Estimating Price Impact and its Components: Evidence from HFT Around Earnings Announcements

Thomas J. George and Moazzam A. Khoja*

This draft: March 25, 2026

First draft: November, 29, 2023

Abstract

A simple approach to estimating price impact and its components is provided that extends easily to multiple classifications of trades. Temporary and permanent components for all classifications are estimated jointly using OLS with only transaction prices and trades as inputs. Hypothesis testing is conducted using the delta method. The state space approach used in prior work fails on transaction price data due to violations of its Normality assumptions. The OLS approach avoids these problems entirely and can be used with classifications that exist in regulatory or institutional data to test hypotheses about price impacts for transactions that are classified by trader ID, trader type, routing path, or clearing mechanism. It can also be used without trade classifications to estimate measures of illiquidity for asset pricing research. We illustrate the approach in the former setting by analyzing price impacts around earnings announcements for the various possible counterparty pairings between high-frequency traders (HFTs) and non-HFTs identified in a special sample provided by Nasdaq. Permanent components are larger when an HFT is the liquidity provider, and temporary components are smaller (and negative) when an HFT is the liquidity demander. These features of price impact are related in the cross-section, suggesting that HFTs secure more attractive pricing on their liquidity demanding trades after absorbing more informative trades as liquidity providers. The OLS approach is fast to execute on millisecond transactions data and performs well in the rich, eight-parameter HFT/non-HFT setting.

*Acknowledgements: We are very grateful to the Nasdaq Office of the Chief Economist for providing data and answering questions, and to Terry Hendershott, Albert Menkveld, and seminar participants at the University of Houston and the University of Texas at San Antonio for helpful comments and suggestions. Any errors are the responsibility of the authors. George is from the C. T. Bauer College of Business, University of Houston. Khoja is from the Carlos Alvarez College of Business, University of Texas at San Antonio. Email: tom-george@uh.edu and moazzam.khoja@utsa.edu. The authors have not received external research support and do not have financial relationships with entities external to their universities. The authors have no conflicts of interest to report.

1 Introduction

The price impact of a transaction is an important cost of trading securities, and the components of price impact are often regarded as helpful metrics for understanding how securities markets perform. In the market microstructure literature, the permanent component of price impact captures the severity of adverse selection and the contribution of trading to price discovery. The temporary component captures the compensation liquidity demanders pay to cover liquidity providers' direct costs of market making, including managing an inventory, any profit or rents earned by the liquidity provider, and frictions in the trading process such as discreteness in prices. Understanding price impact and its components is also important to industry professionals. Traders with discretion between liquidity seeking and liquidity providing orders can use estimates of price impact to optimize their strategies between order types, trading platforms, and across securities that are close substitutes. Market managers and regulators monitor the performance of markets at least in part by analyzing price and trading dynamics. They may be interested in understanding how the components of price impact are associated with particular market clearing mechanisms or order types (e.g., batch vs continuous, visible vs hidden), routing arrangements (e.g., payment for order flow vs market-exposed), or trader identities (e.g., particular liquidity providers or demanders).[1] A large body of asset pricing research also relies on the notion of price impact or components thereof as the basis for measuring the cost, risk, and return premium associated with illiquidity.[2]

In this paper, we provide a simple way to estimate price impact and its components. The approach can be used unconditionally, with all trades assumed to have the same price impacts, or conditionally with trades belonging to different categories for testing hypotheses about differences in price impacts or components between the categories. For example, trades could be categorized by clearing mechanism, route to market, trader type, or by "ours" versus "others" for an institution benchmarking its own execution quality against the rest of the market. The approach is based on the rather general model of Hendershott and Menkveld

[1] Glosten and Milgrom (1985), Kyle (1985), and Easley and O'Hara (1987) are classic references for the role of adverse selection in securities trading. Ho and Stoll (1980), Grossman and Miller (1988), Menkveld (2013), and Hendershott and Menkveld (2014) focus on dealer inventory and price dynamics. Glosten (1989) and Kyle (1989) show that dealer monopoly power results in an additional markup to the temporary component. Rounding due to price discreteness is modeled in Glosten and Harris (1988). Batch and continuous trading mechanisms are compared in Stoll and Whaley (1990) and George and Hwang (2001). Visible and hidden orders are compared in Boulatov and George (2013), Bloomfield et al. (2015), Foley and Putniņš (2016), Edwards et al. (2021). Hasbrouck (1991a), Hasbrouck (1991b), and Hasbrouck (1993) synthesize these ideas into a general empirical framework that captures how information and frictions affect price dynamics.

[2] A few examples are Amihud (2002) and Bollen and Whaley (2004) who measure illiquidity as a total price impact, Brennan and Subrahmanyam (1996), Chordia et al. (2009) and Huh (2014) use measures of the permanent component, and Pástor and Stambaugh (2003) and Hasbrouck (2009) use measures of the temporary component.

(2014) in the framework of Hasbrouck (1991a, 1991b, 1993). The estimation only requires transaction prices, signed traded quantities, and (if desired) trade category assignments as inputs. Bid, ask and midquotes are not needed, nor are assumptions that midquotes are reasonable proxies for true security values. The components can be estimated for all categories jointly with OLS, which is fast enough to be useful for large-scale applications in microstructure and asset pricing research. The components can optionally be estimated using GMM, which further supplies overidentifying restrictions. Either way, hypothesis testing for the components or functions thereof can be done easily with the delta method.

A key aspect of our analysis is using individual transactions and *transaction prices* rather than midquotes. This enables us to capture the impact of trading on the prices that agents actually pay and receive. This approach embraces bid-ask bounce, which contributes to the temporary component of price impact that traders experience. Using midquotes has the advantage of making the data less noisy, but it smooths out variation that is part of the trading experience. Wanting to use transaction prices is an important motivation to seek a simple approach to estimation, one that does not rely on specific distributional assumptions or require numerical search to estimate parameters. As we show in Section 5, the state space approach that prior work relies upon fails on transaction price data precisely because these distributional assumptions are violated.

We illustrate the approach with data that correspond to a rich set of transaction categories based upon the counterparties' identity types and roles in the transaction. The identity types are high-frequency trader (HFT) and non-HFT, and the roles are liquidity provider and liquidity demander. These attributes generate four categories of counterparty pairings for which we estimate temporary and permanent price impacts, or eight components in all. The data are provided by the Nasdaq market covering 120 securities over the years 2008-09. The Nasdaq HFT data are special because Nasdaq itself provides the identity types of HFT and non-HFT, their roles as liquidity demander or liquidity provider, and the sign (buy/sell) of the liquidity seeking order. While the sample period is from the earlier days of HFT in equities trading, this is one of the few (perhaps the only) data sets readily available to academics that identifies trader types with their roles in individual transactions. The rich detail in classifying trades demands that the approach estimates eight price impact components jointly. These data serve as an ideal sample to assess the performance of the approach and to illustrate concretely how it can be used by regulators, market managers, and institutions

whose proprietary data do contain information about trade routing, individual trader identities, or their own trading, as interesting dimensions along which to estimate price impact components. The approach can also be used straightforwardly without classifying trades into categories to estimate unconditional measures as inputs for asset pricing tests.

Our empirical illustration focuses on earnings announcements. The eight price impact components for the HFT and non-HFT counterparty pairings are estimated jointly at the security-day level, for the day before and the day after earnings announcements for securities and dates covered by the Nasdaq sample. Estimates for the same weekdays one week later are also obtained for comparison. Hypothesis tests are conducted for whether the components differ by counterparty pairing at the security-day level. Cross-sectional regressions examine how price impacts are affected by proximity to earnings announcements and by attributes of the announcements such as the earnings surprise and dispersion in analysts' forecasts, and by the number of analysts covering the firm.

We find that the permanent and temporary components of price impact differ in how they depend on the roles played by the counterparties. The permanent component of price impact depends primarily on the identity type of the liquidity provider (*permanent-provider* for ease of reference). Transactions in which an HFT is the liquidity provider are associated with larger permanent price impact per dollar value traded than transactions in which a non-HFT provides liquidity, regardless of whether the liquidity seeker is an HFT or a non-HFT. This finding is stronger around earnings with greater surprises. The interpretation is that HFTs absorb orders that pose greater adverse selection risk than non-HFTs absorb when trading as liquidity providers, and especially so when earnings convey a greater surprise. The identity of the liquidity demander matters to a lesser degree, with larger permanent components associated with trades where the liquidity demander is again an HFT. Taken together, these findings indicate that HFTs' trading contributes more to price discovery than trading of the same value by non-HFTs *in both trading roles*. Our evidence from transaction price impacts aligns with the conclusion of studies using other approaches that HFT contributes to market quality [e.g.,Carrion (2013), Brogaard et al. (2014), Chordia et al. (2018), and Chordia and Miao (2020)].

The temporary component of price impact depends primarily on the identity of the liquidity demander or taker (*temporary-taker* for ease of reference). Transactions in which a non-HFT is the liquidity demander

are associated with greater temporary price impact than when the liquidity demander is an HFT, whether the liquidity provider is an HFT or not. In fact, temporary price impacts tend to be *negative* when the liquidity demander is an HFT, again regardless of whether the liquidity provider is an HFT or not. Thus, the traditional notion that demanders of liquidity *pay* a price concession to liquidity providers is evident in trades initiated by non-HFT liquidity demanders, but not by HFTs. Instead, HFTs *earn* profit on average from the price impact of their liquidity demanding trades. The liquidity providers in those trades are not compensated by transaction prices for direct costs, and the adverse selection they absorb is underpriced.[3] This finding could be partly related to HFTs hitting stale orders of non-HFTs, leading to the impression that HFTs' liquidity demanding orders are "toxic" to non-HFTs. But the estimates are negative for trades involving HFT liquidity providers as well. Presumably, HFTs are deliberate in the pricing and duration of their limit orders, which implies that at least some of the aggressive pricing is deliberate and intended to attract counterparties.

Two noteworthy elements in these results are that HFTs as a group both absorb and impose more adverse selection than non-HFTs, and that HFTs profit from the temporary component of price impacts with their liquidity demanding trades. This leads us to consider whether the two elements are related, which could mean that HFTs learn from informative trades they absorb as liquidity providers and re-trade on what they learn in their liquidity demanding trades. Cross-sectional regressions support the hypothesis of a relation between these elements. The particulars are different for the days around the earnings announcement and the week after, however. In the week following earnings, the profit that HFTs earn on liquidity demanding trades is positively related to adverse selection absorbed from non-HFTs on the prior day and dominated by adverse selection absorbed from other HFTs on the same day. Around the earnings announcement, the profit that HFTs earn on liquidity demanding trades is again related to adverse selection absorbed from non-HFTs and HFTs on the prior day, but also from non-HFTs on the same day. Close to earnings, the source of profitable information shifts toward the trades that HFTs absorb from non-HFTs. Taken together, the regressions suggest that HFTs profit from their liquidity demanding trades by reusing information gleaned from their liquidity providing trades, and that information gleaned from non-HFTs is more important in close proximity to earnings announcements than in the week after. Indeed, the increase in liquidity demanding

[3]Liquidity provider losses measured only from transaction prices omit rebates for which they qualify under Nasdaq rules. Rebates are discussed in the Results section.

activity around earnings announcements is mostly attributable to non-HFTs. These findings support theories in which HFTs process information gleaned from the trading environment for (re)use in their own profitable trading [e.g.,Biais et al. (2015), Foucault et al. (2016), Huang and Yueshen (2021)].

Our contribution to the literature is twofold. First, we provide a simple approach to jointly estimating and formally comparing the components of price impacts that can be used easily when transactions are classified into several different categories. Estimation requires only transaction prices and trades, is done with OLS, and does not rely on Normality assumptions that are likely to be inconsistent with the dynamics of individual transaction prices and trades. We show that the state space approach, which is the natural alternative to our method and has been used in prior work with midquote data, fails to reliably estimate the model's parameters when applied to transaction price data, primarily due to violations of the Normality assumptions underlying the Kalman filter. We compare the OLS estimates with estimates from a corresponding "full-information" GMM estimation. The OLS approach performs well, including matching the inferences of the GMM-based hypothesis tests. We show that the model our estimates are based upon nests several prominent empirical models of price dynamics that have been used in the microstructure and illiquidity-in-asset-pricing literatures. Our approach can be used without trade categories to estimate the parameters in those models from transactions data. Because OLS estimation is fast, our approach is well suited for use in studies that require estimates of components of price impact (illiquidity) from intraday data for large panels of securities. And because the procedure is easy and hypothesis testing can be done at the security-day (or finer) level, our approach is also suited to narrower case studies of market or trading-strategy performance that could interest regulators, market managers, and practitioners.

Second, our findings confirm and extend what is known about HFT participation in markets.[4] The papers to which ours is closest are Carrion (2013), which studies the profitability of HFTs' trades against benchmarks that represent naive trading strategies, and Brogaard et al. (2014), which studies *midquote* dynamics using trades reported in the Nasdaq HFT data that we analyze. Their conclusions that HFTs absorb adverse selection and trade in the direction of future price changes are supported in our results, which build up to those conclusions from the permanent and temporary components of transaction price

[4]See the excellent review articles Chordia et al. (2013), Biais et al. (2014), O'Hara (2015),Menkveld (2016) and the more recent studies Chordia et al. (2018) and Chordia and Miao (2020) that examine HFT and price efficiency around macroeconomic news and earnings announcements, respectively.

changes estimated jointly for all the individual counterparty pairings. The novelty of our findings relate to estimating components from transactions data, which capture the profit and loss experienced by traders. In particular, the negative temporary price impacts of HFT-initiated trades in transactions data imply that HFTs actually earn profit by taking liquidity from orders that underprice adverse selection. In the cross-section of earnings announcements, such underpricing is greater for the liquidity *taken* by HFTs when HFTs absorb more informative trades as liquidity providers. Our results suggest that absorbing adverse selection is a learning opportunity for HFTs that supports profitable trade initiations.

The next section reviews the Hasbrouck (1991a, 1991b, 1993) framework and the refinements introduced by Hendershott and Menkveld (2014). The regression approach to estimating the model is provided in Section 3. Section 4 describes the data. Results are presented in Section 5, beginning with a comparison of the state space approach to OLS. Section 6 offers a brief conclusion.

2 Model

The model that we estimate is developed by Menkveld (2013) and Hendershott and Menkveld (2014) in the framework of Hasbrouck (1991a, 1991b, 1993). In Hasbrouck's framework, observed security prices are the sum of two unobserved components that reflect the impact of information and market frictions, respectively. The specification is

$$p_t = m_t + s_t, \tag{1}$$

where p_t is the observable security price, m_t is the security's unobservable true economic value (both in logarithms), and s_t is drawn from an unobservable mean-zero stationary stochastic process. The true value is modeled as a random walk because innovations to a random walk are permanent in accordance with the martingale property of (theoretical) security prices. The mean-zero stationary component models market-microstructure frictions by capturing deviations of observed prices from true values, which are transitory and die out over time.[5]

The variances of these components have interpretations as metrics of market quality. The variance of the

[5]The unobserved components model avoids strong assumptions that can accompany working directly with quotes and spreads, such as the assumption that the bid-ask midquote is a good proxy for the true security value. Studies of liquidity provision that fit an unobserved components model to midquotes find temporary components in midquote dynamics, which implies that midquote changes are not ideal proxies for changes in the true price.

random walk's innovation is a measure of the rate of information flow into prices and includes information conveyed by the execution of orders to buy and sell, and the arrival of public news. The variance of the stationary component is a measure of noise or friction in the trading process. The stationary component can include the price concession that liquidity demanders pay to liquidity suppliers for execution services, including profit, a return on capital, and compensation for bearing inventory risk; and it also includes other temporary effects on prices such as rounding due to fixed tick sizes and price discreteness. In this framework, transactions can impact prices through the permanent component by conveying information and through the temporary component by leading the price to deviate temporarily from its true value.

Though the innovation to the random walk and s_t are unobservable, Hasbrouck shows that their variances can be identified from the parameter estimates of a vector-autoregression (VAR) involving observables—price changes and non-price variables that describe the trading process, such as trade indicators (e.g., +1 for buys and -1 for sells) or signed order quantities. Hasbrouck also shows how the contribution of an observable variable to the variance of the random walk innovation can be identified from the VAR estimates as the variance of the infinite impulse response associated with the observable variable. Such measures can be used to assess the degree to which information flow is attributable to specific variables that describe the trading process.

To create specific model parameters that capture the impacts of observable variables on both components of price changes, Menkveld (2013) introduces observables *directly* into the equations for s_t and m_t. Hendershott and Menkveld (2014) introduce the additional feature of auto-regression in the transitory component. Their model is as follows:

$$s_t = \phi \; s_{t-1} + \psi \; x_t + \eta_t \tag{2}$$

$$m_t = m_{t-1} + \kappa \; \epsilon_t + \xi_t \tag{3}$$

$$p_t = m_t + s_t, \tag{4}$$

where x_t is an observable non-price variable, $\epsilon_t \equiv x_t - E_{t-1}[x_t]$ is the unexpected value of x_t, and η_t and ξ_t are mutually uncorrelated iid disturbances. The ψ parameter captures the impact of the non-price variable on the temporary component of prices. The κ parameter captures the impact of the unexpected non-price

variable on the permanent change in prices. The unexpected value of x_t is required in Equation (3) to preserve the martingale property of m_t [see Foster and Viswanathan (1993)]. Compared to Equation (1), Equations (2) - (4) add the minimal structure needed to create specific model parameters that capture the price impacts of interest. Henceforth, we refer to Equations (2) - (4), coupled with a (V)AR for x_t, as the HHM model (for Hasbrouck, Hendershott, and Menkveld).

The autoregressive term in the equation for s_t is important because, without it, the dynamic properties of the temporary component are constrained to mimic those of x_t. Even though x_t and ψ can be vectors, this would be a serious constraint because x_t contains limited information about the trading process. For example, if x_t represents signed transaction quantities and p_t represents transaction prices, as in our analysis, then ϕ can capture an important aspect of inventory management. If liquidity providers reposition their quotes to elicit liquidity seeking orders that reverse prior shocks to inventories, then trade reversals will lead to reversals in the temporary component, and ϕ will be negative. The AR(1) is a parsimonious way to accommodate these dynamics, and our estimates confirm that negative values of ϕ are an important feature in fitting transaction price dynamics in our sample.

Altogether, the HHM model strikes a balance between additional structure connecting the non-price variables to both components of prices, and flexibility in allowing the temporary component to reverse or persist for reasons that are not attributable to contemporaneous variation in the modeled non-price variables. Many familiar structural microstructure models of price dynamics and models used to generate measures of illiquidity for asset pricing research are nested within the HHM model, as we illustrate in subsection 3.1 below.

The selection of the x_t variable(s) and the way that prices are measured determine the interpretations of κ and ψ. Menkveld (2013) uses midquotes sampled at five-minute intervals as the price variable, and the inventory of a dominant HFT as the non-price trading variable, to study whether unexpected inventory contributes to price discovery (κ) and whether the inventory of the HFT is a determinant of the positioning of midquotes in relation to the true security value (ψ). Likewise, Hendershott and Menkveld (2014) study NYSE specialist inventories and the positioning of midquotes. Their price variable is the daily closing midquote, and the key non-price variable x_t is the daily change in specialist inventory. Brogaard et al. (2014), estimate HHM models using the Nasdaq HFT data. Their price variable is also the midquote, sampled at one-

second intervals. The x_t is the net signed quantities aggregated within those intervals traded by either liquidity providers (HFTs and non-HFTs), liquidity demanders (HFTs and non-HFTs), or HFTs only (both liquidity provided and demanded), each in a separate estimation. They also report some estimations using midquotes and quantities sampled at the transaction frequency. In each estimation, κ and ψ are interpreted as contributions of the particular measure of net quantity traded to permanent changes in the midquote and the transitory repositioning of the midquote in relation to the true security value.

Estimating this model is not straightforward because m_t, s_t and ϵ_t are unobservable. The three studies just cited estimate the surprises ϵ_t as residuals from a time-series regression of x_t on its own lags. The residuals are then used as observables in a state-space estimation [see Hamilton (2020), p. 372 - 408] (a maximum-likelihood estimation where the Kalman Filter is used recursively to define forecast errors).[6] Thus, their estimations involve a two-step sequence. The first step is to fit a (V)AR model to the inventory or trades variable(s) to generate residuals that proxy for true innovations. The second step is the maximum likelihood state-space estimation using midquotes, the inventory or trades variable(s), and the VAR residuals as observables. The equations of the Kalman Filter and the likelihood function both rely on the assumption that the midquote and aggregated inventory or trades data are jointly Normal.

In our study, x_t is a *vector* of signed dollar values of individual trades whose four elements correspond to the distinct possible counterparty pairings in individual transactions between HFTs and non-HFTs trading as liquidity demanders and suppliers. Our price variable is the transaction price, to the millisecond. Fitting the model using our definition of x_t and *transaction* prices implies that the four pairs of κ and ψ are the price impacts that reflect the actual wealth transfers between traders in the various possible counterparty pairings of HFTs and non-HFTs. This interpretation is different from when midquotes are used because the effects of all frictions are present in transaction prices. Some and perhaps most of the dealer compensation and frictional variation is missing when midquotes are used as the price variable because the deviations of transaction prices from midquotes that are driven by dealer compensation and other frictions are left out.

Our estimation would be challenging for a state-space approach because we aim to estimate four sets

[6] Menkveld et al. (2007) is the first of which we are aware to use a state-space approach to fit a model in the framework of Hasbrouck.

of κ and ψ parameters jointly, and the parameter space for a numerical search would be large. In addition, joint Normality assumptions are inappropriate for transaction prices and quantities, which are unlikely to satisfy such assumptions.

In the next section, we provide an estimation method that is suited to transaction prices and quantities, and that also easily accommodates higher dimensional κ and ψ vectors. Specifically, κ and ψ can be estimated using OLS regressions without using residuals as proxies for disturbances, and hypothesis testing for κ, ψ and functions thereof can be conducted using the delta method.

3 Regression Approach

Let x_t be a $(k \times 1)$ vector of observable variables that describe the dynamics of trading. In our empirical work, x_t contains signed dollar quantities traded for the $k = 4$ distinct possible pairings of HFTs and non-HFTs as transaction counterparties. The position in the x_t vector that is non-zero indicates the nature of the counterparty pairing, and the sign of the entry indicates whether the liquidity demander is a buyer (positive) or a seller (negative). For example, if an HFT liquidity demander buys \$20,000 worth of shares from an HFT liquidity supplier, then $x_t = (+2, 0, 0, 0)'$. If instead an HFT liquidity demander sells \$4,000 worth of shares to a non-HFT liquidity supplier, then $x_t = (0, -0.4, 0, 0)'$. The third element of x_t indicates transactions in which a non-HFT liquidity demander is paired with an HFT liquidity supplier, and the fourth element indicates transactions in which two non-HFTs are paired. Transaction values are in units of \$10,000 to facilitate reporting κ and ψ in basis points per \$10,000 traded.

As in studies that use state-space estimation where a VAR is first fitted to obtain residuals, we assume that x_t follows a vector autoregression of known order n:

$$x_t = B_1 x_{t-1} + B_2 x_{t-2} + \cdots + B_n x_{t-n} + \epsilon_t, \tag{5}$$

where each B_j is a $(k \times k)$ matrix of VAR parameters and ϵ_t is a $(k \times 1)$ iid vector of disturbances. Our approach also requires fitting the VAR, but it does not use the residuals as independent variables in a later estimation. For now, we assume that the individual elements of the vector x_t are observable to market participants. We revisit this assumption in subsection 3.2 below.

Proposition 1 shows that the parameters κ and ψ are identified in a simple way as functions of the coefficients of independent regressions—i.e., the VAR, and a regression whose dependent variable is the first difference of (log) prices. The proposition allows s_t to have a general ARMA(p, q) structure, which includes the simpler AR(1) specification in Equation (2) as a special case. The content of the proposition is that the HHM model parameters are exactly identified by only the contemporaneous and first-order-lag regression coefficients, provided that those coefficients are estimated with higher-order lagged returns and trading variables included as controls in the regression.

Proposition 1. *Suppose that the VAR(n) in Equation* (5) *describes the dynamics of a* $(k \times 1)$ *vector of trading variables* x_t *and the dynamics of prices are described by the HHM model in which* s_t *is ARMA*(p, q) *plus* $\psi' x_t$*; i.e.,*

$$s_t = \phi_1 \, s_{t-1} + \cdots + \phi_p \, s_{t-p} + \psi' x_t + \eta_t + \theta_1 \eta_{t-1} + \cdots + \theta_q \eta_{t-q} \tag{6}$$

$$m_t = m_{t-1} + \kappa' \epsilon_t + \xi_t \tag{7}$$

$$p_t = m_t + s_t, \tag{8}$$

where the disturbances ξ_t*,* η_t*, and* ϵ_t *are mutually uncorrelated at all lags. Under these conditions, the model parameters* ϕ_1*,* κ *and* ψ *can be written as*

$$\phi_1 = a_1 \tag{9}$$

$$\kappa' = (C' + D_1') \left[(1 - a_1) I_k - B_1 \right]^{-1} \tag{10}$$

$$\psi' = C' - (C' + D_1') \left[(1 - a_1) I_k - B_1 \right]^{-1}, \tag{11}$$

where I_k *is the* $(k \times k)$ *identity matrix,* B_1 *is the* $(k \times k)$ *first-order-lag coefficient matrix of the VAR, and* a_1*,* C' *and* D_1' *are coefficient vectors in the regression:*

$$dp_t = a_1 \, dp_{t-1} + C' x_t + D_1' x_{t-1} + \sum_{j=2}^{p} a_j \, dp_{t-j} + \sum_{s=2}^{n+p} D_s' x_{t-s} + u_t, \tag{12}$$

where dp_t *is the first difference of (log) prices,* a_1 *is a scalar, and* C' *and* D_1' *are* $(1 \times k)$ *row vectors. The disturbance* u_t *is autoregressive of order* $\max\{p, q\}$ *and is uncorrelated with* ϵ_t *in Equation* (5) *at all lags.*

Proposition 1 provides a simple recipe for estimation and hypothesis testing with the HHM model. Estimates of the κ and ψ vectors can be obtained directly from estimating the parameters of Equations (5) and (12) separately by OLS and plugging a subset of those estimates into Equations (9) - (11). Because the regression disturbances in (5) and (12) are uncorrelated, a variance-covariance matrix of their coefficient estimates can be constructed in a block-diagonal manner from the variance-covariance matrices of the separate estimations of the coefficients of (5) and (12). Newey-West with lag order $\max\{p, q\}$ can be used to account for the serial covariance in u_t when constructing the block that relates to the coefficient estimates of Equation (12). The delta method can be used with the constructed variance-covariance matrix to test hypotheses about ϕ_1, κ, ψ, and functions thereof.

Since ϕ_1, κ, and ψ are identified using only the coefficients on the lowest order lags in (5) and (12), the formulas in Equations (9) - (11) are invariant to the researcher's choices of the VAR lag length and the ARMA orders of s_t. This is an advantage in programming because the VAR length can be allowed to differ by security and day, while the code for computing the estimates and hypothesis tests is always the same. The lag and order choices do matter for the specifications of the regressions (5) and (12), however. The higher-order lags of variables can be thought of as control variables that are required to obtain consistent estimates of the coefficients on the lower-order lags that are used to identify and estimate ϕ_1, κ, and ψ. The matrix inverse in Equations (10) and (11) will exist unless $1 - a_1$ happens to be an eigenvalue of B_1, an occurrence that did not appear in our nearly 3050 estimations.

With this approach, estimation and hypothesis testing can be done without the joint Normality assumptions that are essential for correct specification of the Kalman Filter and the likelihood function in state-space estimation. This is especially helpful in working with transactions data, which are not likely to satisfy such assumptions. In addition, regardless of the dimension of κ and ψ, estimation of the two regressions occurs in one step using OLS, which is fast because a search-based optimization is not required. Inferences can be conducted at the firm-day level (or even subintervals of the day) and justified in the usual way. The ability to conduct inferences at the narrow level of the estimation is helpful for analyzing unique events where a cross-sectional distribution of estimates is not available. The CETV earnings announcement that we study below is such an example.

The value in handling higher dimensional κ and ψ is that the parameters for all categories of trading

variables can be estimated jointly. The importance of this is evident from the presence of the matrix B_1 in the formulas for κ and ψ. As long as B_1 is not diagonal, the identification of the permanent and temporary components of price impact for a given category of trades will depend on the cross-serial covariances between those trades and trades in another category. If the trades in different categories are motivated by similar economic forces, the true B_1 matrix is likely not to be diagonal. Failing to estimate the price impact parameters jointly when the cross-serial covariances are non-zero will yield inconsistent estimates.[7]

The approach in Proposition 1 treats the regression coefficients of variables at lags greater than one as nuisance parameters since they are not essential for identification. To see why, note that Equations (6) - (8) can be differenced and combined to yield

$$dp_t = \phi_1\, dp_{t-1} + (\psi' + \kappa')x_t - [\psi' + \kappa'(\phi_1 I_k + B_1)]\, x_{t-1} + h.o.t + u_t. \tag{13}$$

The symbol $h.o.t$ represents linear terms in dp and x of lag orders two and greater. The coefficients on those terms are non-linear functions of the VAR and HHM structural parameters (see the explicit version of (13) appearing as Equation (36) in Appendix A). The leading coefficients in Equation (13) along with B_1 from the VAR are sufficient to identify the HHM structural parameters.

Although the coefficients in $h.o.t$ are not used, the variables themselves are essential to include in the regression, even, for example, if one is interested only in estimating the total price impact $\kappa + \psi$ (and not the separate components). In Proposition 1, the vector C by itself identifies the total price impact (i.e., $C = \psi + \kappa$), but the regression from which C is estimated must include the lagged variables as controls in order for C to take the value $\psi + \kappa$. In this respect, Proposition 1 cautions against the common practice of estimating the total price impact by regressing returns on a contemporaneous trade variable alone (i.e., estimating an empirical "Kyle lambda"). If the HHM model with $n > 0$ describes the true return generating process, then omitting the controls will lead to inconsistent estimates because the regression cannot properly isolate the impact of the current trade and its innovation on the price. The controls are needed in order to separate the impact of the current trade from the effect of lagged trades on the current return that exist because of serial correlation in both trades and the temporary component of price changes.

[7] Our approach estimates price impact parameters jointly for classifications of trades that are given. Hendershott et al. (2022) can also be thought of as estimating price impact parameters jointly, but for classifications that are inferred from the data via maximum-likelihood estimation of a detailed structural model. Their trade classifications relate to agents who exhibit varying degrees of inattention to trading, which they show can explain puzzling low frequency patterns in equity returns.

Though not required for identification, the higher-order coefficients can be exploited as a source of overidentifying restrictions in a GMM approach to estimating the HHM model. This is because there are more higher-order coefficients that can be estimated than structural parameters that define them. The VAR and Equation (43) in Appendix A can be estimated together as a system by selecting values for κ, ψ, ϕs, and Bs to minimize a GMM objective function in a search-based optimization. As with the OLS approach suggested in the Proposition, GMM does not require joint Normality assumptions. The parameters of the HHM model are overidentified by the usual regression orthogonality conditions for Equations (5) and (43), so the GMM J statistic can be used to test whether the HHM model is well-specified. We use GMM estimates and inferences below as benchmarks for assessing the performance of the OLS approach in Proposition 1 and two alternative approaches. The numerical optimization takes considerable time, which makes GMM infeasible for large scale applications that are typical in microstructure and illiquidity-in-asset-pricing research.

3.1 Nesting Certain Structural Models

In this subsection, we show how several familiar structural models that are used to estimate price impact or illiquidity are nested within the HHM structure. This discussion is helpful in several ways. It provides intuition for why κ and ψ are identified by contemporaneous and lag-one dependence, and it shows how trading variables that are non-linear in quantity traded can fit within the HHM structure. In addition, the connections to the nested models highlight the generality of the HHM model and offer some clear avenues for future research. In these applications, trades are not classified into separate categories as in our empirical analysis below with the Nasdaq HFT data.

Rearranging Equation (13) yields:

$$dp_t = \phi_1 \left(dp_{t-1} - \kappa' x_{t-1} \right) + \psi'(x_t - x_{t-1}) + \kappa'(x_t - B_1 x_{t-1}) + h.o.t + u_t. \tag{14}$$

Following the taxonomy of models in Huh (2014), we begin with two classic market microstructure models of price dynamics. In our notation, the basic Glosten and Harris (1988) model is,

$$dp_t = \psi(S_t - S_{t-1}) + \kappa Q_t + u_t, \tag{15}$$

where Q_t is the signed quantity traded at time t and upper-case S_t is an indicator variable +1 or -1 for the

sign of Q_t . The Foster and Viswanathan (1993) model is

$$dp_t = \psi(S_t - S_{t-1}) + \kappa(Q_t - E_{t-1}[Q_t]) + u_t. \tag{16}$$

These models are nested in Equation (13) by defining $x_t = (S_t, Q_t)'$ and adding the "loadings constraints" that $\phi_1 = 0$, and that ψ and κ are non-zero on only the S_t and Q_t components of x_t, respectively. Under these constraints, the temporary component of price changes is fixed with respect to trade size and the permanent component is variable (scales linearly) with trade size. If, in addition, $(S_t, Q_t)'$ is AR(1), then the second element of the 2×1 vector $B_1 x_{t-1}$ is exactly the expectation at $t-1$ of Q_t.[8]

Note that S_t is a non-linear function of Q_t, but both models are easily nested in the HHM structure because dp_t depends linearly on S_t. A related example is Sadka (2006), who adds the *unexpected* sign S_t, computed from a Normal CDF of a forecast of Q_t, as another predictor of price changes. This variable is highly non-linear in Q_t, but because dp_t is linear in this variable, Sadka's model can be nested in the HHM structure by including an additional element to the vector x_t.

The Glosten-Harris and Foster-Viswanathan models capture the economics that are essential to the HHM model, all of which are reflected in contemporaneous and lag-one relations. It is therefore probably not surprising that the key parameters in the HHM model are identified by coefficients for the same lags as are needed to estimate the classic models. But the approaches are not identical. Glosten and Harris (1988) show that the "loadings constraints" described above are necessary for identification of κ and ψ when estimating a single equation model. However, nesting the classic models within HHM makes it evident that the loadings constraints are not necessary for identification when additional information from the dynamics of x_t is incorporated in the estimation. Adding the B_1 parameter(s) is just enough to preserve identification while relaxing the loadings constraints. Provided that $(1 - \phi_1)$ is not an eigenvalue of B_1, the temporary and permanent components of prices can *both* be modeled as having fixed *and* variable parts and estimated using the recipe in Proposition 1. Exponents of Q_t could also be added as variables to the x_t vector to allow the permanent and temporary components of price impact to be polynomial functions of signed trades, for

[8] Adding more constraints nests the equilibrium pricing equations in Glosten and Milgrom (1985) and Kyle (1985). In those models, transaction prices are equal to the market makers' conditional expected value of the security and the transitory component is non-existent, $\psi = 0$. In Glosten-Milgrom, the permanent component is fixed because transactions are unit sized, so κ is constrained to zero on the Q_t element in x_t. In Kyle, prices depend linearly on signed order flow and the permanent component is variable, so κ is constrained to zero for the S_t element in x_t.

example. All these variations are easily accommodated in the HHM model and estimated via Proposition 1. Our empirical work below follows the earlier HHM papers in modeling only a variable part to *both* the temporary and permanent components. This specification will capture temporary price concessions associated with profit or market-making costs that scale with trade size.

Some asset pricing studies that incorporate illiquidity use measures that are motivated heuristically, such as the Amihud (2002) measure, while other studies use measures that are model-based. Those that are model-based can differ in their focus. For example, Brennan and Subrahmanyam (1996) and Huh (2014) measure illiquidity using intraday estimates of κ from the classic models above, which places their focus on illiquidity associated with permanent price impacts. In contrast, Pástor and Stambaugh (2003) estimate illiquidity as the temporary price impact of signed volume. They define illiquidity as γ in a regression involving daily returns and *lagged* daily volume signed by the stock's net-of-market return:

$$dp_t = \phi_o + \phi_1 dp_{t-1} + \gamma\, Q_{t-1} + u_t. \tag{17}$$

The regression is designed so that γ captures the price impact of yesterday's signed volume that reverses today. The Pástor-Sambaugh measure is also nested in the HHM model. By reference to the coefficient on x_{t-1} in Equation (13), γ is a combination of all the HHM parameters. Its partial derivative with respect to ψ is indeed positive. However, if contemporaneous Q_t is excluded from the estimation as in Equation (17), there is no assurance that γ will be equal to its counterpart in Equation (13). Even if the Q_t process is AR(0), both Q_t and Q_{t-1} must be included in the regression. Thus, if the HHM model describes the data generating process at the daily frequency, then Proposition 1 indicates that the Pástor-Stambaugh measure could be improved as a proxy for ψ by including contemporaneous Q_t in Equation (17) when estimating γ. Hasbrouck (2009) also proposes a temporary component measure based on ψ in Equation (15) with κ constrained to zero. Under the assumption that the S_t are iid at the daily frequency, the ψ parameter is then estimated along with trade signs that are inferred from price changes rather than constructed from signed trading volume as in Pástor and Stambaugh.

In light of these contrasting approaches, Proposition 1 suggests that the HHM model can be used to estimate illiquidity parameters at large scale with OLS to *compare* the performance of, say, κ, ψ, and $\kappa + \psi$ as measures of illiquidity in asset pricing tests. As noted above, the impact of higher orders of trades can

be estimated easily too, which could be used to investigate whether, for example, price impact is convex in order size and the illiquidity of entering and exiting especially large positions is priced. Looking deeper into these questions is likely to be a fruitful avenue for future research. Huh (2014) conducts a large-scale study of illiquidity premia in the cross-section of equity returns and finds that intraday estimates of κ using even the simpler classical models strongly dominate the best performing daily measures [e.g., Amihud (2002) and Pástor and Stambaugh (2003)]. Extending Huh's study using the HHM model as suggested in Proposition 1 can provide greater specificity to whether permanent, temporary, or both types of price impacts are priced as illiquidity costs or risks in equity markets.

The HHM model structure also applies outside of equity markets research. Bollen and Whaley (2004) estimate the impact of measures of buying pressure on option implied volatilities. They use heuristic economic arguments to motivate a model specification that turns out to be a constrained version of Equation (13), with implied volatilities replacing prices:

$$\Delta\sigma_t = \phi_o + \phi_1 \Delta\sigma_{t-1} + \beta' x_t + \text{controls}_t + u_t, \tag{18}$$

where σ_t is option implied volatililty on day t and x_t is a vector of demand pressure measures. Bollen and Whaley are interested in estimating a total contemporaneous price impact, and indeed their β' maps to $\kappa'+\psi'$ in Equation (13). The approach in Proposition 1 suggests that by including lags of the demand pressure measures, and estimating a VAR to capture their dynamics, one can further identify the *separate* permanent and temporary components of buying pressure on implied volatility in Bollen and Whaley's experiment. Such measures, perhaps estimated using intraday data, could then be used as inputs to study further whether illiquidity premia exist in options markets [see Christofferson, et al. (2018) for such a study where intraday realized spreads are used as the measure of illiquidity].

3.2 Observability of x_t to Market Participants

The discussion so far assumes that the individual elements of x_t are observable to market participants in real time. Otherwise, market participants would be unable to compute the vector ϵ_t in real time, and Equation (7) would not be a martingale with respect to the market's beliefs. One can imagine that in applications of this estimation approach to blockchain markets, trader (or wallet) identities may be observable

in real time. However, in the Nasdaq HFT data, market participants were aware of the trades, but not the identity types of the counterparties, in real time. The procedure in Proposition 1 requires a slight modification in this case.

The model that traders perceive is one in which a *univariate* series of signed dollar order quantities $\hat{x}_t$ is observed, and the unconditional price impact parameters are *scalars* $\hat{\kappa}$ and $\hat{\psi}$. Therefore, the equation for return dynamics is

$$dp_t = \phi_1 \, dp_{t-1} + (\hat{\psi} + \hat{\kappa})\hat{x}_t - \left[\hat{\psi} + \hat{\kappa}(\phi_1 + b_1)\right] \hat{x}_{t-1} + \widehat{h.o.t} + u_t, \tag{19}$$

where b_1 is a scalar first-order autoregressive parameter from a univariate AR model for $\hat{x}_t$, and $\widehat{h.o.t}$ are linear terms in dp and $\hat{x}$ of lag orders two and greater. Nevertheless, as researchers, we can do the equivalent of adding dummy variables as interactions with the regressors to estimate and test for differences in coefficients conditional on the data categories identified by the dummies.

In particular, as researchers and in hindsight we know to which of the four counterparty bins each $\hat{x}_t$ belongs. So we can define (4×1) vectors x_t whose elements are each identified with a particular counterparty pairing. For example, if the transaction at time t is between two HFTs, then the first element of x_t is assigned the value of $\hat{x}_t$ and the other elements of x_t are assigned values of zero; likewise for the other three elements of x_t and counterparty pairings. Thus, $\hat{x}_t$ is the sum of the elements of x_t, and Equation (19) is equivalent to

$$dp_t = \phi_1 \, dp_{t-1} + (\hat{\psi} + \hat{\kappa})(x_{t1} + \cdots + x_{t4}) - \left[\hat{\psi} + \hat{\kappa}(\phi_1 + b_1)\right] (x_{t-1,1} + \cdots + x_{t-1,4}) + \widehat{h.o.t} + u_t. \tag{20}$$

We can then replace the scalars $\hat{\kappa}$ and $\hat{\psi}$ with analogs of dummy slope coefficients to accommodate different price impacts by counterparty pairing:

$$\begin{aligned} dp_t =& \phi_1 \, dp_{t-1} + (\psi_1 + \kappa_1)x_{t1} + \cdots + (\psi_4 + \kappa_4)x_{t4} \\ & - \left[\psi_1 + \kappa_1(\phi_1 + b_1)\right] x_{t-1,1} + \cdots + \left[\psi_4 + \kappa_4(\phi_1 + b_1)\right] x_{t-1,4} + \widehat{h.o.t} + u_t. \end{aligned} \tag{21}$$

Writing this in vector notation yields

$$dp_t = \phi_1 \, dp_{t-1} + (\psi' + \kappa')x_t - \left[\psi' + \kappa'(\phi_1 + b_1)\right] x_{t-1} + \widehat{h.o.t} + u_t, \tag{22}$$

which is the same as Equation (13) except that the identity matrix I_k is not needed, the scalar b_1 replaces the matrix B_1, and $\widehat{h.o.t}$ includes lags of $\hat{x}_t$ rather than lags of the individual elements of the vector x_t. In Equations (10) and (11) for the model parameters the scalar b_1 replaces the matrix B_1. Our estimations with the Nasdaq HFT data use this modified procedure. The modified version of the full equation for return dynamics that includes the higher-order terms, and which we use for the GMM estimations, is given in Equation (43) in Appendix A.

4 Data

The data we analyze are provided by Nasdaq.[9] We use these data because the classifications are extensive and result in eight parameters to be estimated jointly. This setting challenges the performance of the exactly identified approach in Proposition 1. We compare its estimates and inferences to those based on overidentifying restrictions that should be more efficient when the parameter space has high dimension. If the simpler approach performs well in this case, it ought to perform well in applications that do not require classifications of trades, such as estimating illiquidity components for asset pricing tests. [10]

The Nasdaq HFT data are also used in Carrion (2013) and Brogaard et al. (2014), which provide excellent descriptions of the data and how it came into existence. We describe key aspects for our study.

The data contain all transactions executed on Nasdaq from January 1, 2008, to December 31, 2009 for 120 stocks divided equally between small, medium and large categories by market capitalization. Each record contains the following fields:

1. Timestamp in milliseconds
2. Stock symbol
3. Number of shares
4. Buy/sell indicator of the liquidity seeking side of the trade
5. Transacted price
6. Trader pairing (i.e., HH, HN, NH, NN)

[9]We thank the Office of the Chief Economist for kindly providing the data.

[10]While the Nasdaq HFT data are special in providing the actual classifications of trades by HFTs and non-HFTs, one could also use the approach with trades classified by a proxy. In the HFT setting, see for example, Chordia and Miao (2020), who use the Hasbrouck and Saar (2013) measure of strategic runs of fast message traffic as a proxy for low-latency trading activity, and runs that end with a transaction as trades in which liquidity is provided by a low-latency trader.

The trader pairing variable identifies the types of traders who are counterparties to the transaction (H=HFT or N=non-HFT), with the trader type that was seeking liquidity listed first. Nasdaq identifies 26 traders as HFTs based on whether the trader has characteristics that are usually attributed to HFTs, such as a high frequency of trades, orders, and cancellations, and low inventory accumulation. A few institutions with large brokerage operations are not identified as HFTs despite having HFT capabilities. Sixty of the stocks are listed on Nasdaq and 60 are listed on the NYSE, and many trade on other venues. Only trades that are executed on Nasdaq are included in the sample.

Most of the transaction records have distinct millisecond time stamps, which enable us to identify the sequence of distinct transactions within each day. When transactions are distinct, the x_t vector of signed trades for millisecond t has only a single non-zero entry among its four elements. The location of the entry corresponds to the trader pairing for the transaction, the sign of the entry indicates whether a purchase (positive) or a sale (negative) was on the liquidity demanding side of the trade, and the magnitude of the entry is the dollar value of the trade in $10,000 units.

There are some instances in which more than one transaction has the same millisecond time stamp. Nasdaq advised us that transactions with the same time stamp are not necessarily written to the database in chronological order within the millisecond. In these instances, we compute the dollar value of the trades within each trader pairing, and sign the trade indicator for that pairing according to whether the *net* liquidity taken was a buy or sell. For example, if a $30,000 HH buy and a $10,000 HH sell have the same timestamp, the net HH transaction is a $20,000 buy, which we code as +2 in the first element of x_t. If there are multiple types of pairings within the same millisecond (e.g., HH and HN transactions with the same timestamp), then the x_t vector will have multiple entries, one for each pairing. The magnitude of each entry is again the dollar value of the net trade in $10,000 units, and the sign of each entry is positive or negative depending on whether the *net* liquidity taken in that pairing was a buy or sell.[11]

For the stocks and time period covered by the Nasdaq sample, we draw the dates and times of earnings announcements and analysts' earnings forecasts from I/B/E/S. The vast majority of earnings announcements occur during the overnight period between the market close and open. We exclude daytime announcements

[11]The ratio of the mean (median) number of milliseconds with more than one transaction to milliseconds with at least one transaction is only 0.08 (0.13) across stock-days.

in order to have uniform pre-announcement and post-announcement days throughout the sample. We record the number of analysts forecasting earnings as an indication of the intersection of supply and demand in the market for public information about a company's earnings. We use this number to stratify the sample into quartiles.

We compute the earnings surprise for each announcement as the difference between realized earnings per share and the median analyst estimate prior to the earnings announcement reported on IBES, divided by the opening stock price on the day preceding the overnight announcement. For announcements with two or more analyst forecasts, we compute dispersion in analysts' forecasts as the difference between the highest and lowest analyst estimates divided by the median estimate.

Our estimation and testing focus on prices and trades from four days anchored by each earnings announcement. The Pre-day is the trading day prior to the overnight announcement and the Post-day is the trading day following the announcement. The Week-After-Pre (WA-Pre) day is one week later than the Pre-day on the same weekday, and the WA-Post day is likewise one week later than the Post-day on the same weekday. We include all days with 50 or more transactions total and at least 10 trades for each of the four counterparty pairings HH, HN, NH, NN. After these constraints, we have 3047 security-days relating to 823 earnings announcements belonging to 120 distinct stocks. Cutoffs for the analyst coverage quartiles are based on the distribution of coverage for these security-days.

A separate set of structural parameters ϕ, ψ, and κ is estimated for each security-day using prices and trades during daytime trading hours (8:30-4:00pm Eastern Time). We exclude the opening transaction from all security-days because the first trade of the day is known to have different characteristics than transactions during the rest of the day [e.g., Amihud and Mendelson (1987), Stoll and Whaley (1990)]. Table **??** reports summary statistics for the sample. For the item listed in each row, the columns report statistics describing the distribution of daily values for that item across security-days. For example, the median across security-days of the daily dollar volume of trades in which both counterparties are HFTs (i.e., the HH category) is \$761,782. HH is the category with the smallest median and mean dollar volume across security-days, and NN is the category with the largest. However, there is a high degree of overlap between the distributions of all categories. For example, the 75th percentile for HH is larger than the median of NN. The 25th percentile of NN is larger than the median of HH, but substantially smaller than the 75th percentile of HH. As noted

earlier, counterparty types are not reported to traders in transaction records. The high degree of overlap would make it difficult for traders in real time to infer counterparty types from transaction sizes.

Table **??** also reports distributions of the daily numbers of transactions in each counterparty category. The mean and median numbers of transactions are smallest for HH and largest for NN. The range of values constituting the middle 98% of each category is impressive. HH ranges from 10 to 13,442, and NN ranges from 102 to 15,677. These distributions also exhibit substantial overlap across categories of counterparty pairings.

Items concerning the scale of transaction returns are also reported in the table. Transaction returns are centered on zero, as is evident by the mean of means and median of median transaction returns both being zero to the tenth of a basis point. Transaction returns are also small in magnitude. Across security-days, the mean and median of the mean daily *absolute* transaction returns are 4.5bp and 3.1bp respectively; and the 99th percentile is 20.4bp. The mean and median of the median daily absolute transaction returns are smaller at 1.8bp and 0.0bp, respectively; with a 99th percentile of 14.8bp.

By construction, the Nasdaq sample is diverse in market capitalization, ranging from about $218 million to $325 billion as of December 2007. The mean and median are $27.5 billon and $2.58 billion, respectively. The number of analysts forecasting earnings ranges widely—from zero to 34—with a mean and median of 12.6 and 11, respectively.

5 Results

We begin with the results of attempting to estimate the model using state space methods for a subsample of 40 earnings announcements (i.e., 160 stock-days). It turns out that the state space approach fails to identify separately the temporary and permanent components of price impact for 51 of the stock days, and fails to produce estimates and valid standard errors for 26 of the days, even when OLS estimates are used as starting values. More focused investigation points to non-Normality of the data as the culprit for both failures, which strongly suggests that having an alternative to state-space estimation, and especially when using transaction prices, is a valuable resource for microstructure researchers. Then we turn to evaluating the quality of the OLS approach by comparing its estimates and inferences to those of GMM, which we

take as the gold standard because it incorporates all the restrictions of the model. The main finding is that estimates and hypothesis tests that follow the approach in Proposition 1 compare favorably to GMM and perform better than other OLS based approaches that take shortcuts vis-a-vis the approach in Proposition 1. Finally, we present detailed results for the OLS approach in Proposition 1 for the specific days around earnings announcements and some regression tests that characterize cross-sectional and temporal patterns in the components of price impacts around earnings. We also conduct an event study of the announcement in our sample with the largest earnings surprise.

5.1 State Space Estimation

Since we motivate the OLS approach by observing that the assumptions underlying the state-space (SS) approach are likely to be seriously violated in transaction price data, we begin by attempting to estimate the model's parameters using a SS approach that follows the steps in Brogaard et al. (2014). For each stock-day, we estimate a VAR as in Equation (5) whose order is determined by the AIC, and residuals from the VAR along with signed trades and log price changes are used as data to estimate ϕ_1 and the κ and ψ parameters using maximum likelihood with the Kalman filter. The ϕ_1 parameter is constrained to lie between -0.99 and 0.99 in the optimization, and diffuse priors are used for the disturbance variances of the permanent and temporary components. A difference with Brogaard et al. (2014) is that we estimate all eight κ and ψ parameters jointly, whereas they group trades together which reduces the dimension of the parameter space.

Because these optimizations are time consuming, we focus on the days surrounding 40 randomly chosen earnings announcements in our sample (160 stock-days in total). We use OLS estimates obtained as described in Proposition 1 (henceforth, OLS-Prop estimates) as starting values for the nine parameters of interest. The model converges and standard errors are valid (i.e., the Hessian is non-singular at the estimates) for 134 days. For the other 26 stock-days, the negative log-likelihood is flat around a candidate optimum and not able to identify a minimum. Perhaps more concerning is that the estimate of ϕ_1 is equal to 0.99 for 51 of the 134 days for which there is a solution. Since 0.99 is the upper boundary of the parameter space, this outcome indicates that the optimizer is unable to find an interior solution and is instead being pushed to the boundary, where the temporary and permanent components are indistinguishable. As far as the data are concerned, a

temporary component in *transaction* price changes does not exist, which seems unreasonable a priori given the vast microstructure literature that documents its existence with simple estimation approaches. This leaves us with 83 stock-days with useful estimates, which is only about half of the number of stock-days we started with.

Details of the results for these 83 estimations are discussed below. However, before we get there and to better understand the sources of these failures, we conduct a series of focused experiments on five stock-days drawn from the set of problematic cases, progressively replacing real data with simulated Normal draws to isolate what it is about transaction price data that causes the SS approach to break down. Table 2 reports the outcomes of the experiments, with only meaningful details for brevity. Panel C of Table 2 reports the results of the first experiment, a Monte Carlo simulation in which all variables are drawn from Normal distributions. The signed trades x_t and their innovations ϵ_t are drawn as independent standard Normals, and the disturbances ξ_t and η_t are drawn as independent Normals whose standard deviations are set to the median standard deviation of transaction returns in the sample. The OLS-Prop estimates of ϕ_1, and the κ and ψ vectors are used as the 'true' parameter values in generating the simulated data. No VAR is estimated, so this experiment focuses on the Kalman filter step in isolation. SS estimation was performed on the simulated data with zeros as starting values, then with OLS-Prop estimates as starting values. With both sets of starting values, the SS estimator converges and provides valid standard errors (non-singular Hessian) for all five stock-days. The 'true' parameter values are recovered almost exactly, which confirms that the failures observed with the real data are not attributable to programming errors or numerical problems relating to the OLS-Prop estimates.

The second experiment is reported in Panel A where real data are used for the estimation, and starting values of zeros and OLS-Prop estimates, alternatively, are used. Stock-days S2, S4 and S5 converge to the boundary where $\phi = 0.99$ when zero starting values are used, just as when the OLS starting values are used. S1 and S3 did not converge with Hessian-based standard errors with OLS starting values. However, when the outer product of gradients is permitted as a substitute for the Hessian to approximate the standard errors, the estimates for S1 and S3 are $\phi = 0.99$ and $\phi = -0.31$, respectively. For S3 the Hessian is singular even at an interior solution, which means that the identification problem is not confined to cases where the optimizer is pushed to the boundary of the parameter space. It can also happen that the likelihood is

uninformative in the vicinity of reasonable parameter values. When zeros are used as starting values, all five stock-days converge to the boundary. These findings suggest that the failures of SS in the larger sample to identify the temporary and permanent components separately are not related to the starting values used in the estimation, and that flat spots in the log-likelihood function are not limited to boundary solutions.

The third experiment, reported in Panel B, combines real trading data with simulated Normal disturbances, which allows the real trading activity data to introduce dynamics that are possibly contrary to the joint Normality assumptions underlying the SS approach. In this experiment, we estimate the VAR for trading activity using real data to generate observations of trading innovations. We then use these innovations and actual trades to generate a temporary component series and a permanent component series. The OLS-Prop estimates are used as the 'true' parameter values and the disturbances ξ_t and η_t are drawn as independent Normal random variables with mean zero and standard deviations that are again equal to the median standard deviation of transaction returns in the sample. When zeros are used as starting values, four of the five stock-days converge to the boundary with $\phi = 0.99$, though S3 converges to an interior solution with $\phi = -0.098$. However, when OLS-Prop starting values are used, all five stock-days converge to plausible interior values of ϕ_1 with small standard errors, and all five match the OLS-Prop values to the second or third decimal place. The κ and ψ parameters are also recovered accurately. Interestingly the estimates of ϕ_1 for S3 are nearly identical when both zeros and the OLS-Prop starting values are used, suggesting that it is possible, though not likely, for SS to find the "true" values even when starting at zero. The results in this panel suggest that real trading activity introduces non-Normal dynamics into the data that prevents SS from recovering accurate estimates unless starting values are sufficiently close to the solution. if the structural disturbances are Normal, SS has a chance of working well, but only reliably when started from values that are sufficiently close to the solution. In practice, a researcher would not have access to such starting values without first running OLS-Prop or something like it, making SS either infeasible or merely a refinement of methods that explicitly do not rely on Normality assumptions.

Our final experiment is reported in Panel D, where the return and four trading activity series are transformed into univariate Normal series by applying the inverse of the Normal CDF to "plotting positions" based on a ranking of each original series.[12] The SS procedure including the VAR and using OLS-Prop initial

[12]The plotting positions we use are $p_i = (r_i - 3/8)/(n + 1/4)$ where r_i is the rank of observation i and n is the number of observations. This procedure is attributable to Blom (1958).

values is then applied to the transformed data, just as it was with the original sample. All five stock-days converge to interior solutions with valid Hessian-based standard errors. The estimate of S3 came close to the boundary with $\phi = 0.973$ and a large standard error. The remaining four stock-days produced plausible estimates ranging from -0.255 to 0.639 with small standard errors, which are consistent with identifying separately permanent and temporary components of price changes. Although the parameter estimates are not economically meaningful after the transformation, the improvement in SS behavior relative to Panel A is informative. The fact that four of five stock-days produced well-behaved interior solutions after the transformation strongly implicates non-Normality of the original transaction price and trading data as the primary source of SS failures. The case of S3 suggests that univariate Normality of the marginals alone may not be sufficient, and that violations of the multivariate Normality assumption underlying the Kalman filter may also play a role.

The failure of the state-space approach documented here is consistent with Allena and Chordia (2025), who show that the rounding induced by discrete tick sizes destroys the Gaussian structure that the Kalman filter requires, rendering it unsuitable for transaction-level data. They provide an estimation method that is Bayesian and numerical but does not use the Kalman filter for recovering the latent true price purged of tick-rounding effects. Our objective differs from theirs. Rather than purging frictions from observed prices, we aim to measure them since the bid-ask bounce, dealer compensation, and rounding that Allen and Chordia's framework filters out are elements of the temporary component (and traders' experience) that we wish to estimate. For our purpose, OLS is preferable because it requires no structural or distributional assumptions about the frictions. This is important because our Panel B results indicate that non-Normality intrinsic to the underlying trade dynamics would remain even after applying a rounding correction.

Returning to the larger sample of SS estimates, Panel A of Table 11 in Appendix B reports the performance of the SS approach to GMM for the 83 stock-days with valid SS estimates and $\hat{\phi}_1 < 0.99$. The SS estimates perform poorly in absolute terms and relative to the OLS estimates considered below. For ϕ_1, the average SS estimate is 0.38 compared to -0.12 for GMM, and the match rate with GMM inferences is 67% when the GMM estimate is positive and only 10% when it is negative. Corresponding match rates are all above 80% for the OLS estimates in the full sample. The κ parameters are particularly poorly estimated by SS, with match rates of only 14% to 21% when the GMM estimate is positive and average estimates that

are a fraction of the GMM averages. The ψ parameters are estimated somewhat better in matching GMM inferences, but the point estimates differ substantially. Taken together, these results confirm that even when SS produces a valid solution on transaction price data, the estimates are still unreliable.

5.2 Assessing the Quality of OLS Estimates

5.2.1 Comparison of OLS and GMM

Having established that the SS estimation is unreliable for transaction price data, we now turn to evaluating the performance of three OLS-based approaches by comparing their estimates and inferences to GMM, which we regard as the gold standard. For ease of exposition, we refer to these as "diagnostic comparisons". Estimates and tests are conducted at the security-day level for all three OLS approaches and GMM, and performance is evaluated by comparing the distribution of estimates and inferences across security-days. To ensure uniformity of the underlying assumptions, GMM and the two OLS approaches that use a VAR are all based upon an HHM model with ($n = 7$, $p = 1$ and $q = 0$). The lag length of seven was chosen after visual inspection of the correlograms of a random selection of security-days before any estimation was attempted. This choice turned out to be very close to the overall sample median of eight when lag lengths are selected by the AIC separately on each security-day, as reported in the next section.

For the GMM estimation of κ, ψ and ϕ_1, the moment conditions are the least-squares normal equations corresponding to an AR(7) for $\hat{x}_t$ (to which Equation (5) reduces in the case where the individual elements are not observable by market participants) and to Equation (43) in Appendix A, altogether in one pass. We regard the GMM estimates as a "full information" benchmark in the diagnostic comparisons because it imposes the constraints that are implied by the model on *all* the regression coefficients in Equations (5) and (43). In contrast, only the exactly-identifying information from the contemporaneous and first-order-lag regression coefficients are used in the OLS approach of Proposition 1 and the two other OLS-based alternatives considered here.

The GMM estimations require a numerical optimization, which takes considerable processing time. There are 25 orthogonality conditions (regressors) and only 19 parameters that define the regression coefficients, so the system of moment conditions is overidentified. The overidentifying restrictions do have some bite.

The GMM J statistic rejects the null in 25.9% of security-days in the sample. Nevertheless, the HHM model seems to fit the data well for a large majority of security-days.

The first OLS approach is that of Proposition 1, which estimates Equations (5) and (43) in Appendix A separately by OLS. Inferences are based on a block-diagonal variance-covariance matrix with a Newey-West adjustment as described following Proposition 1. We refer to this approach as "OLS-Prop."

The second OLS approach is a two-step procedure, which first estimates a VAR model for trades by OLS, then uses its residuals as proxies for the ϵ_t vector in a second-stage regression for returns. The second-stage equation corresponds to Equation (32) in Appendix A, with ϵ_t treated as an observable variable and the lagged ϵ_ts absorbed into the disturbance. Price impact parameters correspond to distinct coefficients in this regression. This approach does not include the higher order lags of trades as controls because they are not needed if residuals from the first-stage are regarded as observationally equivalent to ϵ_t—the controls appear in Proposition 1 when substituting for ϵ_t because it is unobservable. Hypothesis tests with this method ignore the estimation error involved in the first step. Inferences are based on Newey-West standard errors from the second-stage equation. Comparing the results from this approach to those of Proposition 1 provides evidence on whether the errors-in-variables arising from using a proxy for ϵ_t, and ignoring the first stage estimation error in computing standard errors, matter to the estimates and inferences. We refer to this method as "OLS-Resid".

The third OLS approach assumes that $E_{t-1}[x_t] = 0$ for all t, and does not use either the time-series model residuals or the trades beyond lag-one as controls in the returns equation. This approach treats x_t entirely as a surprise, and there is no role for fitting Equation (5). Price impact parameters are computed from the formulas in Proposition 1 with $B_1 = 0$. Inferences are based on standard errors with no adjustment for lagged dependence in residuals. Comparing the results from this approach to the others provides evidence on the importance of modeling the market's conditional expectations. As we show below, this approach exhibits large distortions in the estimates and tests. We refer to this method as "OLS-Zero".

Detailed results of the four estimation methods for all the parameters appear in Table 11 in Appendix B. Altogether, OLS-Prop matches the estimates and inferences of GMM rather well. The overidentifying restrictions GMM brings to the estimation do not change the estimates and inferences very much from those

of OLS-Prop. OLS-Resid is a close second best. OLS-Zero does poorly, especially with the ψ parameters, which are biased upward and toward rejecting in the positive direction. The differences between the OLS-based methods is whether and how conditional expectations of x_t are incorporated into the estimation. The diagnostic results indicate that care taken in modeling expectations of x_t matters to the quality of the estimates and inferences.

The lag length was held fixed at seven across these estimations. However, the choice of lag length is itself an important aspect of modeling the conditional expectations of x_t. We do not know what model market participants use to form beliefs, and the appropriate lag length could vary by security and by day. Searching over lag lengths and the parameter space using GMM is infeasible for a panel of security-days as would be typical in microstructure or asset pricing research. However, allowing such flexibility in selecting lag lengths is easy with a method based on OLS. We exploit this flexibility in obtaining the estimates for our main results.

5.3 OLS-Prop with AIC

This section presents our main results; it has six subsections. The first three provide detailed results for each of the key sets of structural parameters—ϕ_1, κs and ψs. The estimation is done using OLS-Prop as described in the previous section. The difference here is that lag lengths are determined by the AIC for each security-day. These are followed by a fourth subsection that reports descriptive statistics for trading volume by trade classification and type of day. The fifth subsection uses cross-sectional regression to analyze parameter differences that relate to hypotheses about HFT contributions to market quality. The final subsection examines sets of parameters relating to the same earnings announcement (and week-after days) for evidence that HFTs use information learned from liquidity providing trades to enhance the profitability of their liquidity demanding trades.

5.3.1 ϕ_1 Estimates

Table 3 reports estimates and tests for ϕ_1. The format of this table is a template for the tables that follow. Under the heading for a particular counterparty pairing, there are four rows of statistics that relate to the four classifications of days around earnings announcements. Those labeled "Pre" are estimates for days prior to the overnights when earnings announcements occur, those labeled "Post" are for post-announcement

days, and those labeled "WA Pre" and "WA Post" are estimates from the same weekday one week after the Pre and Post days, respectively. For each day classification, we report the fractions of the sample for which the estimates of ϕ_1 are positive and significant, and negative and significant, followed by the mean and median of the estimates.[13] The fifth row reports statistics for the pooled set of estimates that includes all four day classifications.

For the full sample, the average (median) ϕ_1 parameter estimate is -0.11 (-0.10). Negative estimates reject the null of equality with zero at a rate of 64% for the full sample in Panel A, and 42% and 91% for the low- and high-analyst-coverage quartiles in Panels B and C, respectively. Rejections with positive estimates are rare at 3% or less. Modeling autoregression in s_t is evidently important to proper specification of the model for price dynamics for most security-days.

The magnitudes of the mean and median estimates for the high-coverage quartile are about 1.5 times those of the low-coverage quartile. The estimates are slightly larger in magnitude on the Post day than other days for the sample as a whole for both high- and low-coverage stocks. Negative estimates of ϕ_1 indicate reversals in the temporary component of prices that are unrelated to contemporaneous trading, which is consistent with liquidity providers quoting prices to elicit trades that reverse past shocks to inventory. The estimates and tests suggest that liquidity providers manage inventory actively on all four days, and more aggressively on the Post day and for high-coverage stocks. This accords with our observation below that the number, size and volume of transactions are highest on the Post day.

Panel D of Table 3 reports statistics for the distribution of lags selected by the AIC across the security-days in our sample. As noted above, the full-sample median is eight, and the average is 10.4 lags. Larger lag lengths are selected for high-coverage (median of 16) than for low-coverage stocks (median of four).

5.3.2 Kappas and their Differences

Table 4 reports estimates of the permanent components of price impact, the κs, and Table 5 reports key κ differences. The results in Table 4 Panel A for the full sample show that between 66% and 92% of the κ estimates are positive and significant across the four different counterparty pairings and days in relation to the earnings announcement. The inferences almost uniformly favor positive estimates as very few rejections

[13]Rejection rates are based on asymptotic two-sided tests with $\alpha = 0.05$ using the delta method and the block-diagonal Newey-West variance-covariance matrix of regression parameter estimates described following Proposition 1.

are associated with negative estimates. The average and median estimates are highest for κ_{HH} and lowest for κ_{NN}. The estimates are in units of basis points per \$10,000 value traded. The transactions that generate the greatest permanent price impacts are those that occur between two HFTs, with an overall average permanent price impact of 14.55bp (median 2.35bp), which translates to \$14.55 on a \$10,000 trade. The overall average is about half that size at 6.39bp (median 1.39bp) for transactions between two non-HFTs.

Panels B and C report corresponding results for the subsets of the sample in the lowest and highest quartiles of analyst coverage, respectively. In Panel B, the rejection rates are lower than in the full sample, but rejections are still nearly uniformly positive. The estimates for low-coverage stocks are much larger than in the full sample. The overall mean and median are 26.56bp and 9.08bp for κ_{HH}, and 12.82bp and 5.92bp for κ_{NN}. The rejection rates are higher for high-coverage stocks in Panel C than in the full sample, with most above 95% for positive estimates. However, the average (median) estimates are much smaller for high coverage stocks, for example 1.05bp (0.61bp) for κ_{HH}. Across the parameters, the κ_{HH} estimates are again the largest and the κ_{NN} estimates are the smallest within both the low- and high-coverage subsamples, suggesting that HFTs generally both *pose and absorb* greater adverse selection in trading with each other than occurs when non-HFTs trade with each other. This hypothesis and others like it are tested formally below.

The distributions of estimates are tight for the high-coverage quartile in Panel C, making it easy to see where there are differences between the κ parameters; e.g., the means and medians of κ_{NH} are clearly larger than those of κ_{NN} for all four days. For the low-coverage quartile in Panel B, the rejection rates vary more and the patterns are less clear, making it difficult to say how different counterparty pairings affect permanent price impacts. For this, it is helpful to have formal hypothesis tests at the security-day level. Since the estimation approach of Proposition 1 allows us to conduct hypothesis tests in this manner, we can formally test whether the apparent differences are statistically reliable.

To that end, Table 5 reports the results of hypothesis tests involving *differences* between parameter pairs. We focus on four differences. The HH-NH (i.e., "HH-minus-NH") difference tests whether the identity of the liquidity demander matters when the liquidity supplier is an HFT; and the HN-NN difference does the same when the liquidity supplier is not an HFT. These pairings can help us understand how the identity of the liquidity demander affects the permanent price impact, while holding the liquidity supplier's

identity type constant. Likewise, the differences HH-HN and NH-NN hold the identity type of the liquidity demander constant to isolate the association between the liquidity supplier's identity and the permanent price impact.

The left half of Panel A reports tests for differences in κs attributable to the identity of the liquidity *demander*. The upper-left set of tests indicates that the differences $\kappa_{HH} - \kappa_{NH}$ tend to be significant and positive. 22% of security-days in the full sample reject the null of equality to zero in favor of a positive difference, while only 5% reject in favor of a negative difference. The overall average (median) difference is 3.06bp (0.16bp). The interpretation is that transactions initiated by HFTs have greater permanent price impact and therefore impose greater adverse selection on HFT liquidity providers than transactions initiated by non-HFTs.

The lower-left set of tests indicate that the differences $\kappa_{HN} - \kappa_{NN}$ also tend to be significant and positive across days. This difference relates to trades in which a non-HFT is the liquidity provider. In this case also, HFTs' liquidity demanding trades have greater permanent price impact than trades in which non-HFTs demand liquidity. The overall average (median) for the full sample is 2.49bp (0.11bp), which are similar to the estimates obtained when the liquidity provider is an HFT.

The results on the left side of Panel A thus indicate that the identity type of the liquidity demander matters to the permanent price impacts of trades. Trades initiated by HFTs pose greater adverse-selection than trades initiated by non-HFTs regardless of the identity type of the liquidity provider, and the day. The difference is elevated on Pre-days (and especially for low-coverage stocks) for trades in which the liquidity provider is also an HFT.

The right half of Panel A holds the liquidity demander constant and tests for differences by the identity type of the liquidity *provider*. The key finding is that the rejections rates in the positive direction (42% and 46%) here are about double those in the left half of the panel, and the negative rejection rates are also smaller. The average and median differences are larger too, almost by a factor of two compared to those in the left half of the panel. The implications are, first, that the identity of the liquidity provider has a stronger association with permanent price impact than the identity of the liquidity demander. Second, HFT liquidity providers absorb more adverse selection than non-HFT liquidity providers, regardless of the identity of the

liquidity demander.

Price priority implies that HFTs have the most aggressive quotes for the trades to which they provide liquidity; otherwise, they would not have been paired as counterparties to those trades. We can therefore conclude that HFT liquidity providers tend to offer more aggressive quotes than non-HFTs to trades that pose the greatest adverse-selection risk. The fact that non-HFT liquidity providers are not competitive for the trades with larger permanent price impacts implies that, in a relative sense, HFTs step up and shield non-HFTs from adverse-selection risk. Evidence below suggests that HFTs benefit from doing so.

5.3.3 Psis and their Differences

Table 6 reports estimates and hypothesis tests for the ψ parameters. The rather surprising finding is the high frequency with which negative ψ estimates reject the null of equality to zero compared to positive estimates when the liquidity *demander* in the transaction is an HFT. This means that HFTs actually *capture,* rather than pay, a price concession when they execute liquidity *seeking* trades. This finding is contrary to the traditional notion that liquidity demanders *pay* for immediacy with price concessions that are greater for larger orders.

These results appear on the left side of Panel A. When the liquidity provider is also an HFT, the negative (positive) rejection rate is 33% (8%). When the liquidity provider is a non-HFT, the negative (positive) rejection rate is 54% (2%). These percentages are for all days. The overall pattern suggests that when HFTs trade as liquidity demanders, the counterparty orders that supply liquidity with which they are paired are priced aggressively. Those orders could be stale, or they could be deliberately priced aggressively to obtain execution. The fact that this happens when the liquidity providers are other HFTs, and that the point estimates are even larger in magnitude for those trades, suggests that staleness is not the full explanation. Presumably, HFTs have the capability to change or withdraw orders quickly and are diligent about cancelling orders that are stale. The findings suggest that offering liquidity on favorable terms is common for *both* HFTs and non-HFTs, and that HFTs are adept at identifying and trading against such offers as liquidity demanders.

Panels B and C report corresponding results for the low- and high-coverage quartiles separately. The same patterns are apparent in these quartiles as for the full sample. For all days, the average and median

magnitudes of the price concession *captured* by HFT liquidity demanders are larger for low-coverage stocks at between 2.83bp and 1.81bp, versus 0.26bp to 0.07bp for the high-coverage stocks.

The results are different for transactions in which non-HFTs demand liquidity from HFTs, depicted in the upper right quadrant of the panel. In this case, the rejections favor positive estimates over negative estimates of ψ. In other words, non-HFTs tend to *pay* price concessions to HFT liquidity providers. Nevertheless, the magnitudes are rather modest—for the full-sample and all days, the positive (negative) rejection rates are 20% (11%). The average and median estimates are quite small at 0.27bp and 0.02bp, respectively. The corresponding statistics are somewhat larger at 1.28bp and 0.52bp for the low-coverage quartile, and they are tiny at 0.042bp and 0.016bp for the high-coverage quartile.

The frequency of positive versus negative rejections is less when the liquidity provider is a non-HFT at 12% and 11% for the full sample, and the average and medians are less positive at -0.52bp and 0.00bp, respectively. A similar pattern exists for both low- and high-coverage stocks. In these transactions, there appears to be no dominant direction in which the temporary component of price impact relates to the value of the transaction, and the magnitudes are very small on average. When non-HFTs trade with each other, it is just as likely that the liquidity demander pays as receives a price concession.

In the HHM model, $\psi = 0$ is the price concession at which the liquidity provider just covers the cost of adverse-selection. When $\psi < 0$, the liquidity provider underprices the adverse selection associated with the trade, and the liquidity demander captures profit from the trade execution. However, during the time of our sample, Nasdaq fees and rebates added costs to demanding liquidity, and added revenue to liquidity providers who qualified for rebates. Fees and rebates are not reflected in the transaction prices reported in the data, but their existence can affect those prices. Receiving rebates reduces the price concessions that liquidity providers must embed in their limit prices to cover costs. Therefore, negative estimates for ψ do not necessarily imply that the liquidity provider suffers trading losses overall. Because fees and rebates favor active liquidity providers such as HFTs, we expect that HFTs can offer more aggressive pricing than non-HFTs while not suffering trading losses. Brogaard et al. (2014) report that in the time period of the Nasdaq HFT data, fees ranged from \$0.0025 to \$0.00295, and rebates ranged from \$0.0025 to \$0.0028 per share. These correspond to about 1bp or less of the transaction value for a stock whose price is \$30/share

or more.[14]

The ψ estimates suggest that the temporary component of price impact has more to do with the identity of the liquidity demander than the liquidity provider. Formal tests of pairwise differences are reported in Table 7. The left sides of the panels focus on the identity of the liquidity demander. The rejection rates and difference estimates in the upper left of panel A for the full sample indicate that non-HFT liquidity demanders pay greater price concessions than HFT liquidity demanders when the liquidity provider is an HFT. The rejection rates are 33% vs 5% and the average difference is -3.24bp for the full sample. The rejection rates are stronger (45% vs 2%), but the average -1.06bp and median -0.30bp differences are smaller in magnitude, when the liquidity provider is a non-HFT (lower-left). These patterns hold for the low- and high-coverage quartiles as well.

The statistics reported on the right side of the panel focus on the identity of the liquidity provider. The rejection rates in the upper right, where the liquidity demander is an HFT, favor negative rejections only slightly for the sample as a whole (17% vs 11%) and in the high-coverage quartile (26% vs 21%), and rejection rates are nearly a tie for the low-coverage quartile (7% vs 8%). The average and median differences for the sample as a whole are also small at -1.38bp and -0.05bp. The identity of the liquidity provider has only a limited association with the price concession that HFTs earn when trading as liquidity demanders; i.e., HFT liquidity demanders capture profit from HFT and non-HFT liquidity providers alike. This reinforces our conjecture that (non-HFT) stale limit orders are not the entire explanation for why HFTs earn price concessions when they trade as liquidity demanders.

The identity of the liquidity provider also has a limited impact on the price concessions that *non*-HFTs pay when seeking liquidity. The lower right of the panel pertains to when the liquidity demander is a non-HFT. The estimates and rejection rates favor the inference that non-HFTs pay greater price concessions when they trade against HFT liquidity providers than non-HFT liquidity providers. However, the strength of the difference is weak here too. For the sample as a whole, the rejection rates are 16% vs 8%, and the mean and median differences are 0.80bp and 0.03bp. The inferences for the high- and low-coverage quartiles are consistent with the full sample. The average and median differences are largest for the low-coverage quartile at 2.46bp and 0.53bp, respectively. These results suggest that HFTs and non-HFTs receive similar

[14]The current fee and rebate schedule is available online at http://nasdaqtrader.com/Trader.aspx?id=PriceListTrading2.

compensation in the form of price concessions when providing liquidity to non-HFTs, despite any advantage HFTs have over non-HFTs in qualifying for rebates that could enable them to offer more aggressive pricing on their liquidity providing orders.

Summarizing, the ψ estimates tend to be negative for trades initiated by HFTs, regardless of whether the liquidity provider is an HFT or non-HFT. The estimates tend to be positive when HFTs absorb orders initiated by non-HFTs. The difference in ψ between trades initiated by HFTs versus non-HFTs suggests that HFTs are more adept than non-HFTs at targeting aggressively priced orders, regardless of whether the liquidity provider is another HFT or a non-HFT. Thus, while the results for κs above suggest that HFTs contribute to price discovery to a greater degree than non-HFTs (by initiating trades that inform market prices and by offering liquidity to trades that are more informative to market prices), the results for ψs suggest that HFTs also profit from initiating trades *selectively* against orders that are aggressively priced. We examine the possible relation between these aspects of HFTs' trading further below.

5.3.4 Trading Volume by Day

The parameters κ and ψ are the price impacts per \$10,000 traded in a given category on a given day. Even if these estimates are similar for a given category or day, the total impact could be very different if trading volume is different across categories or days. Indeed, the distribution of trading volume is different across categories and days.

Table 8 reports means and medians of three measures of trading volume by day within each counterparty pairing category. The measures are the dollar volume of individual transactions, the number of transactions, and total dollar volume for the day in thousands of shares.

For the full sample in Panel A, the number of transactions and the transaction sizes are both elevated on the Pre-day and especially so on the Post-day and for trades initiated by non-HFTs compared to the same measures on the week-after days. This pattern is evident in the means and medians of both the number of transactions and the transaction size on the Post-day. The differences on the Post-day are the largest. The average total dollar volume for the NH pairing on the Post-day (Pre-day) is greater than the WA-Post (WA-Pre) day by a factor of 2.01 (1.35). Likewise for the NN pairing, average total volume is greater on the Post-day (Pre-day) versus the WA-Post (WA-Pre) day by a factor of 2.02 (1.30).

Trading volume initiated by HFTs that is absorbed by non-HFT liquidity providers is also greater on the Post-day (Pre-day), but by a factor of 1.70 (1.21) in relation to the WA-Post (WA-Pre) day. The scale factors are similar for the HH pairing.

The elevated trading volume around earnings announcements means that the economic significance of the κ and ψ differences reported earlier are larger around earnings announcements, and especially on the Post-day. We showed earlier that HFTs absorb greater adverse selection than non-HFTs by comparing κs that are estimated on a per \$10,000 transaction basis. Pairing that finding with the 2-times elevated scale in dollar trading volume initiated by non-HFTs doubles the dollar value of the divergence between adverse selection absorbed by HFTs versus non-HFTs around earnings announcements versus the week after. Though smaller at 1.7, the scale of trading initiated by HFTs is also elevated on the Post-day, implying that the total profit HFTs capture by targeting attractively priced orders that provide liquidity is also higher on the Post-day.

The patterns in trading volume for the high-coverage quartile in Panel C are similar to the full sample, with the surge in dollar volume on the Post-day being strong for trades initiated by both non-HFTs and HFTs. The figures in Panel B for the lowest coverage quartile indicate that trading volume is higher on the Post-day for these stocks as well. However, the increase is driven by a larger number of trades and not by the dollar values of transactions being larger.

5.3.5 Regressions for Key Differences

Two interesting results in the subsections above are that when HFTs provide liquidity they absorb trades that pose greater adverse selection than non-HFTs, and that HFTs are adept at targeting attractively priced limit orders when they trade as liquidity demanders. In this section, we use regression to examine whether those findings depend on the timing of the trades in relation to earnings announcements, the number of analysts following, and attributes of the earnings announcements such as the dispersion in analyst forecasts and the magnitude of earnings surprises.

The first three columns of Table 8 report regressions of $\kappa_{NH} - \kappa_{NN}$ for all security-days. This difference represents the incremental adverse selection absorbed by HFTs over non-HFTs when providing liquidity to non-HFTs. Positive coefficients indicate that HFTs provide liquidity to trades that pose a greater adverse

selection risk to the market than the trades to which non-HFTs provide liquidity.

The independent variables in the first regression are indicators for whether the stock is in the lowest or highest quartile of analyst coverage and indicators for whether the day is the pre-announcement day or the post-announcement day. The regression in column (2) adds the earnings surprise and analyst forecast dispersion as independent variables then, finally, interactions between those variables and the indicators are added in column (3). In columns (2) and (3), the sample is limited to earnings announcements with two or more analysts in order to compute forecast dispersion.

The interpretation of the intercept is as a benchmark for an "unremarkable" stock-day. It is the average value of $\kappa_{NH} - \kappa_{NN}$ for stocks in the middle two quartiles of analyst coverage, on the week-after days following an earnings announcement with a zero earnings surprise and zero analyst dispersion. Since each earnings announcement contributes four security-day observations to the sample, standard errors are clustered by earnings announcement to account for possible correlation between disturbances that relate to each earnings announcement.

The intercepts are positive and significant in all three regressions, and largest in column (1) that includes the announcements with zero or only one analyst forecast. The point estimates indicate that on "unremarkable" stock days, HFTs absorb orders that move prices permanently by 2.7bp to 4.8bp *more* than non-HFTs when providing liquidity to non-HFTs. The significant negative coefficients on the high-coverage quartile dummy are close in magnitude to the intercepts, suggesting there is effectively no difference in κs for these stocks. Indeed, the sum of the intercept and high-coverage coefficient is not significant, as reported in the third row from the bottom of the table. However, the coefficients on the low-coverage dummy are positive, significant, and even larger than the intercepts. Adding these coefficients to the intercepts indicates that for low-coverage stocks, HFTs absorb orders whose permanent price impact is nearly 7.8bp to 8.6bp *greater* than those absorbed by non-HFTs when providing liquidity to non-HFTs.

The coefficient on the absolute value of the earnings surprise is also positive and significant. Interestingly, the dummies for the pre- and post-announcement days are not significant, meaning that the extra adverse selection absorbed by HFTs over non-HFTs *per dollar traded* is no different on earnings days. We observed above that trading is heavier on those days, so the aggregate adverse selection absorbed by HFTs versus non-

HFTs is larger on those days. Taken together, these findings suggest that comparing HFTs and non-HFTs as providers of liquidity to non-HFT liquidity demanders, HFTs absorb orders that pose greater adverse selection risk in stocks with low and moderate analyst coverage; and they absorb no less than non-HFTs for stocks in the highest coverage quartile. In addition, there is no evidence that HFTs back away from absorbing adverse selection around earnings announcements. In fact, they absorb even more per-dollar traded around announcements with larger absolute surprises.

The dependent variable in columns (4) - (6) is $\kappa_{HH} - \kappa_{HN}$, which focuses on trades initiated by HFTs rather than non-HFTs as examined in columns (1) - (3). We noted in Table 4 that the liquidity demanding orders of HFTs pose adverse selection risk to the market. The dependent variable in columns (4) - (6) is the incremental adverse selection absorbed by HFTs over non-HFTs when providing liquidity to HFTs. Negative values of this difference and coefficients in the regression indicate that the adverse selection posed by HFTs is absorbed to a greater degree by non-HFT liquidity providers than by HFT liquidity providers. Positive values indicate that other HFTs absorb more than non-HFTs.

In column (4) where all announcements are included, the intercept is positive and significant. The coefficient on the high-coverage dummy is negative, significant, and again about the same magnitude as the intercept, suggesting that HFTs absorb more adverse selection than non-HFTs when providing liquidity to other HFTs for stocks outside the highest quartile of coverage. The low-coverage dummy is not significant; and its standard error is large enough that its sum with the intercept is not significant. In columns (5) and (6) where the sample is restricted to announcements with two or more earnings forecasts, the intercept is not significant. While the coefficient on the high-quartile dummy is negative and significant, the sum of this coefficient and the intercept is not significant. Thus, outside of the announcements with little or no analyst coverage in column (4), there is no difference in the adverse selection absorbed by HFTs versus non-HFTs from orders initiated by HFTs. In particular, it does *not* appear that HFTs impose more adverse selection on non-HFT liquidity providers than on their fellow HFT liquidity providers. If anything, the reverse is true for low-coverage stocks.

Taken together, the regression results in Table 8 suggest that HFTs contribute to the liquidity environment by having a greater incidence than non-HFTs of absorbing adverse selection associated with trades initiated by non-HFTs and other HFTs alike. This effect is greater in scale around earnings announcements because

the dollar volume of trading is heavier.

Table 9 turns to the temporary components. In columns (1) - (3), the dependent variable is $\psi_{HN} - \psi_{NN}$, which is the differential price concession that HFTs pay over non-HFTs when trading with a non-HFT liquidity provider. We observed earlier that the average difference is negative in the full sample, indicating that HFTs are more adept than non-HFTs at selecting liquidity providing orders that are priced to yield low or negative price concessions. Negative signs of regression coefficients mean that a regressor is associated with HFT liquidity demanders obtaining more attractive prices than those paid by non-HFTs who demand liquidity.

The intercepts in columns (2) and (3) are a negative and significant 1.6bp indicating that HFTs capture a greater discount than non-HFTs in the "unremarkable" benchmark case. In those columns, the discount is smaller for stocks in the high-coverage quartile by 1.3bp yielding a negligible and insignificant difference for high-coverage stocks. However, the discount in the low-coverage quartile is a significant additional 3.0bp in columns (2) and (3), for a total of approximately 4.6bp. None of the other coefficients in columns (2) and (3) are significant, indicating that whatever advantage HFTs have in obtaining attractive prices from non-HFTs outside the quartile of high-coverage stocks is not associated with earnings days or the surprise and dispersion attributes of the earnings announcements.

The findings are different in column (1). Although the intercept and low-coverage coefficients are both negative, they are not significant. The high-coverage coefficient is positive 0.95bp and significant, suggesting that for the full sample that includes very low coverage announcements, HFTs do not capture a statistically significant discount in price concessions, and in fact pay an insignificant premium price concession to non-HFT liquidity providers of 0.34bp (intercept plus coefficient) for high-coverage stocks. The differences between the results for columns (2) and (3) and those for column (1) suggest that HFTs' advantage even for low-coverage stocks evident in columns (2) and (3) disappears when the very-lowest-coverage stocks are included in the analysis in column (1). HFTs have an advantage in some subsets of the cross-section, but not uniformly.

In columns (4) - (6), the dependent variable is $\psi_{HH} - \psi_{HN}$, which focuses on whether price concessions for trades initiated by HFTs differ depending on whether the liquidity provider is an HFT or a non-HFT.

A negative value for this difference and coefficients in the regression mean that HFTs' discounted price concessions come at the greater expense of non-HFT liquidity providers than HFT liquidity providers.

For the full sample in column (4), the intercept is a significant and negative 3.0bp, and the high-coverage quartile coefficient is a significant and positive 3.3bp. This suggests that HFTs capture larger discounts when non-HFT (vs HFT) liquidity providers absorb their trades for stocks outside the highest-coverage quartile. If greater rebates for HFT versus non-HFT liquidity providers were an important source of aggressive pricing, discounts would be larger for HFT liquidity providers and the intercept would be positive. Within the highest-coverage quartile, the price concessions HFTs pay are about the same for both HFT and non-HFT liquidity providers (intercept plus coefficient). The announcement day dummy is also negative and marginally significant indicating that the price concession discount is larger on earnings days. The additional regressors in columns (5) and (6) are not significant, and seem only to dilute the significance of the intercept.

Taken together, the regression results indicate that HFTs absorb more informative trades when providing liquidity (than do non-HFTs), and they initiate trades that target more attractively priced orders (than do non-HFTs). While we cannot observe the information that HFTs learn from their trading, these findings are consistent with the possibility that HFTs use information learned from their liquidity providing trades to profit from their liquidity demanding trades. Having built infrastructure for which the marginal cost of trading is very low, HFTs could be better able than non-HFTs to infer information from their liquidity providing trades to (re)trade further on the information until it becomes more fully reflected in prices. This possibility is examined next.

5.3.6 CETV Event and Cross-Sectional Regressions

We begin this section with an example. All except one of the earnings surprises in our sample range from -29% to 13%. The exception is the February 24, 2009 earnings announcement of Central European Media Enterprises Ltd. (ticker CETV), whose earnings surprise was -145%. The global financial crisis decimated TV ad spending by CETV's clients in late 2008 and early 2009, which depressed revenue and earnings, and caused a suspension in company guidance beyond the following quarter.[15]

[15] See the earnings call transcript at https://seekingalpha.com/article/134262-central-european-media-enterprises-ltd-q1-2009-earnings-call-transcript. On March 23, 2009, Time Warner purchased a 31% stake, and by 2013 owned a controlling interest in CETV.

Figure 1 is a time series plot of the total price impacts, $\kappa + \psi$, for three counterparty pairings beginning on day -16 and ending on day 36 relative to the CETV announcement, which occurs between days 0 and 1. The green line that lies mostly below the others is the total price impact for trades in which non-HFTs are both the demander and provider of liquidity (NN). The blue line, mostly in the middle, represents transactions in which an HFT demands liquidity from a non-HFT liquidity provider (HN). The red line that lies mostly above the others is the total price impact for trades initiated by non-HFTs and absorbed by HFT liquidity providers (NH). There is economically significant variation in all three lines over this time period. The total price impact between non-HFT counterparties (in green) varies between zero and 50bp, while the price impact for trades where non-HFTs demand liquidity from HFTs (in red) varies from 10bp to 140bp. Variation in the blue and green lines tends to coincide. However, the largest peaks in the red line appear to lead the largest peaks in the blue line.

Figure 2 plots only the permanent component of price impact, κ, for the same set of counterparty pairings. The magnitudes are larger, but the shapes of the lines are similar to those in Figure 1 indicating that the permanent component dominates in determining the total price impact, and that the temporary component often contributes negatively to the total. Figure 3 plots only the temporary component. On most days, the temporary component associated with trades between non-HFTs (in green) is positive, meaning that non-HFTs pay for liquidity provided by other non-HFTs. In contrast, the temporary component is often negative for the blue and red lines. The red line is large and negative in the approximately 10 days following the earnings announcement. When trading as liquidity providers, HFTs were effectively paying non-HFTs to *take* liquidity. Figures 2 and 3 together suggest that HFTs were underpricing the elevated adverse selection they absorbed from non-HFTs in the days following the earnings announcement. Figure 4 reproduces the red line plot of ψ_{NH} in Figure 3 with its 95% confidence bands. While zero is within the confidence bands for most days, the negative spikes on 10 of the days in the graph are significantly different from zero.

In all three figures, peaks in the red line appear to precede peaks in the blue line. The interpretation of this pattern in relation to Figure 2 is that periods in which HFTs absorb a high degree of adverse selection when supplying liquidity to non-HFTs (the red line) are followed by periods when HFTs impose greater adverse selection in their liquidity demanding trades with non-HFT liquidity suppliers (the blue line).

Redeploying information gleaned from the trading with others is a way for HFTs to benefit from absorbing adverse selection—like a self-made rebate.

In this exceptional CETV case, the effect appears *across* calendar days. To examine whether the effect itself and a lag across days are general features of HFT trading around earnings announcements, Table 10 reports regressions for the full sample that test whether κ_{NH} leads ψ_{HN} (i.e., "red leads blue") between the two days that straddle the earnings announcement (Pre and Post), and separately between the two corresponding days in the week after the announcement (WA Pre and WA Post). Each earnings announcement appears as a single observation in this analysis, so the reason that led us to cluster standard errors by earnings announcement in the regressions in Table 8 is not present here. However, as a clear outlier, the 2/24/2009 earnings announcement of CETV described above is excluded from this analysis.

We consider the "normal" pair of days in the week after announcements first. Column (1) examines whether ψ_{HN} on the WA-Post day is related to κ_{NH} on the WA-Pre day. The ψ_{HN} on the WA-Pre day is included as a control to account for persistence in ψ_{HN}. The estimates indicate very strong and significant persistence, with a coefficient of 0.76. The coefficient on the WA-Pre day κ_{NH} is negative and significant, implying that HFTs profit more (i.e., ψ_{HN} is lower) on days following those in which they absorb greater adverse selection (i.e., "red leads blue"). The adjusted R^2 is 0.53.

The regression in column (2) adds κ_{HH} on the WA-Pre day as a regressor to account for learning from trades absorbed from other HFTs. We also include κ_{NH} and κ_{HH} on the WA-Post day itself to capture the contemporaneous re-use of information gleaned from informative trades on the same day. Persistence in ψ_{HN} is again strong in this regression, with a coefficient of 0.68. However, the coefficient on the WA-Pre day κ_{NH} that was significant in column (1) is not significant here. Instead, the contemporaneous κ_{HH} is negative and significant, and larger in magnitude than the coefficient on κ_{NH} in column (1). After accounting for persistence, the significant incremental source contributing to profitable liquidity demanding trades by HFTs on the WA-Post day are informative trades absorbed from other HFTs on the same day. The significant coefficient in column (1) and the similarity of the adjusted R^2 at 0.60 suggests that the WA-Pre day κ_{NH} is acting as a proxy for, and is perhaps a determinant of, κ_{HH} on the WA-Post day.

The intercept not being significant suggests that after accounting for persistence, the incremental

profitability of liquidity-demanding trades would trend toward zero were it not for information gleaned from trades with other HFTs (i.e., without positive κ_{HH} on the WA-Post day). In other words, on average across the WA day pairs, absorbing adverse selection from other HFTs seems essential to HFTs profiting from their liquidity demanding trades. These findings are consistent with the hypothesis that, on "normal" days, HFTs profit from the reuse of information gleaned from their trading with other-HFTs.

Columns (3) and (4) report the same regressions as columns (1) and (2) but for the days on either side of earnings announcements. The results are different. In column (3) the coefficient on κ_{NH} is again negative and highly significant, and it remains so in column (4) after adding the other κ regressors—i.e., "red leads blue" strongly in the two-days around earnings announcements. In fact, the coefficients on all the κs are negative and significant, *except* the contemporaneous κ_{HH}; and the ones with coefficients of larger magnitude and significance are those relating to trades absorbed from *non*-HFTs. The trades that HFTs absorb from non-HFTs on both the Pre and Post days, contribute most to the profitability of liquidity demanding trades by HFTs on the Post Day. Compared to the week-after results in columns (1) and (2), it seems that around earnings announcements, HFTs learn more from absorbing trades initiated by non-HFTs than by other HFTs; the trades initiated by other HFTs on the Pre day contribute also, but less.

Despite these stronger regression relations with adverse selection absorbed from trading with others, the reliability with which HFTs turn adverse selection absorbed into price concessions captured from liquidity demanding trades is much less around earnings announcements than on the "normal" week-after days. The adjusted R^2 in the column (4) regression is only 0.23 versus 0.60 for the WA days.

Unlike the WA days, there is negative persistence in ψ_{HN} from the Pre to the Post day, which is highly significant. This means that Post days with the most profitable liquidity demanding HFT trades (controlling for the other price impacts) follow Pre days having the least profitable liquidity demanding trades. We do not have a simple explanation for this finding. Perhaps all we can say from the contrast with the strong positive persistence exhibited in the "normal" WA days is that earnings announcements disrupt the usual dynamics of learning and profitability of HFT trading. The algorithms used by HFTs for trading may be in transition from a pre-earnings environment to a new post-earnings environment. The reduction in the adjusted R^2 is also consistent with the interpretation that earnings announcements are disruptive to how HFTs infer and trade on information.

Taken together, the results in Table 10 support the hypothesis that HFTs profit in their liquidity demanding trades for having learned from the adverse selection they absorb when providing liquidity. However, the counterparties from whom they learn are different around earnings announcements than in the week after. Around earnings announcements, the important source of the information tends to be trades initiated by non-HFTs, whereas trades initiated by other HFTs are important on the "normal" week-after days. This is consistent with non-HFTs being the main source of adverse selection and informative trades when information asymmetries concern significant fundamental information such as earnings and the other information that is announced in connection with earnings.

6 Conclusion

We provide a simple approach to estimating price impact and its components based on a model attributable to Hasbrouck (1991a, 1991b, 1993) and Hendershott and Menkveld (2014) that is rather general and descriptive of many securities market settings. The approach can easily handle situations in which transactions are categorized in some way, such as by trader type, order type, trading platform, or trade routing. Estimates of the temporary and permanent components for the different categories can be obtained jointly in a consistent framework, and hypothesis testing can be done using the delta method. The approach does not require that the data conform to particular distributional assumptions. We show that the state space approach used in prior work fails to satisfy Normality assumptions when applied to transaction price data, and that the simpler approach provides reliable estimates where state space estimation does not. The new approach uses OLS, which is fast enough for generating estimates of illiquidity for asset pricing research from intraday data to use in place of simpler constructs based on closing prices or quotes and daily volume.

We apply the approach to millisecond-level transactions data supplied by Nasdaq to trades categorized by counterparty pairings between HFTs and non-HFTs and their roles as liquidity demander or liquidity supplier. The permanent and temporary components for the four possible pairings generate eight distinct price impacts, which we estimate jointly for the days bracketing earnings announcements and the same weekdays one week later. This dataset and exercise mirror the type of data and questions that regulators and market operators could be interested in analyzing with regulatory or proprietary data that identifies

categories of trades in various ways.

The OLS approach performs well when compared to a full-information GMM approach, and it yields some interesting economic results. The permanent and temporary price impacts depend on counterparty pairings in opposite ways. The permanent component depends most strongly on the identity of the liquidity provider, and tends to be large when the liquidity provider is an HFT and small when the liquidity provider is a non-HFT. HFTs absorb liquidity seeking trades that pose greater adverse-selection risk than the liquidity seeking trades absorbed by non-HFTs. The temporary component depends most strongly on the identity of the liquidity demander. It tends to be positive when the liquidity demander is a non-HFT and *negative* when the liquidity demander is an HFT. HFTs capture rather than pay a price concession in their liquidity seeking trades. HFTs appear to identify instances in which liquidity is provided at attractive prices by both non-HFTs and other HFTs. These attributes of HFT trading are related in the cross-section of earnings announcements. HFTs capture larger price concessions from liquidity seeking orders following (and contemporaneous with) days when they absorb more informative trades as liquidity providers. This association suggests that HFTs both intermediate and leverage the process by which information is incorporated into market prices. HFTs provide liquidity to informative trades and reuse the information in subsequent trades that they initiate.

7 Appendix A

Proof of Proposition 1: To simplify notation, the proof considers the case in which s_t follows an ARMA(2,1) plus $\psi' x_t$. This case is sufficient to infer the implications for a general ARMA(p, q). In the ARMA(2,1) case, Equations (6) -(8) in the text are:

$$s_t = \phi_1 s_{t-1} + \phi_2 s_{t-2} + \psi' x_t + \eta_t + \theta \eta_{t-1} \tag{23}$$

$$m_t = m_{t-1} + \kappa' \epsilon_t + \xi_t \tag{24}$$

$$p_t = m_t + s_t, \tag{25}$$

where x_t and ϵ_t are $(k \times 1)$ vectors. Writing these equations in first differences yields

$$ds_t = \phi_1 \, ds_{t-1} + \phi_2 \, ds_{t-2} + \psi' dx_t + \eta_t - (1-\theta)\eta_{t-1} - \theta \eta_{t-2} \tag{26}$$

$$dm_t = \kappa' \epsilon_t + \xi_t \tag{27}$$

$$dp_t = \kappa' \epsilon_t + ds_t + \xi_t, \tag{28}$$

where $dx_t \equiv x_t - x_{t-1}$ and where we have substituted from Equation (27) for the dm_t term into Equation (28). Next, we substitute for lags of ds_t in terms of observables. Lagging Equation (28) by one, then by two, and multiplying by the corresponding ϕ_j yields

$$\phi_1 ds_{t-1} = \phi_1 \, dp_{t-1} - \phi_1 \, \kappa' \epsilon_{t-1} - \phi_1 \, \xi_{t-1} \tag{29}$$

$$\phi_2 ds_{t-2} = \phi_2 \, dp_{t-2} - \phi_2 \, \kappa' \epsilon_{t-2} - \phi_2 \, \xi_{t-2}. \tag{30}$$

Substituting these into Equation (26) yields

$$ds_t = \phi_1 dp_{t-1} + \phi_2 dp_{t-2} - \phi_1 \kappa' \epsilon_{t-1} - \phi_2 \kappa' \epsilon_{t-2} - \phi_1 \xi_{t-1} - \phi_2 \xi_{t-2} + \psi' dx_t + \eta_t - (1-\theta)\eta_{t-1} - \theta \eta_{t-2}. \tag{31}$$

Substituting this into Equation (28) yields

$$dp_t = \phi_1 \, dp_{t-1} + \phi_2 \, dp_{t-2} + \psi' dx_t + \kappa'(\epsilon_t - \phi_1 \epsilon_{t-1} - \phi_2 \epsilon_{t-2}) + \eta_t^*, \tag{32}$$

where $\eta_t^* = (\xi_t + \eta_t) - (\phi_1 \xi_{t-1} + (1-\theta)\eta_{t-1}) - (\phi_2 \xi_{t-2} + \theta \eta_{t-2})$ is an AR(2) disturbance. In general, if s_t follows an ARMA(p,q) process, then η_t^* is autoregressive of order $\max\{p, q\}$. Mutual uncorrelatedness of ξ, η and ϵ at all lags implies that η_t^* is uncorrelated with ϵ_t at all lags as well.

Using the VAR model, Equation (5), we write

$$\epsilon_t - \phi_1\epsilon_{t-1} - \phi_2\epsilon_{t-2} = \left(x_t - \sum_{s=1}^{n} B_s x_{t-s}\right) - \phi_1\left(x_{t-1} - \sum_{s=1}^{n} B_s x_{t-s-1}\right) \tag{33}$$
$$- \phi_2\left(x_{t-2} - \sum_{s=1}^{n} B_s x_{t-s-2}\right)$$
$$= x_t - (\phi_1 I_k + B_1)x_{t-1} - (\phi_2 I_k + B_2 - \phi_1 B_1)x_{t-2} \tag{34}$$
$$- \sum_{s=3}^{n}(B_s - \phi_1 B_{s-1} - \phi_2 B_{s-2})x_{t-s}$$
$$+ (\phi_1 B_n + \phi_2 B_{n-1})x_{t-n-1} + \phi_2 B_n x_{t-n-2},$$

where I_k is the $k \times k$ identity matrix. Substituting Equation (34) into Equation (32) yields

$$dp_t = \phi_1\, dp_{t-1} + \phi_2 dp_{t-2} + \psi'(x_t - x_{t-1}) + \kappa'\Big[x_t - (\phi_1 I_k + B_1)x_{t-1} - (\phi_2 I_k + B_2 - \phi_1 B_1)x_{t-2} \tag{35}$$
$$- \sum_{s=3}^{n}(B_s - \phi_1 B_{s-1} - \phi_2 B_{s-2})x_{t-s} + (\phi_1 B_n + \phi_2 B_{n-1})x_{t-n-1} + \phi_2 B_n x_{t-n-2}\Big] + \eta_t^*$$
$$= \phi_1\, dp_{t-1} + (\psi' + \kappa')x_t - [\psi' + \kappa'(\phi_1 I_k + B_1)]\, x_{t-1} \tag{36}$$
$$+ \phi_2\, dp_{t-2} - \kappa'(\phi_2 I_k + B_2 - \phi_1 B_1)x_{t-2} - \sum_{s=3}^{n}\kappa'(B_s - \phi_1 B_{s-1} - \phi_2 B_{s-2})x_{t-s}$$
$$+ \kappa'(\phi_1 B_n + \phi_2 B_{n-1})x_{t-n-1} + \phi_2\kappa' B_n x_{t-n-2} + \eta_t^*.$$

It is clear from inspection that the highest-order lags of dp and x on the right-hand side of Equation (36) are determined as p and $n + p$, respectively, where p is the AR order of the ARMA specification for s_t and n is the lag order of the VAR.

The rest of the proof involves matching the coefficients in Equation (12) in the text and Equation (36) here. Matching the first terms of both equations yields $\phi_1 = a_1$. Matching the second and third terms yields

$$C' = \psi' + \kappa' \tag{37}$$
$$D_1' = -\psi' - \phi_1\kappa' - \kappa' B_1 = -\psi' - a_1\kappa' - \kappa' B_1, \tag{38}$$

which imply that $\psi' = C' - \kappa'$, and therefore

$$D_1' = \kappa' - C' - a_1\kappa' - \kappa' B_1 \tag{39}$$
$$D_1' + C' = \kappa'[(1 - a_1)I_k - B_1] \tag{40}$$
$$\kappa' = (D_1' + C')[(1 - a_1)I_k - B_1]^{-1}, \tag{41}$$

and

$$\psi' = C' - (D_1' + C')[(1 - a_1)I_k - B_1]^{-1}. \tag{42}$$

Finally, noting that $u_t = \eta_t^*$ establishes that u_t is autoregressive of order $\max\{p, q+1\}$ and uncorrelated with ϵ_t at all lags. $\square$

Additional Equations: When the individual elements of x_t are not observable to market participants, the equation for price dynamics in Equation (36) is modified as follows:

$$\begin{aligned} dp_t = {} & \phi_1\, dp_{t-1} + (\psi' + \kappa')x_t - [\psi' + \kappa'(\phi_1 + b_1)]\, x_{t-1} \\ & + \phi_2\, dp_{t-2} - \hat{\kappa}(\phi_2 + b_2 - \phi_1 b_1)\hat{x}_{t-2} - \sum_{s=3}^{n} \hat{\kappa}(b_s - \phi_1 b_{s-1} - \phi_2 b_{s-2})\hat{x}_{t-s} \\ & + \hat{\kappa}(\phi_1 b_n + \phi_2 b_{n-1})\hat{x}_{t-n-1} + \phi_2 \hat{\kappa} b_n \hat{x}_{t-n-2} + \eta_t^*. \end{aligned} \tag{43}$$

If $s_t = \phi_1 s_{t-1} + \psi' x_t + \eta_t$ then the disturbance to the regression in Proposition 1 (denoted in the proof by η_t^*) given by

$$u_t = (\xi_t + \eta_t) - (\phi_1 \xi_{t-1} + \eta_{t-1}), \tag{44}$$

whose variance and first-order serial covariance are

$$\mathrm{Var}[u_t] = (1 + \phi_1^2)\sigma_\xi^2 + 2\sigma_\eta^2 \tag{45}$$

$$\mathrm{Cov}[u_t, u_{t-1}] = -\phi_1 \sigma_\xi^2 - \sigma_\eta^2. \tag{46}$$

Solving for σ_η^2 and σ_ξ^2 yields:

$$\sigma_\eta^2 = \frac{\phi_1 \mathrm{Var}[u_t] + (1 + \phi_1)\,\mathrm{Cov}[u_t, u_{t-1}]}{2\phi_1 - (1 + \phi_1^2)} \tag{47}$$

$$\sigma_\xi^2 = \frac{\mathrm{Var}[u_t] - 2\sigma_\eta^2}{1 + \phi_1^2}. \tag{48}$$

These formulas can be used to estimate σ_η^2 and σ_ξ^2 based on moments of the regression residuals and the regression estimate of ϕ_1 (i.e., a_1 in Proposition 1).

8 Appendix B

Panel A of Table 11 reports comparisons across estimation methods for the ϕ_1 parameter, which captures dynamics of the temporary component of price changes that are not related to the variables in x_t. The first two columns report the percentages of the security-days in the sample for which the estimate is positive or negative and the hypothesis that $\phi_1 = 0$ is rejected at the 5% level. Only 1%-2% of the security-days reject in favor $\phi_1 > 0$, while 62%-65% reject in favor of $\phi_1 < 0$, meaning that s_t tends to be of opposite sign to s_{t-1}. The rejection rates are nearly uniform across methods, except that OLS-Zero rejects slightly less often than the other methods when the estimate of ϕ_1 is negative. The average and median estimates appear in columns three and four, and they are similar across the methods at -0.10 and -0.11.

Columns five and six report the percentages of security-days for which the inference concerning the hypothesis that $\phi_1 = 0$ by each method matches the inference drawn using the "full-information" GMM alternative. When the GMM estimate is positive, the OLS-Prop and OLS-Resid approaches match GMM for 86% of security-days versus an 82% match rate for OLS-Zero. This is a small set of estimates, however, because the vast majority of ϕ_1 estimates are negative. When the GMM estimates are negative, the OLS-Prop and OLS-Resid inferences match GMM 92% of the time, while OLS-Zero matches 86% of the time. The root mean square errors (RMSE) and root median square errors (RMedSE) between the GMM estimates and the other methods are reported in the last two columns. These numbers are similar across all three methods. The RMSEs are about a third of the magnitude of the mean ϕ_1 estimate, and the RMedSEs are tiny at 1/10 the size of the RMSEs.

Panel B of Table 11 reports results for all four of the κ parameters using the same format as Panel A. In all four counterparty pairings, the vast majority (68%-92%) of rejections occur when the estimate of κ is positive, while less than 1% occur when the estimate is negative. This means that permanent price changes tend strongly to occur in the same direction as the order that demands liquidity. The units are in basis points, or equivalently, dollars per \$10,000 in traded value. The estimates are largest for κ_{HH} with GMM mean 14.85bp and median 2.29bp, and smallest for κ_{NN} with GMM mean 7.13bp and median 1.43bp. Thus, when a non-HFT supplies liquidity to another non-HFT, the adverse-selection cost borne by the liquidity demander is \$7.13 per \$10,000 value traded on average.

The rejection rates and mean and median estimates (across stock days) in the first four columns are similar for three of the four estimation methods; the exception is OLS-Zero where the estimates of κ are smaller. Comparing the OLS methods to GMM at the security-day level, the next four columns show that the OLS-Prop method is uniformly a better match to the GMM estimates and inferences than are the other methods. The other two methods are close behind except for the κ_{HH} parameter where the RMSE and RMedSE are considerably higher for OLS-Resid and OLS-Zero compared to OLS-Prop. For κ_{HN}, the match percentage when the GMM estimate is negative is also low for OLS-Zero. Overall, the results in Panel B suggest that OLS-Prop does well at matching GMM inferences, and better than the other methods in matching GMM's point estimates, especially for κ_{HH}.[16]

[16]The small percentages reported for all methods for κ_{NN} in matching GMM when the GMM estimate is less than zero

Panel C reports corresponding statistics for the ψ parameters. For most security-days the hypothesis that $\psi_{HN} = 0$ is rejected when the estimate is *negative*. The GMM rejection frequency is 56% when negative versus only 2% when the estimate is positive. The percentages are similar for OLS-Prop and OLS-Resid. This means that HFTs capture, rather than pay, a temporary price concession that scales with quantity when they execute liquidity *seeking* trades against non-HFT liquidity providers. The average and median compensation is 2.65bp and 0.32bp, respectively.

The results are different for counterparty pairings in which the liquidity demander is not an HFT. When non-HFTs demand liquidity from other non-HFTs, the ψ_{NN} GMM estimate is just as likely to be positive as negative (both at 14%) when rejecting the hypothesis that $\psi_{NN} = 0$. However, when the liquidity provider is an HFT, the GMM rejections favor the inference that non-HFTs *pay* a price concession to the HFT liquidity provider. The rejection rate is 20% when the estimate of ψ_{NH} is positive and 12% when the estimate is negative. These findings support the traditional notion that liquidity demanders typically pay a price concession, but only when the liquidity provider is an HFT.

This asymmetry appears in the main results in the paper, where we describe it in greater detail. The focus here is comparing OLS-Prop to GMM and the other OLS-based approaches to estimation. As above with the κ parameters, the OLS-Prop estimates and inferences for the ψ parameters perform the best in matching the GMM inferences and estimates. The performance of the other OLS methods lag behind to a greater degree with the ψs than with the κs above. For all four ψ parameters, the OLS-Zero approach is biased upward in its estimates and tends to reject too often (too rarely) when the GMM estimate is positive (negative). This is especially true for ψ_{NH} and ψ_{NN} where the OLS-Zero estimates and inferences overstate the magnitude and significance of the price concessions non-HFTs pay when seeking liquidity. The same direction of bias seems to exist but to a lesser extent with OLS-Resid. For example, the ψ_{NH} positive-estimate rejection rates are higher at 40% for OLS-Zero and 25% for OLS-Resid, than for GMM and OLS-Prop, which are both at 20%. The negative-estimate rejection rates are smaller for OLS-Resid and OLS-Zero than for GMM and OLS-Prop. The average and median point estimates are likewise greater for OLS-Resid and OLS-Zero (positive means of 0.38bp and 1.94bp, respectively) than they are for GMM and OLS-Prop (negative means of -0.86bp and -0.20bp, respectively).

Altogether, OLS-Prop matches the estimates and inferences of GMM rather well. The overidentifying restrictions GMM brings to the estimation do not change the estimates and inferences very much from those of OLS-Prop. OLS-Resid is a close second best, and OLS-Zero does poorly, especially with the ψ parameters, which are biased upward and toward rejecting in the positive direction. The differences between the OLS-based methods is whether and how conditional expectations of x_t are incorporated into the estimation. The results suggest that care taken in modeling expectations of x_t matters to the quality of the estimates and inferences.

reflects only a few security-days. Very few days have negative GMM estimates for any of the four κ parameters.

Table 1: Summary Statistics

The table reports summary statistics for all four stock days defined in the paper (i.e., Pre, Post, WA-Pre and WA-Post) for 823 earnings announcements from 120 distinct stocks in our sample that spans 2008-2009. Each row corresponds to a statistic that is computed from Nasdaq daytime trading hours for each stock-day. For each row, the numbers reported across the columns correspond to the distribution of the statistic across stock-days. Statistics include the dollar volume of transactions within each counterparty pairing: HH, HN, NH, and NN, where H represents HFT and N represents non-HFTs. The first letter identifies the liquidity demander and the second letter identifies the liquidity provider. Also included are statistics on the transaction returns, the number of transactions by counterparty pairing, equity market capitalization at the end of December 2007, and analyst coverage taken from I/B/E/S.

	Median	Mean	Standard deviation	1st percentile	25th percentile	75th percentile	99th percentile
Dollar volume HH ($)	761,782	15,948,604	54,888,290	6,753	109,326	11,722,029	241,017,598
Dollar volume HN	4,043,612	29,415,150	96,184,331	41,269	847,259	24,731,626	467,411,748
Dollar volume NH	1,856,180	29,676,621	91,006,027	23,888	292,077	21,581,772	408,913,358
Dollar volume NN	7,016,318	36,589,210	103,331,394	161,090	1,705,066	32,708,586	475,522,937
Dollar volume All	13,857,100	111,629,585	332,655,730	265,547	3,266,132	95,893,555	1,612,497,009
Number of transactions HH	298	1,542	3,131	10	49	1,890	13,442
Number of transactions HN	1,026	2,407	3,630	42	303	3,237	16,879
Number of transactions NH	562	2,036	3,720	18	121	2,510	16,308
Number of transactions NN	1,350	2,516	3,280	102	573	3,289	15,677
Number of transactions All	3,362	8,500	13,166	242	1,119	11,286	60,840
Mean transaction return	0.000%	0.000%	0.005%	-0.015%	0.000%	0.001%	0.015%
Median transaction return	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Std. dev. of transaction returns	0.062%	0.092%	0.103%	0.012%	0.029%	0.117%	0.422%
Mean absolute transaction return	0.031%	0.045%	0.045%	0.006%	0.014%	0.059%	0.204%
Median absolute transaction return	0.000%	0.018%	0.031%	0.000%	0.000%	0.027%	0.148%
Std. dev. of absolute transaction returns	0.053%	0.079%	0.093%	0.010%	0.025%	0.101%	0.391%
Market capitalization (millions $)	2,579	27,530	56,574	218	1,442	29,339	324,786
Number of Analysts	11.00	12.60	7.79	0.00	7.00	18.00	34.00

Table 2: State Space Estimation Results: Focused Diagnostic Experiments

The table reports state space estimation outcomes for five stock-days selected from cases in which the state space estimator failed on the full data (either non-convergence or $\hat{\phi}_1 = 0.99$). The state space model is estimated by maximum likelihood using the Kalman filter. The parameter ϕ_1 governs persistence of the temporary component s_t; values near unity indicate failure to distinguish the stationary and permanent components. "Valid" indicates that the estimation produced parameter estimates and valid standard errors.

		Stock-Day				
Experiment	**Outcome**	**S1**	**S2**	**S3**	**S4**	**S5**
Panel A: Real Data						
Real data, OLS starting values	Valid	No[b]	Yes	No[c]	Yes	Yes
	$\hat{\phi}_1$	0.99	0.99	−0.31	0.99	0.99
Real data, zero starting values	Valid	Yes	Yes	Yes	Yes	Yes
	$\hat{\phi}_1$	0.99	0.99	0.99	0.99	0.99
Panel B: Semi-Simulation (real trades x_t, simulated Normal disturbances)						
Semi-simulation, zero starting values	Valid	Yes	Yes	Yes	Yes	Yes
	$\hat{\phi}_1$	0.99	0.99	−0.098[a]	0.99	0.99
Semi-simulation, OLS starting values	Valid	Yes	Yes	Yes	Yes	Yes
	$\hat{\phi}_1$	−0.074	−0.125	−0.011	−0.032	−0.072
	(s.e.)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)
Panel C: Monte Carlo (all variables simulated with Normal distributions)						
Monte Carlo, Normal disturbances	Valid	Yes	Yes	Yes	Yes	Yes
	$\hat{\phi}_1 \approx$ input	Yes	Yes	Yes	Yes	Yes
Panel D: Rank-Transformed Data (univariate Normal marginals)						
Rank-transformed, OLS starting values	Valid	Yes	Yes	Yes	Yes	Yes
	$\hat{\phi}_1$	0.639	−0.255	0.973	0.089	0.177
	(s.e.)	(0.118)	(0.032)	(986)	(0.012)	(0.124)

[a] Standard error = 0.131.

[b] Estimation converged but the Hessian is singular, preventing computation of standard errors. $\hat{\phi}_1 = 0.99$ indicates failure to identify the temporary component.

[c] Estimation converged but the Hessian is singular, preventing computation of standard errors. Unlike S1, $\hat{\phi}_1 = -0.31$ is a reasonable value, suggesting the likelihood surface is flat in the neighborhood of a valid solution rather than at a boundary.

Table 3: Phi Parameter Estimates and AIC Autoregression Order

The HHM model is estimated for each security-day in the sample. Panel A reports average and median estimates of ϕ_1 by day in proximity to the earnings announcement (Pre and Post bracket the announcement, WA Pre and WA Post are the week after). Also reported are rejection rates for security-day hypothesis tests of $\phi_1 = 0$, where the > 0 and < 0 columns report the percentage of days with rejections in which the parameter estimate is positive and negative, respectively. Numbers in the row labeled ALL are computed using all security-days. Panels B and C report the same statistics for the subsamples of stocks in the lowest and highest quartiles by analyst coverage, respectively. Panel D reports statistics for the distribution of auto-regression lag orders selected by the AIC for the full sample and for the low and high quartiles of analyst coverage.

Panel A: Full Sample

	% rej >0	% rej <0	ave	median
Pre	1%	63%	-0.10	-0.10
Post	1%	68%	-0.12	-0.11
WA Pre	1%	64%	-0.11	-0.10
WA Post	1%	62%	-0.11	-0.10
ALL	1%	64%	-0.11	-0.10

Panel B: Low Coverage Quartile

	% rej >0	% rej <0	ave	median
Pre	1%	40%	-0.08	-0.07
Post	3%	45%	-0.09	-0.09
WA Pre	1%	42%	-0.09	-0.08
WA Post	1%	41%	-0.08	-0.08
ALL	2%	42%	-0.09	-0.08

Panel C: High Coverage Quartile

	% rej >0	% rej <0	ave	median
Pre	1%	89%	-0.14	-0.13
Post	0%	96%	-0.16	-0.15
WA Pre	1%	89%	-0.15	-0.13
WA Post	0%	88%	-0.15	-0.14
ALL	0%	91%	-0.15	-0.14

Panel D: AIC Autoregression Order

	ave	median	25th %tile	75th %tile
Full Sample	10.4	8	4	17
Low Coverage Quartile	6.3	4	2	8
High Coverage Quartile	15.6	16	10	22

Table 4: Kappa Estimates

The HHM model parameters are estimated jointly for each security-day in the sample. Panel A reports average and median estimates of the κs for all counterparty pairings by day in proximity to the earnings announcement (Pre and Post bracket the announcement, WA Pre and WA Post are the week after). Particular counterparty pairings are identified by HH, HN, NH, and NN, where e.g., HN indicates that an HFT takes liquidity from a non-HFT. Also reported are rejection rates for security-day hypothesis tests of $\kappa_{xy} = 0$, where the > 0 and < 0 columns report the percentage of days with rejections in which the parameter estimate is positive and negative, respectively. Numbers in the row labeled ALL are computed using all security-days. Panels B and C report the same statistics for the subsamples of stocks in the lowest and highest quartiles by analyst coverage, respectively.

Panel A - Full Sample

Kappa HH	% rej >0	% rej <0	ave	median	Kappa NH	% rej >0	% rej <0	ave	median
Pre	68%	0%	17.72	2.37	Pre	75%	0%	11.02	2.05
Post	66%	0%	14.76	2.68	Post	81%	0%	13.39	3.32
WA Pre	68%	0%	12.01	2.08	WA Pre	71%	1%	10.92	1.72
WA Post	69%	0%	13.65	2.27	WA Post	73%	0%	10.43	1.62
ALL	68%	0%	14.55	2.35	ALL	75%	0%	11.49	2.09
Kappa HN	% rej >0	% rej <0	ave	median	Kappa NN	% rej >0	% rej <0	ave	median
Pre	85%	0%	9.21	1.65	Pre	90%	0%	6.15	1.33
Post	84%	0%	10.03	1.75	Post	91%	0%	7.96	1.39
WA Pre	85%	0%	7.45	1.44	WA Pre	92%	0%	5.89	1.39
WA Post	85%	0%	8.68	1.55	WA Post	89%	0%	5.38	1.41
ALL	85%	0%	8.88	1.59	ALL	90%	0%	6.39	1.39

Panel B - Low Coverage Quartile

Kappa HH	% rej >0	% rej <0	ave	median	Kappa NH	% rej >0	% rej <0	ave	median
Pre	37%	0%	32.69	12.23	Pre	47%	1%	19.67	8.70
Post	34%	0%	30.66	9.99	Post	60%	0%	25.46	12.24
WA Pre	34%	0%	16.51	7.08	WA Pre	47%	2%	19.88	7.99
WA Post	36%	0%	25.32	8.44	WA Post	44%	0%	22.15	7.56
ALL	35%	0%	26.56	9.08	ALL	50%	1%	22.03	9.15
Kappa HN	% rej >0	% rej <0	ave	median	Kappa NN	% rej >0	% rej <0	ave	median
Pre	74%	0%	18.47	7.03	Pre	81%	0%	12.00	5.95
Post	70%	0%	21.60	6.05	Post	87%	0%	16.57	5.81
WA Pre	76%	0%	14.08	6.28	WA Pre	85%	0%	11.29	5.99
WA Post	76%	0%	19.03	6.78	WA Post	81%	0%	10.32	5.83
ALL	73%	0%	18.50	6.59	ALL	83%	0%	12.82	5.92

Panel C - High Coverage Quartile

Kappa HH	% rej >0	% rej <0	ave	median	Kappa NH	% rej >0	% rej <0	ave	median
Pre	97%	0%	1.13	0.64	Pre	98%	0%	0.89	0.57
Post	97%	0%	1.03	0.61	Post	100%	0%	0.95	0.56
WA Pre	96%	0%	1.03	0.57	WA Pre	96%	0%	0.77	0.49
WA Post	95%	0%	1.02	0.57	WA Post	98%	0%	0.84	0.55
ALL	96%	0%	1.05	0.61	ALL	98%	0%	0.86	0.55
Kappa HN	% rej >0	% rej <0	ave	median	Kappa NN	% rej >0	% rej <0	ave	median
Pre	93%	0%	0.56	0.28	Pre	97%	0%	0.38	0.19
Post	94%	0%	0.44	0.20	Post	97%	0%	0.32	0.17
WA Pre	91%	0%	0.61	0.25	WA Pre	97%	0%	0.45	0.19
WA Post	91%	0%	0.54	0.23	WA Post	95%	0%	0.46	0.20
ALL	92%	0%	0.54	0.24	ALL	97%	0%	0.40	0.19

Table 5: Kappa Differences

The HHM model parameters are estimated jointly for each security-day in the sample. Panel A reports average and median estimates of κ differences between selected counterparty pairings by day in proximity to the earnings announcement (Pre and Post bracket the announcement, WA Pre and WA Post are the week after). Particular counterparty pairings are identified by HH, HN, NH, and NN, where e.g., HN indicates that an HFT takes liquidity from a non-HFT. Also reported are rejection rates for security-day hypothesis tests of $\kappa_{xy} - \kappa_{wz} = 0$, where the > 0 and < 0 columns report the percentage of days with rejections in which the parameter estimate is positive and negative, respectively. Numbers in the row labeled ALL are computed using all security-days. Panels B and C report the same statistics for the subsamples of stocks in the lowest and highest quartiles by analyst coverage, respectively.

Panel A - Full Sample

KappaHH - Kappa NH	% rej >0	% rej <0	ave	median	KappaHH - KappaHN	% rej >0	% rej <0	ave	median
Pre	23%	4%	6.70	0.20	Pre	42%	3%	8.51	0.65
Post	20%	6%	1.37	0.08	Post	46%	2%	4.72	0.73
WA Pre	22%	5%	1.09	0.16	WA Pre	38%	3%	4.55	0.52
Wa Post	22%	4%	3.21	0.21	Wa Post	40%	2%	4.96	0.49
ALL	22%	5%	3.06	0.16	ALL	42%	2%	5.67	0.61
KappaHN - KappaNN	% rej >0	% rej <0	ave	median	KappaNH - KappaNN	% rej >0	% rej <0	ave	median
Pre	25%	6%	3.06	0.12	Pre	44%	4%	4.87	0.56
Post	24%	6%	2.07	0.10	Post	58%	2%	5.43	1.05
WA Pre	22%	8%	1.57	0.11	WA Pre	39%	4%	5.03	0.36
Wa Post	23%	6%	3.30	0.09	Wa Post	41%	4%	5.05	0.40
ALL	23%	7%	2.49	0.11	ALL	46%	4%	5.11	0.56

Panel B - Low Coverage Quartile

KappaHH - Kappa NH	% rej >0	% rej <0	ave	median	KappaHH - KappaHN	% rej >0	% rej <0	ave	median
Pre	10%	1%	13.02	2.17	Pre	13%	2%	14.23	3.07
Post	6%	4%	5.20	0.13	Post	14%	2%	9.06	3.55
WA Pre	6%	5%	-3.37	0.05	WA Pre	11%	3%	2.42	1.02
Wa Post	8%	6%	3.17	1.39	Wa Post	14%	2%	6.29	1.23
ALL	8%	4%	4.53	0.69	ALL	13%	2%	8.06	2.18
KappaHN - KappaNN	% rej >0	% rej <0	ave	median	KappaNH - KappaNN	% rej >0	% rej <0	ave	median
Pre	16%	6%	6.46	0.94	Pre	22%	7%	7.67	2.02
Post	15%	6%	5.03	0.51	Post	32%	3%	8.89	5.22
WA Pre	15%	11%	2.80	0.56	WA Pre	18%	4%	8.59	2.37
Wa Post	17%	6%	8.71	0.74	Wa Post	21%	8%	11.83	2.43
ALL	16%	7%	5.69	0.71	ALL	24%	5%	9.21	2.97

Panel C - High Coverage Quartile

KappaHH - Kappa NH	% rej >0	% rej <0	ave	median	KappaHH - KappaHN	% rej >0	% rej <0	ave	median
Pre	38%	4%	0.24	0.09	Pre	76%	3%	0.57	0.31
Post	39%	5%	0.08	0.07	Post	85%	1%	0.58	0.40
WA Pre	34%	6%	0.26	0.06	WA Pre	71%	2%	0.42	0.25
Wa Post	31%	7%	0.18	0.08	Wa Post	69%	2%	0.48	0.30
ALL	36%	6%	0.19	0.07	ALL	75%	2%	0.51	0.31
KappaHN - KappaNN	% rej >0	% rej <0	ave	median	KappaNH - KappaNN	% rej >0	% rej <0	ave	median
Pre	32%	7%	0.18	0.03	Pre	72%	3%	0.51	0.24
Post	31%	9%	0.12	0.02	Post	86%	0%	0.63	0.35
WA Pre	26%	10%	0.16	0.03	WA Pre	66%	2%	0.33	0.18
Wa Post	24%	9%	0.08	0.02	Wa Post	66%	1%	0.38	0.22
ALL	28%	9%	0.14	0.03	ALL	73%	1%	0.46	0.24

Table 6: Psi Estimates

The HHM model parameters are estimated jointly for each security-day in the sample. Panel A reports average and median estimates of the ψs for all counterparty pairings by day in proximity to the earnings announcement (Pre and Post bracket the announcement, WA Pre and WA Post are the week after). Particular counterparty pairings are identified by HH, HN, NH, and NN, where e.g., HN indicates that an HFT takes liquidity from a non-HFT. Also reported are rejection rates for security-day hypothesis tests of $\psi_{xy} = 0$, where the > 0 and < 0 columns report the percentage of days with rejections in which the parameter estimate is positive and negative, respectively. Numbers in the row labeled ALL are computed using all security-days. Panels B and C report the same statistics for the subsamples of stocks in the lowest and highest quartiles by analyst coverage, respectively.

Panel A - Full Sample

Psi HH	% rej >0	% rej <0	ave	median	Psi NH	% rej >0	% rej <0	ave	median
Pre	6%	37%	-8.73	-0.46	Pre	17%	14%	0.04	0.01
Post	12%	24%	0.17	-0.16	Post	20%	13%	0.59	0.02
WA Pre	7%	36%	-3.05	-0.45	WA Pre	21%	8%	-0.63	0.02
WA Post	6%	37%	-0.55	-0.39	WA Post	23%	8%	1.06	0.03
ALL	8%	33%	-2.96	-0.35	ALL	20%	11%	0.27	0.02
Psi HN	% rej >0	% rej <0	ave	median	Psi NN	% rej >0	% rej <0	ave	median
Pre	1%	63%	-3.20	-0.40	Pre	10%	13%	-0.62	-0.010
Post	4%	36%	-1.76	-0.22	Post	13%	10%	-1.11	0.003
WA Pre	2%	59%	-1.26	-0.35	WA Pre	11%	10%	-0.24	0.002
WA Post	2%	58%	-0.09	-0.33	WA Post	12%	9%	-0.05	0.002
ALL	2%	54%	-1.58	-0.32	ALL	12%	11%	-0.52	0.000

Panel B - Low Coverage Quartile

Psi HH	% rej >0	% rej <0	ave	median	Psi NH	% rej >0	% rej <0	ave	median
Pre	2%	20%	-15.35	-3.79	Pre	12%	6%	0.26	0.28
Post	7%	9%	3.20	-2.19	Post	13%	6%	2.87	0.88
WA Pre	5%	16%	-2.46	-1.86	WA Pre	12%	4%	-1.39	0.14
WA Post	6%	17%	3.31	-2.26	WA Post	17%	4%	2.98	0.82
ALL	5%	15%	-2.43	-2.38	ALL	13%	5%	1.28	0.52
Psi HN	% rej >0	% rej <0	ave	median	Psi NN	% rej >0	% rej <0	ave	median
Pre	0%	50%	-5.80	-3.00	Pre	6%	9%	-1.29	-0.06
Post	4%	28%	-3.88	-1.42	Post	9%	8%	-2.63	-0.03
WA Pre	4%	43%	-4.28	-1.64	WA Pre	10%	7%	-0.55	0.07
WA Post	3%	35%	2.99	-2.02	WA Post	9%	5%	0.17	0.04
ALL	3%	39%	-2.83	-1.81	ALL	8%	7%	-1.18	-0.001

Panel C - High Coverage Quartile

Psi HH	% rej >0	% rej <0	ave	median	Psi NH	% rej >0	% rej <0	ave	median
Pre	12%	52%	-0.318	-0.104	Pre	29%	20%	0.022	0.003
Post	24%	37%	-0.128	-0.027	Post	37%	19%	0.042	0.015
WA Pre	12%	50%	-0.327	-0.084	WA Pre	35%	10%	0.066	0.024
WA Post	9%	49%	-0.275	-0.076	WA Post	33%	8%	0.039	0.023
ALL	14%	47%	-0.26	-0.07	ALL	34%	14%	0.042	0.016
Psi HN	% rej >0	% rej <0	ave	median	Psi NN	% rej >0	% rej <0	ave	median
Pre	1%	79%	-0.24	-0.09	Pre	17%	19%	-0.018	-0.005
Post	5%	44%	-0.10	-0.02	Post	18%	12%	0.005	0.003
WA Pre	0%	77%	-0.27	-0.08	WA Pre	18%	12%	-0.009	0.003
WA Post	0%	73%	-0.20	-0.08	WA Post	17%	13%	-0.005	0.001
ALL	1%	68%	-0.20	-0.07	ALL	18%	14%	-0.007	0.001

Table 7: Psi Differences

The HHM model parameters are estimated jointly for each security-day in the sample. Panel A reports average and median estimates of ψ differences between selected counterparty pairings by day in proximity to the earnings announcement (Pre and Post bracket the announcement, WA Pre and WA Post are the week after). Particular counterparty pairings are identified by HH, HN, NH, and NN, where e.g., HN indicates that an HFT takes liquidity from a non-HFT. Also reported are rejection rates for security-day hypothesis tests of $\psi_{xy} - \psi_{wz} = 0$, where the > 0 and < 0 columns report the percentage of days with rejections in which the parameter estimate is positive and negative, respectively. Numbers in the row labeled ALL are computed using all security-days. Panels B and C report the same statistics for the subsamples of stocks in the lowest and highest quartiles by analyst coverage, respectively.

Panel A - Full Sample

PsiHH - PsiNH	% rej >0	% rej <0	ave	median	PsiHH - PsiHN	% rej >0	% rej <0	ave	median
Pre	4%	36%	-8.78	-0.50	Pre	10%	18%	-5.53	-0.08
Post	9%	26%	-0.42	-0.18	Post	13%	14%	1.93	-0.01
WA Pre	3%	35%	-2.42	-0.50	WA Pre	11%	18%	-1.80	-0.05
Wa Post	3%	37%	-1.60	-0.49	Wa Post	11%	16%	-0.46	-0.05
ALL	5%	33%	-3.24	-0.39	ALL	11%	17%	-1.38	-0.05
PsiHN - PsiNN	% rej >0	% rej <0	ave	median	PsiNH - PsiNN	% rej >0	% rej <0	ave	median
Pre	1%	49%	-2.58	-0.40	Pre	15%	10%	0.67	0.02
Post	3%	34%	-0.64	-0.20	Post	17%	10%	1.70	0.02
WA Pre	1%	49%	-1.02	-0.30	WA Pre	17%	5%	-0.40	0.04
Wa Post	1%	50%	-0.04	-0.36	Wa Post	16%	6%	1.11	0.04
ALL	2%	45%	-1.06	-0.30	ALL	16%	8%	0.80	0.03

Panel B - Low Coverage Quartile

PsiHH - PsiNH	% rej >0	% rej <0	ave	median	PsiHH - PsiHN	% rej >0	% rej <0	ave	median
Pre	1%	15%	-15.61	-4.46	Pre	7%	8%	-9.56	-1.26
Post	4%	11%	0.33	-2.16	Post	8%	4%	7.08	-0.13
WA Pre	3%	12%	-1.07	-2.32	WA Pre	8%	7%	1.82	-0.05
Wa Post	4%	16%	0.34	-3.54	Wa Post	8%	8%	0.32	-0.54
ALL	3%	13%	-3.71	-2.92	ALL	8%	7%	0.40	-0.36
PsiHN - PsiNN	% rej >0	% rej <0	ave	median	PsiNH - PsiNN	% rej >0	% rej <0	ave	median
Pre	2%	36%	-4.51	-2.75	Pre	12%	5%	1.55	0.52
Post	3%	25%	-1.25	-1.49	Post	11%	4%	5.50	0.64
WA Pre	2%	37%	-3.73	-1.93	WA Pre	10%	3%	-0.84	0.24
Wa Post	2%	32%	2.82	-2.15	Wa Post	13%	4%	2.81	0.60
ALL	2%	32%	-1.65	-2.01	ALL	11%	4%	2.46	0.53

Panel C - High Coverage Quartile

PsiHH - PsiNH	% rej >0	% rej <0	ave	median	PsiHH - PsiHN	% rej >0	% rej <0	ave	median
Pre	6%	57%	-0.34	-0.11	Pre	18%	31%	-0.08	-0.01
Post	19%	41%	-0.17	-0.03	Post	24%	23%	-0.02	0.00
WA Pre	4%	55%	-0.39	-0.12	WA Pre	21%	28%	-0.06	-0.01
Wa Post	2%	56%	-0.31	-0.11	Wa Post	22%	22%	-0.08	-0.02
ALL	8%	52%	-0.30	-0.10	ALL	21%	26%	-0.06	-0.01
PsiHN - PsiNN	% rej >0	% rej <0	ave	median	PsiNH - PsiNN	% rej >0	% rej <0	ave	median
Pre	1%	65%	-0.22	-0.08	Pre	23%	16%	0.04	0.01
Post	4%	44%	-0.11	-0.03	Post	29%	13%	0.04	0.02
WA Pre	0%	61%	-0.26	-0.09	WA Pre	29%	7%	0.08	0.02
Wa Post	0%	67%	-0.19	-0.08	Wa Post	27%	5%	0.04	0.02
ALL	1%	59%	-0.20	-0.06	ALL	27%	10%	0.05	0.02

Table 8: Summary Statistics on Transactions and Volumes

The table reports trading statistics by counterparty pairing and by day in proximity to the earnings announcement. The Pre and Post days bracket the announcement, the WA Pre and WA Post days are the corresponding weekdays the week after the announcement. Particular counterparty pairings are identified by HH, HN, NH, and NN, where e.g., HN indicates that an HFT takes liquidity from a non-HFT. For each stock-day, dollar volume is computed as the within-day sum of the product of transaction prices and quantities for a given counterparty pairing. Dollar volume per transaction is dollar volume divided by the number of transactions in that day with that counterparty pairing. For each day type and counterparty pairing, means and medians are computed across stock-days. Mean and median dollar volume (full-day) in the right-most columns are reported in thousands of dollars.

Panel A: Full Sample

	HH					
	Dollar Volume Per Transaction		Number of Transactions		Dollar Volume ($thousands)	
day	Mean	Median	Mean	Median	Mean	Median
Pre	$5,280	$3,659	1526	316	$14,978	$802
Post	$5,428	$3,831	1973	318	$22,665	$878
Week-After Pre	$5,323	$3,715	1306	297	$12,617	$667
Week-After Post	$5,278	$3,627	1318	260	$12,859	$638
			HN			
Pre	$6,404	$4,580	2393	1001	$28,673	$3,731
Post	$6,765	$4,876	3017	1198	$40,407	$4,744
Week-After Pre	$6,575	$4,699	2087	947	$23,727	$3,954
Week-After Post	$6,530	$4,634	2069	931	$23,756	$3,688
			NH			
Pre	$6,999	$4,614	2028	558	$29,390	$1,655
Post	$7,199	$4,571	2789	747	$44,136	$2,734
Week-After Pre	$6,813	$4,712	1612	478	$21,743	$1,496
Week-After Post	$6,752	$4,594	1635	451	$21,942	$1,456
			NN			
Pre	$8,085	$5,481	2430	1277	$35,440	$6,967
Post	$8,726	$5,858	3461	2000	$54,711	$10,828
Week-After Pre	$7,928	$5,678	2052	1172	$27,325	$6,036
Week-After Post	$7,854	$5,556	2015	1090	$27,002	$5,597

Panel B: Lowest Quartile

	Dollar Volume Per Transaction		Number of Transactions		Dollar Volume Traded ($thousands)	
day	Mean	Median	Mean	Median	Mean	Median
			HH			
Pre	$2,778	$2,304	102	40	$245	$85
Post	$2,779	$2,422	112	43	$313	$118
Week-After Pre	$2,783	$2,421	88	29	$217	$92
Week-After Post	$2,671	$2,274	87	31	$204	$86
			HN			
Pre	$3,409	$2,924	325	236	$1,105	$720
Post	$3,673	$3,241	418	287	$1,702	$946
Week-After Pre	$3,676	$3,176	315	229	$1,346	$836
Week-After Post	$3,724	$3,087	292	232	$1,170	$829
			NH			
Pre	$2,944	$2,540	176	85	$528	$212
Post	$3,081	$2,760	267	122	$915	$298
Week-After Pre	$3,092	$2,556	173	71	$562	$197
Week-After Post	$3,014	$2,635	151	65	$459	$200
			NN			
Pre	$3,856	$3,528	514	380	$2,033	$1,424
Post	$4,081	$3,693	916	639	$4,256	$2,123
Week-After Pre	$3,966	$3,492	552	421	$2,588	$1,404
Week-AFter Post	$3,805	$3,308	486	387	$1,980	$1,314

Panel C: Highest Quartile

	HH					
	Dollar Volume Per Transaction		Number of Transactions		Dollar VolumeTraded ($thousands)	
day	Mean	Median	Mean	Median	Mean	Median
Pre	$8,630	$6,367	3280	2386	$36,888	$13,230
Post	$9,440	$6,692	4739	3268	$62,164	$20,476
Week-After Pre	$8,667	$6,402	2747	1848	$30,158	$12,132
Week-After Post	$8,587	$6,474	2808	1917	$30,972	$12,420
			HN			
Pre	$10,500	$7,552	5132	3814	$73,374	$30,318
Post	$11,644	$8,527	7001	5498	$111,642	$43,243
Week-After Pre	$10,601	$7,982	4379	3428	$58,925	$26,680
Week-After Post	$10,472	$7,991	4356	3254	$58,847	$25,114
			NH			
Pre	$11,825	$8,269	4368	3076	$69,821	$27,079
Post	$12,826	$8,924	6484	4837	$113,865	$41,157
Week-After Pre	$11,359	$8,237	3385	2399	$50,927	$22,048
Week-After Post	$11,152	$8,212	3466	2530	$51,372	$21,259
			NN			
Pre	$13,630	$9,946	5028	3973	$88,698	$38,783
Post	$15,628	$11,620	7321	5709	$145,697	$70,101
Week-After Pre	$13,211	$10,343	4119	3521	$67,188	$36,072
Week-After Post	$12,980	$10,110	4068	3287	$65,531	$32,837

Table 8: Regressions - Kappa Parameter Differences

Differences between κs of selected counterparty pairings for each security-day are regressed on day and coverage dummies and attributes of the earnings announcement associated with each κ difference. Particular counterparty pairings are identified by HH, HN, NH, and NN, where e.g., HN indicates that an HFT takes liquidity from a non-HFT. Each earnings announcement contributes up to four security-days to the sample. Standard errors are clustered by earnings announcement.

	(1) Kappa NH-NN	(2) Kappa NH-NN	(3) Kappa NH-NN	(4) Kappa HH-HN	(5) Kappa HH-HN	(6) Kappa HH-HN
Lowest Quartile Analyst Coverage=1	3.84**	4.745***	4.769***	.8264	5.334	5.363
	(2.26)	(2.61)	(2.62)	(0.17)	(1.15)	(1.15)
Highest Quartile Analyst Coverage=1	-4.451***	-3.072***	-3.063***	-6.405***	-3.376***	-3.365***
	(-6.46)	(-4.72)	(-4.69)	(-4.53)	(-3.35)	(-3.32)
Pre Day=1	.01509	.5743	.02963	3.818	2.364	1.012
	(0.01)	(0.48)	(0.03)	(1.28)	(0.92)	(0.41)
Post Day=1	.2787	1.246	1.463	-.08203	1.427	1.963
	(0.18)	(0.87)	(1.03)	(-0.02)	(0.38)	(0.51)
Abs(Earning Surprise)		141.2***	167.0**		267.7	264.6
		(2.62)	(2.07)		(1.52)	(0.90)
Abs(Analyst Dispersion)		53.15	23.04		43.79	17.89
		(1.23)	(0.41)		(0.46)	(0.13)
Pre Day=1 × Abs(Earning Surprise)			49.66			152.4
			(0.42)			(0.76)
Post Day=1 × Abs(Earning Surprise)			-137.3			-123.5
			(-1.17)			(-0.34)
Pre Day=1 × Abs(Analyst Dispersion)			38.64			74.29
			(0.51)			(0.41)
Post Day=1 × Abs(Analyst Dispersion)			71.59			20.18
			(0.90)			(0.11)
Constant	4.839***	2.674***	2.75***	5.981***	2.364	2.559
	(5.15)	(2.98)	(3.10)	(3.24)	(1.41)	(1.52)
Const + Low Covg Coeff	8.679***	7.419***	7.519***	6.8074	7.698	7.922*
Const + High Covg Coeff	0.388	-0.398	-0.313	-0.424	-1.012	-0.806
Observations	3042	2823	2823	3042	2823	2823
R^2	0.010	0.023	0.025	0.002	0.008	0.009

t statistics in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 9: Regressions - Psi Parameter Differences

Differences between ψs of selected counterparty pairings for each security-day are regressed on day and coverage dummies and attributes of the earnings announcement associated with each ψ difference. Particular counterparty pairings are identified by HH, HN, NH, and NN, where e.g., HN indicates that an HFT takes liquidity from a non-HFT. Each earnings announcement contributes up to four security-days to the sample. Standard errors are clustered by earnings announcement.

	(1)	(2)	(3)	(4)	(5)	(6)
	Psi HN-NN	Psi HN-NN	Psi HN-NN	Psi HH-HN	Psi HH-HN	Psi HH-HN
Lowest Quartile Analyst Coverage=1	-.5517	-3.013***	-3.019***	3.647	-2.053	-2.044
	(-0.22)	(-3.10)	(-3.10)	(0.78)	(-0.73)	(-0.72)
Highest Quartile Analyst Coverage=1	.9457***	1.263***	1.262***	3.283***	1.858**	1.876**
	(2.60)	(4.29)	(4.27)	(2.69)	(2.20)	(2.19)
Pre Day=1	-2.044	-.4821	-.2318	-4.407*	-2.932*	-.05829
	(-1.48)	(-0.73)	(-0.38)	(-1.89)	(-1.65)	(-0.03)
Post Day=1	-.06938	.8756	.8328	2.925	.6709	2.37
	(-0.07)	(1.50)	(1.41)	(1.11)	(0.33)	(0.98)
Abs(Earning Surprise)		13.88	10.97		-79.34	-74.46
		(0.69)	(0.35)		(-0.60)	(-0.29)
Abs(Analyst Dispersion)		7.64	16.98		11.49	168.8
		(0.40)	(0.49)		(0.12)	(0.93)
Pre Day=1 × Abs(Earning Surprise)			-22.78			-104.1
			(-0.58)			(-0.45)
Post Day=1 × Abs(Earning Surprise)			30.65			74.41
			(1.01)			(0.25)
Pre Day=1 × Abs(Analyst Dispersion)			-17.67			-314.1
			(-0.46)			(-1.46)
Post Day=1 × Abs(Analyst Dispersion)			-16.91			-284.2
			(-0.46)			(-1.30)
Constant	-.6107	-1.608***	-1.659***	-2.971**	-1.267	-2.452
	(-0.83)	(-3.52)	(-3.43)	(-2.03)	(-1.05)	(-1.46)
Const + Low Covg Coeff	-1.1624	-4.621***	-4.678***	0.676	-3.32	-4.496
Const + High Covg Coeff	0.335	-0.345	-0.397	0.312	0.591	-0.576
Observations	3042	2823	2823	3042	2823	2823
R^2	0.001	0.017	0.017	0.002	0.003	0.009

t statistics in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 10: Redeploying Information Gleaned from Absorbing Adverse Selection

The temporary component of price impact of transactions where HFTs take liquidity from non-HFTs (ψ_{HN}) is regressed on its prior-day value and contemporaneous and prior-day values of κ_{NH} and κ_{HH}, which measure the permanent price impact (informativeness) of trades absorbed by HFTs providing liquidity. Columns (1) and (2) analyze sample days in the week after each earnings announcement; columns (3) and (4) analyze the days bracketing each earnings announcement. Each announcement contributes at most one observation to each regression.

	(1)	(2)	(3)	(4)
	PsiHN (WA Post)	PsiHN (WA Post)	PsiHN (Post)	PsiHN (Post)
PsiHN (WA Pre)	.7579***	.6809***		
	(27.09)	(24.88)		
KappaNH (WA Pre)	-.05679***	.006134		
	(-4.45)	(0.40)		
KappaNH (WA Post)		.001178		
		(0.07)		
KappaHH (WA Pre)		-.003796		
		(-0.43)		
KappaHH (WA Post)		-.07187***		
		(-8.88)		
PsiHN (Pre)			-.1171***	-.2359***
			(-4.00)	(-7.56)
KappaNH (Pre)			-.1156***	-.06423***
			(-10.49)	(-5.25)
KappaNH (Post)				-.09716***
				(-7.30)
KappaHH (Pre)				-.02656***
				(-3.86)
KappaHH (Post)				.004249
				(0.32)
Constant	-.4202	-.1407	-.6566***	-.3636*
	(-1.57)	(-0.55)	(-2.94)	(-1.66)
Observations	666	666	666	666
R^2	0.533	0.600	0.143	0.235

t statistics in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 11: Appendix B - GMM vs State Space and OLS

The HMM model is estimated for each security-day for the full sample in four different ways. Panels B - D report results for different parameter types comparing OLS-based approaches to GMM. The rows under each parameter heading correspond to the estimation approach. Average and median estimates and rejection rates computed under each approach appear in the left-most columns. Rejection rates correspond to the null hypothesis that the parameter is equal to zero. Columns to the right compare each OLS-based approach to GMM by reporting percentages of security-days for which the inference of the OLS-based approach match those of GMM. Also reported are root-mean-squared differences and root-median-squared differences between the OLS-based estimate and the GMM estimate. Panel A reports results for the state space (SS) estimator and GMM for the subset of 83 stock-days with valid SS estimates and $\hat{\phi}_1 < 0.99$. Panel E reports the rejection rate across all security-days of the GMM J test of overidentifying restrictions. All models are estimated with 7 lags.

Panel A - State Space (SS) vs GMM: 83 stock-days with valid SS estimates and $\phi_1 < 0.99$

Phi 1	% rej >0	% rej <0	ave	median	% match GMM>0	% match GMM<0	RMSE vs GMM	RMedSE vs GMM
GMM	1%	79%	-0.12	-0.11				
SS	42%	1%	0.38	0.05	66.7%	9.5%	0.69	0.08
Psi HH								
GMM	9%	47%	-2.48	-0.14				
SS	10%	10%	-0.26	0.00	50.0%	37.7%	9.37	0.14
Psi HN								
GMM	1%	60%	-1.33	-0.08				
SS	8%	2%	-0.58	0.00	84.6%	25.0%	4.73	0.05
Psi NH								
GMM	19%	17%	-0.13	0.00				
SS	8%	6%	0.26	0.00	59.0%	63.2%	7.16	0.10
Psi NN								
GMM	8%	13%	-0.42	0.00				
SS	10%	1%	-0.03	0.00	83.8%	62.5%	3.20	0.02

Table 11: Appendix B - GMM vs State Space and OLS

Kappa HH								
GMM	82%	3%	5.38	0.65				
SS	11%	7%	1.02	0.12	14.3%	57.1%	14.14	0.39
Kappa HN								
GMM	86%	0%	3.37	0.38				
SS	6%	8%	0.52	0.02	17.3%	100%[a]	7.84	0.21
Kappa NH								
GMM	87%	0%	5.21	0.63				
SS	12%	6%	0.95	0.05	20.8%	100%[a]	13.62	0.40
Kappa NN								
GMM	95%	0%	3.58	0.29				
SS	11%	4%	0.32	0.02	14.3%	na[b]	8.21	0.14

[a] The 100% match rate is a small-sample artifact (2 and 5 observations respectively) where the GMM estimate is negative. In every such instance neither estimator produces a significant result, so all observations are classified as matches.

[b] No GMM estimates are negative for Kappa NN in this subsample.

Table 11: Appendix B - GMM vs State Space and OLS

Panel B - ϕ_1

Phi 1	% rej >0	% rej <0	ave	median	% match GMM>0	% match GMM<0	RMSE vs GMM	RMedSE vs GMM
GMM	1.5%	64%	-0.10	-0.10				
OLS Prop	0.9%	64%	-0.11	-0.10	86%	92%	0.03	0.004
OLS resid	0.9%	65%	-0.11	-0.11	86%	92%	0.04	0.005
OLS zero	1.2%	62%	-0.11	-0.10	82%	86%	0.03	0.004

Table 11: Appendix B - GMM vs State Space and OLS

Panel C - Kappas

Kappa HH	% rej >0	% rej <0	ave	median	% match GMM >0	% match GMM <0	RMSE vs GMM	RMedSE vs GMM
GMM	68%	0.5%	14.85	2.29				
OLS Prop	68%	0.3%	14.74	2.39	95%	94%	13.99	0.04
OLS resid	69%	0.3%	13.67	2.31	93%	92%	22.21	0.11
OLS zero	68%	0.3%	13.43	1.98	94%	93%	39.04	0.21
Kappa HN								
GMM	85%	0.1%	8.96	1.59				
OLS Prop	85%	0.03%	8.49	1.59	94%	83%	13.16	0.02
OLS resid	83%	0.1%	8.68	1.67	92%	80%	14.58	0.05
OLS zero	88%	0.03%	7.36	1.44	93%	69%	13.31	0.09
Kappa NH								
GMM	76%	0.3%	10.84	1.99				
OLS Prop	76%	0.3%	11.24	2.08	94%	88%	14.24	0.04
OLS resid	77%	0.2%	10.69	1.99	92%	87%	15.42	0.08
OLS zero	75%	0.2%	9.81	1.77	93%	88%	15.30	0.17
Kappa NN								
GMM	92%	0.1%	7.13	1.43				
OLS Prop	90%	0.03%	6.25	1.38	94%	33%	16.66	0.03
OLS resid	91%	0.03%	6.42	1.41	95%	27%	17.42	0.04
OLS zero	92%	0.03%	5.43	1.22	94%	27%	16.90	0.06

Table 11: Appendix B - GMM vs State Space and OLS

Panel D - Psis

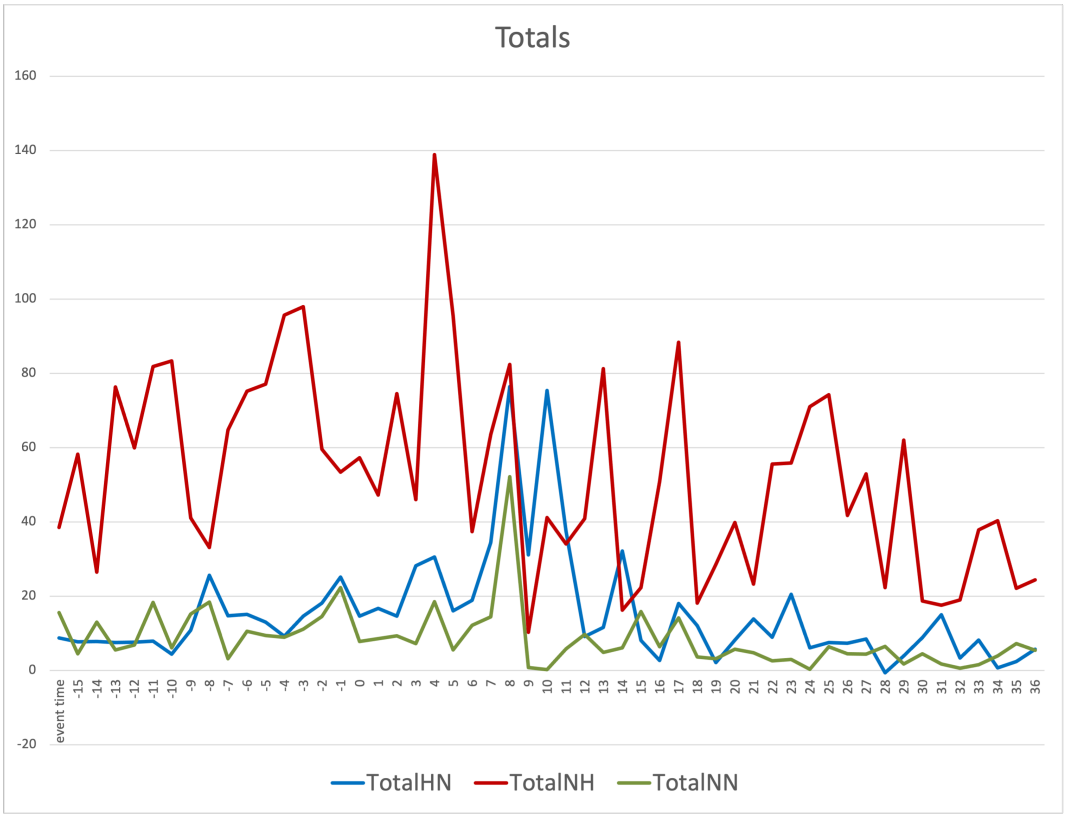
Psi HH	% rej >0	% rej <0	ave	median	% match GMM >0	% match GMM <0	RMSE vs GMM	RMedSE vs GMM
GMM	8.2%	35%	-4.66	-0.37				
OLS Prop	7.7%	33%	-4.11	-0.35	95%	93%	12.58	0.03
OLS resid	7.8%	32%	-3.29	-0.32	90%	87%	18.77	0.10
OLS zero	13%	24%	-2.33	-0.13	82%	81%	37.29	0.21
Psi HN								
GMM	2.0%	56%	-2.65	-0.32				
OLS Prop	2.0%	53%	-2.16	-0.31	92%	92%	11.40	0.02
OLS resid	1.4%	52%	-2.34	-0.36	89%	87%	12.51	0.04
OLS zero	3.7%	44%	-0.84	-0.18	80%	78%	12.31	0.09
Psi NH								
GMM	20%	12%	-0.86	0.02				
OLS Prop	20%	11%	-0.20	0.02	92%	92%	15.09	0.02
OLS resid	25%	8.4%	0.38	0.04	83%	86%	16.96	0.07
OLS zero	40%	3.7%	1.94	0.18	69%	72%	16.65	0.17
Psi NN								
GMM	14%	14%	-1.35	-0.001				
OLS Prop	12%	10%	-0.37	0.001	90%	88%	16.11	0.01
OLS resid	11%	9.9%	-0.53	0.001	86%	84%	17.19	0.02
OLS zero	30%	6.3%	0.54	0.06	66%	78%	17.17	0.06

Table 11: Appendix B - GMM vs State Space and OLS

Panel E - GMM Specification Tests	
J-stat % rej	25.9%

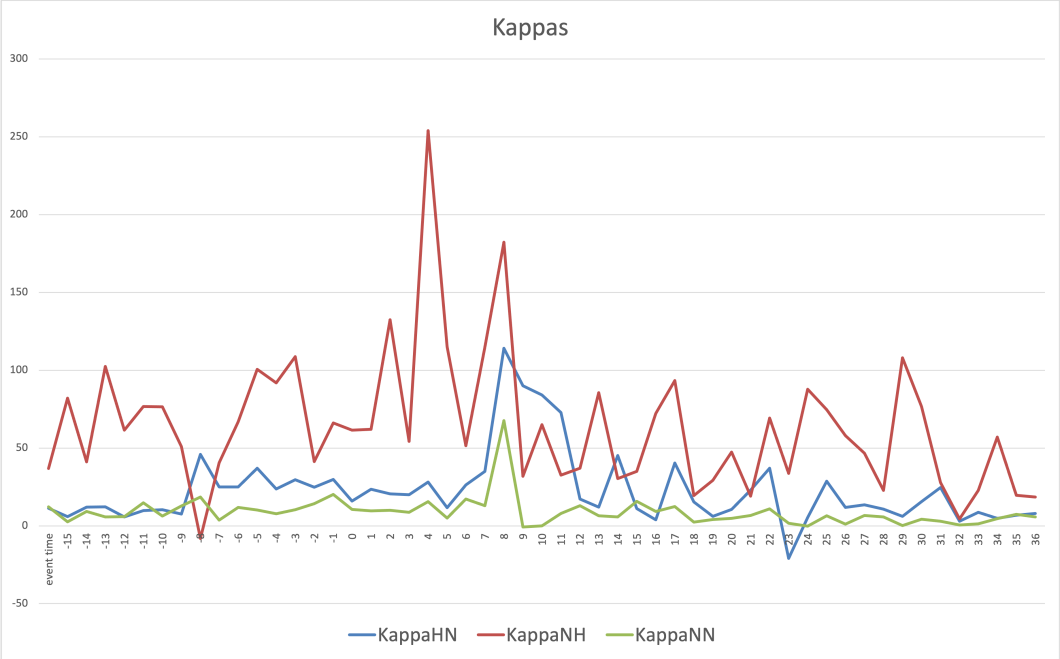
FIGURES

Figure 1



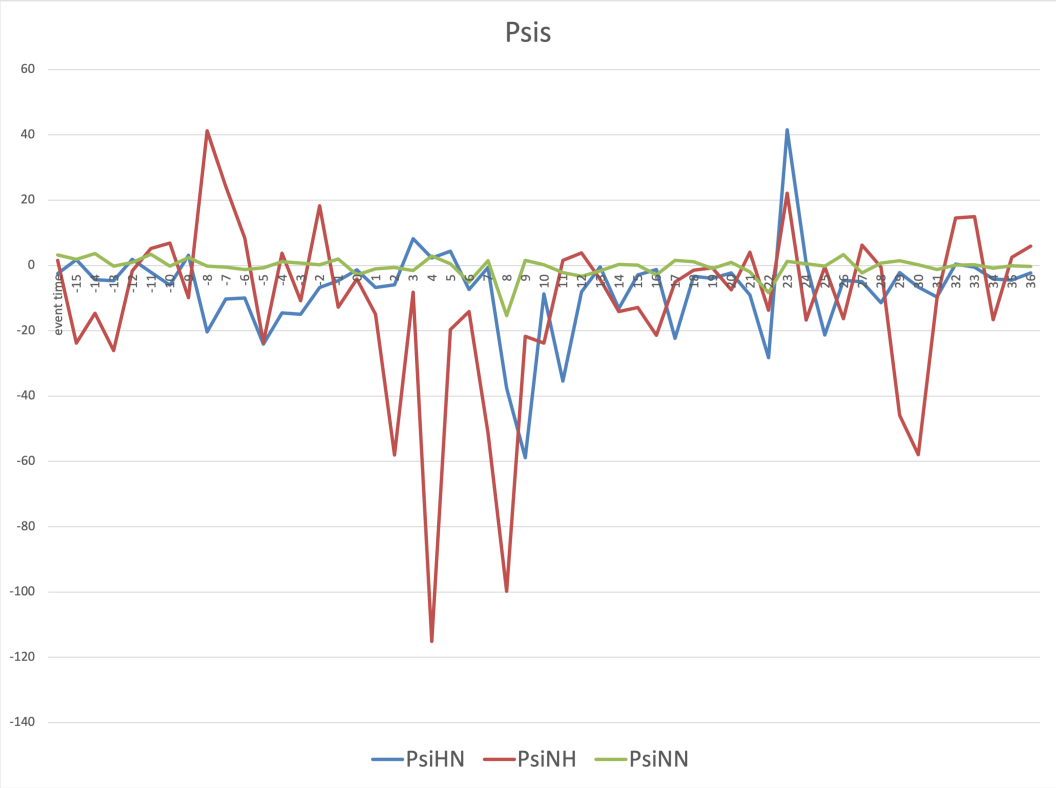
Total price impacts, $\kappa+\psi$,for various counterparty pairings on the days prior and subsequent to the February 24, 2009 (day 0) earnings announcement by Central European TV.

Figure 2



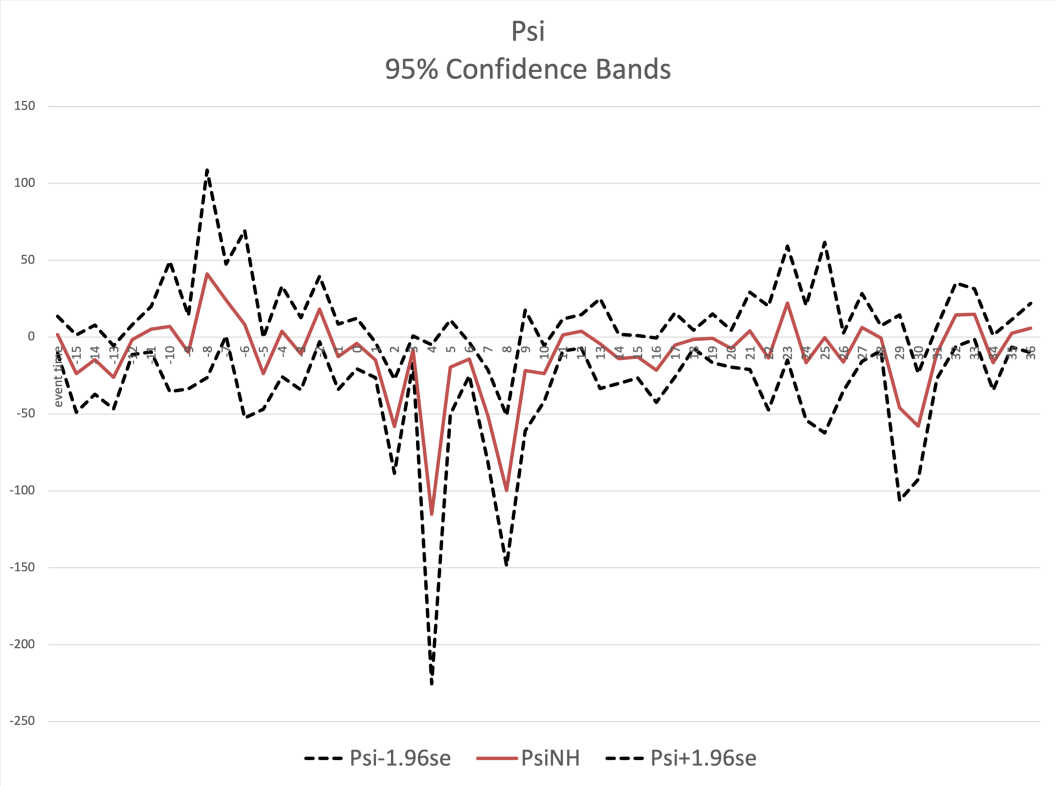
Permanent component of price impacts, κ, for various counterparty pairings on the days prior and subsequent to the February 24, 2009 (day 0) earnings announcement by Central European TV.

Figure 3



Temporary component of price impacts, ψ, for various counterparty pairings on the days prior and subsequent to the February 24, 2009 (day 0) earnings announcement by Central European TV.

Figure 4



The solid line is the temporary component of price impacts, ψ_{NH}, for transactions initiated by non-HFTs and absorbed by HFTs on the days prior and subsequent to the February 24, 2009 (day 0) earnings announcement by Central European TV. The dashed lines are 95% confidence bands.

References

Allena, R. and Chordia, T. (2025). True liquidity and fundamental prices: Us tick size pilot. *Journal of Financial Economics, forthcoming*.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56.

Amihud, Y. and Mendelson, H. (1987). Trading mechanisms and stock returns: An empirical investigation. *The Journal of Finance*, 42(3):533–553.

Biais, B., Foucault, T., et al. (2014). Hft and market quality. *Bankers, Markets & Investors*, 128(1):5–19.

Biais, B., Foucault, T., and Moinas, S. (2015). Equilibrium fast trading. *Journal of Financial Economics*, 116(2):292–313.

Blom, G. (1958). *Statistical Estimates and Transformed Beta Variables*. John Wiley & Sons, New York.

Bloomfield, R., O'Hara, M., and Saar, G. (2015). Hidden liquidity: Some new light on dark trading. *The Journal of Finance*, 70(5):2227–2274.

Bollen, N. P. and Whaley, R. E. (2004). Does net buying pressure affect the shape of implied volatility functions? *The Journal of Finance*, 59(2):711–753.

Boulatov, A. and George, T. J. (2013). Hidden and displayed liquidity in securities markets with informed liquidity providers. *The Review of Financial Studies*, 26(8):2096–2137.

Brennan, M. J. and Subrahmanyam, A. (1996). Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of Financial Economics*, 41(3):441–464.

Brogaard, J., Hendershott, T., and Riordan, R. (2014). High-frequency trading and price discovery. *The Review of Financial Studies*, 27(8):2267–2306.

Carrion, A. (2013). Very fast money: High-frequency trading on the nasdaq. *Journal of Financial Markets*, 16(4):680–711.

Chordia, T., Goyal, A., Lehmann, B. N., and Saar, G. (2013). High-frequency trading. *Johnson Graduate School Research Paper Series*.

Chordia, T., Green, T. C., and Kottimukkalur, B. (2018). Rent seeking by low-latency traders: evidence from trading on macroeconomic announcements. *The Review of Financial Studies*.

Chordia, T., Huh, S.-W., and Subrahmanyam, A. (2009). Theory-based illiquidity and asset pricing. *The Review of Financial Studies*, 22(9):3629–3668.

Chordia, T. and Miao, B. (2020). Market efficiency in real time. *The Journal of Accounting and Economics*, (70):1–27.

Easley, D. and O'Hara, M. (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics*, 19(1):69–90.

Edwards, A. K., Hughes, P., Ritter, J., Vegella, P., and Zhang, H. (2021). The effect of hidden liquidity: Evidence from an exogenous shock. *Available at SSRN 3766512*.

Foley, S. and Putniņš, T. J. (2016). Should we be afraid of the dark? dark trading and market quality. *Journal of Financial Economics*, 122(3):456–481.

Foster, F. D. and Viswanathan, S. (1993). Variations in trading volume, return volatility, and trading costs: Evidence on recent price formation models. *The Journal of Finance*, 48(1):187–211.

Foucault, T., Hombert, J., and Roşu, I. (2016). News trading and speed. *The Journal of Finance*, 71(1):335–382.

George, T. J. and Hwang, C.-Y. (2001). Information flow and pricing errors: A unified approach to estimation and testing. *The Review of Financial Studies*, 14(4):979–1020.

Glosten, L. R. (1989). Insider trading, liquidity, and the role of the monopolist specialist. *Journal of Business*, pages 211–235.

Glosten, L. R. and Harris, L. E. (1988). Estimating the components of the bid/ask spread. *Journal of Financial Economics*, 21(1):123–142.

Glosten, L. R. and Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1):71–100.

Grossman, S. J. and Miller, M. H. (1988). Liquidity and market structure. *the Journal of Finance*, 43(3):617–633.

Hamilton, J. D. (2020). *Time series analysis*. Princeton University Press.

Hasbrouck, J. (1991a). Measuring the information content of stock trades. *The Journal of Finance*, 46(1):179–207.

Hasbrouck, J. (1991b). The summary informativeness of stock trades: An econometric analysis. *The Review of Financial Studies*, 4(3):571–595.

Hasbrouck, J. (1993). Assessing the quality of a security market: A new approach to transaction-cost measurement. *The Review of Financial Studies*, 6(1):191–212.

Hasbrouck, J. (2009). Trading costs and returns for us equities: Estimating effective costs from daily data. *The Journal of Finance*, 64(3):1445–1477.

Hasbrouck, J. and Saar, G. (2013). Low-latency trading. *Journal of Financial Markets*, 16(4):646–679.

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

Hendershott, T., Menkveld, A. J., Pratz, R., and Seasholes, M. (2022). Asset price dynamics with limited attention. *The Review of Financial Studies*, 35:962–1008.

Ho, T. and Stoll, H. R. (1980). On dealer markets under competition. *The Journal of Finance*, 35(2):259–267.

Huang, S. and Yueshen, B. Z. (2021). Speed acquisition. *Management Science*, 67(6):3492–3518.

Huh, S.-W. (2014). Price impact and asset pricing. *Journal of Financial Markets*, 19:1–38.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, pages 1315–1335.

Kyle, A. S. (1989). Informed speculation with imperfect competition. *The Review of Economic Studies*, 56(3):317–355.

Menkveld, A. J. (2013). High frequency trading and the new market makers. *Journal of Financial Markets*, 16(4):712–740.

Menkveld, A. J. (2016). The economics of high-frequency trading: Taking stock. *Annual Review of Financial Economics*, 8:1–24.

Menkveld, A. J., Koopman, S. J., and Lucas, A. (2007). Modeling around-the-clock price discovery for cross-listed stocks using state space methods. *Journal of Business & Economic Statistics*, 25(2):213–225.

O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics*, 116(2):257–270.

Pástor, L. and Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3):642–685.

Sadka, R. (2006). Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk. *Journal of Financial Economics*, 80(2):309–349.

Stoll, H. R. and Whaley, R. E. (1990). Stock market structure and volatility. *The Review of Financial Studies*, 3(1):37–71.