(preliminary and incomplete)

(tentative title)

Dynamic Trade Informativeness*

Bart Zhou Yueshen† Jinyuan Zhang‡

this version: February 1, 2018

* This paper benefits tremendously from discussions with Albert Menkveld, Vincent van Kervel, and Marcin Zamojski. There are no competing financial interests that might be perceived to influence the analysis, the discussion, and/or the results of this article.

† INSEAD; b@yueshen.me; INSEAD, 1 Ayer Rajah Avenue, Singapore 138676.

‡ INSEAD; jinyuan.zhang@insead.edu; Boulevard de Constance, Fontainebleau 77300, France.

(preliminary and incomplete)

(tentative title)

Dynamic Trade Informativeness

Abstract

This paper develops a structural model to examine price dynamics. The innovation lies in that trades' permanent price impact can be time-varying—dynamic trade informativeness. A distribution-free filtering technique pins the real-world data to the model. The filtered series demonstrate that the time-variation of trade informativeness accounts for a quarter of efficient price innovation; capture the intraday pattern of information asymmetry; improve the explanatory power of current trades for future returns; zoom in on informed trading around intraday events; and gauge informed investors' patience. The framework contributes to the better utilization of high-frequency trading data.

Keywords: trade informativeness, price impact, filtering, high-frequency data

(There are no competing financial interests that might be perceived to influence the analysis, the discussion, and/or the results of this article.)

1 Introduction

This paper studies the estimation of trade informativeness. Broadly speaking, a trade moves price in two ways: There is both a permanent price impact and a transitory one. The permanent component reflects a trade's information content (as, intuitively, the true information persists), while the transitory component reflects short-run pricing error (arising from price pressure, inventory cost, etc.) that dies out in the long-run. The terminology "trade informativeness" emphasizes our focus on the *permanent* price impact.

A bulk volume of the literature has developed various econometric methods to gauge trade informativeness (see Hasbrouck, 2007 for a review). More recently, the state space method pioneered by Menkveld, Koopman, and Lucas (2007) has seen a number of applications. A common feature in the extant literature is that the trade informativeness is treated as a constant, capturing the *average* permanent price impact of the estimation sample (typically a trading day). Our main contribution in this paper is to identify the *high-frequency dynamics* of trade informativeness.

Trade informativeness can indeed be time-varying, for a number of reasons. Fundamentally, the information-to-noise ratio might vary over trading hours, possibly due to investors coordination (Admati and Pfleiderer, 1988), time-varying volatility of noise trading (Back and Pedersen, 1998; and Collin-Dufresne and Fos, 2016), or simply because intraday corporate events or news that add to or resolve information asymmetry. It could also be driven by informed investors' strategic behavior due to risk aversion (Baruch, 2002), learning and competition (Holden and Subrahmanyam, 1992; Foster and Viswanathan, 1996; and Back, Cao, and Willard, 2000), and discounting or random horizon (Back and Baruch, 2004; and Caldentey and Stacchetti, 2010). Further, the number of informed traders who are active in the market might change (Back, Crotty, and Li, 2015; Wang and Yang, 2017; Banerjee and Breon-Drish, 2017). Such theory-predicted patterns, to the best of our knowledge, have not been empirically examined, perhaps because of the lack of an econometric framework to account for time-varying trade informativeness.

To account for this dynamic feature, we develop a structural model by extending the state space used in, e.g., Brogaard, Hendershott, and Riordan (2014). In essense, we treat the possibly time-varying trade informativeness as a hidden state variable. Then the main econometric objective becomes to dynamically filter out this hidden state based on observed data (trades and prices). To this end, we also develop a novel filtering technique following the generalized *autoregressive* method of moment, GaMM, first proposed by Creal, Koopman, Lucas, and Zamojski (2016). Section 2 discusses in detail of our methodology.

A notable advantage of our technique is that it does not require distribution assumptions, as opposed to, e.g., Kalman filter which requires normality in the data. The distribution-free property is particularly important when the model is applied to real-world trading data, where heavy / fat tails (and skewness) are often seen. In a simulation horserace shown in Appendix B, we demonstrate that our filtering technique outperforms Kalman filter when normality is lost or when the model is misspecified.

We then apply our structural model and the filtering technique to the U.S. equity trading data. The result, presented in Section 3, strongly supports the view that trade informativeness is indeed time-varying: While an order flow of size $10,000 on average moves the price by about 2.76 bps, from one trade to another, this (permanent) impact can change considerably, with a standard deviation of 0.73 bps. Under our structural model, this time-variation translates to about 24% of the innovation in the efficient price. Put alternatively, failure to account for the dynamics of trade informativeness underestimates the true information content in order flows by about a quarter.

Four applications illustrate the usefulness of our method logy. First, we examine the intraday pattern of trade informativeness by averaging across all the filtered intraday series of trade informativeness (of each stock, each day). We find over a typical trading day, trade informativeness appears monotone decreasing. It is the highest at the opening (about 20% above intraday average), suggesting the most severe adverse-selection, and the lowest at the closing (10% below). The decreasing pattern is in contrast to the well-known U-shape of volume (Admati and Pfleiderer, 1988), the reverse J-shape of bid-ask spread (McInish and Wood, 1992), and the hump-shape of depth (Ahn, Bae, and Chan, 2001). In particular, the fact that trade informativeness is at its lowest at the closing reveals that the widened spread and the reduced depth in the meantime are largely accounting for transitory price impact—price pressure, inventory cost, etc. The pattern is also consistent with the prediction that trade informativeness follows a supermartingale (decreasing on average), as seen in Baruch (2002) due to risk aversion and in Back and Baruch (2004) and Caldentey and Stacchetti (2010) due to random horizon.[1]

The filtered intraday pattern of trade informativeness can help investors make better execution decisions in real-time trading (as our filtering steps allow feeding in new observations as they become available). For uninformed investors, it is advisable to trade at times with low trade informativeness to avoid adverse-selection costs.

Second, to the extent that our filtered trade informativeness is capturing the *permanent* price impact, it should also help explain long-run, future price returns. We demonstrate that this is indeed the case by comparing the explanatory powers of two nested regression models. In the benchmark, only the current order flow is used to explain future returns. In the extension, we include also the current time-varying component (from our filtering) of trade informativeness. The latter model significantly improve the explanatory power of current trades for future returns by tripling or even quadrupling the adjusted-R^2. This improvement persists even if we consider a relatively far future return of ten minutes, affirming that the filtered time-varying component of trade informativeness indeed corresponds to the "permanent" price impact.

Third, our filtered dynamic trade informativeness can be used for intraday, high-frequency analyses, which can help understand the nature of intraday events and the connection with (informed) investors' behavior. For such an illustrative purpose, we conduct an event study to examine the dynamics of trade informativeness around scheduled Fed announcements. Our findings show that

[1] The decreasing pattern does *not* deny the mechanisms that might push trade informativeness to be a *super*martingale (e.g., Collin-Dufresne and Fos, 2016). It simply suggests that informed investors' incentives to trade early are empirically strong.

trade informativeness begins to pick up about two to three minutes ahead of, but then plummets within the minute immediately before the scheduled announcements. After the announcement, trade informativeness does not rise until about ten seconds later and then remain elevated for another six to seven minutes. The patterns are robust after controlling usual liquidity measures. We then discuss narratives consistent with the pattern in the paper.

Fourth, the filtered trade informativeness also identifies the (relatively more) informed side of each and every trade. Indeed, every trade has a buyer and a seller. If the efficient price is moved up (down) by the trade, the buyer (the seller) is more likely to be the informed. This attribute—the (relatively more) informed side of a trade—stands separately from the trade's aggressiveness, i.e., whether the buyer or the seller is using the market (usually determined by algorithms like Lee and Ready, 1991). The separation of trade informativeness from aggressiveness makes it possible to identify *executed* informed limit orders; that is, an aggressive buy (sell) that actually moves the efficient price down (up). The literature has long recognized that limit orders can be informed; see, e.g., Foucault (1999) and Kaniel and Liu (2006). More recently, Brogaard, Hendershott, and Riordan (2016) and Fleming, Mizrach, and Nguyen (2017) analyze the price discovery contribution by limit order *quotes*. The difference from their approach is that our analysis focuses on the limit order side of each and every *trade*.[2]

Along this line, we design a structural estimator for informed investors' patience, simply by counting the proportion of informed passive trades. Applied to days around earnings announcements, we find indeed that such patience drops significantly. Key to this patience measures is the identification of the more informed side for each and every trade. Conventional frameworks provides no such identification for individual trades, as they only generate average estimates.

The amount of transaction-level data has been overwhelming since the rise of algorithmic and

[2] The conventional term "(permanent) price impact" is often implicitly associated with order flows—the aggressive market orders. As a trade's (relatively more) informed side can be separate from its aggressive side, we favor the term "trade informativeness" for its neutrality: The interpretation is entertained that such informativenss can arise from either market or limit orders.

high-frequency trading. A large stock can easily see tens of thousands of trades on an average trading day. Using such rich intraday data as input, the conventional econometric frameworks however only outputs an overview, averaging out the trends and patterns in the dynamics. In contrast, the current econometric framework improves the utilization of high-frequency data in that it also maitains the data's high-frequency nature. To this extent, we believe this paper add empirical tools that help bridge some of the gaps in market microstructure research in the high-frequency trading world (O'Hara, 2015). We conclude with some potential directions for future research along this line in Section 4.

2 A structural model for dynamic trade informativeness

Consider a dataset where trades are sequentially indexed by $k \in \{1, 2, \dots\}$, each occurring at time t_k with signed size y_k (> 0 for buy and < 0 for sell market order). Denote by $p(t)$ the prevailing log-midquote price by time t. Hence, $p(t_k)$ is observed *immediately before* the k-th trade occurs and we write $p_{k-1} := p(t_k)$. This way, the observed information *before* the k-th trade can be conveniently summarized as $\mathcal{F}_{k-1} := \{p_j, y_j\}_{j \leq k-1}$. This section develops a structural model that can be cast on such a trading dataset to filter out, analyze, and understand the dynamic informativeness of each and every trade. (Our model discussion and applications are based on event time but can be analogously done for clock time.)

2.1 The canonical framework

We begin by revisiting the canonical approach to model price dynamics. See, e.g., Hasbrouck (2007) for a review; and more recent applications include Menkveld, Koopman, and Lucas (2007), Hendershott and Menkveld (2014), and Brogaard, Hendershott, and Riordan (2014). The observed price p_k is decomposed into two unobservable components:

$$p_k = m_k + s_k \tag{1}$$

where m_k is the (semi-strong) efficient price and s_k is the pricing error. Specifically, the efficient price m_k is modeled as a random walk, so that its innovation will persist, while the pricing error s_k is modeled as a stationary process, so that its innovation will decay to the long-run mean zero (hence the name "pricing error"). The dynamics of the efficient price and of the pricing error are further linearly linked to trading $\{y_k\}$:

$$\Delta m_k = m_k - m_{k-1} = \lambda y_k^* + \mu_k \tag{2}$$

$$(1 - \phi(L))s_k = \psi(L)y_k + \nu_k \tag{3}$$

where $\{\mu_k\}$ and $\{\nu_k\}$ are some independent white noises. The lag polynomials $\phi(L)$ ensures that s_k is stationary, so that it mean-reverts to zero in the long-run, satisfying the definition of "pricing error". The structure $\psi(L)$ captures the transitory price impact of trades. The order flow innovation, $\{y_k^*\}$, is another white noise, satisfying $(1 - A(L))y_k = y_k^*$, so that the efficient price m_k indeed follows a martingale: $\mathbb{E}_{k-1}[\Delta m_k] = 0$.

A focal point of such a system is the parameter λ, which governs, on average, how order flow innovation $\{y_k^*\}$ moves the efficient price $\{m_k\}$. The usual interpretation is that the larger is λ, the more (private) information the order flow has (Kyle, 1985). As such, λ is often labeled as (permanent) price impact, order flow informativeness, information asymmetry, or simply illiquidity.

A limitation of the above structural model is that it only informs empiricists of the *average* λ over the sample period. Below we generalize the canonical framework to allow time-varying λ_k and we shall refer to it as "the trade informativeness" of the k-th trade (Section 2.2). We also discuss the unique economic insights unearthed through the generalization (Section 2.3). Finally, we propose a new estimation technique (Section 2.4).

2.2 Enriching trade informativeness

To begin with, consider the following general structure of how the efficient price moves from one trade to another:

$$\Delta m_k = m_k - m_{k-1} = f_k(y_k^*),$$

where the possibly time-varying function $f_k(\cdot)$ describes the role of the order flow surprise y_k^*. Instead of assuming a specific form of $f_k(\cdot)$ (or providing a microfoundation), we turn to a generic first-order linear approximation by expanding $f_k(\cdot)$ around $y_k^* = 0$: $f_k(y_k^*) \approx f_k(0) + f_k^{'}(0) y_k^*$. We relabel the two components as $\mu_k := f_k(0)$ and $\lambda_k := f_k^{'}(0)$, thus, arriving at the familiar structure

$$\Delta m_k = \lambda_k y_k^* + \mu_k.$$

Compared to equation (2), the only difference is the (possibly) time-varying λ_k. The usual interpretation holds. The efficient price increment during the time interval $[t_k, t_{k+1})$ has a trade-related component, driven by the order flow surprise y_k^* at t_k, and a non-trade component μ_k summarizing everything else. The trade informativeness, λ_k, captures the per-unit impact of the k-th order flow surprise.

There are some conditions implied by the above structure. First, through the linearization, the possibly time-varying terms $f_k(0)$ and $f_k^{'}(0)$ are by construction orthogonal to y_k^*. That is, $\text{cov}_{k-1}[\lambda_k, y_k^*] = \text{cov}_{k-1}[\mu_k, y_k^*] = 0$, where the subscript $k-1$ emphasizes the covariance is based on all the observables before the k-th trade, i.e., conditional on $\mathcal{F}_{k-1} = \{p_j, y_j\}_{j \leq k-1}$. Second, the martingale property of m_k implies that

$$\mathbb{E}_{k-1}[\Delta m_k] = \mathbb{E}_{k-1}[\mu_k] + \mathbb{E}_{k-1}[\lambda_k y_k^*] = \mathbb{E}_{k-1}[\lambda_k]\mathbb{E}_{k-1}[y_k^*] = 0;$$

and

$$\mathbb{E}_{k-1}[\Delta m_k \Delta m_{k+j}] = \mathbb{E}_{k-1}[\mu_k \mu_{k+j}] = 0, \ \forall j \geq 1.$$

As such, μ_k has to be a white noise process.

Apart from $\text{cov}_{k-1}[\lambda_k, y_k^*] = 0$, the implied conditions on λ_k are rather mild, in fact, too mild that some additional structure is needed for identification. For this purpose, we further impose an autoregressive structure on λ_k with long-run mean $\bar{\lambda}$:

$$(1 - \alpha(L))(\lambda_k - \bar{\lambda}) = \varepsilon_k,$$

where ε_k is a white noise innovation, uncorrelated with the innovations y_k^* and μ_k;and $\alpha(L)$ is some lag polynomial capturing the persistence of λ_k.[3]

Summing up the discussion above, we arrive at the following structural model:

$$\begin{aligned}
&\text{observed price:} && p_k = m_k + s_k \\
&\text{efficient price:} && m_k - m_{t-1} = \lambda_k y_k^* + \mu_k \\
&\text{pricing error:} && (1 - \phi(L))s_k = \psi(L)y_k + v_k \\
&\text{trade informativeness:} && (1 - \alpha(L))(\lambda_k - \bar{\lambda}) = \varepsilon_k
\end{aligned} \tag{4}$$

where the order flow and its innovation satisfies $(1 - A(L))y_k = y_k^*$. We assume that the four innovation series—$\{\mu_k\}$, $\{v_k\}$, $\{\varepsilon_k\}$, and $\{y_k^*\}$—are white noises with zero means and are pair-wise uncorrelated. We write their respective (finite) variance as σ_μ^2, σ_v^2, σ_ε^2, and σ_y^2. The system degenerates to the canonical framework if λ_k is not time-varying; e.g., by forcing $\sigma_\varepsilon^2 = 0$.

2.3 Why dynamic trade informativeness is useful

The main econometric objective is to estimate the structural model and filter out the hidden process $\{\lambda_k\}$ from the observed trade and price series. Before proceeding to the estimation details, we dedicate some discussions to the potential economic insights that can only be obtained from the dynamic trade informativeness λ_k.

[3] The identification of λ_k requires certain persistence. If λ_k behaves like a white noise (e.g., with $\alpha(L) = 0$ and $\bar{\lambda} = 0$), then the product $\lambda_k y_k^*$ will also be a white noise, indistinguishable from μ_k, and the identification will not be possible. Note that our specification does not rule out such degenerate case and we let data inform us about the persistence in λ_k.

The time-varying nature of trade informativeness. It is well-known in the theory literature that trade informativeness (price impact) may well be time-varying. We provide a brief, incomplete review. Holden and Subrahmanyam (1992) show that because of imperfect competition, identically informed investors trade more aggressively early on in the trading period, generating a decaying trade informativeness over time. Foster and Viswanathan (1996) and Back, Cao, and Willard (2000) extend the analysis by allowing informed investors to have correlated dispersed private signals and they demonstrate non-monotone time-series patterns for trade informativeness. Even in the case of monopolist informed trader, the equilibrium trade informativeness can be a supermartingale if the discount rate is non-zero, because the investor competes with his impatient self (see, e.g., Back and Baruch, 2004; Caldentey and Stacchetti, 2010). With an activist who can dynamically affect the company's operation, the model by Back et al. (2016) features a stochastic trade informativeness process in equilibrium. Back and Pedersen (1998) find that trade informativeness can be a martingale when noise trading is stochastic. Collin-Dufresne and Fos (2016) extend the analysis to let noise trading have stochastic volatility and find trade informativeness can be a submartingale, increasing in expectation over time.

Under the conventional approach, the constant λ is unable to capture the above theory-predicted dynamic patterns. Researchers can only make a statement of the average trade informativeness. One may opt to break the trading hour into blocks and estimate λ for each block (e.g., Lee, Mucklow, and Ready, 1993) but such an approach faces the tradeoff between frequency (block length) and estimation accuracy (number of observations available in each block). Our structural model, together with the proposed filtering technique (see below), fills this gap. As we demonstrate later in Section 3.3, our filtered $\{\lambda_k\}$ does a better job in explaining future price changes.

Intraday events. Another use of the dynamics of trade informativeness is to analyze intraday events. For example, one can compare scheduled and unscheduled corporate announcements; different reactions after different macro news; and differences in the cross-section of securities. Reversely, given a salient structural change in the dynamics of λ_k, one can check back if it is driven by the trading behavior of some specific traders or trading groups, e.g., algorithmic / high-frequency traders, corporate insiders, mutual funds, etc.

In reality, the presence of informed trading during trading hours is uncertain and even if they exist, their informativeness might change through, e.g., dynamic information acquisition. To the best of our knowledge, previously these questions are only studied theoretically. Back, Crotty, and Li (2015) study a trading environment where a strategic investor always exists but may or may not have private information, while Wang and Yang (2017) study the situation where an informed investor may or may not exist. Banerjee and Breon-Drish (2017) look at an investor's strategic information acquisition timing decision. The richness of dynamic λ_k also enables researchers to empirically examine the implication of these intraday events—the entry of informed investor and their acquisition of information.

From a practice point of view, liquidity-driven investors can benefit from knowing the intraday pattern of trade informativeness in the real time, as they can avoid periods of time where trade informativeness (adverse-selection) is high. This can be easily achieved by filtering the trade informativeness using our model (4) and the filtering techinques discussed in Section 2.4. Notably, the predict-update recursion (equations 7 and 8) can be computed by feeding from real-time observations of trades and prices. Researchers with data of, e.g., institution execution can examine the $\{\lambda_{k|k}\}$-implied optimal execution time and study the potential savings in implementation shortfalls.

Distinguishing trade informativeness v.s. trade aggressiveness. The market microstructure literature often assumes that the aggressive side (market order) is more informed than the passive side (limit order). Such a conventional dichotomy goes back as early as to Copeland and Galai (1983) and is reinforced in the seminal papers like Kyle (1985) and Glosten and Milgrom (1985). In the context of a limit order market, however, it is very possible that the passive side is, at least sometimes, more informed (Foucault, 1999).

Indeed, every trade always involves two investors, a buyer and a seller, but the more informed one needs not to also be the more aggressive one—a trade's informativeness and aggressiveness are two different, though possibly correlated, attributes. Our structural model (4) allows the separation of the two and lets data inform us of the relation between them. To see this, the trade-related component in the efficient price m_k is $\lambda_k y_k^*$ for the k-th trade. Therefore, if $\lambda_k y_k^* > 0$ (< 0), the buy (sell) side of the trade is (relatively more) informed. Separately, the k-th trade is an aggressive buy (sell) if $y_k > 0$ (< 0), as often determined by algorithms like Lee and Ready (1991). To compare, the conventional approach (2) does not admit such separation as no statement about individual trades' (relatively more) informed side can be made. (The constant λ estimate only states about the *average* trade informativeness, not about individual trades.)

Empirically, estimates of the static λ of model (2) are typically positive, suggesting aggressiveness and informativeness are *on average* aligned: A positive (negative) surprise order flow moves the efficient price up (down), on average. However, this does not reject that the informed side and the aggressive side of individual trades can *sometimes* differ. For example, Collin-Dufresne and Fos (2015) argue that the use of limit orders could have contributed to the surprising finding that the estimated static λ is lower on days when informed investors (activists) heavily trade.

The separation of informativeness and aggressiveness allows us to study the information content in the passive side of trades, i.e., limit orders. These are trades where $\text{sign}(\lambda_k y_k^*) \neq \text{sign}(y_k)$. For example, if $\lambda_k y_k^* > 0$ but $y_k < 0$, the trade is initiated by an aggressive sell market order, but it has a positive impact on the efficient price, suggesting the buy limit order is more informed. Previous literature has also empirically studied the information content of limit orders, with the focus on the best bid and ask quotes (see Kaniel and Liu, 2006, Brogaard, Hendershott, and Riordan, 2016, and Fleming, Mizrach, and Nguyen, 2017). Complementarily, our approach focuses on the *trade* informativeness, letting the data determine, for each trade, whether the passive side (limit order) is (relatively) more informative than the aggressive side (market order).

Informed traders' patience. The theory of limit order market teaches us that investors, when allowed to choose order types, face the tradeoff between immediacy and trading cost: A market order guarantees immediate execution (no waiting cost), but incurs transaction cost (e.g., bid-ask spread). On the other hand, a limit order has better execution price (saving the spread) but is subject to uncertain execution (waiting cost). See Parlour and Seppi (2008) for a review.

Inspired by this tradeoff theory, one can use the filtered λ_k to construct empirical measures for informed investors' willingness to use limit orders—their "patience". For example, if the informed are very impatient, using only market orders, the proportion of trades that satisfy $\text{sign}(\lambda_k y_k^*) = \text{sign}(y_k)$ will be close to 100%; vice versa. Such a structural measure can be applied to different economic settings to empirically study the relationship between informed investors' (im)patience and, to name a few, maker/taker fee structure; characteristics of different sources of information; funding and market liquidity; etc.

2.4 Estimating the structural model

To estimate the structural model, the literature typically begins by estimating the order flow innovation y_k^* as the residual of $(1 - A(L))y_k = y_k^*$. We follow this approach. The structural model (4) can then be estimated by treating $\{y_k^*, y_k\}$ as known time-varying parameters, together with the observation series $\{p_k\}$.

The conventional approach is to impose some specific distribution on the innovation terms, $\{\mu_k, \nu_k, \varepsilon_k\}$, and then estimate the model (4) by maximum likelihood. For example, assuming normality, Menkveld (2013), Brogaard, Hendershott, and Riordan (2014), and Hendershott and Menkveld (2014) employ Kalman filter to the data. In principle, the same can be done to the current structural model, assuming the innovation terms are jointly normal. However, real-world trading data typically feature heavy tails, high kurtosis, and possibly skewness.[4] Rather then making explicit assumptions to account for possible non-normality, we would like to remain agnostic and

[4] Before implementing the structural model in the real data, we actually conjectured, a priori, that the trade informativeness series $\{\lambda_k\}$ could be right-skewed. Our conjecture was based on the interpretation that a negative λ_k could indicate a (more) informed limit order in the trade. All else equal, we expected an extremely informed investor would be more likely to trade aggressively on his private information, due to considerations like timeliness, competition, and waiting cost. Therefore, extremely negative λ_k would be rarer than positive extreme values, skewing the distribution to the right. Our estimation, summarized later in Table 1, confirms this conjecture and thus rejects the normality needed to implement a standard Kalman filter.

let data speak. To this end, we develop a *distribution-free* econometric technique.

Our approach has two steps. First, we estimate the structure of the pricing error, governed by $\phi(L)$ and $\psi(L)$. Second, we use these static parameters to help filter out the time-varying λ_k.

Estimating the pricing error structure with generalized method of moments, GMM. Yueshen (2016) notes that the structure, $\phi(L)$ and $\psi(L)$, of the pricing error s_k can be estimated by GMM without specifying the distribution of the innovation terms. We briefly review the method. To begin with, rearrange the price change between two trades Δp_k as

$$\begin{aligned}\Delta p_k &= (1-L)(m_k+s_k) = \lambda_k y_k^* + \mu_k + \frac{1-L}{1-\phi(L)}(\psi(L)y_k + \nu_k) \\ &= \hat{y}_k + \lambda_k y_k^* + \mu_k + \frac{1-L}{1-\phi(L)}\nu_k, \end{aligned} \tag{5}$$

where the second line introduces a short-hand notation of $\hat{y}_k := (1-\phi(L))^{-1}(1-L)\psi(L)y_k$. While $\hat{y}_k$ is not directly observed, it can be approximated via Taylor expansion. For example, when $\phi(L) = \phi \in (-1,1)$ and $\psi(L) = \psi$ (as in, e.g., Menkveld, 2013), $\hat{y}_k = \psi y_k - (1-\phi)\psi \sum_{j=1}^{\infty} \phi^{j-1} y_{k-j}$ and can be easily constructed recursively. By subtracting $\hat{y}_k$ from Δp_k, we obtain a residual term that is uncorrelated with all past order flow innovations y_{k-j}^* $(j \geq 1)$. This leads to sufficiently many moment conditions to identify $\phi(L)$ and $\psi(L)$:

$$\mathbb{E}\Big[(\Delta p_k - \hat{y}_k) y_{k-j}^*\Big] = 0, \text{ for } j \geq 1. \tag{6}$$

For example, in Menkveld (2013) and Brogaard, Hendershott, and Riordan (2014), $\phi(L) = \phi$ and $\psi(L) = \psi$. The two unknowns can be pinned down by letting $j \in \{1, 2, \ldots\}$ in the above.

Recursively filtering λ_k for each trade. Having estimated $\phi(L)$ and $\psi(L)$, next we turn to filtering $\{\lambda_k\}$. In doing so, we will also estimate the unconditional trade informativeness $\bar{\lambda}$ and the autoregressive structure $\alpha(L)$. For exposition simplicity, we focus on a special case of $\alpha(L) = \alpha \in (-1,1)$, i.e., letting $\{\lambda_k\}$ follow an AR(1) process.

Our objective is to "best guess" the unobservable true λ_k. It is useful to distinguish the information set based on which we make our best guesses. Given all the information before the next trade, i.e., conditional on $\mathcal{F}_k = \{p_j, y_j\}_{j\le k}$, our best guess—our *prediction*—of the next trade's informativeness, λ_{k+1}, is written as $\lambda_{k+1|k} := \mathbb{E}_k[\lambda_{k+1}]$. Upon seeing the next pair of trade and price, our best guess is *updated* to $\lambda_{k+1|k+1} := \mathbb{E}_{k+1}[\lambda_{k+1}]$. The terminology of "prediction" v.s. "update" follows the filtering literature (see, e.g., Durbin and Koopman, 2012). Under the AR(1) structure of $\{\lambda_k\}$, the "prediction step" becomes

$$\lambda_{k+1|k} = \mathbb{E}_k\left[(1-\alpha)\bar{\lambda} + \alpha\lambda_k + \varepsilon_{k+1}\right] = (1-\alpha)\bar{\lambda} + \alpha\lambda_{k|k}, \tag{7}$$

a weighted average between the long-run mean $\bar{\lambda}$ and the current "best guess" $\lambda_{k|k}$.

In order to compute (filter out) our new best guess $\lambda_{k+1|k+1}$, we need to complement the above prediction with an "update step", making use of the new observation $\{p_{k+1}, y_{k+1}\}$. This step can be generically expressed as (see, e.g., Creal, Koopman, and Lucas, 2013)

$$\lambda_{k+1|k+1} = \lambda_{k+1|k} + \beta\mathcal{S}_{k+1} \tag{8}$$

where $\mathcal{S}_{k+1}$ represents some "surprise" due to the new observation $\{p_{k+1}, y_{k+1}\}$ and β some scaling coefficient (to be estimated). Under Kalman filter, for example, this $\mathcal{S}_{k+1}$ would be the conditional expectation (given the new observation) of the innovation ε_k, scaled by the Hessian of the log-likelihood function. Since no specific distributions are assumed, we are unable to obtain such a conditional expectation. We turn to some moment conditions instead.

Our approach is inspired by the generalized *autoregressive* method of moments, or GaMM, proposed by Creal, Koopman, Lucas, and Zamojski (2016). Assume we already have a prediction $\lambda_{k+1|k}$ and we want to update it to $\lambda_{k+1|k+1}$ based on the newly observed $\{p_{k+1}, y_{k+1}\}$ (from which we also observe Δp_{k+1}, y^*_{k+1}, $\hat{y}_{k+1}$, etc.). To do so, we first define a function of

$$g_{k+1}(\lambda) := \left(\Delta p_{k+1} - \hat{y}_{k+1} - \lambda y^*_{k+1}\right) y^*_{k+1}. \tag{9}$$

Note that $g_{k+1}(\cdot)$ is monotone decreasing. Evaluated at the prediction $\lambda_{k+1|k}$, $g_{k+1}(\cdot)$ satisfies

$$\begin{aligned}\mathbb{E}_k[g_{k+1}(\lambda_{k+1|k})] &= \mathbb{E}_k\left[\left(\mu_{k+1} + \frac{1-L}{1-\phi(L)}\nu_{k+1}\right)y^*_{k+1} + \left(\lambda_{k+1} - \lambda_{k+1|k}\right)\left(y^*_{k+1}\right)^2\right] \\ &= \underbrace{\mathbb{E}_k\left[\left(\mu_{k+1} + \frac{1-L}{1-\phi(L)}\nu_{k+1}\right)y^*_{k+1}\right]}_{=0} + \underbrace{\mathbb{E}_k\left[\lambda_{k+1} - \lambda_{k+1|k}\right]}_{=0}\mathbb{E}_k\left[\left(y^*_{k+1}\right)^2\right] = 0,\end{aligned}$$

where the first equality follows the expression of Δp_{k+1} from Equation (5); and the second equality uses the structural assumption that y^*_{k+1} is orthogonal of λ_{k+1}.

We have obtained a (conditional) moment condition of $\mathbb{E}_k[g_{k+1}] = 0$, where we omit the argument $\lambda_{k+1|k}$ for brevity. That is, immediately before the $(k+1)$-th trade, under $\mathcal{F}_k$, $\mathbb{E}_k[g_{k+1}] = 0$ holds with the prediction $\lambda_{k+1|k}$. After observing $\mathcal{F}_{k+1}$, the realized value of g_{k+1} thus evaluates how good the prediction $\lambda_{k+1|k}$ was: If it turns out that $g_{k+1} > 0$ (< 0), one can conclude that $\lambda_{k+1|k}$ was too small (too large), because $g_{k+1}(\cdot)$ is monotone decreasing. That is, the sign of the realized g_{k+1} gives an indication of how one can improve the prediction of $\lambda_{k+1|k}$, i.e., how to *update* to $\lambda_{k+1|k+1}$ using the new observation. This is the core idea of Creal et al. (2016): For a possibly time-varying variable (e.g., λ_k), the *one-step ahead realization* of a conditional moment condition (g_{k+1}) essentially measures the misfit of (i.e., the "surprise" to) the current prediction ($\lambda_{k+1|k}$).

The misfit only informs us of the direction (and a rough magnitude) of how to update $\lambda_{k+1|k}$ to $\lambda_{k+1|k+1}$. To optimally quantify the update from the prediction, we take a scaled steepest descent by letting

(10) $$\mathcal{S}_{k+1} := \left(\mathbb{E}_k\left[\frac{\partial g_{k+1}}{\partial \lambda_{k+1|k}}\right]\right)^{-1} g_{k+1},$$

That is, we scale the misfit g_{k+1} is with the gradient of the GMM objective function

$$\max_{\lambda_{k+1|k}} \left(\mathbb{E}_k[g_{k+1}\left(\lambda_{k+1|k}\right)]\right)^2$$

with respect to $\lambda_{k+1|k}$. The idea is the same as Newton-Raphson method in root-finding. Creal et al. (2016) show that the updating step (8) with surprise (10) is locally optimal. In fact, taking both the prediction and the updating steps together (Equations 7 and 8), we obtain a dynamic structure for our "best guess":

$$\lambda_{k+1|k} = (1-\alpha)\bar{\lambda} + \alpha\lambda_{k|k-1} + \alpha\beta\mathcal{S}_k,$$

which is referred to as "GaMM(1,1)" by Creal et al. (2016).

To filter out $\lambda_{k+1|k}$ dynamics and to estimate the static parameters ($\bar{\lambda}$, α, and β), we follow Creal et al. (2016) and instrument the misfit g_{k+1} with $[1;\lambda_{k|k-1};\mathcal{S}_k]$, yielding the following set of unconditional moment conditions (where $\otimes$ denotes Kronecker product):

$$\mathbb{E}\big[g_{k+1}\otimes[1;\lambda_{k|k-1};\mathcal{S}_k]\big] = 0. \tag{11}$$

Specifically, $\mathbb{E}[g_{k+1}] = 0$ pins down $\bar{\lambda}$ as in the standard GMM (Yueshen, 2016). The unconditional moments $\mathbb{E}[g_{k+1}\lambda_{k|k-1}] = 0$ and $\mathbb{E}[g_{k+1}\mathcal{S}_k] = 0$ exploit the zero autocorrelation in g_{k+1} (because of the white noise $\{y_k^*\}$) by optimizing over the GaMM coefficients α and β. These moment conditions readily fit into the standard GMM framework (Hansen, 1982) and the usual asymptotic normality of the estimators holds.

Higher moments. The GMM conditions (11) only identify the dynamic $\{\lambda_{k+1|k}\}$ and the static parameters $\{\bar{\lambda},\alpha,\beta\}$ (hence also $\{\lambda_{k|k}\}$ following Equation 8). In Appendix A, we exploit several second moment conditions so that our approach also estimates the variances of the innovations: σ_μ^2, σ_ν^2, and σ_ε^2. Third moments or higher can be estimated in a similar fashion.

Outperforming Kalman filter. We illustrate the superiority of our proposed estimation method over standard Kalman filter via simulation. The results are summarized in Appendix B, where we run a horserace between the two approaches via simulated data. When the innovation terms are non-normal or when there is model misspecification, our approach outperforms Kalman filter as judged by the root-mean-squared error of the filtered estimates. (Of course, when the innovations are normally distributed and the model is correctly specified, Kalman filter outperforms.)

3 Applications

We estimate our structural model using the U.S. equity intraday trading data. Our sample includes 300 randomly selected stocks from S&P 1500 index over the year 2014. Price and trade information during trading hours (9:30 to 16:00, Eastern Standard Time) are collected for each stock-day from Monthly Trade and Quote (TAQ) database. To avoid bid-ask bounces, the log-transformed national best bid-offer (NBBO) midquote price immediately before the k-th trade is used for p_{k-1}. (The last prevailing midquote by 16:00 is used for the last p_k.) Trades are signed using the algorithm proposed by Ellis, Michaely, and O'Hara (2000). Our results are robust to alternative signing algorithms. To facilitate comparison across stocks, we measure the order flow $\{y_k\}$ in \$10,000. We also use the algorithm by Holden and Jacobsen (2014) to alleviate the potential data issues associated with Monthly TAQ data.

For each stock-day, we estimate the state space model (4) using the method described in Section 2.4. The lag polynomials are chosen to be $\phi(L) = \phi L$, $\psi(L) = \psi$, and $\alpha(L) = \alpha L$. The order flow innovation y_k^* is constructed from an autoregressive model of y_k with ten lags. This parsimonious specification closely follows Brogaard, Hendershott, and Riordan (2014).

Lastly, we complement our intraday data with daily stock information collected from the Center for Research in Security Price (CRSP). Specifically, we sort our 300 stocks into three size groups—small, medium, and large—according to their daily average market capitalization during October to December 2013, three months ahead of the above sample period. Each group has 100 stocks.

3.1 Estimation result

There are 248 trading days in 2014. With the 300-stock cross-section, our sample has a maximum of 248×300 stock-day observations. The number of valid estimates falls short of this maximum because the numerical optimization does not converge for some stock-days. The overall convergence rate is about 97.4%, with a breakdown of 97.8%, 98.8%, and 95.7% respectively for large, medium,

	Unit	All	Large	Medium	Small	Percentile in the full sample 5%	25%	50%	75%	95%
(a) Structural parameters for trade informativeness λ_k										
$\bar{\lambda}$	bps/$10,000	2.76	0.31	1.22	6.85	0.12	0.37	0.88	2.96	11.12
		(11.81)	(19.64)	(15.60)	(13.56)					
α		0.13	0.15	0.08	0.15	-0.14	-0.01	0.08	0.21	0.67
		(32.35)	(15.99)	(32.95)	(34.37)					
β		-0.34	-0.29	-0.29	-0.46	-1.37	-0.38	-0.18	-0.07	-0.01
		(-53.97)	(-51.45)	(-54.10)	(-38.43)					
(b) Structural parameters for pricing error s_k										
ψ	bps/$10,000	-1.90	-0.22	-0.82	-4.69	-7.62	-1.93	-0.56	-0.24	-0.08
		(-11.84)	(-19.11)	(-15.38)	(-13.70)					
ϕ		0.50	0.52	0.49	0.47	0.19	0.42	0.52	0.62	0.74
		(92.29)	(52.35)	(55.95)	(59.46)					
(c) Innovation sizes										
$\sqrt{\text{var}[\varepsilon_k]}$	bps/$10,000	0.61	0.10	0.31	1.44	0.01	0.05	0.17	0.52	2.51
		(12.62)	(12.98)	(21.09)	(13.28)					
$\sqrt{\text{var}[\mu_k]}$	bps	1.60	0.63	1.30	2.90	0.01	0.63	1.04	2.03	4.78
		(24.02)	(32.92)	(30.53)	(228.00)					
$\sqrt{\text{var}[\nu_k]}$	bps	0.70	0.22	0.38	1.53	0.00	0.01	0.28	0.73	3.10
		(14.52)	(18.58)	(18.35)	(14.50)					
Count		72,501	24,264	24,504	23,733					

Table 1: Estimated parameters of the structural model. This table reports the summary statistics of the estimated parameters of the structural model:

$$\begin{aligned} \text{observed price:} \quad & p_k = m_k + s_k \\ \text{efficient price:} \quad & m_k = m_{k-1} + \lambda_k y_k^* + \mu_k \\ \text{pricing error:} \quad & s_k = \phi s_{k-1} + \psi y_k + \nu_k \\ \text{trade informativeness:} \quad & \lambda_k = (1-\alpha)\bar{\lambda} + \alpha\lambda_{k-1} + \varepsilon_k \end{aligned}$$

where in filtering the hidden λ_k we assume a GaMM(1,1) structure (Creal et al., 2016):

$$\begin{aligned} \text{prediction:} \quad & \lambda_{k+1|k} = (1-\alpha)\bar{\lambda} + \alpha\lambda_{k|k}; \\ \text{update:} \quad & \lambda_{k+1|k+1} = \lambda_{k+1|k} + \beta\mathcal{S}_{k+1}. \end{aligned}$$

The reported estimates include the structural parameters for (a) the trade informativeness $\{\lambda_k\}$; (b) the pricing error $\{s_k\}$; and (c) the standard deviations of innovations. The averages (across all stock-days) of the estimates are reported for the full sample and for each size tercile. The t-statistics, reported in brackets, are calculated using two-way clustered (stock and day) standard errors. Selected percentiles of the estimates in the full sample are reported.

and small stocks. The resulting sample sees 72,501 stock-day observations.

Table 1 provides summary statistics for the static structural parameters. Our focus is on trade informativeness λ_k. The estimates of the unconditional mean, $\bar{\lambda}$, suggests that on average, a surprise buy (sell) of size \$10,000 moves the efficient price up (down) by about $\bar{\lambda} = 2.76$ basis points (bps). Large stocks see much smaller price impact (0.31 bps) than small stocks (6.85 bps). The magnitude is comparable with the estimates by Brogaard, Hendershott, and Riordan (2014).

The innovation of our approach is that we allow and estimate the dynamics of λ_k. This can be seen via the estimate of the standard deviation of ε_k in Panel (c). On average, from one trade to another, the innovation in trade informativeness has a standard deviation of 0.61 bps/\$10,000, roughly a quarter of the long-run mean $\bar{\lambda}$. Such trade-to-trade fluctuation mean-reverts rather quickly, as the autoregressive coefficient appears small ($\alpha = 0.13$).

3.2 Intraday pattern of trade informativeness

To further examine the dynamics of trade informativeness, Table 2 presents the statistical moments of the filtered $\{\lambda_{k|k}\}$, our best guesses for the hidden $\{\lambda_k\}$. The statistical moments are first calculated for each intraday series of $\{\lambda_{k|k}\}$ and then aggregated across the stock-days. Unsurprisingly, the filtered $\lambda_{k|k}$ has unconditional mean very close to the long-run average $\bar{\lambda}$ (2.80 v.s. 2.76). The unconditional mean of $|\lambda_{k|k}|$ is about 3.54 bps/\$10,000. This suggests that conventional methods, treating λ as a constant, underestimate trade informativeness (price impact) by about 27% ($\approx 3.54/2.80 - 1$), because positive and negative λ_k might average out. Note that our structural model 4 does *not* require λ_k to be positive. A negative λ_k simply means that the trade's permanent price impact has opposite sign of the order flow innovation y_k^*. From the summary statistics we can see there are around 4% of such trades in the sample.

In terms of higher moments, the standard deviation of $\lambda_{k|k}$ averages at 0.73 bps/\$10,000.[5]

[5] The number 0.73 slightly defers from $\sigma_\varepsilon/\sqrt{1-\alpha^2}$ implied by the AR(1) structure of $\{\lambda_k\}$ because $\sigma_\varepsilon = 0.63$ and $\alpha = 0.13$ are the averages across the sample. Due to the nonlinearity, the evaluation at the sample averages in general differs from the average of evaluations across the sample.

	Unit	All	Large	Medium	Small	Percentile in the full sample 5%	25%	50%	75%	95%
$\mathbb{E}[\lambda_{k\|k}]$	bps/$10,000	2.80	0.32	1.23	6.95	0.12	0.37	0.88	2.99	11.35
		(11.74)	(19.82)	(15.65)	(13.46)					
$\mathbb{E}[\|\lambda_{k\|k}\|]$	bps/$10,000	3.54	0.35	1.38	9.03	0.13	0.40	0.97	3.43	15.12
		(11.03)	(20.10)	(14.86)	(12.66)					
$\sqrt{\text{var}_k\left[\lambda_{k\|k}\right]}$	bps/$10,000	0.73	0.16	0.50	2.30	0.02	0.09	0.24	0.59	2.63
		(13.10)	(9.92)	(20.26)	(12.50)					
skew$[\lambda_{k\|k}]$		1.60	1.91	1.50	1.38	-0.14	0.70	1.15	1.88	5.68
		(18.92)	(12.27)	(10.44)	(10.65)					
kurt$[\lambda_{k\|k}]$		13.47	15.05	12.46	12.88	4.74	6.92	9.09	13.18	42.08
		(19.35)	(11.25)	(10.94)	(11.84)					
$\mathbb{P}(\lambda_{k\|k} < 0)$		0.04	0.03	0.04	0.07	0.00	0.00	0.01	0.05	0.20
		(32.06)	(25.33)	(26.99)	(23.31)					
Count		72,501	24,264	24,504	23,733					

Table 2: Intraday dynamics of trade informativeness. This table summarizes the statistical moments of the filtered $\lambda_{k|k}$, filtered for the structural model (4) using the method proposed in Section 2.4. These statistics are calculated first for each intraday $\{\lambda_{k|k}\}$ series and then averaged across stock-days, for the full sample and for each size tercile. The t-statistics reported in brackets are calculated using two-way clustered (stock and day) standard errors. Selected percentiles in the full sample are also reported.

Notably, there is moderate positive skewness (1.60) and severe kurtosis (13.47). These statistical moments demonstrate the importance of *not* assuming normality, justifying our distribution-free estimation technique.

From the efficient price structure described in (4), $\Delta m_k = \bar{\lambda} y_k^* + (\lambda_k - \bar{\lambda}) y_k^* + \mu_k$. Replacing the hidden λ_k with out best guess, we can compute the contribution of the time-varying trade informativeness $(\lambda_{k|k} - \bar{\lambda}) y_k^*$ to the efficient price innovation as

$$\frac{\mathbb{E}\left[(\lambda_{k|k} - \bar{\lambda}) y_k^*\right]}{\bar{\lambda}^2 \mathbb{E}\left[(y_k^*)^2\right] + \mathbb{E}\left[(\lambda_{k|k} - \bar{\lambda}) y_k^*\right] + \text{var}[\mu_k]}.$$

We compute this ratio for each stock-day and it averages at a striking 24.1%. That is, about a quarter of the efficient price innovation can be attributed to the time-varying trade informativeness. The contribution is significant both statistically and economically.

We also explore the intraday dynamic pattern of trade informativeness. For each stock-day, we take snapshots of $\lambda_{k|k}$ at the end of each minute, obtaining a time-series of 390 observations (six and half trading hours). These intraday series are then averaged (equal weighted) across large, medium, and small stocks and plotted in Panel (a), (b), and (c) in Figure 1 respectively.

Despite the salient difference in magnitude, it can be seen that across all stocks trades are on average more informed at the early hours and it decays gradually over the trading day. The downward pattern differs from the well-known U-shapes of volume (Admati and Pfleiderer, 1988), the reverse J-shape of bid-ask spread (McInish and Wood, 1992), and the hump-shape of depth (Ahn, Bae, and Chan, 2001). The peak at market opening is consistent with the idea that uncertainty, news, and information asymmetry build up overnight. The fact that trade informativeness is at its lowest at the closing suggests that the lack of liquidity (large spread and the low depth) is due more to transitory price pressure like inventory cost than information. The pattern is also consistent with the prediction that trade informativeness follows a supermartingale. Baruch (2002) show that this happens because of informed investors' risk aversion, while Back and Baruch (2004) and Caldentey and Stacchetti (2010) yield similar predictions to random horizon (or discounting).

Because the magnitudes are very different across the terciles, we also construct a "standardized" average series across full sample and plot it in Panel (d) as a robustness check. Each intraday series of $\lambda_{k|k}$ is first scaled by its daily mean and then aggregated across all stock-days. This way, the graphed level of $\lambda_{k|k}$ can be read in percentage terms, relative to the unconditional mean (2.76 bps/\$10,000). The qualitative pattern remains and sees that in the early morning, trade informativeness is roughly about 120-130% of the unconditional mean. Toward the closing, trades become less informed, only about 85% the daily average.

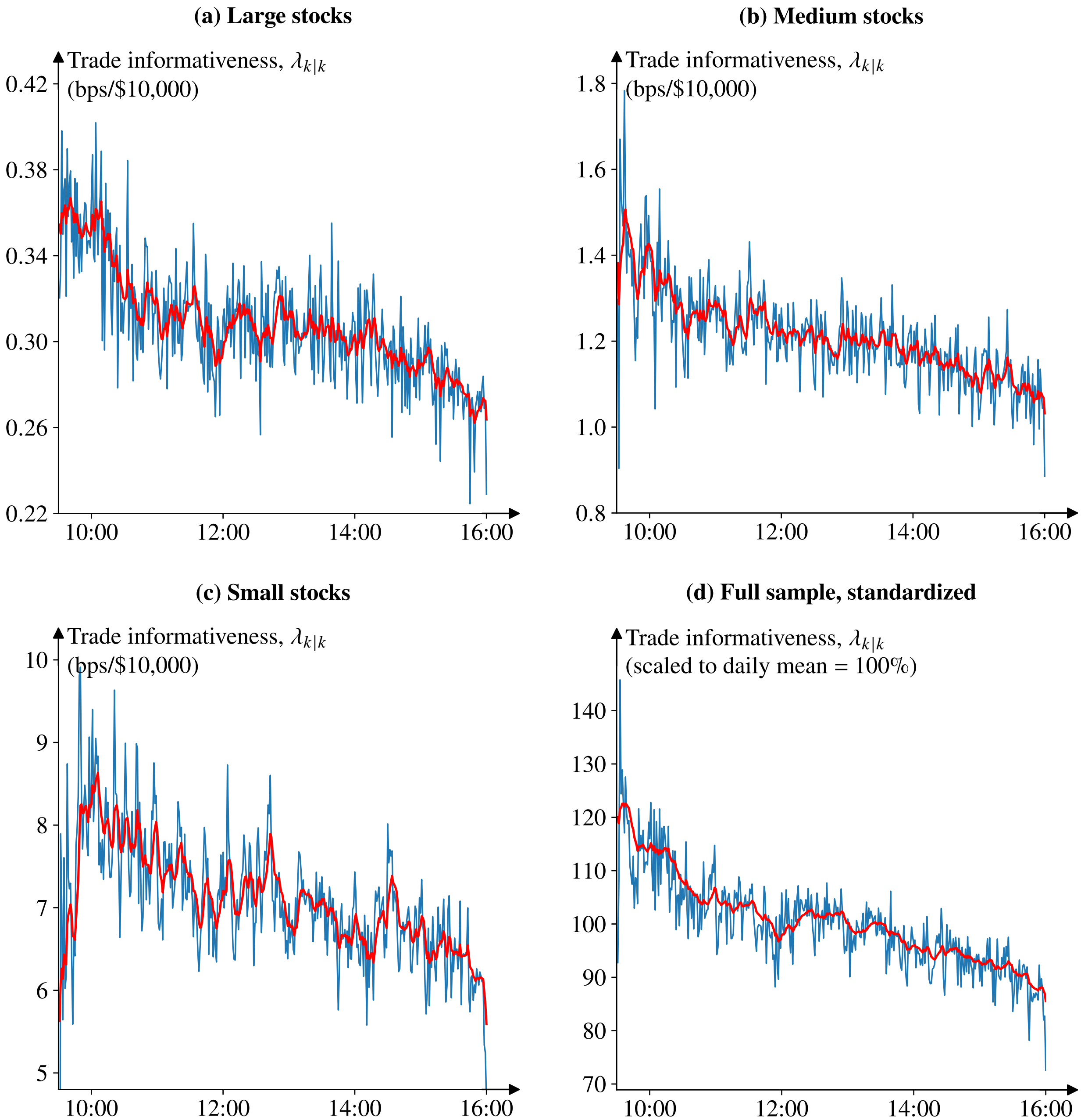

Figure 1: Intraday pattern of trade informativeness. This figure plots the intraday pattern of filtered trade informativeness, $\lambda_{k|k}$. Panel (a) through (c) show for large, medium, and small stocks, respectively. Panel (d) plots the average of the full sample by standardizing each intraday series to have its average at 100%. In all panels, the thin line is the cross-sample average snapshotted at every minute and the bold line is the smoothed trend.

3.3 Explaining future returns

Consider a surprise order flow y_k^* at time t_k. Under the structure (4), it has a permanent price impact of $\lambda_k y_k^*$ (in efficient price m_k) and a transitory price impact of ψy_k^* (in pricing error s_k). Over a relatively long period, the transitory price impact dies out as the pricing error mean reverts to zero, while the permanent price impact persists. The literature therefore often estimates the permanent price impact by regressing future price return on current order flow, for example:

$$\Delta p(t_k, t_k + w) \sim \bar{\lambda} y_k^* + \text{residual}$$

where $\Delta p(t_k, t_k + w)$ denotes the price return from time t_k to $t_k + w$ with a (large) window of size w. (A conventional choice is $w = 5$ minutes; e.g., Collin-Dufresne and Fos, 2015.) The regression coefficient of y_k^* reflects the *average* permanent price impact (trade informativeness), $\bar{\lambda}$.

Such regressions, however, often have relatively low explanatory power. To see this, using the structural model (4), the future price return can be decomposed into:

$$\Delta p(t_k, t_k + w) = \bar{\lambda} y_k^* + \underbrace{(\lambda_k - \bar{\lambda}) y_k^* + \mu_k + \Delta m(t_{k+1}, t_k + w)}_{\text{uncorrelated with } y_k^*} + \underbrace{\Delta s(t_k, t_k + w)}_{\text{asymptotically uncorrelated with } y_k^*}.$$

Only the first part, $\bar{\lambda} y_k^*$, is explained by the above simple linear regression. The other parts are (asymptotically, for large w) uncorrelated with y_k^*, thus serving as the "residual noise" in the above regression. They can have large variance, especially when the window w is large, thus lowering the explanatory power of the above regression.

Under the current approach, the filtered time-varying trade informativeness $\lambda_{k|k}$ could help reduce the "noise" in the above regression. Notably, part of the residual term, $(\lambda_k - \bar{\lambda}) y_k^*$, can be further explained, thanks to the filtered $\lambda_{k|k}$. To see this, we rewrite the above return as

$$\begin{aligned} \Delta p(t_k, t_k + w) =& \bar{\lambda} y_k^* + (\lambda_{k|k} - \bar{\lambda}) y_k^* \\ & + \underbrace{(\lambda_k - \lambda_{k|k}) y_k^* + \mu_k + \Delta m(t_{k+1}, t_k + w)}_{\text{uncorrelated with } y_k^*} + \underbrace{\Delta s(t_k, t_k + w)}_{\text{asymptotically uncorrelated with } y_k^*}. \end{aligned}$$

	Window size w for future return $\Delta p(t_k, t_k + w)$						
	1 sec.	5 sec.	10 sec.	30 sec.	1 min.	5 min.	10 min.
(a) $\boldsymbol{\Delta p(t_k, t_k + w) = b_1 \bar{\lambda} y_k^*}$ **+ residual**							
Average adjusted R^2, %	1.15	1.10	1.02	0.80	0.63	0.26	0.16
(b) $\boldsymbol{\Delta p(t_k, t_k + w) = b_1 \bar{\lambda} y_k^* + b_2 \cdot (\lambda_{k\|k} - \bar{\lambda}) y_k^*}$ **+ residual**							
Average adjusted R^2, %	6.65	6.04	5.45	4.10	3.14	1.18	0.70
Adjusted R^2 difference with **(a)**, %	5.50***	4.94***	4.43***	3.30***	2.51***	0.92***	0.54***
Proportion of adjusted R^2 from (b) > (a), %	99.2	99.2	99.1	99.1	99.0	97.6	94.9
Rejection rate at 1% for $b_2 = 0$, %	98.2	98.2	98.1	97.9	97.5	89.7	74.7

Table 3: Explanatory power of a current trade for future returns. Two models (a) and (b) are estimated for all stock-days and for all return window sizes w. Panel (a) reports the adjusted R^2, averaged across stock-days, of only using the unconditional average trade informativeness $\bar{\lambda}$ to explain future returns. Panel (b) adds to (a) the time-variation, as proxied by the filtered trade informativeness $\lambda_{k|k}$. The superscripts "***" in Panel (b) indicate that the corresponding improvement in R^2 is statistically significant over (a) under a one-sided t-test with 1% confidence. Also reported in (b) are the proportion of stock-days that see an increase in the R^2 and the rejection rates of the null hypothesis that the added explanatory variable is statistically insignificant.

That is, the time-variation in trade informativeness, $\lambda_k - \bar{\lambda}$, which is previously treated as some residual noise, can now be captured. To the extent that our filtered $\lambda_{k|k}$ is indeed tracking the true λ_k well, we hypothesize that the following regression

$$\Delta p(t_k, t_k + w) \sim \bar{\lambda} y_k^* + (\lambda_{k|k} - \bar{\lambda}) y_k^* + \text{residual}$$

will deliver significantly improved explanatory power.

For each stock-day, we regress the future return $\{\Delta p(t_k, t_k + w)\}$ on two different sets of explanatory variables: (a) $\{\bar{\lambda} y_k^*\}$; and (b) $\{\bar{\lambda} y_k^*, (\lambda_{k|k} - \bar{\lambda}) y_k^*\}$. Model (a) only accounts for the average trade informativeness $\bar{\lambda} y_k^*$. Model (b) adds to it the time-variation proxied by $(\lambda_{k|k} - \bar{\lambda}) y_k^*$. We then compare the explanatory power of the models for different window sizes w, ranging from 1 second up to 10 minutes.

Table 3 summarizes the results. For all window sizes, the explanatory power, adjusted R^2, of Model (b) is significantly higher over Model (a). The economic magnitude is also striking: After including the time-variation $\lambda_{k|k} - \bar{\lambda}$, the improvement of adjusted R^2 is roughly three-to-five times higher than the benchmark (a).[6] In fact, across the stock-days, almost all see an increase in the explanatory power. We also report the rejection rate of the null hypothesis that the added explanatory variable $(\lambda_{k|k} - \bar{\lambda})y_k^*$ is insignificant. It can be seen that this is rejected by most of the stock-days and the rejection rate only moderately declines as the forward-looking window extends beyond one minute.

As a robustness check, we redo the above regression exercise by using the predicted $\lambda_{k|k-1}$ (as opposed to the updated $\lambda_{k|k}$). The improvement in the explanatory power remains statistically significant. However, the economic magnitude of the improvement is less salient, which is expected as $\lambda_{k|k-1}$ accounts for less information than $\lambda_{k|k}$.

We conclude this exercise with two observations. First, the results confirm that there is indeed significant, economically meaningful time-variation in trade informativeness $\{\lambda_k\}$, for otherwise the added explanatory variable $(\lambda_{k|k} - \bar{\lambda})y_k^*$ would not be significant. Second, by the same argument, the filtered $\{\lambda_{k|k}\}$ does track the true, yet hidden, $\{\lambda_k\}$.

3.4 An event study: scheduled Fed announcements

This exercise studies the scheduled announcements by the U.S. Fed in 2014, as an example of how the filtered high-frequency trade informativeness can be used to address economic questions. All scheduled Fed announcements in 2014 are collected from Bloomberg. Conveniently, all of them were made at 14:00 Eastern Time and there were no other macro news announced at 14:00. Out of the 242 trading days, 58 had such announcements, including releases of Fed meeting minutes, budget statements, pace of Quantitative Easing, summary of economic projections, amendments, etc. These 58 days serve as the treatment group, while the rest as controls.

[6] In absolute terms, the adjusted R^2 is still relatively small as it only accounts for less than 10% of the total price return. This suggests that most of the price variation originates from activities like quote updates and revisions. See Brogaard, Hendershott, and Riordan (2016) for price discovery from such non-trade activities.

We ask whether trades become more or less informed approaching, upon, and after Fed announcements at 14:00. Ex ante, the pattern of λ_k is unclear to us. For example, assuming there is no information shock before 14:00, investors' trading should remain the same with or without a scheduled announcement. Yet, risk-averse uninformed investors who trade for liquidity reasons might temporarily refrain from trading, either because they are afraid of information leakage or because they are slow and unable to react to the news timely. This would imply that those who remain trading in minutes leading to the announcement are relatively more informed and hence have higher trade informativeness λ_k. Likewise, competing arguments for the pattern of λ_k after the announcement exist. On the one hand, the public news should provide a benchmark for reassessing the asset value, hence reducing information asymmetry and lowering λ_k. On the other, investors can draw different interpretations from the same news (see, e.g., Eyster, Rabin, and Vayanos, 2017; Vives and Yang, 2017), more trading on such information ensues, and trade informativeness λ_k increases. An empirical examination of the pattern of our filtered $\lambda_{k|k}$ can help disentangle the above effects and quantify the magnitude and the duration.

We construct per-second snapshots of the filtered $\lambda_{k|k}$ for all stock-days. Since our estimation is event-based, within each stock-day, we aggregate all trades and their informativeness within the same second and average them to get a snapshot for that second. If there is no trade in a specific second, we linearly interpolate from two nearest available snapshots. Then for each second around 14:00, we compute two cross-sectional averages, one for all stock-days with Fed announcement (treatment) and one for without (control). The two resulting time series are plotted in Figure 2.

Panel (a) shows a fifteen-minute window from 13:55 to 14:10. It can be seen that on no-announcement days, the average trade informativeness $\lambda_{k|k}$ (dashed line) hovers around the unconditional mean (about 2.76 bps/\$10,000) in this period. In contrast, on days with announcement (solid line), $\lambda_{k|k}$ starts to pick up at about 13:57, minutes before the scheduled announcement. At about 13:59, the difference with no-announcement days becomes statistically significant, as indicated by the boxes (95% confidence) and the circles (99% confidence). However, approaching the announcement at 14:00, $\lambda_{k|k}$ quickly reverts to the long-run average level and continuing the trend, drops below the no-announcement level, *after* 14:00. As zoomed in on by Panel (b), it is not until about 30 seconds after the announcement that $\lambda_{k|k}$ trends up again. It then remains significantly higher than no-announcement days for about another seven minutes and eventually dies down, becoming indistinguishable from no-announcement days after 14:07.

The pattern suggests that trades turn *more* informed roughly two-and-half minutes before the announcement. This could be driven by risk-averse uninformed investors' withdrawal from trading, leaving (relatively) more informed investors in the market. After the announcement, the pattern suggests that investors do derive different private signals and trade on those. The elevated information asymmetry ($\lambda_{k|k}$) lasts about six-and-halve minutes. In total, during the nine minutes from 13:57:30 to 14:06:30, trade informativeness is on average 0.18 bps/$10,000 higher on with-announcement days than on without. The difference is statistically significant and economically meaningful, as compared to an unconditional average of 2.76 bps/$10,000 ($0.17/2.76 \approx 6.5\%$).

A natural question is whether the elevated trade informativeness a result of liquidity reduction—wider bid-ask spread and lower order book depth—around the announcement. Indeed, if spread becomes wider and book thinner, a same trade could move price much more. However, such deterioration of liquidity would only have transitory effects on prices, i.e., through the pricing error s_k. In contrast, by construction, the filtered trade informativeness $\lambda_{k|k}$ captures the persistent effect of trades (i.e. the permanent price impact through the efficient price m_k). Conceptually, we would argue that trade informativeness λ_k and liquidity measures (spread and depth) capture (possibly correlated) different aspects of trading.

Nonetheless, as a robustness check that the patterns shown in Figure 2 are not driven by illiquidity, we perform the following regression for every second $t \in$[13:55,14:10] by pooling the second t-observations for all stock i and day d:

$$\lambda_{k|k}(i,d,t) \sim \text{Fed}(d) + \text{no-Fed}(d) + \text{spread}(i,d,t) + \text{depth}(i,d,t) + \text{stock fixed effects} + \text{residual}.$$

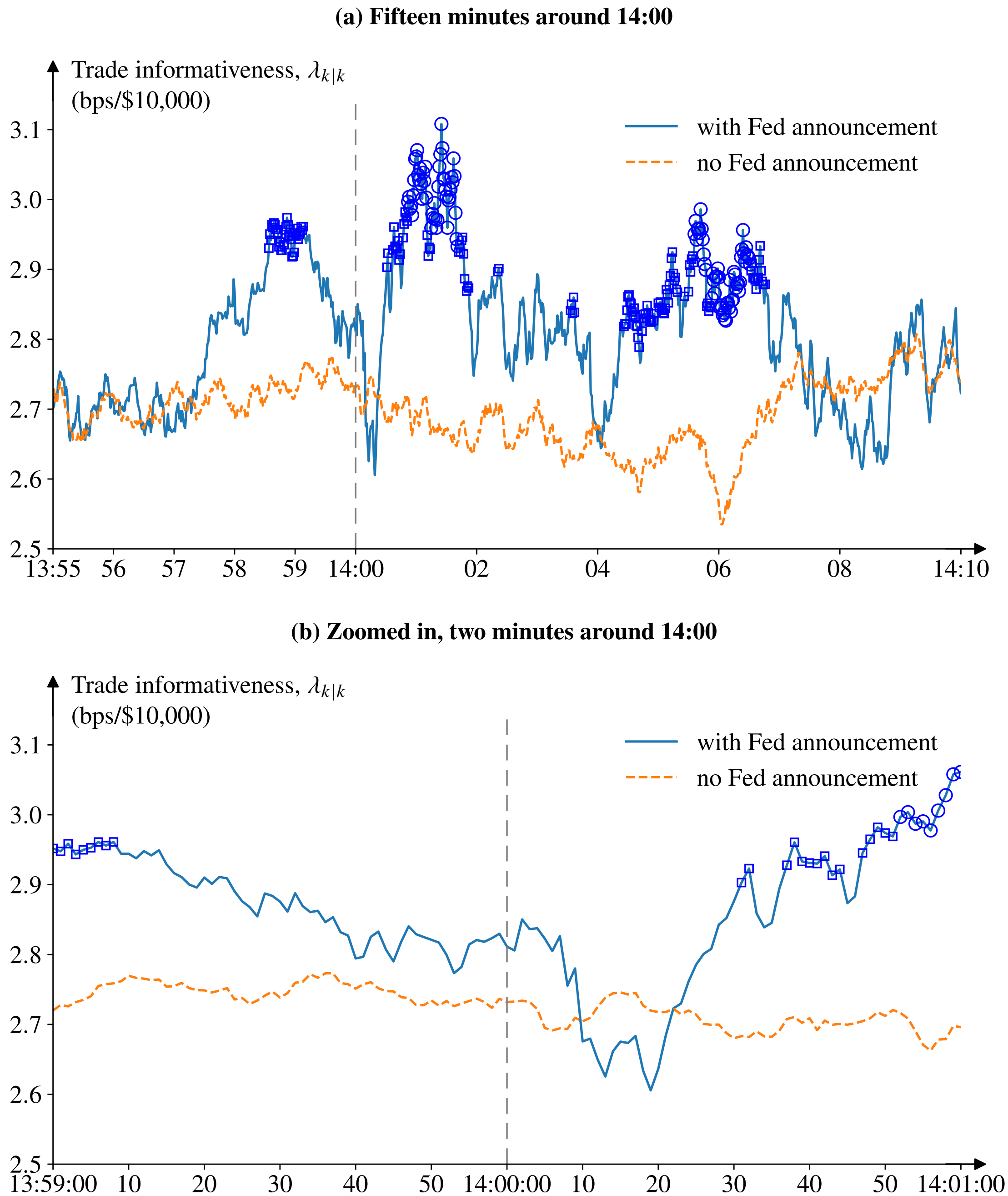

Figure 2: Trade informativeness around 14:00, with v.s. without Fed announcements. The first two panels plot the filtered trade informativeness $\lambda_{k|k}$ around 14:00 by averaging across all stock-days in 2014, separately for days with Fed announcement (blue-solid line) and for days without (orange-dashed line). The circled and boxed points indicate the difference between the two are statistically significant with 99% and 95% confidence, respectively. Panel (a) shows a fifteen-minute window, while Panel (b) zooms in on the two minutes around 14:00. Panel (c), on the next page, shows the difference series, after controlling for liquidity measures.

(c) Controlling for liquidity

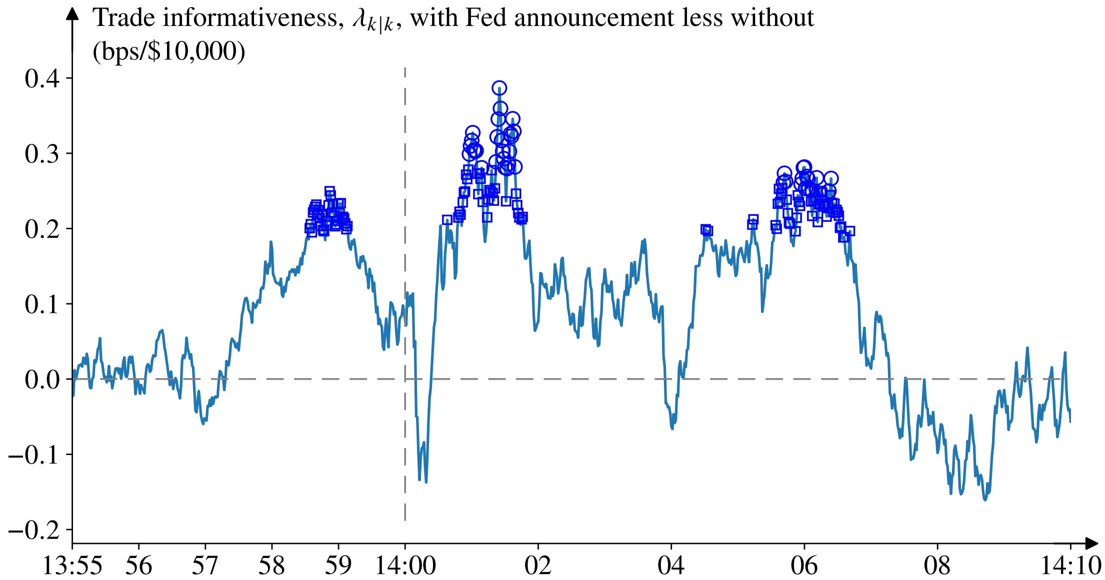

Figure 2: Trade informativeness around 14:00, with v.s. without Fed announcements, continued.

The dummy variables Fed(d) and no-Fed(d) capture whether day d has Fed announcement scheduled at 14:00. The spread(i, d, t) and depth(i, d, t) are stock i's bid-ask spread and depth, respectively, at the end of second t on day d. We measure spread in basis points (relative to the midquote) and depth in the total number of shares available at National Best Bid and Offer (NBBO). Finally, we also control for stock fixed effects. The above regression is separately performed for every second in the fifteen-minute window. For each regression, the difference between the coefficients for Fed(d) and no-Fed(d) shows the effect of the treatment. This resulting difference series is plotted in Panel (c) of Figure 2. We see a pattern similar to the difference between the blue-solid line and the orange-dashed line in Panel (a) of Figure 2, both in terms of magnitude and statistical significance. This evidence strengthens our argument that trade informativeness stands for a different aspect from market (il)liquidity.

3.5 Informativeness v.s. aggressiveness and informed investors' patience

Every trade has a buyer and a seller. In modern limit order markets, the aggressive one initiates the trade with a market order, while the other has his limit order passively hit. Such aggressiveness is often determined by algorithms like Lee and Ready (1991). Conventionally, $y_k > 0$ (< 0) indicates an aggressive buy (sell).

There is a separate aspect: the (relatively more) informed investor of the two. It is possible, but not necessary, that the aggressive one happens also to be the (relatively more) informed one. Put alternatively, the limit order trader can very well have (relatively more) information than the market order trader. Equipped with the structural model (4) and with the estimation technique (GaMM), we use the filtered $\lambda_{k|k}$ to help identify the (relatively more) informed side for *each and every* trade.[7] To do so, recall that the efficient price innovation Δm_k has a trade-related component. For notation simplicity, we write our best guess of this term as

$$\eta_k := \lambda_{k|k} y_k^*.$$

This is the permanent price impact of the k-th trade. If $\eta_k > 0$ is positive, we can say it is the buyer who has pushed the efficient price up and is the (relatively more) informed; and the seller if $\eta_k < 0$.

Clearly, sign$[\eta_k]$ needs not to be the same as sign$[y_k]$—the former distinguishes whether the trade is an informed buy or sell, while the latter indicates an aggressive buy or sell. Our structural model allows the separation of the two. In particular, if sign$[\eta_k] \neq$ sign$[y_k]$, it follows that the passive side—the limit order trader—is the (relatively more) informed. For example, if the k-th trade sees $\eta_k > 0$ and yet $y_k < 0$, it means this market *sell* order drives the efficient price *up*, suggesting the corresponding limit buy order is (relatively more) informed.

Theory of limit order models suggest that informed investors optimally choose between limit orders and market orders by trading off execution and waiting costs (see, e.g., Parlour and Seppi,

[7] Under the canonical framework discussed in Section 2.1, trade informativeness is only an *average* statement $\bar{\lambda}$ for all trades. For an individual trade, however, it is not possible to determine whether the buyer or the seller is (relatively more) informed. See more detailed discussion in Section 2.3.

					Percentile in the full sample				
Empirical frequency of	All	Large	Medium	Small	5%	25%	50%	75%	95%
(a) $\text{sign}[\eta_k] \neq \text{sign}[y_k]$	0.14	0.13	0.13	0.17	0.05	0.09	0.12	0.17	0.30
(b) $\text{sign}[\eta_k] \neq \text{sign}[y_k] \mid \text{sign}\left[y_k^*\right] = \text{sign}[y_k]$	0.05	0.03	0.04	0.07	0.00	0.00	0.01	0.05	0.21
Count	72,501	24,264	24,504	23,733					

Table 4: Informed investors' patience. This table reports two empirical frequencies related to informed investor' patience. The two are first calculated for each stock-day and then averaged across the full sample and respective size terciles. Selected percentiles in the full sample are also reported.

2008 for a review). All else equal, if informed investors mostly use limit orders, we infer they are more patient. We therefore use the empirical frequency

$$\mathbb{P}(\text{sign}[\eta_k] \neq \text{sign}[y_k]) \tag{12}$$

as a proxy for *informed investors' patience*; that is, the frequency of trades where the limit order side is (relatively more) informed. Row (a) of Table 4 reports the statistics. Overall, there are about 14% of the trades that are more informed on the limit order side. This ratio is relatively stable across stock size terciles, ranging from 13% to 17%. Notably, a most patient informed trading day sees more than one-third of such trades (the 95-percentile).

To the best of our knowledge, this is the first structural estimator designed to capture informed investor' patience. This measure, using only the price dynamics and the order flow information, can be coupled with other empirical observations to help strengthen the understanding of informed trading. As an illustration, we consider firms' earnings announcements and examine informed investors' (im)patience with the proposed measure.

We collect from Bloomberg all earnings announcements (quarterly and annual) of the sampled firms in 2014 and find 1,187 such events. With the exception of ten announcements (which we discard from the analysis), all announcements were made off trading hours. We then zoom in on the ten trading days around each announcements, five before and five after. For each stock-day in

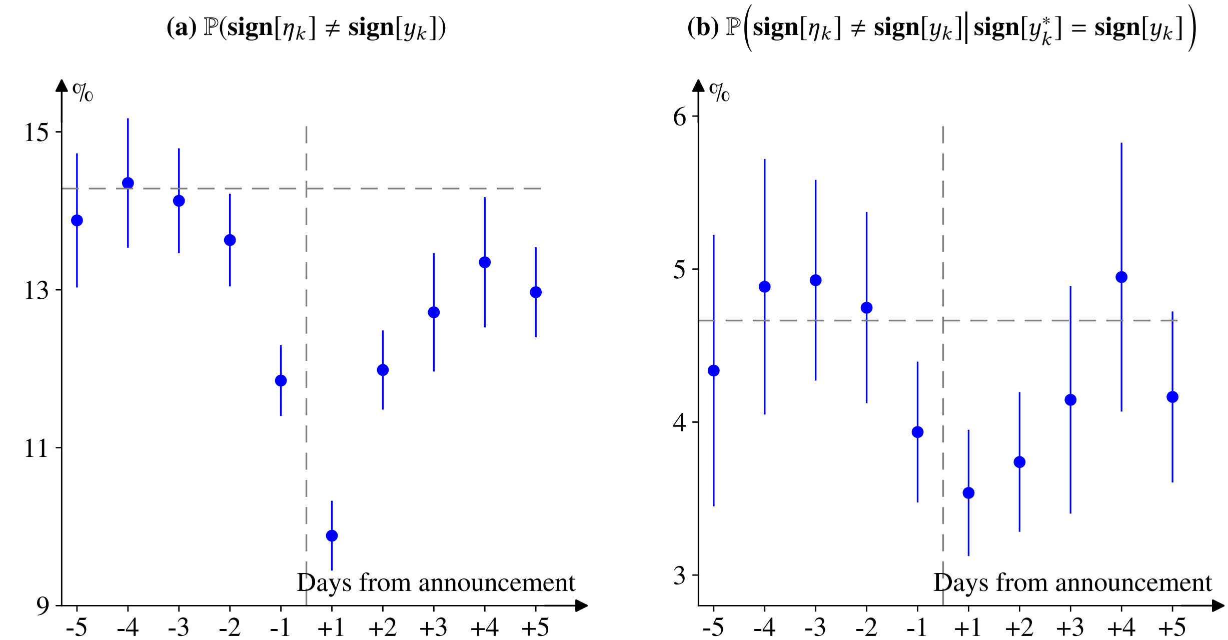

Figure 3: Informed investors' patience around earnings announcements. This figure shows informed investors' patience around earnings announcements. Panel (a) plots an unconditional version defined in equation (12). Panel (b) plots a conditional version defined in equation (13). Each dot indicates the averages and the caps show the 95% confidence bounds on the corresponding day. The horizontal dashed line indicates the average in the full sample (Table 4). The vertical line separates the days before and after the announcement.

this subsample, we then compute the proxy for informed investor' patience as in equation (12).

Panel (a) of Figure 3 demonstrates the pattern by averaging the patience measure across all stock-days in the subsample. Informed trading becomes less and less patient as the earnings day approaches. That is, the proportion of (relatively more) informed limit orders drops. On the day right after the announcement, across the sample, patient informed trades only account for less than 10% of total trades, a one-third reduction from the unconditional average (Row (a) of Table 4). The drop in informed investors' patience is also statistically significant, judged by the 95% confidence bounds.

The evidence supports our proposed structural measure for patience: Facing a scheduled public announcement, informed investors become more impatient and trade more aggressively on their private information, as otherwise the value of their private information will be eroded by the announcement. It is seen that informed investors are least patient right after the announcement. To the extent that information processing is costly and takes time, such extreme impatience can be attributed to those who have parsed the information sooner than others. Given the public information, the others will soon come to the same result. Restricting to market orders to trade on the early-processed information is profitable in such scenarios.

We perform two robustness checks. First, we control for the potential heterogeneity across stocks by standardizing the patience measure (12)—taking out the stock-specific effect and then scaling by the firm-specific standard deviation. We then aggregate the standardized series and redo Panel (a) of Figure 3. The resulting pattern is qualitatively the same as shown.

Second, recall that the order flow innovation is defined as the residual of an autoregressive structure: $A(L)y_k = y_k^*$. Around earnings announcement days, the average size of y_k^* might significantly increase because investors rebalance their portfolio (non-informational) or because information asymmetry intensifies. (However, neither reasons imply informed investors' patience might change.) As y_k^* grows larger, it dominates in y_k and $\text{sign}[y_k^*] = \text{sign}[y_k]$ becomes very likely. This will mechanically drive up our unconditional measure (12), because $\eta_k = \lambda_{k|k} y_k^*$ and because $\lambda_{k|k}$ is mostly positive (Table 1). To control for the effect of order flow innovation size, we consider a conditional version of the patience measure:

$$\begin{aligned} &\mathbb{P}\big(\text{sign}[\eta_k] \neq \text{sign}[y_k] \big| \, \text{sign}\big[y_k^*\big] = \text{sign}[y_k]\big) \\ =&\mathbb{P}\big(\lambda_{k|k} < 0 \big| \, \text{sign}\big[y_k^*\big] = \text{sign}[y_k]\big); \end{aligned} \tag{13}$$

that is, among all trades that have same signs of order flow y_k and its innovation y_k^*, the empirical frequency to see the (relatively more) informed side of the trade differ from the aggressive side. Across all stock-days, this conditional patience is at about 5%, as reported in Row (b) of Table 4. Panel (b) of Figure 3 shows the trend around earnings announcement days. The pattern is consistent with the unconditional patience measure, lending support to the robustness of the result.

4 Concluding remarks

This paper proposes a structural model that permits time-varying trade informativeness and develop a new filtering technique for empirical applications. We argue that the dynamics of trade informativeness will serve as a useful empirical tool. First, trade informativeness is time-varying in nature and, as such, empirically identified dynamic patterns can help assess theoretical models. Second, without a dynamic characterization of trade informativeness, it is hard to examine informed investors' high-frequency trading behavior, both at normal times and upon intraday events like un/scheduled announcements and news. Third, dynamic trade informativeness allows the separation of the (relatively more) informed side from the aggressive side of a trade. Fourth, as trade informativeness and aggressiveness no longer needs to be aligned at all times, a structural measure of informed investors' patience readily follows. We illustrate these uses with examples.

We highlight some notable limitations to the proposed structural model (4). First, trades' transitory price impact ψ are not modeled as time-varying. This can be generalized using the same approach as we have done to trade informativeness λ_k. Second, order flow dynamics $\{y_k, y_k^*\}$ are estimated separately from the system, a convention we inherit from the literature featuring the state space approach. This front can be pushed forward by endogenizing the order flows in a richer state space. Finally, the innovations like μ_k and ν_k are assumed to have constant volatility over time. It is well-known, however, that trade data features time-varying heteroskedasticity. This extension will involve modeling the innovations as GARCH series.

While relaxing these limitations can be challenging, we believe properly accounting for these realistic aspects of trading data will provide better understanding of the nature of financial securities trading. The flexibility of the novel distribution-freee estimation technique can help deal with the complexity in such more realistic state spaces. We leave these extensions for future research.

Appendix

A Estimating the variances

Section 2.4 mainly discusses the filtering of the dynamic $\{\lambda_k\}$. In this appendix, we exploit the second moments to estimate the variances of the innovations: σ_μ^2, σ_ν^2, and σ_ε^2. Construct the following residual series:

$$\xi_k := (1-\phi L)\big(\Delta p_k - \hat{y}_k - \lambda_{k|k-1} y_k^*\big) = (1-\phi L)\Big(\mu_k + \hat{\lambda}_k y_k^*\Big) + (1-L)\nu_k,$$

where $\hat{\lambda}_k := \lambda_k - \lambda_{k|k-1}$ is the one-step ahead prediction error. The autocovariance generating function applied to ξ_k yields

$$\mathrm{cov}[\xi_k, \xi_{k-j}] = \begin{cases} (1+\phi^2)\Big(\sigma_\mu^2 + (y_k^*)^2 \mathrm{var}\big[\hat{\lambda}_k\big]\Big) + 2\sigma_\nu^2, & \text{if } j = 0; \\ -\phi \cdot \Big(\sigma_\mu^2 + (y_k^*)^2 \mathrm{var}\big[\hat{\lambda}_k\big]\Big) - \sigma_\nu^2, & \text{if } |j| = 1; \\ 0, & \text{if } |j| \geq 2. \end{cases}$$

The above moment conditions can identify only two terms: σ_ν^2 and $\sigma_\mu^2 + (y_k^*)^2 \mathrm{var}[\tilde{\lambda}_k]$. It remains to separate σ_μ^2 from $\mathrm{var}[\tilde{\lambda}_k]$ in the second term. To do so, consider $j = 0$ in the above. Since $\{y_k^*\}$ are observed, we have the following two more moment conditions:

$$\mathbb{E}\Big[\xi_k^2 - 2\sigma_\nu^2 - \Big(1+\phi^2\Big)\Big(\sigma_\mu^2 + (y_k^*)^2 \mathrm{var}\big[\hat{\lambda}_k\big]\Big)\Big] = 0$$
$$\mathbb{E}\Big[\Big(\xi_k^2 - 2\sigma_\nu^2 - \Big(1+\phi^2\Big)\Big(\sigma_\mu^2 + (y_k^*)^2 \mathrm{var}\big[\hat{\lambda}_k\big]\Big)\Big)(y_k^*)^2\Big] = 0$$

which essentially pin down σ_μ^2 and $\mathrm{var}\big[\hat{\lambda}_k\big]$ via the regression of $\xi_k^2 \sim \text{constant} + (y_k^*)^2$. Finally, we note that $\mathrm{var}\big[\hat{\lambda}_k\big]$ can be expressed as:

$$\begin{aligned} \mathrm{var}\big[\hat{\lambda}_k\big] &= \mathrm{var}\big[\lambda_k - \lambda_{k|k-1}\big] = \mathbb{E}\Big[\big(\lambda_k - \lambda_{k|k-1}\big)^2\Big] + \big(\mathbb{E}\big[\lambda_k - \lambda_{k|k-1}\big]\big)^2 = \mathbb{E}\Big[\big(\lambda_k - \lambda_{k|k-1}\big)^2\Big] \\ &= \mathbb{E}\big[\lambda_k^2\big] - 2\mathbb{E}\big[\lambda_k \lambda_{k|k-1}\big] + \mathbb{E}\Big[\lambda_{k|k-1}^2\Big] = \mathbb{E}\big[\lambda_k^2\big] - \mathbb{E}\Big[\lambda_{k|k-1}^2\Big] \\ &= \mathbb{E}\big[\lambda_k^2\big] - \mathrm{var}\big[\lambda_{k|k-1}\big] - \big(\mathbb{E}\big[\lambda_{k|k-1}\big]\big)^2 = \mathrm{var}[\lambda_k] - \mathrm{var}\big[\lambda_{k|k-1}\big] = \frac{\sigma_\varepsilon^2}{1-\alpha^2} - \mathrm{var}\big[\lambda_{k|k-1}\big], \end{aligned}$$

where the last equation follows the AR(1) structure of λ_k. We therefore augment the above moment condition with

$$\mathbb{E}\Big[(1-\alpha^2)\Big(\text{var}\big[\hat{\lambda}_k\big] + \big(\lambda_{k|k-1} - \bar{\lambda}\big)^2\Big) - \sigma_\varepsilon^2\Big] = 0,$$

which completes the GMM conditions needed to identify the innovation variances.

B Comparing GaMM with Kalman filter

We perform a horserace between our filtering method (GaMM, as first proposed by Creal et al., 2016) and Kalman filter by simulation. The results demonstrate the superiority of GaMM over Kalman when the model 4 has non-normal innovations or when the model is misspecified. Three specifications are considered:

(a) The model is correctly specified as in (4). The innovations $\{\mu_k, \nu_k, \varepsilon_k\}$ follow an i.i.d. joint normal distribution.

(b) The model is correctly specified as in (4). The innovations $\{\mu_k, \nu_k\}$ follow an i.i.d. bivariate t-distribution (heavy tails), while $\{\varepsilon_k\}$ follows an i.i.d. skewed t-distribution.

(c) The model is misspecified in that λ_k follows an AR(2) process, as opposed to the AR(1) process assumed in (4), while the other structure in (4) hold. The innovations $\{\mu_k, \nu_k, \varepsilon_k\}$ follow the same i.i.d. joint normal distribution as in (a).

In all cases, the static structural parameters are set to the same as the average estimates reported in Table 1: $\psi = -1.90$, $\phi = 0.50$, $\bar{\lambda} = 2.76$, and $\alpha = 0.13$. The innovations have zero mean and the common variances: $\sigma_\mu^2 = 2.60$, $\sigma_\nu^2 = 0.49$, and $\sigma_\varepsilon^2 = 0.40$; and they are pair-wise uncorrelated. In (b), the bivariate t-distribution for $\{\mu_k, \nu_k\}$ has a degree of freedom of 3; and the skewed t-distribution used for $\{\varepsilon_k\}$ has a scale parameter 0.40, a skewness parameter 20, and a degree of freedom of 3 (see Fernández and Steel, 1998). In (c), the AR(1) coefficient is the same $\alpha = 0.13$ and the AR(2) coefficient is set to 0.7.

	Static structural parameters (true value)								
	ψ	ϕ	$\bar{\lambda}$	α	σ_μ^2	σ_ν^2	σ_ε^2	$\{\lambda_k\}$ process	
	(-1.9)	(0.5)	(2.76)	(0.13)	(2.60)	(0.49)	(0.40)	RMSE	$\text{cor}(\lambda_{k\mid k}, \lambda_k)$
(a) Model correctly specified, normal innovations									
GaMM	-1.91	0.50	2.77	0.14	2.73	0.48	0.30	0.74	0.33
Kalman filter	-1.91	0.50	2.77	0.06	2.59	0.50	0.41	0.61	0.35
GaMM less Kalman								0.13	-0.02
								(19.64)	(-14.26)
(b) Model correctly specified, heavy tails in innovations, $\{\varepsilon_k\}$ skewed									
GaMM	-1.90	0.50	2.76	0.11	7.52	1.44	0.38	0.56	0.18
Kalman filter	-1.90	0.50	2.76	0.02	6.81	1.75	0.25	0.57	0.13
GaMM less Kalman								-0.01^{***}	0.05^{***}
								(-3.44)	(17.13)
(c) Model misspecified with true $\{\lambda_k\}$ following AR(2), normal innovations									
GaMM	-1.91	0.50	2.77	0.95	2.85	0.48	0.07	0.84	0.53
Kalman filter	-1.91	0.50	2.77	0.87	2.85	0.50	0.19	0.86	0.52
GaMM less Kalman								-0.02^{***}	0.01^{***}
								(-7.67)	(3.00)

Table B.1: A horserace between GaMM and Kalman filter. This table presents the simulation results for a horserace between GaMM and Kalman filter. Three different data generating processes are considered, each with 1,000 samples of 5,000 observations. For the static structural parameters $\{\psi, \phi, \bar{\lambda}, \alpha, \sigma_\mu^2, \sigma_\nu^2, \sigma_\varepsilon^2\}$, the averages of the point estimates across the 1,000 samples are reported. For the hidden process $\{\lambda_k\}$, the root-mean-squared error (RMSE) of the filtered $\{\lambda_{k|k}\}$ and the correlation between them, again averaged across the 1,000 samples, is reported, together with the difference between GaMM and Kalman filter. The superscripted "***" stands for 1% significance for one-sided t-tests with the null of GaMM outperforming Kalman filter.

For each of the three cases, we simulate 1,000 samples, each with 5,000 observations. We run the horserace by "pretending" the specification (4) is correct. We follow the GaMM steps described in Section 2.4 and in Appendix A to estimate all parameters and obtain the filtered series $\{\lambda_{k|k}\}$. Assuming the innovations are normally distributed, we also estimate all parameters by maximum likelihood and obtain the $\{\lambda_{k|k}\}$ process by Kalman filter.

Table B.1 presents the simulation results. With some exceptions, the static parameters are largely accurately estimated. Notably, even in (a), when the model is correctly specified and all innovations are normally distributed, maximum likelihood following Kalman filter seems to under estimate α. (This bias seems to persist in other model specifications.) In terms of filtering the hidden $\{\lambda_k\}$ process, Kalman filter outperforms GaMM only in (a), when the model is correctly specified *and* when the innovations follow joint normal distribution. This is unsurprising as Kalman filter is the optimal filter in this benchmark case. However, when either of the two requirements fail, GaMM can outperform Kalman filter, as seen in Panel (b) and (c). In both latter cases, the root-mean-squared error of GaMM filtered $\{\lambda_{k|k}\}$ is statistically significantly smaller than that under Kalman filter. In terms of magnitude, the reduced RMSE amounts to about 1-2% (1/57 and 2/86). We opt for the GaMM filtering approach based on the above simulation evidence.

Despite the simulation results above, we still implemented Kalman filter in the three hundred sampled stocks in 2014. However, the overall convergence rate is only about 85%. We suspect this has something to do with the non-normal distribution of $\{\lambda_k\}$ (positive skewness and the high kurtosis; see Table 1). The high numerical convergence rate of GaMM (around 97%) is another reason we favor this approach.

References

Admati, Anat R. and Paul Pfleiderer. 1988. "A Theory of Intraday Trading Patterns: Volume and Price Variability." *The Review of Financial Studies* 1:3–40.

Ahn, Hee-Joon, Kee-Hong Bae, and Kalok Chan. 2001. "Limit Orders, Depth, and Volatility: Evidence from the Stock Exchange of Hong Kong." *The Journal of Finance* 56 (2):767–788.

Back, Kerry and Shmuel Baruch. 2004. "Information in securities markets: Kyle meets Glosten and Milgrom." *Econometrica* 72 (2):433–465.

Back, Kerry, C. Hentry Cao, and Gregory A. Willard. 2000. "Imperfect Competition among Informed Traders." *The Jounral of Finance* 55 (5):2117–2155.

Back, Kerry, Pierre Collin-Dufresne, Vyacheslav Fos, Tao Li, and Alexander Ljungqvist. 2016. "Activism, Strategic Trading, and Liquidity." Working paper.

Back, Kerry, Kevin Crotty, and Tao Li. 2015. "Identifying Information Asymmetry with Endogenous Informed Orders." Working paper.

Back, Kerry and Hal Pedersen. 1998. "Long-lived Information and Intraday Patterns." *Journal of Financial Markets* 1:385–402.

Banerjee, Snehal and Bradyn Breon-Drish. 2017. "Dynamic Information Acquisition and Strategic Trading." Working paper.

Baruch, Shmuel. 2002. "Insider Trading and Risk Aversion." *Journal of Financial Markets* 5:451–464.

Brogaard, Jonathan, Terrance Hendershott, and Ryan Riordan. 2016. "Price Discovery Without Trading: Evidence from Limit Orders." Working paper.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan. 2014. "High Frequency Trading and Price Discovery." *The Review of Financial Studies* 27 (8):2267–2306.

Caldentey, Renè and Ennio Stacchetti. 2010. "Insider Trading with a Random Deadline." *Econometrica* 78 (1):245–283.

Collin-Dufresne, Pierre and Vyacheslav Fos. 2015. "Do Prices Reveal the Presence of Informed Trading?" *The Journal of Finance* 70 (4):1555–1582.

———. 2016. "Insider Trading, Stochastic Liquidity, and Equilibrium Prices." *Econometrica* 84 (4):1441–1475.

Copeland, Thomas E. and Dan Galai. 1983. "Information Effects on the Bid-ask Spread." *The Journal of Finance* 38 (5):1457–1469.

Creal, Drew, Siem Jan Koopman, and André Lucas. 2013. "Generalized Autoregressive Score Models with Applications." *Journal of Applied Econometrics* 28:777–795.

Creal, Drew, Siem Jan Koopman, André Lucas, and Marcin Zamojski. 2016. "Generalized Autoregressive Method of Moments." Working paper, Tinbergen Institute.

Durbin, James and Siem Jan Koopman. 2012. *Time Series Analysis by State Space Methods*. No. 38 in Oxford Statistical Science Series. Oxford, U.K.: Oxford University Press, 2 ed.

Ellis, Katrina, Roni Michaely, and Maureen O'Hara. 2000. "The Accuracy of Trade Classification Rules: Evidence from Nasdaq." *Journal of Financical and Quantitative Analysis* 35 (4):529–551.

Eyster, Erik, Matthew Rabin, and Dimitri Vayanos. 2017. "Financial Markets where Traders Neglect the Informational Content of Prices." *The Journal of Finance* Forthcoming.

Fernández, Carmen and Mark FJ Steel. 1998. "On Bayesian modeling of fat tails and skewness." *Journal of the American Statistical Association* 93 (441):359–371.

Fleming, Michael J., Bruce Mizrach, and Giang Nguyen. 2017. "The Microstructure of a U.S. Treasury ECN: The BrokerTec Platform." *Journal of Financial Markets* Forthcoming.

Foster, F. Douglas and S. Viswanathan. 1996. "Strategic Trading when Agents Forecast the Forecasts of Others." *The Journal of Finance* 51 (4):1437–1478.

Foucault, Thierry. 1999. "Order Flow Composition and Trading Costs in a Dynamic Limit Order Market." *Journal of Financial Markets* 2 (2):99–134.

Glosten, Lawrence R. and Paul R. Milgrom. 1985. "Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Agents." *Journal of Financial Economics* 42 (1):71–100.

Hansen, Lars Peter. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50 (4):1029–1054.

Hasbrouck, Joel. 2007. *Empirical Market Microstructure: the Institutions, Economics, and Econometrics of Securities Trading*. Oxford University Press, New York.

Hendershott, Terrence and Albert J. Menkveld. 2014. "Price pressures." *Journal of Financial Economics* 114 (3):405–423.

Holden, Craig W. and Stacey Jacobsen. 2014. "Liquidity Measurement Problems in Fast, Competitive Markets: Expensive and Cheap Solutions." *The Journal of Finance* 69 (4):1747–1785.

Holden, Craig W. and Avanidhar Subrahmanyam. 1992. "Long-Lived Private Information and Imperfect Competition." *The Journal of Finance* 48 (1):247–270.

Kaniel, Ron and Hong Liu. 2006. "So What Orders Do Informed Traders Use?" *Journal of Business* 79 (4):1867–1913.

Kyle, Albert S. 1985. "Continuous Auctions and Insider Trading." *Econometrica* 53 (6):1315–1336.

Lee, Charles M. and Mark J. Ready. 1991. "Inferring Trade Direction from Intraday Data." *The Journal of Finance* 46 (2):733–746.

Lee, Charles M. C., Belinda Mucklow, and Mark J. Ready. 1993. "Spreads, Depths, and the Impact of Earnings Information: An Intraday Analysis." *The Review of Financial Studies* 6 (2):345–374.

McInish, Thomas H. and Robert A. Wood. 1992. "An Analysis of Intraday Patterns in Bid/Ask Spreads for NYSE Stocks." *The Journal of Finance* 47 (2):753–764.

Menkveld, Albert J. 2013. "High Frequency Trading and the *New Market* Makers." *Journal of Financial Markets* 16 (4):712–740.

Menkveld, Albert J., Siem Jan Koopman, and André Lucas. 2007. "Modelling Round-the-Clock Price Discovery for Cross-Listed Stocks using State Space Methods." *Journal of Business & Economic Statistics* 25 (2):213–225.

O'Hara, Maureen. 2015. "High Frequency Market Microstructure." *Journal of Financial Economics* 116 (2):257–270.

Parlour, Christine A. and Duane J. Seppi. 2008. "Limit Order Markets: A Survey." In *Handbook of Financial Intermediation and Banking*, edited by A.W.A. Boot and A.V. Thakor. Amsterdam, Netherlands: Elsevier Publishing.

Vives, Xavier and Liyan Yang. 2017. "Costly Interpretation of Asset Prices." Working paper.

Wang, Yenan and Ming Yang. 2017. "Insider Trading When There May Not Be an Insider." Working paper.

Yueshen, Bart Zhou. 2016. "Uncertain Market Making." Working paper.