate how the two arrival rates interact with each other and how they move over time. From the estimated dynamics and observations on order flows, we then construct arrival rate forecasts, which in turn predict market liquidity, depth, and potential trading cost in each stock.

Table 1 Summary statistics of trading activities.

Ticker	g (%)	a	SD	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	—

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:

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

The third column (a) reports the regression intercept estimate. The fourth column (SD) 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.

Table 2 Maximum likelihood estimates for model parameters.

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

(Continued overleaf)

3 THE ARRIVAL RATE DYNAMICS

Table 2 reports the parameter estimates and the maximized log likelihood values for each stock. Our focus here is on the dynamics of informed and uninformed order flow rather than directly on the parameter estimates. We first discuss how to construct the dynamics from the parameter estimates. In the next section, we turn our attention to the impact of the dynamics on market liquidity, depth, and trading cost analysis.

To understand how the arrival rates of the two types of trades interact with each other and how they respond to innovations in the order flow, we rewrite the generalized autoregressive process as,

$$\begin{aligned}\tilde{\psi}_t &= \omega + \Phi \tilde{\psi}_{t-1} + \Gamma \tilde{\mathbf{Z}}_t \\ &\doteq \omega + \Phi^* \tilde{\psi}_{t-1} + \Gamma \xi_t.\end{aligned}$$

Table 2 (Continued).

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

Entries are maximum likelihood estimates of model parameters that govern the arrival rate dynamics:

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

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 . The autoregressive matrix is given by $\Phi^* = \Phi + \Gamma$. In the parentheses are standard errors. The last row reports the log-likelihood value (\mathcal{L}).

The second line is obtained via a linear approximation on the expectation of the balanced and imbalanced trades. The term $\Phi^* = \Phi + \Gamma$ captures the first-order persistence of the arrival rate forecasts and $\xi_t \equiv \tilde{Z}_t - \tilde{\psi}_{t-1}$ denotes the forecasting error, or innovation, in trading quantities. Based on this linear approximation, the multiperiod impact of a trade innovation on the arrival rate forecasts is given by

the following impulse response function:

$$\frac{\partial \mathbb{E}_t[\tilde{\psi}_{t+k}^j]}{\partial \xi_t^j} = [(\Phi^*)^k \Gamma]_{ij}, \quad i, j = 1, 2; \quad k = 0, 1, 2, \dots \quad (8)$$

where $[\cdot]_{ij}$ denotes the (i, j) th element of the impulse response matrix and captures the impact of the j th element of the shock ξ^j on the i th element of the arrival rate, $\tilde{\psi}^i$. In this system, the estimates on Γ capture the instantaneous impact of the time- t innovation on the time- t forecast of the next period's arrival rates. In contrast, the autoregressive matrix Φ^* measures the persistence of the arrival rate forecasts and determines to a large degree the multiperiod impact of the trade innovations. The whole picture of dynamics is obtained by a joint analysis of the instantaneous impact Γ , the autoregressive matrix Φ^* , and the whole impulse response function of each element.

3.1 The Instantaneous Impact of Trade Innovations

The instantaneous impact of trade innovations ξ on the arrival rate forecasts $\tilde{\psi}$ is captured by the Γ matrix. Inspecting the estimates of the Γ matrix in Table 2, we find that the estimates for all elements of the matrix are positive for all the 16 stocks. Therefore, shocks to both balanced and imbalanced trades have positive instantaneous impacts on the arrival rate of both informed and uninformed agents. Further inspection shows that the estimates for the Γ_{21} and Γ_{22} elements are larger than the estimates for the Γ_{11} and Γ_{12} estimates, indicating that both trade innovations have a larger impact on the arrival rate forecast of uninformed trades than on the arrival rate forecast of informed trades. As a result, we can more effectively forecast the uninformed arrival rate than the informed.

The elements Γ_{11} and Γ_{21} capture the instantaneous impact of the innovation in trade imbalance $|K|$ on the informed and uninformed arrival forecasts, respectively, holding the number of balanced trades constant. Hence, the positive coefficients imply that given a fixed number of balanced trades, increasing trade imbalances increase the arrival forecasts on both informed and uninformed arrivals, potentially because increasing the trade imbalance in this scenario also increases the total number of trades.

On the other hand, if we hold the total number of trades constant, the instantaneous effect of a *relative* increase in the trade imbalance is captured by $\Gamma_{11} - \Gamma_{12}$ on the informed arrival forecast and by $\Gamma_{21} - \Gamma_{22}$ on the uninformed arrival forecast. We find that the estimates for the difference $\Gamma_{11} - \Gamma_{12}$ remain predominantly positive, with only one exception in Citigroup. Thus, we conclude that a relative increase in the composition of the imbalanced trades also increases the arrival forecasts of informed trades for most stocks. However, the estimates for the difference $\Gamma_{21} - \Gamma_{22}$ have mixed signs negative for seven firms and positive for nine forms. Hence, the impact of a relative increase in the composition of imbalanced trades is ambiguous on the arrival forecast of uninformed trades.

Overall, we find that an absolute increase in either balanced or imbalanced trades increases the forecasts of both informed and uninformed arrivals. So we

Table 3 Stationarity of the dynamic arrival rate processes.

Ticker	Eigenvalue 1	Eigenvalue 2
ASH	0.6473	0.9950
XOM	0.7464	1.0013
DUK	0.6281	0.9958
ENE	0.6821	0.9974
AOL	0.7401	1.0014
MO	0.7080	0.9855
T	0.7347	0.9921
PFE	0.6901	0.9950
LUV	0.6996	0.9979
AMR	0.3310	0.9957
DOW	0.6936	0.9918
C	0.7259	1.0018
JPM	0.5672	0.9979
WMT	0.8115	0.9936
HD	0.6209	0.9956
GE	0.6437	0.9959

Entries report the two eigenvalues of the estimated autocorrelation matrix $\Phi^* = \Phi + \Gamma$ for each of the 16 stocks. The eigenvalues should be less than one for the processes to be stationary.

forecast greater arrival rates for both types of traders following an increase in trade of either type. However, an increase in the relative composition of the imbalanced trades while holding the total number of trades constant has a positive impact on the arrival forecast of informed trades, but an ambiguous impact on the arrival forecast of uninformed trades. So we forecast a greater arrival rate for informed traders following an increase in the share of trades that are imbalanced, but there is no clear effect of the share of imbalanced trades on the forecast of uninformed arrivals.

3.2 The Serial Dependence of Arrival Rate Forecasts

The Φ^* matrix captures the first-order persistence of the vector arrival rate forecasts on informed and uninformed trades. The diagonal terms of Φ^* capture how the current forecast is correlated with the lagged forecast of the same arrival rate. The parameter estimates reported in Table 2 indicate that the diagonal terms of Φ^* are mostly positive, indicating a trend following or herding behavior for both types of arrival rate forecasts. Table 3 reports the eigenvalues of this impact multiplier for the 16 stocks in our sample. Under the linear approximation, both eigenvalues should be less than one for the vector process to be stationary. Given the nonlinearity inherent in the dependence of $\mathbb{E}_{t-1}[\mathbf{Z}_t]$ on ψ_{t-1} , we cannot directly use the eigenvalues to determine the stationarity of the system. Nevertheless, the magnitudes of the eigenvalues give us an approximate picture of the persistence. For all

the 16 stocks, we find that the second eigenvalue of the multiplier matrix is very close to one, demonstrating the extreme persistence of the system.

The dynamics of the vector arrival rate processes is further complicated by the presence of large off-diagonal terms in Φ^* . In particular, the (2, 1)th element of the impact multiplier, Φ_{21}^* , captures the impact of the previous informed arrival rate forecast on the current uninformed arrival rate forecast. For all 16 stocks, the estimates for Φ_{21}^* in Table 2 are all remarkably negative. Thus, a forecasted increase in the arrival rate of informed trades leads to a systematic decrease in our forecasts of the uninformed arrival rate. This forecasting relation is not predicted by traditional microstructure models, which view the only determinant of uninformed trading as the presence of other uninformed traders. The behavior is more in line with models that allow discretionary behaviors for liquidity traders, e.g., Admati and Pfleiderer (1988), Foster and Vishwanathan (1990), and Lei and Wu (2000).

The impact of previous day's uninformed order arrival forecast on today's informed arrival forecast is captured by the (1, 2)th element of impact multiplier, Φ_{12}^* . The estimates on Φ_{12}^* reported in Table 2 are small, and are not consistently positive or negative across the 16 stocks. Hence, the arrival forecasts of informed trades do not depend much on lagged forecasts on the uninformed arrivals. This dynamic behavior is consistent with the hypothesis that informed traders act mainly on information, and do not respond strongly to the activity of uninformed traders.

3.3 The Multiperiod Impact of Trade Innovation

The impulse response function, defined in Equation (8), describes how a shock to one of the state variables will alter the evolution of these variables through time. Such shocks will typically decay over time but in this case there is substantial persistence. The impulse-response function is determined jointly by the instantaneous impact matrix Γ and the impact multiplier Φ^* . In Figure 1, we plot the normalized impulse-response function for the 16 stocks in our sample, computed based on Equation (8). To compare the relative persistence of each of the four elements, we normalize each element of the impulse-response function by the corresponding element in Γ so that all elements of the impulse response are normalized to one at the instantaneous level $k = 0$. The 16 stocks generate very similar persistence patterns. In particular, the arrival rate of uninformed trades (dotted line) is much more persistent than the arrival rate of informed trades (solid line), with one exception on AOL (the fifth panel). The persistence of cross-impacts falls between the two direct impacts.

This persistent behavior of informed and uninformed trades is not unexpected given that many studies have shown volume to be significantly and positively autocorrelated. But this result is at variance with the predictions of microstructure models in which trades are viewed as iid. Perhaps more importantly, the result reveals that trade patterns are predictable across trading days.

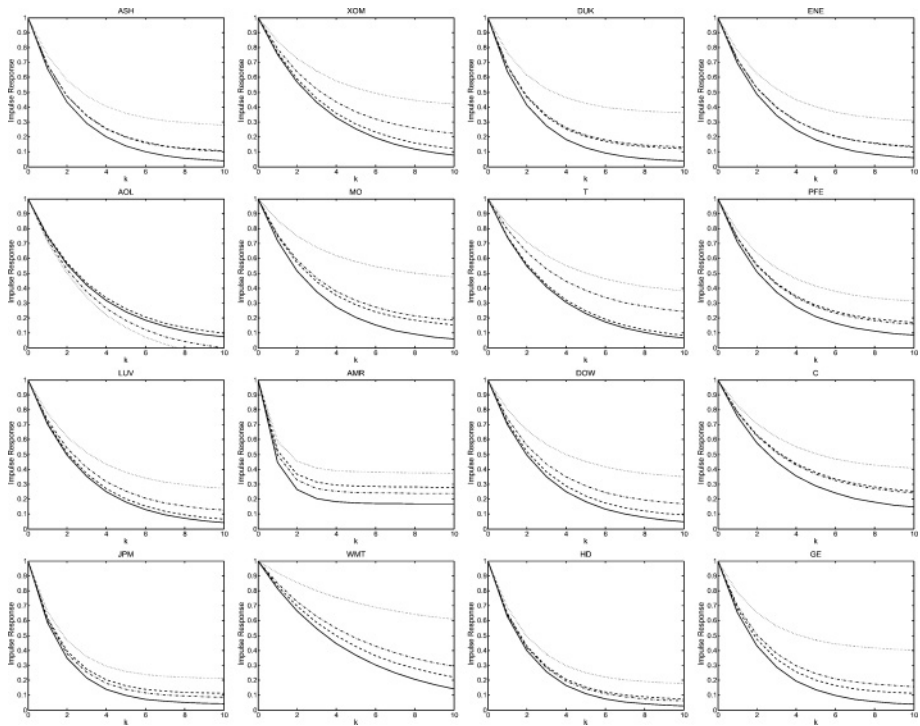


Figure 1 The normalized impulse response function. Lines depict the impulse response functions of the bivariate arrival rate system for the 16 companies. Each panel is for one company. In each panel, the solid line captures the impact of the trade imbalance $|K|$ on the arrival of informed trades, the dashed line captures the impact of the trade imbalance on the arrival of uninformed trades, the dash-dotted line captures the impact of the balanced trade $TT - |K|$ on the informed trade arrival, and the dotted line captures the impact of the balanced trade on the uninformed arrival. For ease of comparison, we normalize all responses at $k = 0$ to one.

3.4 Robustness of Arrival Rate Dynamics with Respect to Model Perturbations

We have also done the estimation with a generalized autoregressive process on the logarithm of the arrival rates instead of the arrival rates themselves. This specification is analogous to the EGARCH model of Nelson (1991). The maximized log likelihood values from the two models are very close to one another, neither model consistently dominating the other model across all stocks. More importantly, parameter estimates from both models imply similar dynamic behaviors for the informed and uninformed arrivals, showing the robustness of the results.² For both models, uninformed trades tend to be highly persistent. Uninformed order arrivals clump together, with high-volume days more likely to follow high-volume days, and conversely. However, an increase in the forecast of informed arrival rate leads

²The estimation results for this alternative specification are available upon request.

to a decline in future forecast of the uninformed arrival rate. The informed arrival rates also exhibit complex patterns, but the forecast of the informed arrival rate depends little on past forecasts of the arrival rates of uninformed trades.

4 FORECASTING MARKET LIQUIDITY AND DEPTH

In addition to providing insights on how the informed and uninformed dynamically interact with each other, the estimation of our dynamic model also generates direct forecasts on the arrival rates of informed and uninformed trades. These forecasts are informative in predicting the market liquidity and market depth. Thus, they are useful not only for academics in better understanding the market microstructure, but also for practitioners in better positioning their trades, and for risk managers seeking to measure the risks of illiquidity.

We also use our dynamic model to generate a time series of the PIN. This variable has been used in many studies to provide insight into the microstructure questions, such as the determinants of bid-ask spreads, and asset pricing questions, such as the determinants of the cost of capital. But all prior work using PIN required an assumption that it was constant over a substantial period of time. So PIN could not be used to provide insight into short-term, transitory changes in information-based trading. Here we show how to use the time series of PINs produced by our dynamic model to investigate the effects of earnings announcements on the variation in information-based trading.

4.1 Market Liquidity and Bid-Ask Spread

Market liquidity is often measured by the bid-ask spread: markets in which the bid-ask spread is small are interpreted as liquid markets. Our model links bid-ask spreads directly to the trade sequence and the arrival rates of informed and uninformed trades. By forecasting the arrival rates, we can predict the dynamics of bid-ask spreads.

We start by analyzing the bid quote in response to a sell order. Under our model, an application of Bayes rule shows that the probabilities of a good and a bad information event conditional on a sell order at time t are given by, respectively,

$$\Pr(\text{good}|\text{sell}_t) = \frac{\Pr_t(\text{good})\varepsilon_{t-1}}{\Pr_t(\text{bad})\mu_{t-1} + \varepsilon_{t-1}}, \quad \Pr(\text{bad}|\text{sell}_t) = \frac{\Pr_t(\text{bad})(\varepsilon_{t-1} + \mu_{t-1})}{\Pr_t(\text{bad})\mu_{t-1} + \varepsilon_{t-1}}, \quad (9)$$

where $\Pr_t(\text{good})$ and $\Pr_t(\text{bad})$ denote the prior probabilities at time t of a good and a bad information event, respectively, and $(\mu_{t-1}, \varepsilon_{t-1})$ denote the time- $(t-1)$ forecast of the arrival rates of informed and uninformed traders at time t . In a competitive market, the bid price must provide the market maker zero expected profit conditional on a trade at the bid, i.e., the arrival of a sell order. Thus, the bid price should be equal to the expected value of the asset conditional on history and on the arrival of a sell order. If we use \bar{V} to denote the expected asset value conditional on good news and \underline{V} the expected value conditional on bad news, we

can derive the bid price as

$$\begin{aligned} \text{bid}_t &= \Pr(\text{good}|\text{sell}_t)\bar{V} + \Pr(\text{bad}|\text{sell}_t)\underline{V} + \Pr(\text{no news}|\text{sell}_t)V^* \\ &= V^* + (\bar{V} - \underline{V}) \frac{\delta \Pr_t(\text{good})\varepsilon_{t-1} - (1 - \delta) \Pr_t(\text{bad})(\varepsilon_{t-1} + \mu_{t-1})}{\Pr_t(\text{bad})\mu_{t-1} + \varepsilon_{t-1}}, \end{aligned} \quad (10)$$

where $\Pr(\text{no news}|\text{sell}_t) = 1 - \Pr(\text{good}|\text{sell}_t) - \Pr(\text{bad}|\text{sell}_t)$ is the probability of no information event and $V^* \equiv (1 - \delta)\bar{V} + \delta\underline{V}$ denotes the unconditional expected value of the asset.

Now, we consider the ask price for a buy order. Again, we can apply the Bayes rule to derive the probabilities of a good and a bad information event conditional on a buy order,

$$\Pr(\text{good}|\text{buy}_t) = \frac{\Pr_t(\text{good})(\varepsilon_{t-1} + \mu_{t-1})}{\Pr_t(\text{good})\mu_{t-1} + \varepsilon_{t-1}}; \quad \Pr(\text{bad}|\text{buy}_t) = \frac{\Pr_t(\text{bad})\varepsilon_{t-1}}{\Pr_t(\text{good})\mu_{t-1} + \varepsilon_{t-1}}. \quad (11)$$

The ask price is the expected value of the asset conditional on this buy order,

$$\text{ask}_t = V^* + (\bar{V} - \underline{V}) \frac{\delta \Pr_t(\text{good})(\varepsilon_{t-1} + \mu_{t-1}) - (1 - \delta) \Pr_t(\text{bad})\varepsilon_{t-1}}{\Pr_t(\text{good})\mu_{t-1} + \varepsilon_{t-1}}. \quad (12)$$

From Equations (10) and (12), we can compute the bid-ask spread as a function of the trade sequence and the arrival rates of informed and uninformed traders. Therefore, our forecasts on the arrival rates lead to direct forecasts on the market liquidity as measured by bid-ask spreads.

For illustration, we consider the special case at the opening of each day t . We start the day with the unconditional probabilities of good and bad information events,

$$\Pr(\text{good}) = (1 - \delta)\alpha, \quad \Pr(\text{bad}) = \delta\alpha. \quad (13)$$

Plugging the unconditional priors in (13) into Equations (10) and (12), we obtain the date- t opening bid-ask spread (OS_t):

$$\text{OS}_t = (\bar{V} - \underline{V})\delta(1 - \delta)\alpha\mu_{t-1} \left[\frac{(\alpha\mu_{t-1} + 2\varepsilon_{t-1})}{((1 - \delta)\alpha\mu_{t-1} + \varepsilon_{t-1})(\delta\alpha\mu_{t-1} + \varepsilon_{t-1})} \right]. \quad (14)$$

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

$$\text{OS}_t = (\bar{V} - \underline{V}) \frac{\alpha\mu_{t-1}}{\alpha\mu_{t-1} + 2\varepsilon_{t-1}} = (\bar{V} - \underline{V})\text{PIN}_{t-1}, \quad (15)$$

where $\text{PIN}_{t-1} \equiv \alpha\mu_{t-1}/(\alpha\mu_{t-1} + 2\varepsilon_{t-1})$ denotes the time- $(t - 1)$ forecasted fraction of informed trades at time t that are based on information. Hence, the opening bid-ask spread is directly linked to the expected trade composition.

Our dynamic model provides conditional expectations of the arrival rates of informed and uninformed trades. We use the arrival rate forecasts to compute

Table 4 Sample properties of the forecasts on proportion of informed trades (PIN).

Ticker	Mean	SD	Auto	Min	Med	Max
ASH	0.157	0.036	0.987	0.063	0.154	0.273
XOM	0.121	0.025	0.951	0.045	0.114	0.213
DUK	0.157	0.026	0.980	0.068	0.158	0.237
ENE	0.149	0.036	0.992	0.065	0.145	0.286
AOL	0.123	0.017	0.714	0.068	0.125	0.165
MO	0.120	0.018	0.838	0.046	0.120	0.179
T	0.110	0.012	0.831	0.053	0.109	0.161
PFE	0.103	0.011	0.973	0.063	0.103	0.150
LUV	0.187	0.043	0.988	0.073	0.187	0.302
AMR	0.153	0.010	0.594	0.113	0.153	0.184
DOW	0.103	0.011	0.808	0.056	0.103	0.158
C	0.162	0.027	0.991	0.085	0.162	0.257
JPM	0.140	0.017	0.838	0.058	0.141	0.198
WMT	0.168	0.048	0.977	0.040	0.159	0.355
HD	0.128	0.036	0.978	0.054	0.114	0.233
GE	0.083	0.011	0.774	0.026	0.083	0.124

Entries report the sample average (Mean), standard deviation (SD), first-order autocorrelation (Auto), minimum (Min), median (Med), and maximum (Max) estimates on the forecasted time series of proportion of informed trades (PIN).

forecasts of the probability of informed trades, PIN. This conditional PIN is interpreted as the forecast of the probability that a trade on the next day will be from an informed agent. Then, we use these conditional PINs to predict market liquidity, exemplified by the opening bid-ask spread, using (14). The summary statistics for the PIN forecasts are reported in Table 4.

Figure 2 plots the time series of the PIN forecasts for each stock. For ease of comparison, we apply the same scale for all panels. We observe an obvious decline in the PIN forecasts over time for several stocks, especially during the last several years of our sample.

A new generation of asset pricing theories ascribe a role to liquidity. Easley, Hvidkjaer, and O'Hara (2002), O'Hara (2003), and Acharya and Pedersen (2005) differ on the measures of liquidity but agree on their importance. A simple measure of illiquidity is PIN, or the probability of informed trading. High values imply wide bid-ask spreads, small market depths, and costly trading by uninformed traders. From Table 4 and Figure 2, it is clear that PIN varies across assets and over time. Although the average level of PIN is substantially different for these 16 stocks, perhaps even more important is the movement in this indicator. For each stock, the PIN estimate varies greatly over time. The minimum PIN estimates for most stocks are in single digits (in percentage points), but the maximum can well be over 30 percentage points.

Table 5 Cross-correlations of the PIN forecasts on different stocks.

	ASH	XOM	DUK	ENE	AOL	MO	T	PFE	LUV	AMR	DOW	C	JPM	WMT	HD	GE
ASH	1.00	0.67	0.75	0.81	0.06	0.20	0.19	0.04	0.77	-0.04	0.04	0.75	0.58	0.52	0.68	0.26
XOM	0.67	1.00	0.68	0.67	0.09	0.19	0.17	0.06	0.61	0.03	0.06	0.58	0.58	0.34	0.61	0.20
DUK	0.75	0.68	1.00	0.77	-0.05	0.22	0.15	0.20	0.72	0.04	0.15	0.79	0.64	0.37	0.50	0.28
ENE	0.81	0.67	0.77	1.00	-0.02	0.11	0.17	-0.00	0.83	-0.09	-0.08	0.80	0.49	0.37	0.58	0.13
AOL	0.06	0.09	-0.05	-0.02	1.00	0.05	0.07	0.03	0.06	0.07	0.07	0.07	0.15	-0.01	0.21	0.05
MO	0.20	0.19	0.22	0.11	0.05	1.00	0.15	0.41	0.09	0.24	0.32	0.01	0.42	0.30	0.09	0.47
T	0.19	0.17	0.15	0.17	0.07	0.15	1.00	0.15	0.13	0.27	0.18	0.21	0.16	0.05	0.13	0.21
PFE	0.04	0.06	0.20	-0.00	0.03	0.41	0.15	1.00	0.01	0.50	0.36	0.11	0.34	0.15	-0.14	0.53
LUV	0.77	0.61	0.72	0.83	0.06	0.09	0.13	0.01	1.00	-0.02	-0.06	0.80	0.45	0.33	0.56	0.15
AMR	-0.04	0.03	0.04	-0.09	0.07	0.24	0.27	0.50	-0.02	1.00	0.32	0.09	0.19	0.02	-0.06	0.38
DOW	0.04	0.06	0.15	-0.08	0.07	0.32	0.18	0.36	-0.06	0.32	1.00	-0.08	0.36	0.15	-0.13	0.46
C	0.75	0.58	0.79	0.80	0.07	0.01	0.21	0.11	0.80	0.09	-0.08	1.00	0.37	0.10	0.48	0.19
JPM	0.58	0.58	0.64	0.49	0.15	0.42	0.16	0.34	0.45	0.19	0.36	0.37	1.00	0.45	0.28	0.44
WMT	0.52	0.34	0.37	0.37	-0.01	0.30	0.05	0.15	0.33	0.02	0.15	0.10	0.45	1.00	0.47	0.38
HD	0.68	0.61	0.50	0.58	0.21	0.09	0.13	-0.14	0.56	-0.06	-0.13	0.48	0.28	0.47	1.00	0.09
GE	0.26	0.20	0.28	0.13	0.05	0.47	0.21	0.53	0.15	0.38	0.46	0.19	0.44	0.38	0.09	1.00

Entries report the cross-correlation estimates of the PIN forecasts underlying 16 different stocks. Each estimate is based on the common sample of the two stocks involved.

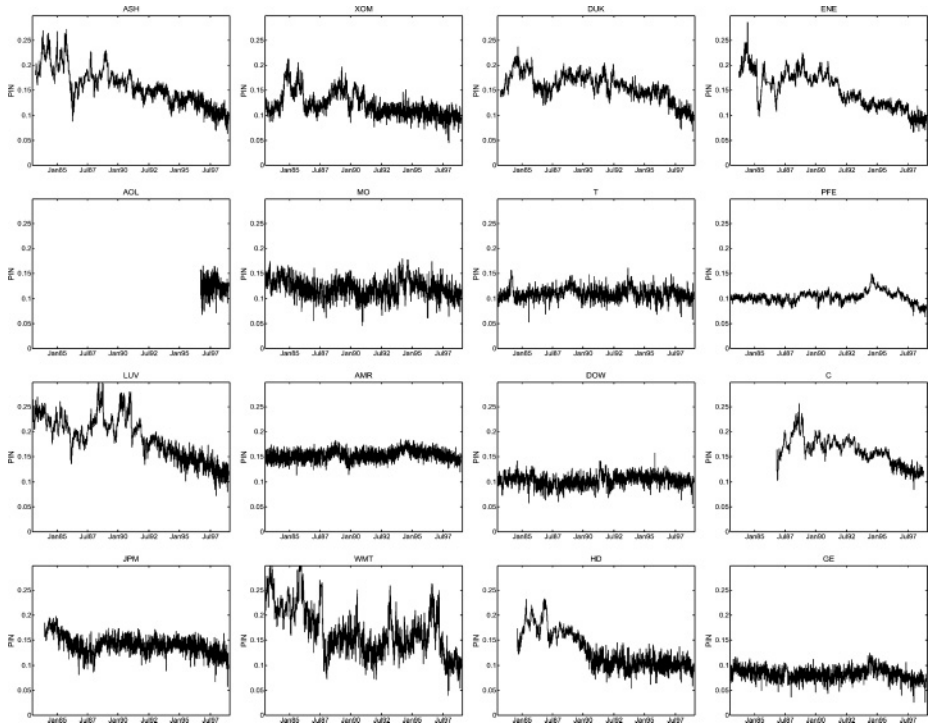


Figure 2 The time series of PIN forecasts. Lines depict the time series of the PIN forecasts from our estimated dynamic model for each stock. PIN denotes the probability of informed trades, defined as the arrival of informed trades over the arrival of total trades. Each panel represents one stock. For ease of comparison, we apply the same scale on all panels.

From an asset pricing point of view, the covariance of illiquidity across assets is also of importance. Just as with the risk of return, diversification can reduce the risk that an investor must sell when an asset is particularly illiquid. Hence, the strength of correlation matters, see for example Hasbrouck and Seppi (2001) and Chordia, Roll, and Subrahmanyam (2000). It is clear from Figure 2 that PIN moves similarly across assets. Table 5 reports the cross-correlation estimates between the PIN time series on different stocks. The correlations are estimated using the common sample of the two stocks involved. The estimates differ greatly across different stock pairs, ranging from -0.14 to 0.83 . Based on the common sample of 14 stocks,³ we perform principal component analysis and plot the normalized eigenvalues of each principal component in Figure 3. The plots show that one principal component explains 37% of the daily variation in the 14 PIN series. This estimate suggests that there is a systematic liquidity factor that underlies the stocks

³We drop AOL and C from the common sample analysis because their sample lengths are much shorter than the other stocks.

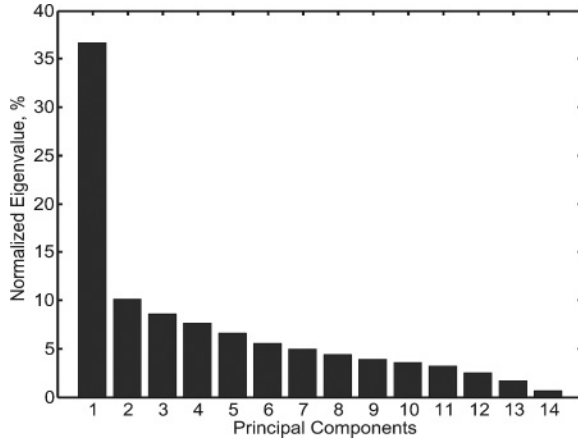


Figure 3 Percentage variation explained by each principal component of the PIN time series on 14 stocks. The length of bars denotes the normalized eigenvalues of the covariance matrix of the daily changes in the 14 time series of PIN estimates from our dynamic model. The normalized eigenvalues can be interpreted as the percentage variation explained by each principal component.

that we estimate. While diversification can remove the idiosyncratic component of the liquidity risk, the systematic liquidity risk in each stock should be priced.

To examine how informative the arrival rate forecasts are in predicting the opening bid-ask spread, we run the following forecasting regression on each stock:

$$\ln \text{OPS}_t = c_0 + c_1 \ln \text{PIN}_{t-1} + c_2 \ln \text{OPS}_{t-1} + c_3 \text{GARCH}_{t-1} + c_4 \ln \text{Volume}_{t-1} + e_t, \quad (16)$$

where OPS denotes the percentage opening bid-ask spread of a stock, defined as

$$\text{OPS} \equiv 2 \frac{\text{ask} - \text{bid}}{\text{ask} + \text{bid}}, \quad (17)$$

where we normalize the bid-ask spread by the average of the bid and ask level. The normalization has two purposes. First, we want to abstract from the impact of the scale of the quote. Second, we use the mid-quote as a proxy for the maximum impact of the information event. The term PIN_{t-1} denotes the time- $(t-1)$ forecast of the proportion of informed trades at time t . In addition to PIN, we also include three control variables: (1) the lagged spread OPS_{t-1} , (2) a standard GARCH(1,1) volatility estimate on the stock returns, GARCH_{t-1} , which measures the time- $(t-1)$ forecast of time- t return volatility, and (3) the aggregate trading volume at time $(t-1)$. We use these control variables to capture variations in the spread that are not explained by the proportion of informed trades. The first variable captures the unexplained persistence of the spread. The second variables captures the contribution of the price data, which can potentially reveal information about the variation in the spread between the upper and lower bounds of the valuation

$(\bar{V} - \underline{V})$. The last variable captures the impact of trade size, which is absent from our model. The significance of the estimates on c_1 indicates how informative our PIN forecasts are in predicting the opening bid-ask spread, on top of the predictions from the three control variables.

Since the estimate for δ is not exactly at $1/2$ for most stocks, in theory we should use a more complicated function of arrival rates as in (14) rather than PIN. Nevertheless, we use PIN for its simplicity and its intuitive interpretation as a measure for expected trade composition. Furthermore, several studies have generated the PIN estimates from the static model (based on either a rolling or a nonoverlapping window) and explored their implications. Using PIN from our dynamic model provides a comparison with these studies.

We estimate the regressions using generalized methods of moments, with the weighting matrix calculated according to Newey and West (1987) with 30 lags. Table 6 reports the slope estimates, their standard errors (in parentheses), and the R -squares of the regressions in (16). The forecasting performance of the PIN forecasts are quite remarkable. The estimates for the c_1 coefficient, which captures the impact of the probability of informed trades, are significantly positive for all but two stocks. The sample average of c_1 over the 16 stocks is 0.253, with an average standard deviation of 0.105. The strong statistical significance of the coefficient estimates are remarkable given that the arrival rate forecasts are obtained from purely trade quantities while the opening bid-ask spread is a price behavior.

The c_2 coefficient estimates on the autoregressive component are also significantly positive for all stocks, indicating that the persistence of the bid-ask spreads cannot be fully explained by the arrival rate forecasts. Furthermore, the coefficient estimates c_3 on the GARCH volatility are on average positive and that on the trading volume are on average negative, suggesting that the opening bid-ask spread is higher if the previous day's volatility is high but trading volume is low. Overall, the regression in (16) exhibits pronounced forecasting power, with an average R -square of 31.2%.

It is important to note that our arrival rates forecasts can be used to forecast the bid-ask spreads under any given trade sequences. Here, we use the specific regressions on the opening bid-ask spreads to illustrate their forecasting power and potential usefulness in forecasting the time-variation in market liquidity.

4.2 Market Depth and Price Impacts of Trade Orders

When a portfolio manager tries to purchase or liquidate a large position by sending consecutive buy or sell orders to the market, the price change induced by this series of orders could be significant. Using our dynamic microstructure framework, we can compute the price impact of this sequence of orders as a function of the arrival rates of informed and uninformed trades. Since we have forecasts of the arrival rates, our dynamic model can also be used to forecast the market depth and the potential cost of loading or unloading a position.

We use a sequence of N consecutive buy orders as an example. Let $\Pr_t^{N-1}(\text{good})$ and $\Pr_t^{N-1}(\text{bad})$ denote the probabilities of a good and a bad information event

Table 6 Forecasting opening bid-ask spread.

Ticker	c_0	c_1	c_2	c_3	c_4	R^2
ASH	−3.234 (0.664)	0.625 (0.061)	0.200 (0.024)	3.576 (2.453)	−0.036 (0.010)	0.221 —
XOM	−1.007 (0.319)	0.520 (0.059)	0.334 (0.033)	−0.205 (0.399)	−0.118 (0.022)	0.274 —
DUK	−5.512 (1.133)	0.446 (0.122)	0.228 (0.025)	9.648 (3.118)	−0.018 (0.013)	0.229 —
ENE	−1.860 (0.364)	0.405 (0.068)	0.352 (0.035)	−2.245 (0.950)	−0.010 (0.012)	0.206 —
AOL	−14.457 (1.539)	−0.886 (0.167)	0.344 (0.079)	27.622 (2.697)	0.162 (0.044)	0.410 —
MO	−3.705 (0.323)	0.124 (0.061)	0.285 (0.028)	−0.751 (0.579)	0.003 (0.011)	0.085 —
T	−0.232 (0.349)	0.341 (0.103)	0.522 (0.026)	−0.289 (0.451)	−0.106 (0.018)	0.325 —
PFE	−4.942 (0.441)	−0.215 (0.102)	0.238 (0.027)	6.878 (0.662)	−0.105 (0.015)	0.292 —
LUV	−2.140 (0.278)	0.578 (0.069)	0.269 (0.026)	−0.239 (0.848)	−0.019 (0.011)	0.264 —
	−1.314 (0.416)	0.082 (0.138)	0.557 (0.032)	−0.119 (0.173)	−0.068 (0.013)	0.339 —
AMR	−0.063 (0.259)	0.418 (0.081)	0.499 (0.025)	0.383 (0.874)	−0.144 (0.018)	0.321 —
DOW	−5.156 (1.515)	0.335 (0.289)	0.277 (0.053)	11.511 (1.970)	−0.045 (0.049)	0.491 —
C	−0.648 (4.531)	0.267 (0.170)	0.444 (0.032)	−1.784 (19.243)	−0.108 (0.014)	0.400 —
JPM	−3.886 (0.356)	0.210 (0.049)	0.453 (0.033)	4.243 (0.987)	0.029 (0.010)	0.318 —
WMT	−0.587 (0.329)	0.531 (0.059)	0.420 (0.030)	−0.302 (1.020)	−0.093 (0.010)	0.610 —
HD	−2.508 (0.317)	0.273 (0.078)	0.377 (0.037)	2.333 (0.822)	−0.065 (0.016)	0.214 —
GE						
Average	−3.203 (0.821)	0.253 (0.105)	0.363 (0.034)	3.766 (2.328)	−0.046 (0.018)	0.312 —

Entries report the estimates (standard deviation in parentheses) of the following forecasting regression on opening percentage bid-ask spread:

$$\ln OPS_t = c_0 + c_1 \ln PIN_{t-1} + c_2 \ln OPS_{t-1} + c_3 \text{GARCH}_{t-1} + c_4 \ln \text{Volume}_{t-1} + e_t,$$

where OPS_t is the opening percentage bid-ask spread at date t , PIN_{t-1} the time- $(t - 1)$ forecast of the proportion of informed trades at time t , GARCH_{t-1} is the time- $(t - 1)$ GARCH(1,1) forecast of time- t volatility for the stock return, and Volume_{t-1} is the aggregate trading volume of the stock on date $t - 1$. PIN is computed based on the arrival rates forecasts, which are obtained based on parameter estimates reported in Table 2. The columns under R^2 reports the R -square of the regression. The last two rows report the sample averages of the estimates and standard errors.

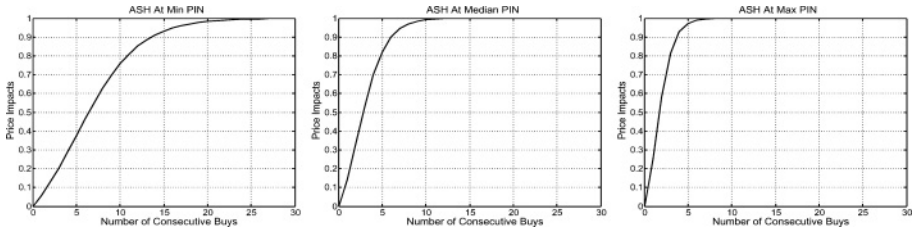


Figure 4 The price impact of consecutive buy orders on Ashland Oil. The lines depict the normalized price impact curves of consecutive buys (γ_t^N/δ), computed based on the arrival rate forecasts on Ashland Oil from our dynamic model on three different dates, when the forecasted proportion of informed trades (PIN) is at the minimum (left panel), median (middle panel), and maximum (right panel), respectively.

conditional on $N - 1$ consecutive buy orders. From (12), we can derive the price impact of N consecutive buys as

$$\text{ask}_t^N = V^* + (\bar{V} - \underline{V})\gamma_t^N,$$

where γ_t^N captures the impact of N consecutive buys:

$$\gamma_t^N = \frac{\delta \Pr_t^{N-1}(\text{good})(\varepsilon_{t-1} + \mu_{t-1}) - (1 - \delta) \Pr_t^{N-1}(\text{bad})\varepsilon_{t-1}}{\Pr_t^{N-1}(\text{good})\mu_{t-1} + \varepsilon_{t-1}}. \quad (18)$$

The probabilities $\Pr_t^{N-1}(\text{good})$ and $\Pr_t^{N-1}(\text{bad})$ can be readily updated via Bayes rule as in (11), starting from 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 value \bar{V} . The speed of convergence governs the depth of the market and is determined by the arrival rate forecasts (μ_{t-1} , ε_{t-1}).

To illustrate how the arrival rate forecasts impact the market depth, we use the first stock of our sample, Ashland Oil, as an example and consider three dates in our sample period when the PIN forecasts on Ashland are at the sample minimum, median, and maximum, respectively. At each of three PIN levels, we use the estimated model parameters on Ashland Oil and the arrival rate forecasts for that date to compute the price impacts of N consecutive buy orders (γ_t^N) according to Equation (18) and then normalize the impacts by their convergent value δ . Figure 4 plots the three normalized price impact curves (γ_t^N/δ) as a function of the number of consecutive buy orders (N) at the three selected PIN levels for Ashland.

All three normalized curves start at zero with zero trade and converge to one as the stock price converges to its upper bound \bar{V} with increasing number of consecutive buy orders. The speeds of convergence are captured by the slope of the curves and are different under different arrival rate forecasts. During the sample period, the minimum forecasted PIN for Ashland is 6.34%. Under this minimum level of forecasted informed trading (left panel), the market maker adjusts the ask quote slowly to the order flow. It takes about 30 consecutive buy orders for the stock

price to converge to its upper bound. The market is therefore deep. In contrast, the maximum PIN forecast for Ashland is 27.27%. At this maximum level of forecasted informed trading (right panel), the market makers adjust the ask price quickly to the sequence of buy orders. The price converges to its upper bound after fewer than 10 consecutive buy orders. The middle panel in Figure 4 shows an intermediate price impact curve when the forecasted proportion of arrival rates are at the median level of 15.37%.

Given the estimated model parameters and the arrival rate forecasts, we can also compute the price impact curve for N consecutive sells and for any sequence of buys and sells. Knowledge of the price impact curve is very important for institutional portfolio managers in analyzing the potential trading cost and in designing strategies for loading or unloading their positions. Our arrival rate forecasts can be used to predict the market depth and trading cost in terms of such price impact curves.

The price impact curve of a sequence of order flows provides the complete picture on the market depth, but it is often useful to summarize the market depth with a more compact measure. For example, Engle and Lange (2001) define a market depth measure VNET, which is designed to capture the net order flow associated with a fixed price movement. The larger this net order flow is for a fixed price movement, the deeper the market is. Based on 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 normalized price impact γ_i^N/δ to exceed half of its maximum impact. Intuitively, the half-life measure provides the portfolio managers an estimate on the maximum number of buy orders he can execute for the price impact to stay within a certain range.

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

Figure 5 depicts three typical time series of our market depth (half-life) forecasts for, from left to right, Ashland, Exxon Mobil, and General Electric, respectively. For all three stocks, the market depth measured by half-life has increased in the nineties.

4.3 Informed Arrivals Before and After Earnings Announcements

We specify the arrival rates of informed and uninformed trades as a vector autoregressive process, in which balanced and imbalanced trades act as noisy signals about the underlying arrival rates. The arrival of informed trades is jointly determined by the arrival of traders and the arrival of information. Large informational

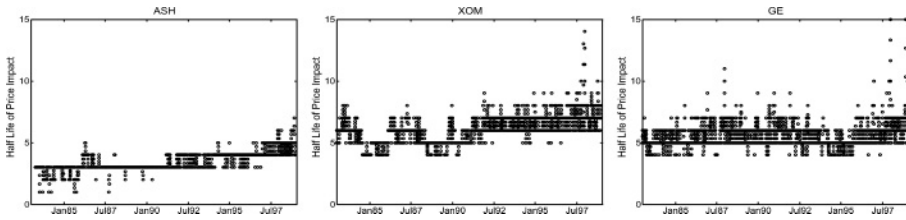


Figure 5 Time-varying forecasts of market depth. Lines depict three typical time series on the half-life ($\tau_{1/2}$) of the price impact of consecutive buy orders, defined as the number of consecutive buys needed for the impact to exceed half of its maximum. The half-life is computed based on our estimated dynamic model for, from left to right, Ashland, Exxon Mobil, and General Electric, respectively.

events, such as the releases of important economic numbers and the announcement of corporate earnings, happen at predetermined calendar dates, generating calendar days effects in the information flow and in the arrival rate of informed trades.

To study whether our arrival rate estimates capture some of the calendar day effects, we take corporate earnings announcements as an example and perform an event study around announcement days. Specifically, we compute the average PIN estimates as a function of the number of business days before and after the earning announcement days for each company. We obtain the announcement dates from the CompuStat. During our sample period, there are altogether 834 earnings announcements for the 16 stocks. Among them, 124 happened on Mondays, 183 happened on Tuesdays, 190 happened on Wednesdays, 229 happened on Thursdays, and 108 happened on Fridays. We do not know the exact timing of the announcement. Since the announcement can happen before the open, after the close, or during the trading hours, our measure of the number of business days before and after the announcement can deviate from the true measure by one day.

The left panel of Figure 6 plots the average PIN estimates as a function of the number of business days before and after the earning announcement date. The solid line denotes the sample average. The two dash-dotted lines represent the one standard deviation bands on the mean estimates. The plot shows that the proportion of informed trades increases as the announcement date approaches and declines after the announcement. The variation is the most significant within a ± 7 business day window.

For comparison, we also compute the average proportion of imbalanced trades, defined as $|B - S|/(B + S)$. If the order imbalances reveal the informed arrival with little noise, we would expect to observe similar average patterns. Figure 6 plots the results on the proportion of imbalanced trades in the right panel. The average proportions of imbalanced trades are higher than the average proportion of informed trade arrivals, potentially indicating that the imbalanced trades contain more noise than the balanced trades about the underlying arrival rates. The average estimates of the proportion of imbalanced trades show large zig-zag variations across different days, with little identifiable systematic patterns. Comparing the

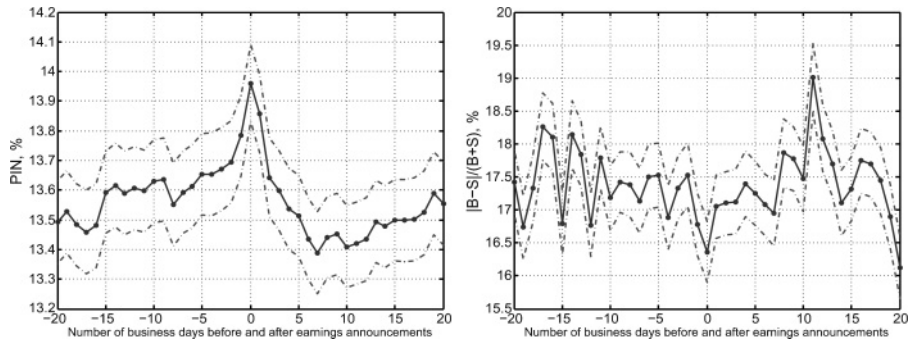


Figure 6 Proportion of informed arrivals and imbalanced trades before and after earnings announcements. The left panel plots the average proportion of informed trade arrival rates as a function of number of days before and after the earning announcement dates. The right panel plots the proportion of imbalanced trades. In both panels, the solid lines denote the sample average over the 834 earning announcement events across the 16 stocks. The dash-dotted lines define a one-standard deviation band.

scales of the two panels reveals that the large noise-induced zig-zag variation in the right panel completely dominates the systematic pattern observed in the left panel. The difference between the highest average PIN estimate at the announcement day and the lowest PIN estimate at seven days after the announcement is 56 basis points. By contrast, the zig-zag pattern in the right panel generates differences between neighboring estimates as high as 153 basis points. Thus, the systematic information about informed and uninformed arrivals is completely buried in the large noise of the raw order flow numbers. The different results from the two panels highlight the virtue of our dynamic specification in extracting useful information from the highly noisy realizations.

In a working paper, Benos and Jochec (2007) divide each quarter in between the earnings announcement days into two subsample periods. Then, they estimate the constant arrival rates in the static model separately for the two periods. They compute the PIN from the two sets of estimates and find that the average PIN estimate is lower before the announcement than after the announcement period. Our result shows that the PIN variation around the announcement dates happen within a very narrow window of ± 7 days before and after the announcement. Thus estimation based on the static model using one and a half-month window is unlikely to reveal actual announcement day effects.

Several other studies have estimated the PIN measure assuming that the arrival rates of informed and uninformed trades are constant over time. Examples include Easley, Kiefer, and O'Hara (1996), Easley et al. (1996), Easley, O'Hara, and Paperman (1998), Easley, Hvidkjaer, and O'Hara (2002, 2005), Dennis and Weston (2001), Easley, O'Hara, and Saar (2001), Vega (2005). To accommodate the time variation of the arrival rates in reality, they resort to repeated calibration over different sample periods. The re-calibration is often performed on a quarterly or annual frequency. In this paper, by specifying a dynamic process for the arrival rates, we can capture

the daily time variation of the arrival rates, and we can do so consistently with a model that is estimated only once over the whole sample period. Our analysis in this section further shows that capturing the daily variation in the arrival rates can be important in, for example, predicting opening bid-ask spreads and revealing announcement day effects.

5 DIAGNOSTIC ANALYSIS

By blending a microstructure framework with a dynamic forecasting specification, we can infer and predict the unobservable arrival rates of informed and uninformed trades based on the observable order flow of buys and sells. Our analysis so far has shown the great promise of using the arrival rate forecasts to predict bid-ask spreads and the price impact of a sequence of order flows, both of which are determining factors for trading cost analysis.

The accuracy of the arrival rate forecasts depends both on the microstructure setup and on the dynamic forecasting specification. The microstructure setup determines the likelihood of a sequence of buys and sells whereas the dynamic specification dictates the dynamics of the arrival rates. In this section, we perform diagnostic analysis on our dynamic microstructure model.

Based on the forecasted arrival rates of informed and uninformed trades, we forecast balanced and imbalanced trades. Therefore, one way to investigate the robustness of our specification is to check for remaining structure in the residuals of these trade forecasts. If our specification captures the data well, we should find minimal structure from the following standardized forecasting residuals $\mathbf{e}_t \equiv [e_{1t}, e_{2t}]^T$:

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

If the model is specified correctly, each of the two standardized forecasting residuals e_{it} should be serially independent and should have zero mean and unit variance. The number of buys and sells is governed by mixtures of Poisson distributions as shown in Equation (5). We can readily compute the conditional mean and variance of the trade quantities \mathbf{Z}_t by simulation.

Table 7 reports the summary statistics for the residuals. For each stock, the first row reports the properties of the standardized residual on the absolute trade imbalance and the second row reports the properties of the standardized residual on the balanced trades. Compared to the summary statistics of the raw trade quantities \mathbf{Z} in Table 1, the forecasting residuals show much less structure. The sample averages of the residuals are very close to zero. The serial dependence (the first-order autocorrelation) is significantly smaller than that in the raw series, and in many cases are no longer significantly different from zero. The cross-correlations between the two residuals are also smaller than those between the two elements of the raw trade quantities.

Nevertheless, we can still discern some remaining structure in the residuals. For example, the standard deviation of the standardized residual of the balanced

Table 7 Summary statistics of the standardized forecasting residuals.

Ticker	Mean	SD	Auto	VR	ρ
ASH	-0.022	0.996	0.014	0.157	-0.111
	0.008	1.872	0.026	0.693	—
XOM	0.176	1.208	0.206	0.182	-0.354
	-0.100	2.934	0.064	0.844	—
DUK	0.073	1.179	0.090	0.184	-0.275
	-0.069	1.845	0.007	0.807	—
ENE	0.044	1.074	0.050	0.254	-0.186
	-0.024	2.081	0.025	0.837	—
AOL	0.215	1.084	-0.005	0.500	0.233
	0.028	6.108	0.104	0.832	—
MO	0.000	1.028	0.145	0.317	-0.026
	-0.101	4.122	0.096	0.865	—
T	0.204	1.182	0.280	0.140	-0.192
	-0.275	3.837	0.097	0.728	—
PFE	0.204	1.076	0.179	0.239	-0.033
	-0.098	3.439	0.095	0.919	—
LUV	-0.124	0.880	0.096	0.342	-0.036
	-0.066	2.584	0.057	0.830	—
AMR	-0.220	0.887	0.028	0.564	0.058
	0.057	3.233	0.042	0.703	—
DOW	0.012	1.001	0.152	0.093	-0.110
	0.018	2.825	0.047	0.573	—
C	0.441	0.896	0.089	0.200	0.068
	-0.337	4.527	0.205	0.773	—
JPM	-0.025	1.038	0.085	0.387	-0.076
	-0.014	2.707	0.072	0.830	—
WMT	0.085	1.156	0.187	0.237	-0.236
	-0.177	3.132	0.089	0.867	—
HD	0.317	1.115	0.285	0.135	-0.089
	-0.142	3.417	0.121	0.825	—
GE	0.103	1.074	0.111	0.415	-0.138
	0.029	3.513	0.083	0.893	—
Average	0.093	1.055	0.124	0.272	-0.094
	-0.079	3.261	0.077	0.801	—

Entries report the summary statistics of standardized forecasting residuals on the absolute trade imbalance $|K|$ and the balanced trade $TT - |K|$:

$$e_{it} = \frac{Z_{it} - \mathbb{E}_{t-1}[Z_{it}]}{\sqrt{Var_{t-1}[Z_{it}]}} \text{ , } i = 1, 2, \text{ with } Z_t = [|K_t|, TT_t - |K_t|].$$

“Mean” is the sample average, “SD” is the standard deviation, “Auto” is the first-order autocorrelation, “VR” is the ratio of the variance of the forecasted trade quantity $\mathbb{E}_{t-1}[Z_{it}]$ versus the variance of the realized quantity Z_t , $VR = Var(\mathbb{E}_{t-1}[Z_{it}]) / Var(Z_{it})$, and “ ρ ” denotes the cross-correlation coefficient between the two residuals. For each stock, the first row reports the properties of the residual on absolute trade imbalance and the second row reports the properties of the residual on balanced trade. The last two rows report the sample averages of the statistics across the 16 stocks.

trade is systematically greater than one for all stocks. We conjecture that the bias is induced by the Poisson arrival assumptions on the informed and uninformed traders. A Poisson distribution is natural and tractable for modeling the arrival of traders, but a key limitation of this assumption is that the mean and variance of a Poisson variable are controlled by a single parameter—the Poisson arrival rate. The biases in the residuals of the balanced trades seem to indicate that the observed distribution of the balanced trades is more dispersed than implied by the mixture of Poisson distributions. If our conjecture is correct, the observed biases in the forecasting residuals can be corrected by choosing a more flexible distribution for the arrival of informed and uninformed traders so that we can disentangle the mean and variance of the trade quantities.

To gauge the forecasting power of our dynamic systems, we also report the percentage of variance (VR) in the two trade quantities explained by our model. The percentage variance VR is defined as the ratio of the variance of the forecasted trade quantities over the variance of the realized trade quantities. Overall, the model explains a larger percentage of the variation in the balanced trades than for the trade imbalance. This is not surprising given the inherent difficulty in forecasting informed arrivals.

As a test on our forecasting dynamics specification, we investigate whether the standardized residuals e_{it} can be forecasted further by additional variables:

$$e_{it} = c_0 + c_1 \hat{Z}_{i(t-1)} + c_2 \text{GARCH}_{t-1} + c_3 \ln \text{Volume}_{t-1} + \eta_{it}, \quad (19)$$

where $\hat{Z}_{i(t-1)}$ denotes the time- $(t-1)$ autoregressive forecasts of Z_{it} , GARCH_{t-1} denotes the time- $(t-1)$ GARCH(1,1) forecasts of the stock return volatility, and Volume_{t-1} is the time- $(t-1)$ aggregate trading volume of the stock. In this regression, we use the first variable to capture the persistence of the order flows that is missed by our dynamic specification in Equation (4). Furthermore, we use the GARCH volatility to capture the information from the price data, and the trading volume to capture the information in the trade size, neither of which is accommodated in our forecasting specification in Equation (4). A significant estimate on c_1 would point to further room for improvement in our dynamics specification. Significant estimates on c_2 and c_3 would show that price and trade size information can also be embedded in the forecasting dynamics for the arrival rates. As in GARCH-type models, the forecasting dynamics specification in Equation (2) can readily be extended to accommodate additional observables as long as they are informative about the unobservable arrival rates. Thus, a diagnostic regression like this can be used to identify additional informational sources for the prediction of the informed and uninformed arrival rates.

We estimate Equation (19) using the generalized methods of moments, where the weighting matrix is calculated according to Newey and West (1987) with 30 lags. Furthermore, to control the scale of the slope coefficients, we standardize each regressor by subtracting its sample mean and further scaling it by its sample standard deviation. Table 8 reports the coefficient estimates and standard errors. The last column under R^2 reports the R -square of the regression. For each stock, the

Table 8 Regressing the standardized forecasting residuals on additional variables.

Ticker	c_0		c_1		c_2		c_3		R^2
ASH	−0.022	(0.020)	0.006	(0.018)	−0.025	(0.016)	−0.033	(0.022)	0.002
	0.007	(0.029)	0.064	(0.055)	0.025	(0.030)	0.038	(0.039)	0.002
XOM	0.176	(0.041)	0.217	(0.033)	−0.018	(0.019)	−0.193	(0.030)	0.050
	−0.103	(0.049)	0.095	(0.076)	−0.036	(0.042)	−0.012	(0.070)	0.001
DUK	0.073	(0.030)	0.086	(0.025)	0.117	(0.030)	−0.035	(0.023)	0.015
	−0.070	(0.028)	0.041	(0.062)	0.011	(0.041)	0.051	(0.038)	0.002
ENE	0.044	(0.029)	0.071	(0.030)	0.003	(0.031)	−0.093	(0.034)	0.007
	−0.028	(0.035)	0.004	(0.081)	−0.007	(0.052)	0.095	(0.058)	0.002
AOL	0.209	(0.064)	−0.042	(0.066)	−0.275	(0.044)	−0.252	(0.057)	0.043
	0.008	(0.246)	0.819	(0.595)	−0.361	(0.421)	−1.041	(0.463)	0.011
MO	−0.001	(0.026)	0.227	(0.022)	−0.015	(0.014)	−0.154	(0.026)	0.039
	−0.107	(0.060)	0.023	(0.143)	0.028	(0.055)	−0.042	(0.122)	0.000
T	0.203	(0.039)	0.302	(0.028)	0.021	(0.021)	−0.122	(0.031)	0.059
	−0.276	(0.061)	0.256	(0.094)	0.125	(0.056)	−0.170	(0.090)	0.003
PFE	0.202	(0.027)	0.271	(0.028)	0.107	(0.030)	−0.068	(0.031)	0.041
	−0.099	(0.058)	−0.000	(0.187)	−0.163	(0.134)	−0.091	(0.099)	0.001
LUV	−0.124	(0.021)	0.105	(0.024)	0.010	(0.014)	0.025	(0.019)	0.019
	−0.068	(0.040)	−0.037	(0.071)	−0.042	(0.038)	0.066	(0.061)	0.001
AMR	−0.220	(0.022)	0.012	(0.020)	−0.002	(0.013)	−0.075	(0.017)	0.006
	0.055	(0.048)	0.176	(0.088)	−0.015	(0.040)	−0.062	(0.071)	0.002
DOW	0.012	(0.022)	0.200	(0.041)	0.001	(0.014)	−0.136	(0.020)	0.048
	0.019	(0.047)	0.004	(0.071)	0.040	(0.038)	0.058	(0.063)	0.001
C	0.437	(0.053)	0.140	(0.041)	−0.053	(0.068)	−0.094	(0.050)	0.020
	−0.347	(0.170)	1.291	(0.364)	1.005	(0.473)	−0.165	(0.544)	0.032
JPM	−0.026	(0.029)	0.087	(0.022)	−0.005	(0.015)	−0.103	(0.027)	0.010
	−0.014	(0.050)	0.086	(0.072)	0.018	(0.034)	0.023	(0.058)	0.001
WMT	0.079	(0.029)	0.206	(0.023)	0.095	(0.028)	0.015	(0.037)	0.036
	−0.183	(0.043)	0.060	(0.097)	−0.097	(0.056)	−0.064	(0.070)	0.001
HD	0.315	(0.031)	0.374	(0.041)	0.051	(0.029)	−0.020	(0.030)	0.100
	−0.148	(0.062)	0.266	(0.105)	0.134	(0.077)	−0.057	(0.093)	0.003
GE	0.098	(0.029)	0.150	(0.035)	0.025	(0.022)	−0.186	(0.026)	0.031
	0.026	(0.053)	0.145	(0.154)	0.078	(0.064)	0.025	(0.105)	0.002
Average	0.091	(0.032)	0.151	(0.031)	0.002	(0.025)	−0.095	(0.030)	0.033
	−0.083	(0.067)	0.206	(0.145)	0.046	(0.103)	−0.084	(0.128)	0.004

Entries report the estimates and standard deviations of the following regression on the standardized forecasting residuals:

$$e_{it} = c_0 + c_1 \widehat{Z}_{it(t-1)} + c_2 \text{GARCH}_{it-1} + c_3 \ln \text{Volume}_{it-1} + \eta_{it},$$

where $\widehat{Z}_{it(t-1)}$ denotes the autoregressive forecasts on Z_{it} , GARCH_{it-1} is the GARCH(1,1) volatility estimate from the stock returns, and Volume_{it-1} is the aggregate trading volume of the stock on date $t - 1$. We standardize each regressor by subtracting its sample mean and further scaling it by its sample standard deviation. The column under R^2 reports the R -square of the regression. For each stock, the first row reports the properties of the residual on absolute trade imbalance and the second row reports the properties of the residual on balanced trade. The last two rows report the sample averages of the statistics across the 16 stocks.

first row reports the regression for the residual on absolute trade imbalance and the second row reports that for the residual on balanced trade. The last two rows report the sample average of the statistics across the 16 stocks. The R -squares of these regressions are very small, with an average of 3.3% for the imbalanced trade residual and an average of 0.4% for the balanced trade. The small R -squares suggest that the three additional variables add little to the forecasting of the order flows. Inspecting the slope coefficient estimates, we find that on average, the autoregressive forecast of the trade imbalance predicts positively and the trading volume predicts negatively on the trade imbalance residual. On the other hand, none of the three variables add significantly to the prediction of the balanced trades. On average, the intercept of the balanced residual is not statistically different from zero, either.

6 CONCLUSION

In this paper, we model the dynamics of trade arrivals in the context of the Easley and O'Hara (1992) microstructure model. We select a sample of 16 actively traded NYSE stocks and use 15 years of daily transaction data to illustrate how we can estimate the arrival dynamics of trades originating from informed and uninformed traders using the daily number of buys and sells for each stock. The model is formulated as a point process for each type of investor where the parameter measuring the trade intensity is conditional on past information. These conditional intensities are modeled as functions of past balanced and unbalanced trades allowing for time trends and dynamic interactions.

The most immediate conclusion is our finding of strong daily autocorrelations in all of the trade series. Not only are trades in each stock strongly autocorrelated, but the number of balanced trades is highly autocorrelated and the number of unbalanced trades is weakly autocorrelated. As a consequence, the model implies that the arrival rate of uninformed trades is highly autocorrelated but that of informed traders is less so. Of particular interest is the observation that the arrival rate of uninformed trades is negatively affected by the past conditional arrival rates of informed trades. We conclude from this evidence that uninformed traders attempt to time their trades to avoid informed traders.

A further intriguing observation is in the dynamics of the informed traders. They show only a weak response to the conditional intensity of uninformed traders, even though they could be expected to hide in the crowd of uninformed trading. This is a natural finding in models with competing informed traders who will trade until profit is extracted from prices and then stop. Thus we conclude that it is the presence of information, rather than variation in the intensity of uninformed trade, that determines the arrival rate of informed traders.

We also show how these insights into the behavior of informed and uninformed traders can be used to predict other characteristics of the market. We calculate forecasts of measures of liquidity such as bid-ask spreads and market depth from the model estimates. We show that these forecasts are informative in predicting variations in market liquidity. We also use our dynamic model to generate a time series of the probability of information-based trade (PIN). We show how to use

this time series of PINs to investigate the effects of earnings announcements on the variation in information-based trading.

There are many potential applications for the market liquidity and depth forecasts that our dynamic microstructure model generates. From the academic perspective, understanding the evolution of liquidity and its interaction with information flow provides insight into the price formation process as well as into the more general area of asset pricing. For example, Easley, Hvidkjaer, and O'Hara (2002) use a static version of PIN forecasts to analyze whether information risk is priced. Using our dynamic model should provide a more consistent analysis between the estimated dynamics underlying the PIN forecasts and the cross-sectional patterns in stock returns. Most recently, Acharya and Pedersen (2005) develop an equilibrium asset pricing model with liquidity risk. They show that a security's required return depends on its expected liquidity as well as on the covariances of its own return and liquidity with market return and liquidity. Our dynamic arrival rate specifications provide evidence on how liquidity risk evolves over time and on its correlations across assets. We believe that our approach should provide a valuable starting point for further tests of the asset pricing implications of dynamic liquidity and a deeper understanding of the origins of liquidity correlations.

From a practical perspective, our model-based market liquidity and depth forecasts should be useful for trading cost analysis, which has emerged as a valuable tool to help portfolio managers handle execution costs in implementing their trading strategies. With performance typically benchmarked to some index, whether a manager is able to outperform is ever more determined by the skill with which he executes trades. Our dynamic microstructure model allows richer forecasts of trading cost from hypothetical trade sequences and should be useful in allowing traders to select strategies to minimize these costs.

Finally, from a risk management point of view, our estimates provide a forward-looking estimate of the sources of liquidation costs of a portfolio position. This should prove a useful input to calculation of the "liquidity risk" of a portfolio.

Received October 24, 2006; revised January 7, 2008; accepted January 10, 2008.

REFERENCES

- Acharya, V. V., and L. H. Pedersen. (2005). "Asset Pricing with Liquidity Risk." *Journal of Financial Economics* 77(2), 375–410.
- Admati, A. R., and P. Pfleiderer. (1988). "A Theory of Intraday Patterns: Volume and Price Variability." *Review of Financial Studies* 1(1), 3–40.
- Benos, E., and M. Jochec. (2007). "Testing the PIN Variable." Working paper, University of Illinois at Urbana-Champaign, IL.
- Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31, 307–327.
- Chordia, T., R. Roll, and A. Subrahmanyam. (2000). "Commonality in Liquidity." *Journal of Financial Economics* 56(1), 3–28.

- Chordia, T., R. Roll, and A. Subrahmanyam. (2001a). "Market Liquidity and Trading Activity." *Journal of Finance* 56(2), 501–530.
- Chordia, T., R. Roll, and A. Subrahmanyam. (2001b). "Trading Activity and Expected Stock Returns." *Journal of Financial Economics* 59(1), 3–28.
- Chordia, T., R. Roll, and A. Subrahmanyam. (2002). "Order Imbalance, Liquidity, and Market Returns." *Journal of Financial Economics* 65(1), 111–130.
- Chordia, T., R. Roll, and A. Subrahmanyam. (2005). "Evidence on the Speed of Convergence to Market Efficiency." *Journal of Financial Economics* 76(2), 271–292.
- Chordia, T., and A. Subrahmanyam. (2004). "Order Imbalance and Individual Stock Returns: Theory and Evidence." *Journal of Financial Economics* 72(3), 485–518.
- Dennis, P. J., and J. Weston. (2001). "Who Is Informed? An Analysis of Stock Ownership and Informed Trading." Working paper, University of Virginia and Rice University.
- Dufour, A., and R. F. Engle. (2000). "Time and the Price Impact of a Trade." *Journal of Finance* 55(6), 2467–2598.
- Easley, D., S. Hvidkjaer, and M. O'Hara. (2002). "Is Information Risk a Determinant of Asset Returns?" *Journal of Finance* 57(5), 2185–2221.
- Easley, D., S. Hvidkjaer, and M. O'Hara. (2005). "Factoring Information into Returns." Working paper, Cornell University and University of Maryland.
- Easley, D., N. M. Kiefer, and M. O'Hara. (1996). "Cream-Skimming or Profit-Sharing? The Curious Role of Purchased Order Flow." *Journal of Finance* 51(3), 811–833.
- Easley, D., N. M. Kiefer, and M. O'Hara. (1997a). "The Information Content of The Trading Process." *Journal of Empirical Finance* 4(2–3), 159–185.
- Easley, D., N. M. Kiefer, and M. O'Hara. (1997b). "One Day in the Life of a Very Common Stock." *Review of Financial Studies* 10(3), 805–835.
- Easley, D., N. M. Kiefer, M. O'Hara, and J. B. Paperman. (1996). "Liquidity, Information, and Infrequently Traded Stocks." *Journal of Finance* 51(4), 1405–1436.
- Easley, D., and M. O'Hara. (1992). "Time and the Process of Security Price Adjustment." *Journal of Finance* 47(2), 577–605.
- Easley, D., M. O'Hara, and J. Paperman. (1998). "Financial Analysts and Information-Based Trade." *Journal of Financial Markets* 1(2), 175–201.
- Easley, D., M. O'Hara, and G. Saar. (2001). "How Stock Splits Affect Trading: A Microstructure Approach." *Journal of Financial and Quantitative Analysis* 36(1), 25–51.
- Engle, R. F. (2000). "The Econometrics of Ultra High Frequency Data." *Econometrica* 68(1), 1–22.
- Engle, R. F., and J. Lange. (2001). "Predicting VNET: A Model of the Dynamics of Market Depth." Manuscript, New York University and Federal Reserve Board.
- Engle, R. F., and J. R. Russell. (1998). "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data." *Econometrica* 66(5), 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(4), 593–624.
- Hasbrouck, J. 1991. "Measuring the Information Content of Stock Trades." *Journal of Finance* 46(1), 179–208.
- Hasbrouck, J., and D. J. Seppi. (2001). "Common Factors in Prices, Order Flows, and Liquidity." *Journal of Financial Economics* 59(3), 383–411.

- Katti, S. K. (1960). "Moments of Absolute Difference and Absolute Deviation of Discrete Distributions." *Annals of Mathematical Statistics* 31(1), 78–85.
- Korajczyk, R. A., and R. Sadka. (2006). "Pricing the Commonality Across Alternative Measures of Liquidity." Working paper, Northwestern University.
- Lee, C. M. C., and M. A. Ready. (1991). "Inferring Trade Direction from Intraday Data." *Journal of Finance* 46(2), 733–746.
- Lei, Q., and G. Wu. (2000). "The Behavior of Uninformed Investors and Time-Varying Informed Trading Activities." Manuscript, University of Michigan.
- Manganelli, S. (2000). "Time, Volume and Price Impact of Trades." Manuscript, European Central Bank, Germany.
- Nelson, D. B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica* 59(2), 347–370.
- Newey, W. K., and K. D. West. (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55(3), 703–708.
- O'Hara, M. (2003). "Presidential Address: Liquidity and Price Discovery." *Journal of Finance* 58(4), 1335–1354.
- Vega, C. (2005). "Stock Price Reaction to Public and Private Information." *Journal of Financial Economics* 82(1), 103–133.