Liquidity, Volume, and Order Imbalance Volatility*

Vincent Bogousslavsky
Carroll School of Management
Boston College
vincent.bogousslavsky@bc.edu

Pierre Collin-Dufresne
Swiss Finance Institute
EPFL
pierre.collin-dufresne@epfl.ch

January 13, 2021

Abstract

We examine the dynamics of liquidity using a comprehensive sample of U.S. stocks in the post-decimalization period. Motivated by a continuous-time inventory model, we compute a high-frequency measure of order imbalance volatility to proxy for the inventory risk faced by liquidity providers. We show that high-frequency order imbalance volatility is an important driver of liquidity and explains the often positive time-series relation between spread and volume for large stocks, which seems to run counter most theoretical models. Furthermore, order imbalance volatility is priced in the cross-section of stock returns.

*We thank our discussants: David Cimon, Itay Goldstein, Andrei Kirilenko, Dmitry Livdan, Bart Yueshen, and Haoxiang Zhu for useful comments. We also thank Pierluigi Balduzzi and conference and seminar participants at the AFA, Bentley University, Boston College, CFM-Imperial workshop, CEPR ESSFM Asset Pricing, EFA, FRIC, IMFC, Laval, LBS Summer Symposium, NFA, and SFI Research Days for helpful comments. A previous version of this paper was circulated under the title "Liquidity, Volume, and Volatility."

Conflict-of-interest disclosure statement

Vincent Bogousslavsky

I have nothing to disclose.

Pierre Collin-Dufresne

I have nothing to disclose.

1 Introduction

We investigate the time-series relation between stocks' trading cost, volume, volatility, and order imbalance volatility. Understanding their joint dynamics is important for asset managers, who need to manage the illiquidity of their portfolios.[1] It is also interesting for academics to distinguish between various determinants of stock illiquidity. In market microstructure theory, trading costs arise primarily to compensate liquidity providers for adverse selection risk (e.g., Kyle (1985), Glosten and Milgrom (1985)) and inventory risk (Stoll (1978b)). If volume is mainly driven by uninformed trading (as in Kyle (1985)), then higher volume should be associated with a lower adverse-selection component of trading costs because it reduces the intermediaries' adverse selection risk.[2] Instead, the effect of higher volume on the inventory-risk component of trading costs depends on whether the higher volume is associated with higher or lower variance of the liquidity provider's inventory.

Indeed, we propose a simple dynamic inventory model that builds on Grossman and Miller (1988) by allowing for stochastic order-flow driven by a continuous time Markov chain. A bid-ask spread arises to compensate the risk-averse liquidity provider (LP) for the risk of holding inventory as she awaits offsetting order flow. In our model, inventory holding period and trade arrival are sources of risk for the LP. This allows to capture rich order flow dynamics and, in particular, separate the effect of volume from that of order imbalance volatility on trading costs. We derive equilibrium price dynamics explicitly and investigate the price impact of order flow, which reflects the LP's cost of providing immediacy.

The model predicts that, holding cumulative order imbalance volatility constant, an increase in volume will lower spreads, as in Demsetz (1968), because it is easier for the LP to find offsetting order flow and thus inventory holding times are shorter and inventory risk is lower. In contrast, holding volume constant, an increase in the cumulative order imbalance volatility leads to an increase in spreads because the LP faces greater inventory risk. The model therefore suggests

[1]Many asset managers set constraints on the size of a stock's position, expressed as a fraction of the stock's average daily volume and thus need to estimate the costs associated with adjusting the portfolio should these constraints bind. Further, Collin-Dufresne, Daniel, and Saglam (2020) show that the optimal portfolio of a long-term investor depends crucially on the joint dynamics of a stock's volatility and trading costs. Intuitively, stocks that become less liquid when their volatility increases typically should be under-weighted in a t-cost optimal portfolio to account for the higher deleveraging costs.

[2]Higher volume can be associated with higher trading costs in adverse selection models if the increase in trading volume reflects an increase in the likelihood of informed trading (Easley and O'Hara (1992)). We discuss in more detail the relation between volume, order imbalance, and spread in adverse selection models of market microstructure in Appendix A.

that, in addition to stock-price volatility, it is necessary to control separately for volume and order imbalance volatility in an empirical analysis of stock liquidity.[3]

Early empirical papers provide cross-sectional evidence that trading costs tend to be higher for low volume and high volatility stocks (e.g., Stoll (1978a)). In the time series, however, trading costs and volume seem to be positively related both at the index level (Chordia, Roll, and Subrahmanyam (2001)) and at the individual stock level (Lee, Mucklow, and Ready (1993)), though the latter study does not control for changes in stock volatility or for the volatility of order imbalance.[4]

We therefore take a systematic look at the time-series relation between trading costs measured by daily effective spread, volume measured by daily turnover, volatility measured by high-frequency realized volatility, and high-frequency order imbalance volatility (HFOIV) measured by the standard deviation of the five-minute share imbalance. We use high-frequency order imbalance to capture the inventory risk of market makers who operate at *high* frequencies. The trend of increased intermediation in modern markets is likely to make the role of imbalance more important. Large institutional investors must split their orders over time to minimize price impact. Furthermore, high-frequency traders, the "new market makers" (Menkveld (2013)), have little capital and closely monitor their inventory to end the day flat, even though together they represent a large fraction of the daily volume.[5]

Our sample covers U.S. stocks post decimalization (from 2002 to 2017). We find that, in pooled regressions, daily effective spreads are negatively related to volume and positively related to volatility. This is consistent with the intuition from Kyle-type adverse selection models. However, for large stocks, effective spreads are generally increasing in volume in the time series even when controlling for volatility. This result holds consistently across our sample period and is robust to using changes or levels in the variables, or vector autoregressions.[6]

[3]In contrast, as we discuss in Appendix A, in the classic continuous-time Kyle (1985)-Back (1992) model, where price impact arises in equilibrium to compensate a risk-neutral market maker against adverse-selection, volume and order imbalance volatility are both driven by the uninformed noise-trader volatility, and thus indistinguishable.

[4]There is considerable empirical evidence that trading volume is positively related to the stochastic stock price volatility. See, e.g., Clark (1973); Tauchen and Pitts (1983); Epps and Epps (1976); Gallant, Rossi, and Tauchen (1992); Andersen (1996). Foster and Viswanathan (1993) provide an early empirical examination of variations in volume, volatility, and trading costs.

[5]Hendershott, Jones, and Menkveld (2011) note the increase in algorithmic trading that represents as much as 73% of trading volume in the U.S. in 2009. The SEC reports that HFT volume in equity markets typically represents 50 percent of the volume or higher (see https://www.sec.gov/rules/concept/2010/34-61358.pdf), a large fraction of which is likely "liquidity provision" strategies.

[6]Johnson (2008) proposes a model to explain the lack of relation between volume and liquidity in the time-series at the aggregate level. In this paper, we find that the relation can be negative. Most of the evidence for a

HFOIV can "reconcile" the empirical behavior of large and small stocks. HFOIV is strongly positively associated with spread, and the relation between volume and effective spread becomes strongly negative once we control for HFOIV, consistent with the model. HFOIV substantially increases the fit of spread regressions on volume and volatility across stocks in both large and small size quintiles. For instance, the median R-squared (across years) increases from 16.63% to 26.19% in level regressions among large stocks. Furthermore, the sensitivities of spread to volatility and volume become similar in magnitude for large and small stocks. Both coefficients also line up more closely with the plus two-third and minus one-third coefficients predicted by the "microstructure invariance hypothesis" of Kyle and Obizhaeva (2016), though the null hypothesis of equality is rejected for most years of the sample.[7]

What drives HFOIV? Controlling for turnover, HFOIV spikes massively on witching days, when options and futures expire. Witching days represent a shock to liquidity trading as arbitrageurs scramble to readjust their positions in all directions (Barclay, Hendershott, and Jones (2008)) and therefore a source of inventory risk for liquidity providers. We find that spreads increase on witching days and that a sizable part of this increase is explained by the increase in HFOIV. In contrast, HFOIV is stable around earnings announcement days, when spreads presumably increase due to an increase in adverse selection risk.

What distinguishes HFOIV from lower-frequency measures? A large literature uses order imbalance at daily and lower frequencies to compute measures of adverse selection risk (e.g., Easley et al. (1996)), though Kim and Stoll (2014) argue that order imbalance is not indicative of private information.[8] We argue that HFOIV is most likely to capture inventory risk, whereas imbalances computed over longer horizons are likely to reflect other factors and therefore provide complementary information. To illustrate, consider a stock that experiences an increase in buy imbalances in the morning followed by an increase in sell imbalances in the afternoon. Daily imbalance is unchanged, whereas HFOIV captures the increased inventory risk for liquidity providers (such as

positive volume-liquidity relation is cross-sectional (e.g., Stoll (2000)). An exception is Barinov (2014), who finds that quarterly turnover is positively related to spread in the cross-section and proposes an explanation based on volatility.

[7] Chordia, Roll, and Subrahmanyam (2002) find that absolute aggregate imbalance is negatively associated with liquidity even when controlling for contemporaneous volume and absolute return. We examine the cross-section of U.S. stocks in the post-decimalization era while they examine variables aggregated from the S&P 500 components over 1988 to 1998.

[8] Back, Crotty, and Li (2018) show that order flow information alone is not enough to identify private information when traders time their trades. Duarte, Hu, and Young (2020) provide a recent overview of these issues.

high-frequency market makers) over the trading day. Empirically, absolute daily order imbalance does not explain the positive spread-volume relation documented above and does not substantially increase the fit of our spread regression.[9]

To gain more insight, we decompose volume, volatility, and HFOIV into common and idiosyncratic components. Intuition suggests that adverse selection risk should be mostly driven by the idiosyncratic component of volatility. Similarly, it is unlikely that the common component of volume or imbalance proxies the likelihood of a firm-specific information event, which could explain a negative volume-liquidity relation as shown by Easley and O'Hara (1992). For small stocks, the idiosyncratic component of volume is significant and negatively related to effective spreads while the idiosyncratic component of volatility is significant and positively related to effective spreads. Common volume and volatility components are only weakly associated with spreads. These findings support Kyle-type adverse selection models. For large stocks, spreads are also positively related to idiosyncratic volatility. However, they are positively related to both idiosyncratic and common components of volume. In addition, both idiosyncratic and common components of HFOIV are economically and statistically significantly positively related with spreads for small and large stocks. A significant common component seems more consistent with inventory models. A significant idiosyncratic component is consistent with inventory risk if market makers have limited risk-bearing capacities and hold concentrated portfolios.

Is HFOIV priced in the cross-section of stock returns? We show that HFOIV predicts the cross-section of weekly returns in our 2002-2017 sample period. Following our model's intuition, we form sequentially-sorted quintile portfolios based on turnover and HFOIV. We then compute value-weighted four-factor (Fama-French-Carhart) alphas with NYSE breakpoints. Controlling for turnover, HFOIV positively predicts returns. Alpha is statistically significant at the level of 1% in four out of five quintiles. Turnover tends to negatively predict returns when controlling for HFOIV but not unconditionally, in contrast to prior work (e.g., Datar, Naik, and Radcliffe (1998)).

In Fama and MacBeth (1973) regressions, HFOIV predicts next-week returns even after controlling for many other liquidity variables. In line with the model, controlling for turnover strengthens

[9] We recognize that our measure could capture some form of adverse selection risk at a high-frequency based on order anticipation rather fundamental information. As discussed in O'Hara (2015), "anything that affects inventory may be thought of as information." However, HFOIV differs from standard adverse selection proxies that are presumed to capture informed trading about fundamentals.

the role of HFOIV. In related prior work, Chordia et al. (2018) compute order imbalance volatility at the monthly level using daily imbalances. They argue that this measure is a proxy for informed trading and is priced. As discussed above, the horizon difference makes the two measures capture different aspects of liquidity and renders them complementary. Our results support the idea that inventory risk is priced.[10] This stands in contrast to many high-frequency liquidity measures, which do not appear to be priced (Lou and Shu (2017)).

This paper is organized as follows. Section 2 presents our theoretical model, which shows that volume and order imbalance volatility can have distinct effects on the inventory risk-component of spreads. Section 3 examines the empirical relation between spread, volume, and volatility, and introduces HFOIV. Section 4 examines what drives HFOIV, and Section 5 examines its predictive power for the cross-section of returns. Section 6 examines how alternative measures of liquidity relate to HFOIV and discusses the measurement of volatility. Section 7 concludes.

2 Volume, order imbalance volatility, and spread in a dynamic inventory model

Liquidity providers face inventory risk. Inventory risk is lower when it is easier for liquidity providers to find an offsetting trade. Hence, as long as volume is not one-sided, a higher volume should be associated with improved liquidity in inventory models (Demsetz (1968)). In contrast, risk-averse liquidity providers require a compensation to absorb one-sided supply shocks (Grossman and Miller (1988), GM). We develop a simple continuous-time inventory model to capture these two distinct effects associated with changes in order flow. Our model is a dynamic stationary version of GM that adds to their framework the stochastic arrival of order flow, which allows to investigate the effect of order imbalance volatility on spreads.[11]

As in GM we consider the liquidity provider (LP) to be a long-lived agent with constant absolute risk-aversion utility, $u(c,t) = -e^{-\beta t - \alpha c_t}$, who is always present in the market and trades continu-

[10] Even if a liquidity provider holds a well diversified portfolio, the common component in order flow (e.g., Hasbrouck and Seppi (2001)) entails an undiversifiable component in order imbalance volatility.

[11] GM is a two-date model, where a buy order arrives at date 1 and a perfectly offsetting sell order arrives at date 2. A competitive risk-averse liquidity provider intermediates by carrying the risky inventory between the two. Instead we consider an infinite horizon framework, where orders arrive at random, exponentially distributed times, which introduces inventory holding-period risk and allows for more complex order flow dynamics.

ously in a stock with price S_t to maximize her expected utility of intertemporal consumption. We assume the LP can also invest at a constant risk-free rate (r).

The LP acts competitively, in that she takes prices as given as in GM, and provides liquidity to incoming buy and sell orders, which arrive at exponentially distributed random times. We assume the total supply of shares equals

$$\theta(N_t) := \sum_{i=1}^{M} \theta_i \mathbf{1}_{\{N_t=i\}}. \tag{1}$$

That is, it switches between M discrete states indexed by $N_t = \{1, 2, \ldots, M\}$ and governed by a continuous-time Markov Chain:

$$dN_t = \sum_{i=1}^{M} \mathbf{1}_{\left\{N_{t^-}=i\right\}} \sum_{j\neq i} (j-i)(dN_{ij}(t) - \lambda_{ij}dt), \tag{2}$$

where $N_{ij}(t)$ are point processes with transition intensities λ_{ij}. Since in equilibrium the LP's inventory must be equal to the total supply, changes in the aggregate supply correspond to trades by the LP who absorbs all the supply shocks. Holding a non-zero stock inventory in between offsetting trades is risky since the stock pays a continuous stochastic dividend δ_t with dynamics given by:

$$d\delta_t = \kappa_\delta(\overline{\delta}(N_t) - \delta_t)dt + \sigma_\delta dZ_t, \tag{3}$$

where Z_t is a standard Brownian motion, and the long-term mean of the fundamental dividend process may also vary with the state, $\overline{\delta}(N_t) := \sum_{i=1}^{M} \overline{\delta}_i \mathbf{1}_{\{N_t=i\}}$.

The equilibrium is derived by solving jointly for (i) the LP's optimal dynamic trading strategy and (ii) the price process that are consistent with the LP holding the total available supply at all times. The full derivation of the model is in Appendix B, where we show that the equilibrium price is a function of both the underlying dividend process and the total supply $S(\delta_t, N_t)$. Specifically, we show its dynamics are of the form:

$$dS_t + \delta_t dt = \mu_t dt + \sigma_t dZ_t + \sum_{i=1}^{M} \mathbf{1}_{\left\{N_{t^-}=i\right\}} \sum_{j\neq i} \eta_{ij}(dN_{ij}(t) - \lambda_{ij}dt), \tag{4}$$

where the stock's expected return μ_t, its diffusion volatility σ_t, and jump volatility η_{ij} are solved explicitly up to a system of coupled non-linear equations that can easily be solved numerically, and in some cases explicitly. We also derive an explicit expression for the average volume (VOL) and unconditional variance of cumulative order imbalance (OIV):

$$VOL = \frac{1}{dt}E[|d\theta_t|] = \sum_{i=1}^{M}\sum_{j\neq i}|\theta_j - \theta_i|\pi_i \sum_{j\neq i}\lambda_{ij}, \tag{5}$$

$$OIV = V[\theta_t] = \sum_{i=1}^{M}\theta_i^2\pi_i - (\sum_{i=1}^{M}\pi_i\theta_i)^2, \tag{6}$$

where $\pi_i = E[\mathbf{1}_{\{N_t=i\}}]$ is the unconditional (stationary) probability of being in a given state i.

OIV is the unconditional variance of the LP's inventory and therefore represents a quantity of risk that affects liquidity in our model since the LP has limited risk-bearing capacity.

2.1 A symmetric model of order flow

To illustrate the predictions of the model for the relation between spreads, volume, and cumulative order imbalance volatility, we focus on a simple symmetric model, where buyers and sellers arrive in a balanced fashion (or the LP systematically waits for a buyer (seller) after having seen a seller (buyer)), and where the dividend process is independent of the order flow (inventory) in (3).

Specifically, we assume there are only three states ($M = 3$), such that the LP's inventory transitions from being long $+\hat{\theta}$ shares to being short $-\hat{\theta}$ shares via a state where she is has zero inventory; that is its inventory dynamics look as follows:[12]

$$\theta_t = -\hat{\theta} \quad \underset{\xrightarrow[\lambda_d]{}}{\xleftarrow{\lambda_i}} \quad \theta_t = 0 \quad \underset{\xleftarrow[\lambda_d]{}}{\xrightarrow{\lambda_i}} \quad \theta_t = +\hat{\theta}$$

With our assumptions, λ_i is the intensity of trades that increase the order imbalance[13] (i.e., buy or sell trades that occur when current inventory is zero), whereas λ_d is the intensity of trades that decrease order imbalance (i.e., buys that arrive when the inventory is positive or sells that arrive when the inventory is negative).

[12]See equation B29 in appendix B for the specific parametrization of the transition probabilities and for further derivations.

[13]In this model, order imbalance is equal to the absolute value of the LP's inventory.

We show in the appendix that the equilibrium price is:

$$S_t = \frac{\delta_t}{r + \kappa_\delta} + \hat{\theta}\,\hat{s}\,(\mathbf{1}_{\{\theta_t = -\hat{\theta}\}} - \mathbf{1}_{\{\theta_t = +\hat{\theta}\}}), \tag{7}$$

The first component is the expected value of the future dividends discounted at the risk-free rate (which is the stock value for a risk-neutral investor). The second component shows that when the LP goes long (short) by $\hat{\theta}$ shares then price drops (increases) by $\hat{\theta}\hat{s}$. The LP essentially buys low and sells high and $\hat{s}$ measures the spread per share earned by the market maker for providing liquidity to arriving buy and sell orders.

We show in the appendix that $\hat{s}$ solves a system of non-linear equations (B30)-(B32) that admits a unique solution, which depends on the parameters $r, \alpha, \hat{\theta}, \sigma, \lambda_d, \lambda_i$, where $\sigma = \frac{\sigma_\delta}{r+\kappa_\delta}$. We also show that the risk premium on the stock has the same sign as the inventory of the market maker. Intuitively, the LP requires a positive risk-premium to hold the risky stock inventory for some random period of time (until an offsetting order arrives). Therefore, the price has to drop for the LP to buy the units and earn a positive expected return, which will be realized when she sells her inventory at a subsequent higher price to an incoming buy order (and vice versa when the LP goes short).

Note that $\hat{s}$ can be interpreted as the ex ante bid-ask (half) spread reflecting the difference between the execution prices of a new buy versus sell order. It can also be measured as the "price impact per share traded" of a trade of size $\hat{\theta}$ (since when a 'client' sells (buys) $\hat{\theta}$ shares to the LP their executed price drops (increases) by $\hat{s}\hat{\theta}$). Of course, this "price impact" is not related to adverse selection but only to inventory risk.[14] We show in the appendix that

$$\lim_{\lambda_d \to \infty} \hat{s} = 0 \le \hat{s} \le \lim_{\lambda_d \to 0} \hat{s} = \alpha\sigma^2. \tag{8}$$

Intuitively, when imbalance decreasing trades occur with infinite frequency, then there is no inventory risk and the spread goes to zero. Conversely, in the limit where there are no imbalance-decreasing trades the spread converges to the buy-and-hold premium. Further, if risk or risk-

[14] We note that in this simple symmetric 3-state example, the ex ante spread is symmetric, in the sense that the price changes in response to a buy versus a sell are equal (in absolute value) to the ex ante half spread. This need not be the case in general, if the model were not symmetric, such as, for example, if the persistence of the long and short inventory states were not equal.

aversion goes to zero then the spread goes to zero irrespective of the trading intensity ($\lim_{\alpha\sigma\to 0} s = 0$), as the LP becomes risk-neutral.

As we illustrate further below, spreads depend in an intricate way on (i) the average of imbalance increasing-and-decreasing trade intensities (which drives volume), (ii) the ratio of these intensities (which drives order imbalance volatility), and (iii) fundamental risk (which drives price volatility).

To investigate the relation between spread, volume and order-imbalance volatility, we show in the appendix that (5) and (6) reduce to

$$VOL = 4\hat{\theta}(\frac{1}{\lambda_i} + \frac{2}{\lambda_d})^{-1}, \tag{9}$$

$$OIV = \frac{\hat{\theta}}{2\lambda_d} VOL. \tag{10}$$

Spread and volume: the Demsetz effect

Suppose first that "imbalance-increasing" and "imbalance-decreasing" order flow arrives at the same rate, that is $\lambda_d = \lambda_i = \lambda$. Then,

$$VOL = \frac{4}{3}\lambda\hat{\theta}, \tag{11}$$

$$OIV = \frac{2}{3}\hat{\theta}^2. \tag{12}$$

In this model, an increase in the trader arrival rate increase trading volume *without* affecting order imbalance volatility, which remains constant. The implication is that, as argued in Demsetz (1968), it becomes less costly for the intermediary to provide immediacy since she can offset an incoming buy order with a sell order faster and thus face lower inventory costs. As we see in the first row of Figure 1, the spread decreases with volume in this case.

Spread and volatility of order imbalance

Suppose now that $\lambda_d \neq \lambda_i$ and that we increase the average time between imbalance-decreasing trades ($\frac{1}{\lambda_d}$) and reduce the average time between imbalance-increasing trades ($\frac{1}{\lambda_i}$) so as to hold the unconditional volume constant (that is, from equation (9) such that $\frac{1}{\lambda_i} + \frac{2}{\lambda_d}$ is constant). In this case, order imbalance volatility increases and average volume is constant. This unambiguously increases the equilibrium spread as we see in the second row of Figure 1.

Comparing the two rows of Figure 1 shows that increasing volume holding order imbalance constant decreases spreads, but instead increasing order imbalance holding volume constant increases spreads. It is also straightforward to change λ_i, λ_d so *as to increase volume and at the same time increase spreads*, because order imbalance volatility increases and its effect dominates.

The intuition for this result is that increasing the trading intensity in the model has two effects. On the one hand, it increases the likelihood of an offsetting trade, which reduces the average holding period of inventory for the liquidity provider. This effect leads to lower spreads. On the other hand, increasing the trading intensity can also increase the variance of the shocks to inventory, which makes liquidity provision riskier and thus increases spreads. Hence, volume does not have an unambiguous effect on spreads unless one controls for the volatility of order imbalance. Both effects are also tied to the risk-bearing capacity of the liquidity provider. More fundamental risk or more risk aversion (as captured by $\alpha\sigma$ in the model) increases the impact on spreads of a change in the variance of order imbalance. It is thus necessary to control for volume, order imbalance volatility, and stock volatility to capture the dynamics of spreads.

We now turn to our empirical investigation.

3 Liquidity, volume, and order imbalance volatility

We examine the relation between spreads, volume, and order imbalance volatility. We discuss our data sources and methodology and then present our empirical results.

3.1 Data

We obtain daily stock data for NYSE, Amex (NYSE American), and NASDAQ common stocks from CRSP. Our focus is on the post-decimalization period. We compute daily liquidity measures over 2002 to 2017 using the Trades and Quotes dataset (TAQ) We apply the corrections and filters for TAQ data proposed by Holden and Jacobsen (2014).[15] To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Observations with a missing CRSP return are excluded. Stocks that are present in CRSP but do not

[15]We rely on the TCLINK macro provided by WRDS to match a TAQ ticker to a CRSP PERMNO. Afterwards, the data is screened for duplicates and obvious matching errors are corrected.

have a single valid TAQ trade in a given month are excluded. The liquidity, volume, and volatility measures (described below) are computed over the regular trading day (9:30am to 4:00pm). Days with early closures are excluded from the analysis.

3.2 Variables and descriptive statistics

We use the percentage effective spread as our primary measure of liquidity. The percentage effective spread of trade t on stock i is defined as

$$\text{Effective Spread}_{i,t} = 2|\ln P_{i,t} - \ln M_{i,t}|,$$

where $P_{i,t}$ is the trade price and $M_{i,t}$ denotes the midpoint of the best quote available immediately preceding the trade. The effective spread over an interval is computed by summing the weighted spread associated with each transaction over the interval, where the weight equals the dollar volume of the transaction over the total dollar volume in the interval.[16]

We use daily intraday turnover as a measure of volume. We focus on intraday turnover rather than total turnover since it is the volume associated with the effective spread. In recent years, a sizable fraction of volume is traded in the closing auction trade, which is executed right after the 4pm close (Bogousslavsky and Muravyev (2021)). Our measure of volatility is realized volatility computed using five-minute midquote returns (e.g., Andersen et al. (2001)).[17] We show in Section 6.2 that realized volatility greatly improves our ability to explain spreads relative to standard volatility measures.

Our model explains why order imbalance volatility affects liquidity but does not specify the frequency at which order imbalance volatility should be measured. Since many liquidity providers manage their inventories at high frequencies, we use a measure of order imbalance volatility computed from high-frequency order imbalance to better capture inventory risk. To do so, we compute share imbalance (as a proportion of shares outstanding) over every five-minute interval of the trad-

[16] Our results hold if we use the dollar effective spread, computed by dollar-weighting or share-weighting $2|P_{i,t}-M_{i,t}|$ over the day. As discussed in section 2.1., our model predictions can be interpreted both with respect to liquidity measured by the ex ante quoted half-spread, as well as by the transaction price change. We focus on the effective spread as it is widely used in the microstructure literature. We look at other liquidity measures including price impact in section 6.1.

[17] To minimize the influence of noisy opening quotes (e.g., Bogousslavsky (2021)), we take the volume-weighted average price over the first five minutes of trading as our opening price.

ing day using the Lee and Ready (1991) algorithm. High-frequency order imbalance volatility (HFOIV) is the standard deviation of the five-minute imbalance, computed over the trading day. If a stock is not traded during a five-minute interval, this interval is not used to compute HFOIV. We contrast HFOIV to lower frequency order imbalance measures in Section 3.4.

Large institutional investors use a combination of market and limit orders (e.g., van Kervel and Menkveld (2019), Korajczyk and Murphy (2019)).[18] The use of limit orders by investors other than liquidity providers affects the interpretation of order imbalance as measured by the Lee and Ready (1991) algorithm. If informed traders use limit orders, then order imbalance may not measure informed order flow well. This is not an issue for us since we interpret HFOIV as a measure of inventory risk. However, if informed traders provide liquidity, this could make the notion of inventory risk less relevant in modern markets. The latter effect could introduce noise in our measured order imbalance. Beyond the fact that high-frequency market makers manage their intraday inventory (Menkveld (2013)), several factors indicate that our measured order imbalance has economic content. First, it is positively autocorrelated, consistent with order flow picking up order splitting strategies of institutions (Toth et al. (2015)). Second, our results below show that HFOIV is strongly associated with liquidity in a way consistent with our theoretical model.

Figure 2 plots the daily cross-sectional median of spread, volume, volatility, and HFOIV over our sample period. Spreads tend to decline over the first part of the sample, then remain stable with large spikes during the financial crisis. Volume increases until the crisis then drops and remains relatively stable. HFOIV spikes in a regular manner. We come back to this seasonal pattern in Section 4.1 to shed light on what drives HFOIV.

Following the microstructure literature, we consider separately large and small stocks. At the beginning of each month, stocks are sorted into quintiles by their average daily market capitalization over the past 250 trading days. On average each quintile contains 540 stocks, with a minimum of 456 and a maximum of 634. We only report results for the bottom size quintile (small stocks) and the top size quintiles (large stocks) since results for the other size quintiles lie in-between these two extremes. Table 1 reports descriptive statistics for our main variables of interest for small and large

[18] In dynamics models of limit order book markets (e.g., Parlour (1998), Foucault (1999), Foucault, Kadan, and Kandel (2005), Goettler, Parlour, and Rajan (2005)), agents endogenously choose between market orders and limit orders.

stocks in even years.[19]

Table 2 reports cross-sectional averages of the individual stocks' time-series correlations for the different variables.[20] As expected, spread and volatility are positively correlated for both small and large stocks. More surprising, spread is positively correlated with volume for large stocks. We show below that this relation is not explained by volatility but is explained by HFOIV.

3.3 Spread, volume, and volatility

To investigate what drives spreads in the time series, we estimate the following panel regression:

$$\log s_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \text{controls} + \epsilon_{i,t}, \tag{13}$$

where the (log) effective spread $s_{i,t}$ is regressed on (log) turnover $\tau_{i,t}$ and (log) volatility $\sigma_{i,t}$ for stock i on day t. The regression includes stock fixed effects since we focus on the time-series relation between spread, volume, and volatility. We include as controls calendar indicators for the day of the week and the month of the year (when the regressions are estimated on a yearly basis), and previous-day market capitalization and price (in logs). The results are similar if we do not include these controls. In all of our specifications, standard errors are double-clustered by date and stock using the method of Thompson (2011).

Equation (13) can be motivated from the invariance of transaction costs hypothesis developed by Kyle and Obizhaeva (2016). Under invariance of transaction costs and additional assumptions, $s_{i,t} \propto \left[\frac{\sigma_{i,t}^2}{P_{i,t}V_{i,t}}\right]^{\frac{1}{3}}$, where V is the share volume and P is the share price. This equation closely maps to our empirical specification since we consider the logarithm of these variables.

Even though we focus on the post-decimalization period, Figure 2 shows that spread and turnover still exhibit trends over parts of the sample period. To deal with nonstationarity, we employ several methods. First, we estimate our regressions over short samples such as month-by-month and year-by-year. As discussed by Lo and Wang (2000), this procedure does not make

[19] The median market capitalization of a small (large) stock is \$0.17 (\$7.09) billion in 2002 and grows to \$0.23 (\$17.81) billion in 2017. The median daily dollar volume of a small (large) stock is \$0.28 (\$41.34) million in 2002 and grows to \$0.61 (\$111.68) million in 2017. Most of the stocks in our sample are traded every day and therefore have a valid effective spread every day. Among stocks in the smallest size quintile, the fraction of missing effective spreads is approximately 1.6%. Among stocks in the top two size quintiles, the fraction of missing effective spreads is negligible.

[20] Correlations between log variables are reported since log transformed variables are used in the analysis. The correlations between raw variables are similar.

the variables stationary but should alleviate the issue and is informative about what happens in the data over time. Furthermore, it is not clear in Figure 2 that spread and turnover exhibit any trend over the second part of the sample. Second, we use percentage changes in the variables. First-differencing helps assuage nonstationarity concerns but makes the results harder to interpret theoretically. The (log) percentage change in daily spread is regressed on the percentage changes in daily turnover and volatility:

$$\Delta s_{i,t} = \alpha_i + \beta_\tau \Delta\tau_{i,t} + \beta_\sigma \Delta\sigma_{i,t} + \text{controls} + u_{i,t}, \tag{14}$$

where $\Delta x_t \equiv \log(\frac{x_t}{x_{t-1}})$, and the controls are the same as before. Last, we estimate vector autoregressions as a robustness check.[21]

Figure 3 reports the month-by-month estimated elasticities for small and large stocks from regression (13). For large stocks, a higher volume is associated with a higher spread, except in the last couple of years of the sample. For small stocks, this relation is consistently negative throughout the sample. For both large and small stocks, a higher volatility is associated with a higher spread (Panel (b)), consistent with theory.[22] However, volatility does not explain the positive spread volume relation for large stocks. Economically, a one (within) standard deviation increase in volume from its mean level leads to a roughly 5-10% increase in spread for large stocks. For small stocks, the spread decreases by around 20%. The average monthly adjusted R^2 is 11.5% for large stocks and 14.1% for small stocks. Figure 3 highlights the importance of separating large stocks from small stocks. When all stocks are pooled together, the conventional intuition holds as a higher volume is associated with a lower spread.

We also estimate (13) and (14) year-by-year. To save space, we do not report the year-by-year results and instead report median coefficients, t-statistics, and adjusted R-squared across years in Table 3. The results for each year are reported in the Internet Appendix for all of our specifications. The median t-statistic for the spread-volume elasticity is 3.41 among large stocks and -26.29 across small stocks. In most years of the sample, there is a statistically significant relation between

[21] We also employ a procedure similar to that of Gallant et al. (1992). For each stock, the spread and turnover series are regressed on a set of calendar and trend control variables. The residuals from this regression (further adjusted using a variance equation) are then employed instead of the raw spread and turnover series. The results are similar and not reported.

[22] We also estimate univariate regressions of spread on volume and of spread on volatility. The results are similar and reported in Figure IA.2 in the Internet Appendix.

spread and volume for large stocks, controlling for volatility. This relation is stronger if we use the specification with changes in the variables.

Reverse causality is a concern in (13). Our specification builds on microstructure theories that suggest that volume and volatility are likely to have exogenous drivers, whereas spreads are mostly endogenous. For large stocks, reverse causality cannot explain the empirical result since it seems implausible for an increase in spread to cause an increase in volume. As a robustness check, we estimate vector autoregressions of spread, volume, and volatility. The results are reported in the Internet Appendix and are consistent with the panel regression results.

The minimum tick size is more likely to bind for large stocks than for small stocks (e.g., Hagströmer (2019)). Since the tick size imposes a lower bound on the quoted spread, the effect of "bad volume" (order imbalance volatility in our model) should be stronger for tick-constrained stocks. In contrast, for small stocks, which tend to have wider spreads, the "good volume" (Demsetz effect in our model) is likely to dominate. Intuitively, the tick size should make "bad volume" more apparent by imposing a lower bound on spreads. In the Internet Appendix (IA.B), we provide consistent evidence. The positive volume-spread relation is stronger among large stocks with low quoted spread than among large stocks with high quoted spread. However, large stocks with high quoted spread also tend to have a positive volume-spread relation. This suggests that "bad volume" can dominate even absent a binding tick size. Crucially, the binding tick size cannot explain why we observe a positive volume-spread relation in the first place (after controlling for volatility). Our model highlights the role of "bad volume" (order imbalance volatility), which we test in the next section.

3.4 High-frequency order imbalance volatility

In the previous section, we find that the relation between spread and volume is complex, with sometimes a positive association between these two variables for large stocks even when controlling for volatility. In our inventory model, a higher volume can be associated with an increase in spread if order imbalance volatility also increases. We update (13) to include HFOIV as defined in

Section 3.2:

$$\log s_{i,t} = \alpha_t + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \beta_{\text{HFOIV}} \log \text{HFOIV}_{i,t} + \text{controls} + \epsilon_{i,t}. \tag{15}$$

Table 3 reports the summary estimation results for small and large stocks across years. (The full set of results is in the Internet Appendix.)

First, the inclusion of HFOIV dramatically improves the explanatory power of the regression. For example, the median adjusted R^2 increases from 16.63% to 26.19% for large stocks. HFOIV is strongly associated with effective spreads at the daily level. Second, the inclusion of HFOIV makes the volume elasticity of spread negative and significant for large stocks, consistent with the idea that a higher volume is beneficial for liquidity. Finally, the inclusion of HFOIV does not reduce the volatility elasticity of spread despite both variables being positively correlated (Table 2). This suggests they complement each other and capture different drivers of liquidity, as discussed in the context of our model in Section 2.1.

HFOIV makes the role of volume consistent across small and large stocks. Table 3 shows a remarkable consistency in the magnitude of the coefficients between small and large stocks. With the inclusion of HFOIV, volatility and volume elasticities are closer to the elasticities predicted by invariance theories. Under invariance of transaction costs and additional assumptions, $s_{i,t} \propto \left[\frac{\sigma_{i,t}^2}{P_{i,t}V_{i,t}}\right]^{\frac{1}{3}}$, where V is the share volume and P is the share price (Kyle and Obizhaeva (2016)). We test whether the coefficients in (15) equal the predicted $-\frac{1}{3}$ and $\frac{2}{3}$ for volume and volatility, respectively. The volatility hypothesis is strongly rejected in all years of the sample. The volume hypothesis, however, cannot always be rejected. Invariance of transaction costs does not explicitly incorporate order imbalance volatility, but we view this evidence as encouraging.

3.4.1 Relation with absolute order imbalance

Does a lower frequency measure of order imbalance volatility explain liquidity as well as high-frequency order imbalance volatility? HFOIV is likely to outperform lower-frequency imbalance measures if intraday imbalances affect liquidity providers' inventory risk. This is a natural assumption as many liquidity providers operate at high frequencies and face intraday risk constraints (Comerton-Forde et al. (2010), Brogaard et al. (2015)). In particular, Brogaard et al. (2015) show a positive relation between high-frequency market makers' absolute inventory and spread at the one-minute frequency.

To illustrate, consider absolute daily order imbalance, a widely-used measure (e.g., Chan and Fong (2000)). Consider a stock that experiences an increase in buy imbalances in the morning followed by an increase in sell imbalances in the afternoon. Daily absolute imbalance does not change, but high-frequency order imbalance volatility increases. This increase captures additional inventory risk faced by liquidity providers as explained above, and therefore our model implies that high-frequency order imbalance volatility should be more strongly related to spread than absolute daily imbalance.

To further illustrate the relation between high- and low-frequency order imbalance and spread, we provide a simple reduced-form model in the Internet Appendix (IA.C). High-frequency order imbalance volatility and absolute daily imbalance should be most similar for thinly-traded securities and for securities with highly-persistent order imbalances. Intuitively, if a stock is traded only once a day or its imbalances are perfectly correlated, then absolute order imbalance convey the same information as HFOIV. The reduced-form model highlights that, in other cases, HFOIV should have stronger explanatory power for spread than absolute order imbalance.

We estimate (15) using absolute order imbalance instead of HFOIV. Table 4 reports median estimate. As a benchmark, the table also reports again the median estimates with HFOIV. Absolute daily order imbalance fails to explain the positive volume-spread sensitivity. The median volume coefficient is about zero in Table 4, which reflects positive coefficients in the first part of the sample followed by negative coefficients in the second half of the sample. Moreover, absolute order imbalance does not meaningfully increase the explanatory power of regressions (13) and (14) for both small and large stocks. Another way to see this is to orthogonalize HFOIV relative to absolute order imbalance (and vice versa) for each stock, and then use the orthogonalized variable in the regression. Table 4 shows that residual HFOIV remains strongly positively associated with spread, whereas residual absolute daily order imbalance tends to be negatively associated with spreads.

In Table IA.12 in the Internet Appendix, we compare the median adjusted R^2 achieved by different measures of order imbalance volatility as we vary the frequency at which we sample order imbalance over the day. The R^2 increases gradually as we increase the frequency. Intervals of 30-minute are good enough to get most of the improvement in explanatory power for small stocks, while intervals of 5-minute achieve the highest explanatory power for large stocks. Finally, the improvement of HFOIV over absolute daily imbalance is stronger for large stocks than for small stocks. This is consistent with the reduced-form model in Section IA.C.[23]

Overall, absolute daily order imbalance does not appear to capture the dynamics of liquidity as well as HFOIV. The better performance of *high-frequency* order imbalance volatility seems more consistent with an interpretation based on inventory risk than on adverse selection related to fundamentals.

4 Order imbalance volatility and liquidity

This section sheds light on the relation between HFOIV and liquidity.

4.1 What drives order imbalance volatility?

We estimate panel regressions of HFOIV on turnover and a set of calendar indicator variables. We also control for lagged price and market capitalization, and include year and stock fixed effects.

The inclusion of calendar indicator variables is motivated by a large literature that finds calendar effects on trading volume and by Figure 2, which shows that the median cross-sectional HFOIV spikes at regular intervals four times a year. The spikes correspond to quadruple witching days. These are days on which index options, stock options, index futures, and single-stock futures expire.[24] We therefore include a calendar indicator variable for the third Friday of each month (stock and index options expiration days), as well as an interaction between this variable and the last month of the quarter (quadruple witching days).

Table 5 shows that HFOIV increases on average by $e^{0.389} - 1 = 47.6\%$ on the third Friday of each month for large stocks, controlling for volume. On the third Friday of end-of-quarter months, HFOIV increases by an additional $e^{0.530} - 1 = 69.9\%$. Small stocks experience smaller but sizable increases in HFOIV. Since Barclay et al. (2008) provide strong evidence that witching days

[23]In Internet Appendix IA.D, we decompose daily volume into "balanced volume" and absolute order imbalance. We expect balanced volume to have a negative relation with spreads. However, in the data balanced volume tends to have a positive relation with spreads. Hence, this simple decomposition does not resolve the positive volume-spread relation. In contrast, once we swap absolute order imbalance with HFOIV, balanced volume becomes consistently negatively associated with spreads.

[24]The spikes are not apparent pre-2005 in Figure 2. One possible explanation is the significant rise in open interest on S&P 500 Futures in 2005 (see Figure 1 in Barclay et al. (2008)).

are associated with informationless liquidity shocks, this result supports a link between inventory effects and HFOIV.

In contrast, HFOIV does not seem to be driven by informational events since it does not increase around earnings announcement days. Except for small stocks, which experience a small increase on the day of the announcement, HFOIV tends to be lower on the day before, of, and after an earnings announcement. The lack of increase in HFOIV ahead of the announcement is consistent with Sarkar and Schwartz (2009), who find that order flow tends to be more two-sided before earnings announcements.

We also include calendar indicator for the first day of the month, the last day of the month, and Russell reconstitution dates. Such days are also often associated with liquidity shocks. There is no evidence of an increase in HFOIV on these dates for small and large stocks. If anything, HFOIV appears to be lower on these days. Table 5 shows that these days also tend to be associated with high volume. For instance, volume is more than 90% higher for small stocks on Russell reconstitution dates, but HFOIV is not significantly greater than zero (controlling for volume). What explains the difference relative to witching days? One possibility is that the nature of the liquidity shock differs. Imbalances related to Russell reconstitution can be better anticipated. In contrast, witching days could be associated with significantly more uncertainty relative to the direction of the imbalance since "arbitrageurs are likely to submit large buy or sell orders in many stocks at the open on the expiration day" (Barclay et al. (2008), p.95).

Based on Table 5, our model suggests that: first, spreads should increase on witching days but not on beginning of month, end of month, and Russell reconstitution dates; second, the increase in spread should be explained by the increase in HFOIV. In line with the first implication, Table 6 shows that spreads increase significantly on witching days for both small and large stocks (Columns 1 and 4). There is only weak evidence of an increase in spread on beginning of month, end of month, and Russell reconstitution dates. Despite the huge increase in intraday volume on Russell reconstitution dates, small stocks do not experience a significant increase in effective spreads on these dates in our sample. Moreover, Table 6 shows that HFOIV explains a sizable fraction of the spread increase on witching days (around 2/3 for large stocks and 1/4 for small stocks). In contrast, absolute daily order imbalance achieves little in explaining the increase in spread on witching days. Finally, there is no evidence that HFOIV helps explain higher spread around earnings announcement days.

Overall, the results support our interpretation of HFOIV as a measure of inventory risk for liquidity providers.

4.2 Commonality analysis

To gain more intuition, we decompose volume, volatility, and order imbalance volatility into common and idiosyncratic components. We expect asymmetric information to affect liquidity via the idiosyncratic component rather than the common component. It seems unlikely that the common component of volume or of order imbalance volatility reflect the likelihood of an information event in a specific stock. Thus, a positive relation between the common components of volume or order imbalance and spreads seems difficult to ascribe to an adverse selection theory of spreads. Instead, idiosyncratic volume or order imbalance volatility could be driven by firm-specific information events that trigger more (one-sided) informed trading and thus could cause a positive relation with spreads as shown in Easley and O'Hara (1992). Alternatively, if idiosyncratic volume is mostly driven by noise trading, then we expect a negative relation with spreads as in Kyle (1985). Similarly, we expect idiosyncratic volatility to be tied to insider information and adverse selection more so than the common component of volatility. Thus, based on adverse selection theories of illiquidity we expect the positive relation between volatility and spreads to be mostly driven by the idiosyncratic component of volatility.

The role of idiosyncratic versus systematic volume, volatility, and order imbalance volatility shocks in inventory theories is more difficult to evaluate. The existence of actively-traded basket securities should make systematic volume and volatility shocks easier to hedge than idiosyncratic shocks for individual liquidity providers. Further, if liquidity providers do not hold well-diversified portfolios, perhaps because they specialize in making markets on a limited number of securities, then idiosyncratic risks should be the primary driver of inventory cost. At the same time, a systematic volume or order imbalance shock consumes liquidity everywhere in the market. If market making capacity is limited, such shocks should matter since the "aggregate" maker maker has to absorb the shock.

We decompose volume into common and idiosyncratic components. For each stock i, we regress daily (log) turnover on a common turnover measure, where the common turnover equals the equal-weighted average daily (log) turnover of stocks in the same size quintile as stock i, excluding stock i. The idiosyncratic component of turnover is given by the residual from this regression, and the common component of turnover by the fitted value. We decompose realized volatility into common and idiosyncratic components as in Patton and Verardo (2012).[25]

We regress spread on common and idiosyncratic components of volume and volatility. Table 7 reports the summary values for level regressions. (Summary values for change regressions are reported in Table IA.17 in the Internet Appendix.) Positive common and idiosyncratic volume elasticities suggest that inventory effects are important drivers of spreads for large stocks. Moreover, the common component of volatility tends to be positive and significant. In contrast, the evidence supports adverse selection theories for small stocks. Idiosyncratic volatility elasticity is large and positive while common volatility elasticity is in general insignificant.[26] Furthermore, idiosyncratic volume elasticity is strongly negative. The standard adverse selection intuition works well for small stocks if we interpret idiosyncratic volume as mostly driven by noise trading. The evidence for small stocks does not support the Easley and O'Hara (1992) theory that higher volume reflects, on average, an increased probability of an information event and thus more adverse selection risk.

The above results neither imply that adverse selection does not matter for large stocks nor that inventory risk does not matter for small stocks, only that inventory effects seem to play an important role for daily liquidity fluctuations. Our results are consistent with Chordia, Roll, and Subrahmanyam (2000), who show that industry and market trading volumes affect individual stocks' spreads. They do not control for volatility in their time-series tests, however.

We decompose HFOIV into common and idiosyncratic components. Each day, we regress the five-minute share imbalance of a stock on the equal-weighted share imbalance of stocks that belong to the same size quintile. Each daily regression has a maximum of 78 observations when a stock is traded in every five-minute interval. Since the average five-minute order imbalance is close to zero,

[25] We use as market return for each stock i the equal-weighted intraday return of stocks that belong to the same size quintile, excluding stock i. We decompose volume and volatility for each stock using the full sample of data. The results are robust to estimating the components on a year by year basis.

[26] Nonsynchronous trading could bias the common component of realized volatility towards zero. As an alternative less susceptible to this issue, we compute for each stock i the equal-weighted daily return of stocks that belong to the same size quintile, excluding stock i. We then regress the return of stock i on the matched quintile return. The common (idiosyncratic) component of volatility is given by the logarithm of the average absolute value of the fitted return (residual) from the regression, where the average is computed over the past five trading days including the current day. The common volatility elasticity of spread is noisy and statistically insignificant for small stocks. Hence, the above result does not appear to be an artifact of nonsynchronous trading.

we do not include an intercept to limit estimation error. Common (idiosyncratic) order imbalance volatility for each stock-day is computed as the standard deviation of the fitted (residual) values and denoted by HFOIV_C (HFOIV_I).

In Table 7, the inclusion of HFOIV makes the sign of volume components negative, in line with Table 3. Both HFOIV_C and HFOIV_I are positive and significant for large and small stocks. For the average large stock, the ratio of standard deviation to mean is roughly 2.1% for HFOIV_C and 1% for HFOIV_I. Hence, while HFOIV_I is larger, HFOIV_C is important as well. A positive and significant HFOIV_I is consistent with both adverse selection and inventory effects, but a positive and significant HFOIV_C seems more supportive of inventory effects. Due to estimation error, these results are likely a lower bound on the importance of HFOIV_C.

5 Order imbalance volatility and the cross-section of stock returns

We examine whether HFOIV is priced in the cross-section of stock returns. First, liquidity providers could be specialized and hold undiversified portfolios. Second, HFOIV could represent a source of undiversifiable risk since order imbalances are correlated across stocks (Hasbrouck and Seppi (2001)). Indeed, Section 4.2 shows that the common component in HFOIV affects spreads.

Our model refers to liquidity provision at a high frequency. Inventory effects are likely to be most relevant at short horizons. In a sample of NYSE intermediary data spanning 1994 to 2005, Hendershott and Menkveld (2014) find inventory half lives of approximately half a day for large stocks and two days for small stocks. As a result, we focus our analysis on weekly returns. We divide our sample period into non-overlapping intervals of five trading days, which gives us 797 weekly return observations. Our main variable of interest is a measure of prior high-frequency order imbalance volatility: an exponentially-weighted moving average of past HFOIV with a half-life of one day. The results are similar if we simply use lagged HFOIV. In some of our specifications, we use effective spread, realized volatility, and absolute order imbalance as controls. To ensure a proper comparison with HFOIV, all of these variables are also computed using an exponentially-weighted moving average with a one-day half life.

We first consider portfolio sorts. Our model implies that we should control for turnover when examining the effect of order imbalance volatility. We therefore perform sequential sorts based on turnover and HFOIV. We measure turnover as the average daily turnover over the previous month. Table 8 reports value-weighted four-factor (Fama-French + momentum) alpha of portfolios built from sequential sorts with NYSE breakpoints. In Panel (a), stocks are first sorted into quintiles based on prior turnover and then are sorted again within each turnover quintile based on prior HFOIV. The results support the idea that HFOIV is priced in the cross-section of stock returns. Within all turnover quintiles, the long-short HFOIV portfolio earns positive alpha. The alpha is statistically significant at the level of 1% for all but one quintile. For example, among stocks with medium turnover, the weekly (five-day) value-weighted alpha is 0.13% with a t-statistic of 3.06.

In Panel (b), the order of the sequential sort is reversed. Among all HFOIV quintiles, stocks with high turnover tend to earn lower alpha than stocks with low turnover. These alphas are statistically significant at the level of 5% for three of the five long-short portfolios. This evidence is consistent with turnover reducing liquidity provider's risk conditional on HFOIV. In untabulated results we find that average turnover is not significantly associated with lower weekly returns in univariate quintile or decile sorts, in contrast to the earlier evidence of Datar et al. (1998). This result is consistent with our model, in which order imbalance volatility and turnover must be disentangled.

We consider several robustness checks. Raw returns are reported in the Internet Appendix and produce mostly similar results. The results are also robust to skipping a full day between the measurement of the predictor variables and the start of the weekly return. Finally, the results are stronger with CRSP breakpoints or equal-weighted portfolios.

Next, we use value-weighted Fama and MacBeth (1973) regressions, which allow us to control for many variables. Table 9 reports the results. HFOIV predicts higher weekly returns (first column). This relation is statistically significant and becomes stronger once we control for turnover (second column). Furthermore, HFOIV remains a strong and statistically significant (with a t-statistic of 3.17) predictor of weekly returns when we control for a host of other liquidity variables (third column). In particular, the regression controls for turnover, market capitalization, lagged return, illiquidity (Amihud (2002)), realized volatility, effective spread, depth as a fraction of volume, and price impact (lambda). We believe this evidence is of particular interest since many high-frequency liquidity and volatility measures do not appear to be priced (Lou and Shu (2017)). Indeed, none of the other liquidity and volatility measures in the third column are statistically significant.

We consider three additional imbalance measures in the last column of Table 9. First, we add the absolute daily order imbalance. Second, we compute the monthly standard deviation of share order imbalance divided by share volume. Chordia et al. (2018) show that this variable predicts future monthly returns. Finally, we compute the absolute daily trade imbalance over the total number of trades. As shown by Aktas et al. (2007), this measure approximates the PIN measure of Easley et al. (1996) at a daily frequency.

The return predictability of HFOIV is not subsumed by other imbalance measures. Absolute order imbalance predicts returns negatively, which corroborates our earlier finding that HFOIV differs from absolute order imbalance. The Chordia et al. (2018) measure and PIN approximation tend to predict returns positively. These measures are associated with adverse selection risk, whereas HFOIV is motivated by inventory risk. We expect these variables to be complementary as they capture different dimensions of liquidity. We note, however, that the Chordia et al. (2018) measure and PIN approximation are statistically insignificant when included separately to the regression without absolute order imbalance. In contrast, HFOIV remains statistically significant in all of our specifications.[27]

In summary, HFOIV predicts future weekly returns in the cross-section even after controlling for other predictors.

6 Additional results

We examine the relation between HFOIV and alternative liquidity measures, and we discuss the measurement of volatility.

6.1 How do other liquidity measures relate to HFOIV?

We consider two standard measures of price impact and a measure of depth. First, we estimate for each stock-day $r_{i,t,k} = \delta_{i,t} + \lambda_{i,t}\sqrt{|\mathrm{OI}^{\$}_{i,t,k}|}\mathrm{sign}(\mathrm{OI}^{\$}_{i,t,k}) + e_{i,t}$, where $r_{i,t,k}$ is the five-minute midquote return for stock i on day t in interval k, and $\mathrm{OI}^{\$}_{i,t,k}$ is the dollar order imbalance (as in Hasbrouck (2009)). Second, we compute a measure of price impact based on Amihud (2002). We compute illiquidity for each stock-day using intraday five-minute midquote returns and dollar volume: $\mathrm{ILLIQ}_{i,t} = \frac{1}{\#\text{traded intervals}} \sum_{k\epsilon\{j|\mathrm{DVOL}_{i,t,j}>0\}} \frac{|r_{i,t,k}|}{\mathrm{DVOL}_{i,t,k}}$. Third, another important dimension

[27] Like for portfolio sorts these results are robust to skipping a day between the measurement interval of the explanatory variables and the following weekly return.

of liquidity is depth. For each stock-day, we compute the average of time-weighted share depth at the best bid and best ask (as a fraction of shares outstanding).

We estimate (15) every year with each alternative liquidity measures as dependent variable. Year-by-year results are reported in Tables IA.21, IA.22, and IA.23 in the Internet Appendix. The first price impact measure, λ, is negatively related to volume, positively related to volatility, and negatively related to HFOIV. This is inconsistent with the spread results. In contrast, ILLIQ is positively related to HFOIV, in line with the spread results. What explains this discrepancy? A negative relation between λ and HFOIV is not surprising. If we assume that order imbalance is symmetric and equally likely to be positive or negative, then $\lambda = \frac{\sigma_r}{\mathrm{E}[|\mathrm{OI}|]}\mathrm{corr}[r, \sqrt{|\mathrm{OI}^{\$}|}\mathrm{sign}(\mathrm{OI}^{\$})]$ for a given stock. Hence, λ is positively (negatively) associated with return volatility (HFOIV) by construction. The second price impact measure, ILLIQ, is positively (negatively) associated with return volatility (volume). ILLIQ is, however, positively related to HFOIV. Hence, a measure of price impact based on volume produces results that are consistent with the spread evidence above, in contrast to a measure based on signed volume. The distinction between the two goes back to the empirical interpretation of noise trading volatility in Kyle-type models that we review in Appendix A.

Depth is negatively associated with HFOIV for large stocks. Hence, volatile imbalances are accompanied by a decrease in liquidity as measured by depth and spread for large stocks.[28] We find a negative relation between depth and HFOIV in each year across all size groups except for small stocks. For these stocks, the relation is not stable over the sample period.

Finally, we decompose effective spread into price impact and realized spread. Price impact is generally associated with adverse selection and equals the signed change in the midquote over a fixed time period following a trade. Realized spread is generally associated with liquidity provision and equals the signed difference between the trade price and the midquote over the same time period. Table IA.24 in the Internet Appendix reports estimates of month-by-month panel regressions of price impact and realized spread on volume, realized volatility, and HFOIV for large stocks in 2017. HFOIV is weakly associated with price impact and strongly associated with realized spread, which

[28]Since we do not observe depth beyond the best quotes, changes in spreads can lead to mechanical changes in depth. Traders can cancel their limit orders at the best ask and replace them with new limit orders at the next level of the ask book. If other orders are unchanged we would observe an increase in depth at the best ask, which wrongly suggests improved liquidity. However, the results are not sensitive to including spread as a control in the regressions.

lends support to the inventory interpretation.

6.2 Measuring volatility

We compare the explanatory power of realized volatility for spread to that of other volatility measures commonly used in the literature such as the absolute return and the average absolute return over the past week. Table IA.25 in the Internet Appendix reports median adjusted R-squared (across years) from estimating (15) and (14) with three different measures of volatility: the absolute daily return, the average absolute daily return over the previous week (including the current day), and the five-minute realized volatility. Realized volatility dramatically improves the explanatory power of the regressions. For instance, in level regressions for small stocks the R^2 increases from about 14% when using the low-frequency measures to about 31% when using realized volatility. The improvement is also marked for large stocks. This highlight the importance of using a "better" measure of volatility to explain the dynamics of spreads, at least in recent samples.

7 Conclusion

We develop a simple continuous-time inventory model to study the dynamics of liquidity. In the model, order imbalance volatility is a key driver of liquidity. Controlling for volume, an increase in order imbalance volatility leads the liquidity provider to widen the spread because of increased inventory risk.

Empirically, we provide new evidence about the time-series relation between daily liquidity, volume, and volatility. For large stocks, volume tends to be positively associated with effective spread. This relation is not explained by volatility and is mostly driven by the common component of volume. This evidence is difficult to explain with adverse selection theories and is more consistent with inventory risk theories.

We compute a measure of high-frequency order imbalance volatility, HFOIV. This measure is strongly associated with effective spread and substantially improves the fit of spread regressions. Consistent with the model, once we control for HFOIV, the relation between volume and spread becomes negative and is consistent across small and large stocks. In line with an inventory risk interpretation, HFOIV spikes on days associated with uninformed rebalancing. HFOIV is priced in the cross-section of weekly returns. This predictability holds for value-weighted returns even after controlling for many other liquidity variables.

Though our evidence suggests that inventory risk matters, high spreads at the beginning of the trading day suggest an important role for adverse selection risk. Our model can accommodate adverse selection by allowing the dividend growth rate to vary with the order flow. An examination of how the interaction between inventory risk and adverse selection affects the spread-volume-order imbalance volatility relation is an interesting avenue for future work.

References

Admati, A. R. and P. Pfleiderer (1988). A Theory of Intraday Patterns: Volume and Price Variability. *Review of Financial Studies 1*(1), 3–40.

Aktas, N., E. De Bodt, F. Declerck, and H. Van Oppens (2007). The pin anomaly around m&a announcements. *Journal of Financial Markets 10*(2), 169–191.

Amihud, Y. (2002). Illiquidity and Stock Returns: Cross-Section and Time-Series Effects. *Journal of Financial Markets 5*(1), 31–56.

Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *Journal of Finance 51*(1), 169–204.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*(1), 43–76.

Back, K. (1992). Insider trading in continuous time. *Review of Financial Studies 5*(3), 387–409.

Back, K., K. Crotty, and T. Li (2018). Identifying information asymmetry in securities markets. *Review of Financial Studies 31*(6), 2277–2325.

Barclay, M. J., T. Hendershott, and C. M. Jones (2008). Order Consolidation, Price Efficiency, and Extreme Liquidity Shocks. *Journal of Financial and Quantitative Analysis 43*(01), 93.

Barinov, A. (2014). Turnover: Liquidity or Uncertainty? *Management Science*.

Bogousslavsky, V. (2021). The cross-section of intraday and overnight returns. *Journal of Financial Economics 141*(1), 172–194.

Bogousslavsky, V. and D. Muravyev (2021). Who trades at the close? Implications for price discovery and liquidity. Working paper.

Brogaard, J., B. Hagströmer, L. Norden, and R. Riordan (2015). Trading fast and slow: Colocation and liquidity. *Review of Financial Studies 28*(12), 3407–3443.

Chan, K. and W. M. Fong (2000). Trade size, order imbalance, and the volatility-volume relation. *Journal of Financial Economics 57*(2), 247–273.

Chordia, T., J. Hu, A. Subrahmanyam, and Q. Tong (2018). Order Flow Volatility and Equity Costs of Capital. *Management Science*.

Chordia, T., R. Roll, and A. Subrahmanyam (2000). Commonality in Liquidity. *Journal of Financial Economics 56*(1), 3–28.

Chordia, T., R. Roll, and A. Subrahmanyam (2001). Market Liquidity and Trading Activity. *Journal of Finance 56*(2), 501–530.

Chordia, T., R. Roll, and A. Subrahmanyam (2002). Order imbalance, liquidity, and market returns. *Journal of Financial Economics 65*(1), 111–130.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 135–155.

Collin-Dufresne, P., K. Daniel, and M. Saglam (2020). Liquidity regimes and optimal dynamic asset allocation. *Journal of Financial Economics 136*(2), 379 – 406.

Collin-Dufresne, P. and V. Fos (2016a). Insider Trading, Stochastic Liquidity and Equilibrium Prices. *Econometrica 84*(4), 1441–1475.

Collin-Dufresne, P. and V. Fos (2016b). Insider trading with a random horizon. *working paper SFI@EPFL*.

Comerton-Forde, C., T. Hendershott, C. M. Jones, P. C. Moulton, and M. S. Seasholes (2010). Time variation in liquidity: The role of market-maker inventories and revenues. *Journal of Finance 65*(1), 295–331.

Datar, V. T., N. Naik, and R. Radcliffe (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets 1*(2), 203–219.

Demsetz, H. (1968). The Cost of Transacting. *Quarterly Journal of Economics*, 33–53.

Duarte, J., E. Hu, and L. Young (2020). A comparison of some structural models of private information arrival. *Journal of Financial Economics 135*(3), 795–815.

Easley, D., N. M. Kiefer, M. O'Hara, and J. B. Paperman (1996). Liquidity, information, and infrequently traded stocks. *Journal of Finance 51*(4), 1405–1436.

Easley, D. and M. O'Hara (1992). Time and the Process of Security Price Adjustment. *Journal of Finance 47*(2), 577–605.

Epps, T. W. and M. L. Epps (1976). The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis. *Econometrica*, 305–321.

Fama, E. F. and J. D. MacBeth (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy 81*(3), 607–636.

Foster, F. D. and S. Viswanathan (1990). A Theory of the Interday Variations in Volume, Variance, and Trading Costs in Securities Markets. *Review of Financial Studies 3*(4), 593–624.

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

Foucault, T. (1999). Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial Markets 2*(2), 99–134.

Foucault, T., O. Kadan, and E. Kandel (2005). Limit Order Book as a Market for Liquidity. *Review of Financial Studies 18*(4), 1171–1217.

Gallant, R. A., P. E. Rossi, and G. Tauchen (1992). Stock Prices and Volume. *Review of Financial*

Studies 5(2), 199—-242.

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

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

Goettler, R. L., C. A. Parlour, and U. Rajan (2005). Equilibrium in a dynamic limit order market. *Journal of Finance 60*(5), 2149–2192.

Grossman, S. J. and M. H. Miller (1988). Liquidity and Market Structure. *Journal of Finance 43*(3), 617–633.

Hagströmer, B. (2019). Bias in the effective bid-ask spread. *Available at SSRN 2939579*.

Hasbrouck, J. (2009). Trading Costs and Returns for U . S . Equities: Estimating Effective Costs from Daily Data. *Journal of Finance 64*(3), 1445–1477.

Hasbrouck, J. and D. J. Seppi (2001). Common factors in prices, order flows, and liquidity. *Journal of Financial Economics 59*(3), 383–411.

Hendershott, T., C. Jones, and A. J. Menkveld (2011). Does algorithmic trading improve liquidity? *Journal of Finance 66*(1), 1–33.

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

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

Johnson, T. (2008). Volume, Liquidity, and Liquidity Risk. *Journal of Financial Economics 87*(2), 388–417.

Kim, S. T. and H. R. Stoll (2014). Are Trading Imbalances Indicative of Private Information? *Journal of Financial Markets 20*, 151–174.

Korajczyk, R. A. and D. Murphy (2019). High-frequency market making to large institutional trades. *Review of Financial Studies 32*(3), 1034–1067.

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica 53*(6), 1315–1335.

Kyle, A. S. and A. A. Obizhaeva (2016). Market Microstructure Invariance: Empirical Hypotheses. *Econometrica 84*(4), 1345–1404.

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

Lee, C. M. C., B. Mucklow, and M. J. Ready (1993). Spread, Depths, and the Impact of Earnings Information: An Intraday Analysis. *Review of Financial Studies 6*(2), 345–374.

Lo, A. W. and J. Wang (2000). Trading Volume : Definitions, Data Analysis, and Implications of Portfolio Theory. *Review of Financial Studies 13*(2), 257–300.

Lou, X. and T. Shu (2017). Price Impact or Trading Volume: Why Is the Amihud (2002) Measure Priced? *Review of Financial Studies* (December).

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

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

Parlour, C. A. (1998). Price dynamics in limit order markets. *Review of Financial Studies 11*(4), 789–816.

Patton, A. J. and M. Verardo (2012). Does Beta Move with News? Firm-specific Information Flows and Learning about Profitability. *Review of Financial Studies 25*(9), 2789–2839.

Sarkar, A. and R. A. Schwartz (2009). Market sidedness: Insights into motives for trade initiation. *Journal of Finance 64*(1), 375–423.

Stoll, H. (2000). Friction. *Journal of Finance 55*(4), 1479–1514.

Stoll, H. R. (1978a). The pricing of security dealer services: An empirical study of nasdaq stocks. *Journal of Finance 33*(4), 1153–1172.

Stoll, H. R. (1978b). The supply of dealer services in securities markets. *Journal of Finance 33*(4), 1133–1151.

Tauchen, G. E. and M. Pitts (1983). The price variability-volume relationship on speculative markets. *Econometrica*, 485–505.

Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics 99*(1), 1–10.

Toth, B., I. Palit, F. Lillo, and J. D. Farmer (2015). Why is equity order flow so persistent? *Journal of Economic Dynamics and Control 51*, 218–239.

van Kervel, V. and A. J. Menkveld (2019). High-frequency trading around large institutional orders. *Journal of Finance 74*(3), 1091–1137.

Appendix A. Volume and order imbalance volatility in adverse selection models of spreads

In this appendix, we discuss how volume, order imbalance volatility, and return volatility are related to spreads in classic adverse selection microstructure models.

Consider the classic continuous-time model of informed trading (Kyle (1985) and Back (1992)). The value of the firm v is drawn from a Normal distribution $v \sim M(0, \sigma_i^2)$. An informed trader, who knows the realization of v, and uninformed noise traders trade continuously by sending their respective order flow (dX_t^i and dX_t^u) to competitive risk-neutral market makers, who offset the net order imbalance $d\theta_t = dX_t^i + dX_t^u$ at a price such that their expected profit is zero given their information set, that is $S_t = E[v \,|\, \mathcal{F}_t^\theta]$. (We denote the information filtration generated by observing the net order imbalance θ_t by $\mathcal{F}_t^\theta$.) The model assumes that noise trader demand is driven by an unpredictable Brownian motion Z_t with volatility σ_u, so that $dX_t^u = \sigma_u dZ_t$. Informed traders maximize their expected profit from trading continuously between 0 and T. In the Kyle-Back equilibrium, price adjusts linearly to total order flow, that is $dS_t = \lambda d\theta_t$, with a constant price impact $\lambda = \frac{\sigma_i}{\sigma_u}$, so that price volatility equals the constant σ_i. Further, the informed agent's optimal order flow is absolutely continuous, $dX_t^i = \frac{1}{\lambda(T-t)}(v - S_t)dt$. A derivation of the continuous time equilibrium as the limit of the discrete time model is in Kyle (1985). A more general proof is in Back (1992) and Collin-Dufresne and Fos (2016a).

Since there are three groups of traders (informed, noise, and market makers), it would seem natural (as in the discrete time model of Admati and Pfleiderer (1988)) to define volume (per unit of time) as one-half the sum of the absolute value of each group's order flow, that is: $\text{VOL} = \frac{1}{2dt}(|dX_t^i| + |dX_t^u| + |d\theta_t|)$. However, in the continuous time limit (as $dt \to 0$) the trading of uninformed traders, which has infinite variation and thus dwarfs that of informed traders, makes this quantity meaningless. Instead, to measure trading volume, a related quantity that can be estimated is $E[VOL^2 \,|\, \mathcal{F}_t^\theta] := \lim_{dt \to 0} \frac{1}{4dt} E[(|dX_t^i| + |dX_t^u| + |d\theta_t|)^2 \,|\, \mathcal{F}_t^\theta] \approx \sigma_u^2$.[29]

Further, in equilibrium, the market makers' net order imbalance θ_t is a σ_u-Brownian motion on its own filtration. To summarize in this model:

$$\text{price impact} = \frac{\sigma_i}{\sigma_u}, \tag{A1}$$

$$\text{price volatility} = \sigma_i, \tag{A2}$$

$$E[\text{VOL}^2] = \sigma_u^2 \tag{A3}$$

$$\frac{1}{dt} Var[d\theta_t] = \sigma_u^2 \tag{A4}$$

In the continuous time Kyle-Back model, volume and order imbalance volatility are driven by noise trading volatility σ_u, as the informed agent is able to hide her trading by smoothing her trades over the entire time horizon. The informed agent's trades are motivated by private information,

[29] Expanding terms and keeping only terms of 'order' dt gives the result.

captured by σ_i, which drives price volatility. All else equal, an increase in noise trading volatility results in a higher volume and a lower price impact. This is intuitive as more noise trading reduces the market maker's adverse selection. For price impact to be positively associated with volume, $\frac{\partial\sigma_i/\sigma_i}{\partial\sigma_u/\sigma_u} > 1$. This condition is difficult to satisfy in most models. To illustrate that this is not specific to continuous time or to risk-neutrality or to ad-hoc noise-trading, we also solved (in unreported results) the simple one-period adverse selection model of Glosten (1989), which extends the Kyle (1985) model to a risk-averse informed trader and replaces noise traders with endowment shocks. In that model, we compute the relation between volume and spread as we move various parameters such as risk-aversion and the variances of the informed signal, of the endowment shock, and of the fundamental. For all comparative statics, the model generates a negative volume-spread relation.

A negative relation between volume and spreads also arises in most dynamic extensions of Kyle's model that generate time-varying volume and volatility by introducing time-varying noise trading volatility (e.g, Admati and Pfleiderer (1988), Collin-Dufresne and Fos (2016a)), or time-varying rate of news arrival (Foster and Viswanathan (1990), Collin-Dufresne and Fos (2016b)). This is because the informed agent's trading is endogenous and it is never optimal to trade so as to increase trading costs. On the other hand, more informed trading is always associated with higher price volatility (as more information is released) in a Kyle-type framework. Thus, adverse selection models typically generate a positive relation between volume or market depth (i.e., inverse price impact) and volatility, if the variation in informed trading is an endogenous response to variation in uninformed noise trading. This is because a higher noise trading volatility increases volume and decreases price impact, but leads to more aggressive informed trading, which increases price volatility (e.g, Admati and Pfleiderer (1988), Collin-Dufresne and Fos (2016a)).

However, volume, volatility, and price impact can all be positively correlated if there exists a direct positive link between volume (or noise trading) and private information. For example, in Easley and O'Hara (1992) the increase in trading volume implies an increase in the likelihood of informed trading which raises price impact and volatility. Similarly, in Collin-Dufresne and Fos (2016b) the increase in trading volume may also raise the likelihood of private information disclosure, which increases the incentive for the informed to trade aggressively, and thus raises price impact and price volatility.

In contrast to the continuous-time Kyle-Back model just discussed, the continuous-time inventory model presented in the main text can have very different implications for the relation between spreads, volume, and order imbalance volatility.

Appendix B. Continuous-time inventory model: Derivations

The risk-averse liquidity provider maximizes

$$\max_{c_t, n_t} \mathrm{E}\left[\int_0^{\infty} -e^{-\beta t - \alpha c_t}\right] \tag{B1}$$

subject to

$$dW_t = (rW_t - c_t)dt + n_t(\mu_t - rS)dt + n_t\sigma_t dZ_t + n_t \sum_{i=1}^{M} \mathbf{1}_{\{N_{t^-}=i\}} \sum_{j\neq i} \eta_{ij}(dN_{ij}(t) - \lambda_{ij}dt). \tag{B2}$$

We conjecture that in equilibrium the stock price is a function of only the dividend and Markov state, that is $S(\delta, N)$ and that the value function is of the form $J(W_t, N_t) = \max_{c,n} \mathrm{E}[\int_t^\infty -e^{-\beta(s-t)-\alpha c_s}ds]$. The HJB equation (assuming the current state is $W, N = i$):

$$0 = \max_{c,n} \left\{ -e^{-\alpha c} + J_W(rW - c + n(\mu_i - \overline{\lambda\eta}_i - rS_t)) + \frac{1}{2}J_{WW}n^2\sigma^2 - \beta J + \sum_{j\neq i} \lambda_{ij}(J(W + n\eta_{ij}, j) - J(W, i)) \right\}$$

where, to simplify notation, we defined the compensator $\overline{\lambda\eta}_i = \sum_{j\neq i} \lambda_{ij}\eta_{ij}$. The FOC are (conditional on being in state $N = i$)

$$J_W = \alpha e^{-\alpha c}, \tag{B3}$$

$$0 = J_W(\mu_i - \overline{\lambda\eta}_i - rS) + J_{WW}n\sigma^2 + \sum_{j\neq i} \lambda_{ij}\eta_{ij}J_W(W + n\eta_{ij}, j). \tag{B4}$$

We guess that the value function is of the form

$$J(W, N) = -\frac{1}{r}e^{-\alpha(rW - b(N))}$$

for some function $b(N) := \sum_{i=1}^{M} b_i \mathbf{1}_{\{N=i\}}$. The first FOC then implies

$$c(W, N = i) = rW - b_i. \tag{B5}$$

The second FOC implies

$$\mu_i - rS = \alpha rn\sigma^2 + \sum_{j\neq i} \lambda_{ij}\eta_{ij}(1 - e^{-\alpha(rn\eta_{ij} - b_j + b_i)}) \tag{B6}$$

Further, the b_i coefficients solve the system of equations ($\forall i, j$):

$$0 = -r + r\alpha(b_i + n(\mu_i - \overline{\lambda\eta}_i - rS)) - \frac{1}{2}r^2\alpha^2n^2\sigma^2 + \beta - \sum_{j\neq i} \lambda_{ij}(e^{-\alpha(rn\eta - b_j + b_i)} - 1). \tag{B7}$$

From equation (B6) we can substitute $\mu_i - rS$ to obtain:

$$0 = -r + r\alpha b_i + \frac{1}{2}r^2\alpha^2n^2\sigma^2 + \beta - \sum_{j\neq i} \lambda_{ij}(e^{-\alpha(rn\eta_{ij} - b_j + b_i)}(1 + r\alpha n\eta_{ij}) - 1). \tag{B8}$$

Now in equilibrium we must have $n_t = \theta(N_t)$. Plugging into the equations we get the system of equations which must be satisfied by $S(\delta, N)$ and the constants b_i for $i = \{1, \ldots, M\}$ in equilibrium:

$$\mu_i - rS = \alpha r \theta_i \sigma^2 + \sum_{j \neq i} \lambda_{ij} \eta_{ij} (1 - e^{-\alpha(r\theta_i \eta_{ij} - b_j + b_i)}), \tag{B9}$$

$$0 = -r + r\alpha b_i + \frac{1}{2} r^2 \alpha^2 \theta_i^2 \sigma^2 + \beta - \sum_{j \neq i} \lambda_{ij} (e^{-\alpha(r\theta_i \eta_{ij} - b_j + b_i)} (1 + r\alpha\theta_i \eta_{ij}) - 1). \tag{B10}$$

Note that μ, σ, η are all obtained from Itô's formula given an expression for $S(\delta, N)$. In fact, to simplify the search for the equilibrium stock price it is helpful to define the *risk-free discounted* value of the dividend:

$$V(\delta_t, N_t) = \mathrm{E}_t \left[\int_t^\infty e^{-r(s-t)ds} \delta_s ds \right].$$

To solve for V note that it satisfies the equation $\mathrm{E}_t[dV(\delta, N_t) + \delta dt] = rV(\delta, N_t)dt$. Then define $V(\delta, N = i) := V^i(\delta)$ and note that we have

$$V_\delta^i \kappa_\delta (\overline{\delta}_i - \delta) + \frac{1}{2} V_{\delta\delta}^i \sigma_\delta^2 + \sum_{j \neq i} \lambda_{ij} (V^j(\delta) - V^i(\delta)) = rV^i(\delta) - \delta. \tag{B11}$$

The solution is of the form

$$V^i(\delta) = \frac{\delta}{r + \kappa_\delta} + v_i,$$

where the constants v_i satisfy the system of equations:

$$\frac{\kappa_\delta \overline{\delta}_i}{r + \kappa_\delta} + \sum_{j \neq i} \lambda_{ij} (v_j - v_i) = rv_i. \tag{B12}$$

The solution obtains in terms of the transition matrix Λ (which has entry $\Lambda_{ij} = \lambda_{ij}\ \forall j \neq i$ and $\Lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}$) and where we define $\overline{\delta}$ to be the column vector of long run means $\overline{\delta}_i$ and I the M-dimensional identity matrix:

$$v = \frac{\kappa_\delta}{r + \kappa_\delta} (rI - \Lambda)^{-1} \overline{\delta}.$$

Now, we decompose the stock price as

$$S(\delta, N) = V(\delta, N) + s(N), \tag{B13}$$

where $s(N) := \sum_{i=1}^M s_i \mathbf{1}_{\{N_t = i\}}$. Then, since $\mathrm{E}_t[dV(\delta, N) + \delta dt] = rV(\delta, N)dt$ and applying Itô we obtain (setting $N_t = i$)

$$\mu_i - rS \equiv \mathrm{E}_t[dS_t/dt + \delta - rS] = \mathrm{E}_t[ds(N_t)/dt - rs(N_t)] = \sum_{j \neq i} \lambda_{ij}(s_j - s_i) - rs_i, \tag{B14}$$

$$\sigma = \frac{\sigma_\delta}{r + \kappa_\delta}, \tag{B15}$$

$$\eta_{ij} = v_j - v_i + s_j - s_i. \tag{B16}$$

Substituting into our system of equilibrium conditions (B9) and (B10) we find that the constants $s_i, b_i \ \forall i \in \{1, M\}$ which characterize the stock price and optimal consumption satisfy the system of equations $\forall i, j \in \{1, M\}$:

$$0 = rs_i + \alpha r \theta_i \sigma^2 + \sum_{j \neq i} \lambda_{ij} \{v_j - v_i - \eta_{ij} e^{-\alpha(r\theta_i \eta_{ij} - b_j + b_i)}\}, \tag{B17}$$

$$0 = -r + r\alpha b_i + \frac{1}{2} r^2 \alpha^2 \theta_i^2 \sigma^2 + \beta + \sum_{j \neq i} \lambda_{ij} \{1 - e^{-\alpha(r\theta_i \eta_{ij} - b_j + b_i)}(1 + r\alpha\theta_i \eta_{ij})\}, \tag{B18}$$

$$\sigma = \frac{\sigma_\delta}{r + \kappa_\delta}, \tag{B19}$$

$$\eta_{ij} = v_j - v_i + s_j - s_i, \tag{B20}$$

$$v = \frac{\kappa_\delta}{r + \kappa_\delta}(rI - \Lambda)^{-1} \bar{\delta}. \tag{B21}$$

The solution of this system characterizes the equilibrium.

Unconditional volume and order imbalance volatility

For the analysis below it is useful to compute the volume and the variance of the order imbalance. The unconditional expected volume of trading (per unit time) is given by

$$\begin{aligned} VOL &= \frac{1}{dt} E[|d\theta_t|] && \text{(B22)} \\ &= \sum_{i=1}^{M} \sum_{j \neq i} |\theta_j - \theta_i| \frac{1}{dt} E[\mathbf{1}_{\{N_{t^-} = i\}} dN_{ij}(t)] && \text{(B23)} \\ &= \sum_{i=1}^{M} \sum_{j \neq i} |\theta_j - \theta_i| \pi_i \sum_{j \neq i} \lambda_{ij}, && \text{(B24)} \end{aligned}$$

where $\pi_i = E[\mathbf{1}_{\{N_t = i\}}]$ is the unconditional (stationary) probability of being in a given state i. (The vector π solves the system of equations: $\Lambda \pi = 0$ and $\pi^\top \mathbf{1} = 1$.)

The unconditional variance of the cumulative order flow process is given by

$$\begin{aligned} OIV &= V[\theta_t] && \text{(B25)} \\ &= \sum_{i=1}^{M}\sum_{j=1}^{M}\theta_i\theta_j Cov(\mathbf{1}_{\{N_t=i\}},\mathbf{1}_{\{N_t=j\}}) && \text{(B26)} \\ &= \sum_{i=1}^{M}\sum_{j=1}^{M}\theta_i\theta_j\pi_i(\mathbf{1}_{\{i=j\}}-\pi_j) && \text{(B27)} \\ &= \sum_{i=1}^{M}\theta_i^2\pi_i-(\sum_{i=1}^{M}\pi_i\theta_i)^2. && \text{(B28)} \end{aligned}$$

Note for the second line that $Cov(\mathbf{1}_{\{N_t=i\}},\mathbf{1}_{\{N_t=j\}}) = E[\mathbf{1}_{\{N_t=i\}}\mathbf{1}_{\{N_t=j\}}] - E[\mathbf{1}_{\{N_t=i\}}]E[\mathbf{1}_{\{N_t=j\}}] = \pi_i\mathbf{1}_{\{i=j\}} - \pi_i\pi_j$.

Spreads in a symmetric model without "adverse selection"

We consider first the symmetric model where buyers and sellers arrive in a balanced fashion (or the market maker systematically waits for a buyer after having seen a seller) and there is no adverse selection in the sense that the fundamental dividend process is independent of the order flow. Specifically we consider the simple model with three states $M = 3$ characterized by

$$\begin{aligned} \lambda_{12} &= \lambda_{32} = \lambda_d \\ \lambda_{21} &= \lambda_{23} = \lambda_i \\ \theta_3 &= -\theta_1 = \theta && \text{(B29)} \\ \theta_2 &= 0 \\ \bar{\delta}_1 &= \bar{\delta}_2 = \bar{\delta}_3 = 0. \end{aligned}$$

So the inventory dynamics of the market maker look as follows:

$$-\theta \quad \underset{\xrightarrow[\lambda_d]{}}{\xleftarrow{\lambda_i}} \quad 0 \quad \underset{\xleftarrow[\lambda_d]{}}{\xrightarrow{\lambda_i}} \quad +\theta.$$

Since the long-run mean is constant and equal to zero across states, the solution to equation B21 is $v_i = 0\,\forall i$. Then it is straightforward to show (by analyzing the system of equation B17-B18) that there exists a unique symmetric solution characterized by

$$\begin{aligned} s_2 &= 0 \\ s_1 &= -s_3 > 0 \\ b_1 &= b_3, \end{aligned}$$

where s_1, b_1, b_2 solve the following system of equations:

$$0 = rs_1 - \alpha r\theta\sigma^2 + \lambda_d s_1 e^{-\alpha(b_1 - b_2 + r\theta s_1)} \tag{B30}$$

$$0 = -r + r\alpha b_1 + \beta + \frac{1}{2}r^2\alpha^2\theta^2\sigma^2 + \lambda_d\{1 - e^{-\alpha(b_1-b_2+r\theta s_1)}(1 + r\alpha\theta s_1)\} \tag{B31}$$

$$0 = -r + r\alpha b_2 + \beta + 2\lambda_i(1 - e^{\alpha(b_1-b_2)}) \tag{B32}$$

$$\sigma = \frac{\sigma_\delta}{r + \kappa_\delta}. \tag{B33}$$

The first equation can be rewritten as

$$s_1 = \frac{\alpha r\theta\sigma^2}{r + \lambda_d e^{-\alpha(b_1-b_2+r\theta s_1)}},$$

which shows that the spread is bounded above and below:

$$\lim_{\lambda_d \to \infty} s_1 = 0 \leq \frac{\alpha r\theta\sigma^2}{r + \lambda_d e^{\alpha(b_2-b_1)}} \leq s_1 \leq \lim_{\lambda_d \to 0} s_1 = \theta\alpha\sigma^2.$$

It is possible to prove that there is a unique solution to the system and that it satisfies $s_1 > 0$ and $b_2 > b_1 = b_3$.[30] In equilibrium the agent consumes more in state 1 and 3 than in state 2, where the uncertainty about the future order flow is highest.

Further, the risk-premium on the stock can be written as:

$$\mu_1 - rS = -(\lambda_d + r)s_1, \tag{B34}$$

$$\mu_2 - rS = 0, \tag{B35}$$

$$\mu_3 - rS = (\lambda_d + r)s_1. \tag{B36}$$

We see that the risk-premium is positively correlated with the inventory of the market maker. When the market maker is long in state 3 the price drops by s_1 so that the risk-premium becomes positive and the market maker is compensated for holding a positive inventory. Conversely, when the market maker is short in state 1 the price rises by s_1 so that the risk-premium becomes negative. This gives rise to an effective bid-ask spread. The spread per unit transacted is $\hat{s} = s_1/\theta$. From above, we have

$$\lim_{\lambda_d \to \infty} \hat{s} = 0 \leq \hat{s} \leq \lim_{\lambda_d \to 0} \hat{s} = \alpha\sigma^2.$$

In this simple model we can easily compute the unconditional volume and order-imbalance

[30] The bounds imply that $s_1 \geq 0$. To prove existence and uniqueness subtract B32 from B31 to get an expression of the form $f(x) = 0$ for $x = b_1 - b_2$. Analyze the variation of the function to show that $f(0) > 0$ and $f'(x) > 0\ \forall x$ and $\lim_{x\to-\infty} f(x) = -\infty$ and conclude that it admits one unique root x_0 such that $f(x_0) = 0$ and that $x_0 < 0$. To show that there is a unique solution $s_1 > 0$ that solves the non-linear equation B30, note that s_1 is the intersection of two continuous decreasing and strictly convex functions $f_1(s) = \frac{1}{s}$, which maps $(0, \infty)$ onto $(\infty, 0)$ and $f_2(s) = \frac{r+\lambda_d e^{-\alpha(b_1-b_2+r\theta s)}}{\alpha r\theta\sigma^2}$ which maps $(0, \infty)$ onto $(\frac{r+\lambda_d e^{-\alpha(b_1-b_2)}}{\alpha r\theta\sigma^2}, \frac{r}{\alpha r\theta\sigma^2})$ and which must therefore cross once and only once at some value $s^* \in (\frac{\alpha r\theta\sigma^2}{r+\lambda_d e^{-\alpha(b_1-b_2)}}, \frac{\alpha r\theta\sigma^2}{r})$.

volatility. First note that the unconditional state probabilities are given by:

$$\pi_1 = \pi_3 = \frac{\lambda_i}{\lambda_d + 2\lambda_i},$$
$$\pi_2 = \frac{\lambda_d}{\lambda_d + 2\lambda_i}.$$

Then we obtain

$$VOL = 2\theta(\pi_1\lambda_d + \pi_2\lambda_i) = 4\theta\frac{\lambda_i\lambda_d}{\lambda_d + 2\lambda_i} = 4\theta(\frac{1}{\lambda_i} + \frac{2}{\lambda_d})^{-1},$$
$$OIV = \theta^2(\pi_1 + \pi_3 - (\pi_1 - \pi_3)^2) = 2\theta^2\frac{\lambda_i}{\lambda_d + 2\lambda_i} = \frac{\theta}{2\lambda_d}VOL.$$

Figure 1. Bid-ask spread, order imbalance volatility (OIV), and volume for different trade intensities.
In the first row trade intensities are chosen so that OIV remains constant. Panel (a) shows the imbalance-increasing trading rate (λ_i) and imbalance-decreasing trading rate (λ_d). Panel (b) shows corresponding volume and OIV. Note that volume increases but OIV remains constant. Panel (c) shows the corresponding bid-ask spread $\hat{s}$. In the second row trade intensities are chosen so that volume remains constant. Panel (d) shows the imbalance-increasing and imbalance-decreasing trading rates. Panel (e) shows corresponding volume and OIV. Note that OIV decreases but volume remains constant. Panel (f) shows the corresponding bid-ask spread $\hat{s}$.

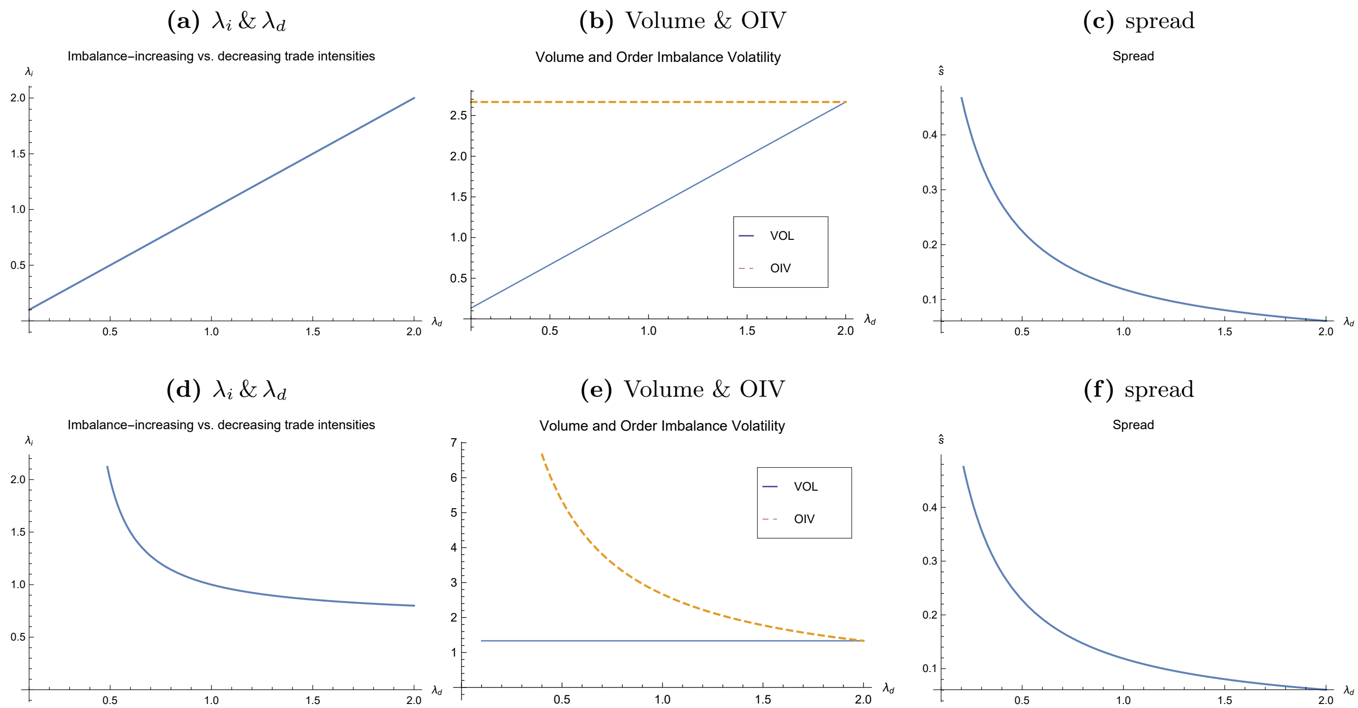
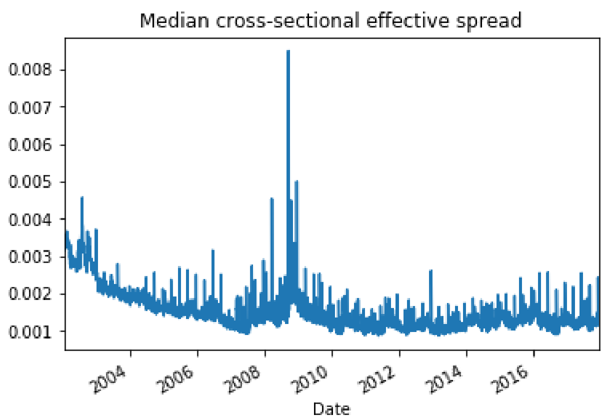
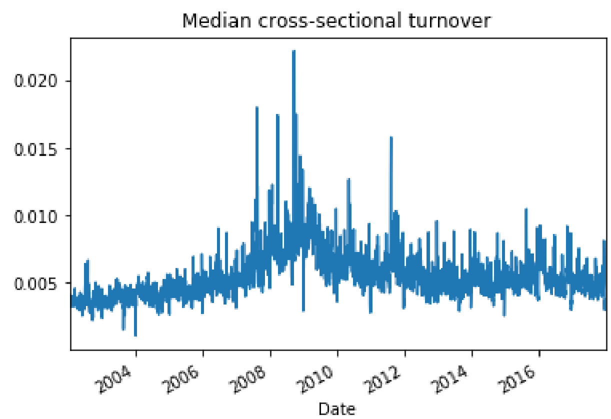


Figure 2. Spreads, volume, volatility, and order imbalance volatility.
Spread is the daily effective spread, volume is the daily intraday turnover, volatility is the realized volatility computed using five-minute intraday midquote returns, and order imbalance volatility is the high-frequency order imbalance volatility computed as the standard deviation of five-minute share imbalance (scaled by total shares outstanding) each day. This figure reports the daily cross-sectional median of each measure over 2002-2017. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading.

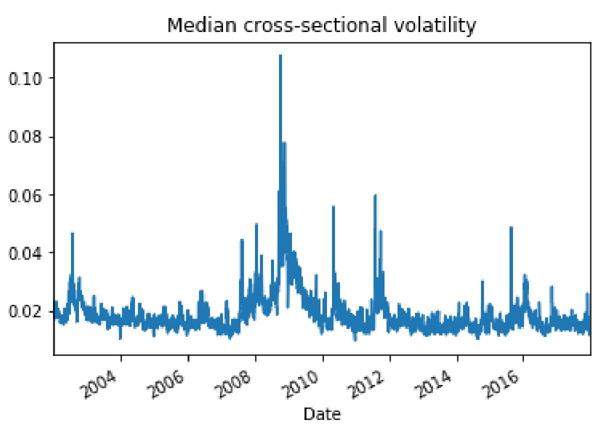
(a) Spread



(b) Turnover



(c) Volatility



(d) Order imbalance volatility

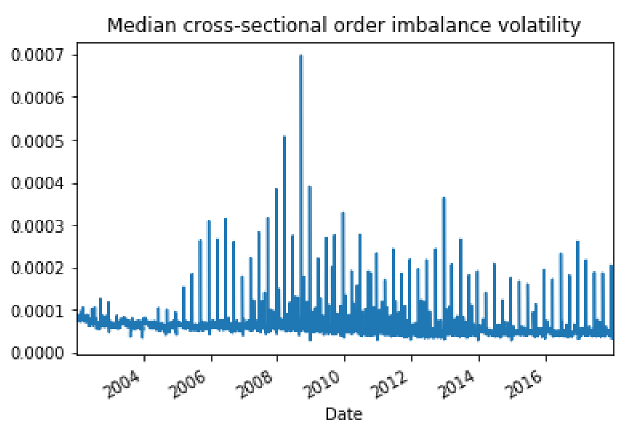
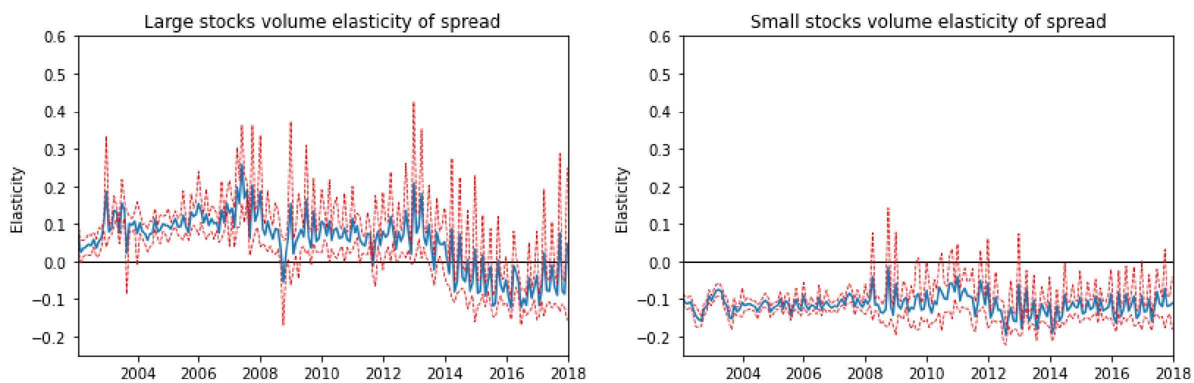


Figure 3. Effective spread regressed on volume and volatility across size quintiles.
Panel regression: $\log s_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover and $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns. Controls are (log) market capitalization, (log) price, and day-of-the-week indicators. The regression includes stock fixed effects and is estimated on a month-by-month basis for stocks in a given size quintile. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each month. Standard errors are double-clustered by date and stock.

(a) Volume elasticity of spread



(b) Volatility elasticity of spread

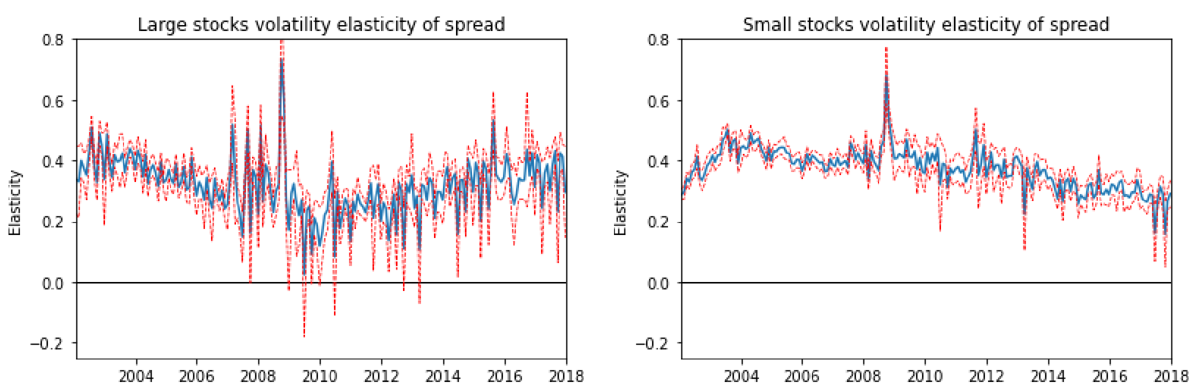


Table 1. Descriptive statistics of daily variables for stocks in the bottom and top size quintiles for a sample of years.

The spread is the percent effective spread (reported in basis points), turnover is the intraday turnover, volatility is the realized volatility computed using five-minute midquote returns, and HFOIV is the high-frequency order imbalance volatility, computed as the standard deviation of five-minute share imbalance (scaled by total shares outstanding). All these variables are computed for each stock on each day. The within standard deviation (σ (within)) is computed as the standard deviation of the deviations from the time-mean of each stock. Spreads are winsorized at 0.05% and 99.95% each year. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading.

		2002	2004	2006	2008	2010	2012	2014	2016
Small stocks									
Spread [bps]	mean	91.53	66.86	58.44	87.78	49.39	59.40	65.21	65.54
	median	73.68	50.94	40.92	49.39	35.55	40.66	46.06	44.53
	σ (within)	51.30	38.47	35.92	73.97	29.57	37.09	43.05	44.88
Turnover [%]	mean	0.40	0.51	0.50	0.53	0.52	0.43	0.49	0.49
	median	0.17	0.20	0.20	0.28	0.29	0.23	0.24	0.25
	σ (within)	0.85	1.40	1.01	0.84	0.95	0.87	0.90	1.30
Volatility [%]	mean	2.59	2.54	2.18	4.03	2.54	2.33	2.38	2.75
	median	2.12	2.22	1.91	3.36	2.32	2.06	2.10	2.35
	σ (within)	1.89	1.55	1.25	3.38	1.26	1.33	1.27	1.57
HFOIV [bps]	mean	1.70	1.46	1.44	1.37	1.02	0.97	1.00	0.94
	median	0.96	0.81	0.76	0.63	0.60	0.54	0.56	0.54
	σ (within)	3.67	3.15	3.64	5.08	1.97	2.18	2.71	3.87
Obs.		125,586	144,105	148,149	129,577	122,615	118,056	128,174	120,994
Large stocks									
Spread [bps]	mean	16.13	8.28	6.68	8.28	5.00	4.65	4.59	4.76
	median	12.66	6.59	5.35	6.21	4.05	3.66	3.44	3.62
	σ (within)	11.75	5.66	4.83	8.33	3.16	3.02	4.49	4.18
Turnover [%]	mean	0.73	0.67	0.76	1.42	1.12	0.91	0.79	0.82
	median	0.47	0.46	0.53	1.03	0.82	0.67	0.58	0.61
	σ (within)	0.71	0.58	0.58	1.22	0.88	0.74	0.62	0.63
Volatility [%]	mean	2.63	1.33	1.31	2.90	1.54	1.25	1.14	1.32
	median	1.94	1.18	1.17	2.26	1.31	1.12	0.99	1.12
	σ (within)	27.99	2.16	0.47	2.12	28.74	0.43	0.47	0.61
HFOIV [bps]	mean	0.74	0.60	0.63	0.69	0.59	0.51	0.49	0.47
	median	0.51	0.41	0.40	0.43	0.38	0.32	0.32	0.31
	σ (within)	0.85	0.61	0.78	0.96	0.74	0.61	0.63	0.59
Obs.		129,987	151,170	158,222	137,587	125,222	121,309	130,998	128,478

Table 2. Correlations among daily variables for stocks in the bottom and top size quintiles. s is the percent effective spread, τ is the intraday turnover, $|\bar{r}|$ is the average absolute return over the past five trading days including the current day, $|r|$ is the absolute daily return, σ is the realized volatility computed using five-minute midquote returns, |OI| is the absolute daily order imbalance as a fraction of shares outstanding, and HFOIV is the high-frequency order imbalance volatility, computed each day as the standard deviation of five-minute share imbalance (scaled by total shares outstanding). All the variables are in logs. The table reports the cross-sectional averages of the individual stocks' time-series correlations. Size quintiles are formed at the beginning of each month based on average daily market capitalization over the past year. Spreads are winsorized at 0.05% and 99.95% each year. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading.

Small stocks

	τ	$\lvert\bar{r}\rvert$	$\lvert r\rvert$	σ	|OI|	HFOIV
s	-0.17	0.22	0.18	0.40	-0.06	-0.00
τ		0.24	0.23	0.32	0.59	0.78
$\lvert\bar{r}\rvert$			0.49	0.47	0.10	0.12
$\lvert r\rvert$				0.41	0.13	0.14
σ					0.12	0.17
|OI|						0.60

Large stocks

	τ	$\lvert\bar{r}\rvert$	$\lvert r\rvert$	σ	|OI|	HFOIV
s	0.15	0.34	0.22	0.51	0.15	0.30
τ		0.41	0.32	0.48	0.40	0.72
$\lvert\bar{r}\rvert$			0.50	0.61	0.14	0.22
$\lvert r\rvert$				0.41	0.13	0.19
σ					0.14	0.26
|OI|						0.48

Table 3. Spread, volume, volatility, and order imbalance volatility.

This table reports median estimate, median t-statistic (in parentheses), and median adjusted R-squared across years. The following panel regression with stock fixed effects is estimated each year for small and large stocks: $\log s_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \beta_{\text{HFOIV}} \log \text{HFOIV}_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is daily intraday turnover, $\sigma_{i,t}$ is realized volatility estimated using five-minute returns over the current day, and HFOIV is high-frequency order imbalance volatility computed as the standard deviation of five-minute share imbalance scaled by total shares outstanding. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression is also estimated with daily changes in the variables: $\Delta s_{i,t} = \alpha_i + \beta_{\Delta\tau}\Delta\tau_{i,t} + \beta_{\Delta\sigma}\Delta\sigma_{i,t} + \beta_{\Delta\text{HFOIV}}\Delta\text{HFOIV}_{i,t} + \text{controls} + \epsilon_{i,t}$, where $\Delta x_t \equiv \log(\frac{x_t}{x_{t-1}})$. Spreads are winsorized at 0.05% and 99.95% each year. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Standard errors are double-clustered by date and stock.

	Median value across years							
	Small stocks				Large stocks			
β_τ	-0.15	-0.33			0.04	-0.28		
	(-26.29)	(-33.60)			(3.41)	(-19.71)		
β_σ	0.42	0.46			0.35	0.46		
	(37.00)	(45.05)			(19.01)	(30.95)		
β_{HFOIV}		0.20				0.29		
		(16.34)				(19.47)		
$\beta_{\Delta\tau}$			-0.08	-0.24			0.13	-0.23
			(-13.62)	(-26.46)			(7.66)	(-19.71)
$\beta_{\Delta\sigma}$			0.32	0.36			0.30	0.39
			(30.02)	(36.59)			(17.14)	(31.14)
$\beta_{\Delta\text{HFOIV}}$				0.17				0.29
				(16.34)				(22.56)
$R^2(\%)$	27.70	31.46	9.25	13.15	16.63	26.19	7.99	19.02

Table 4. Order imbalance volatility and absolute order imbalance.
This table reports median estimate, median t-statistic (in parentheses), and median adjusted R-squared across years. The following panel regression with stock fixed effects is estimated each year for small and large stocks: $\log s_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \beta_v \log v_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is daily intraday turnover, and $\sigma_{i,t}$ is realized volatility estimated using five-minute returns over the current day. $v_{i,t}$ is either high-frequency order imbalance volatility computed as the standard deviation of five-minute share imbalance scaled by total shares outstanding (HFOIV), absolute daily share imbalance scaled by total shares outstanding (|OI|), order imbalance volatility orthogonalized relative to absolute daily order imbalance ($\text{HFOIV}^{\perp|\text{OI}|}$), or absolute order imbalance orthogonalized relative to order imbalance volatility ($|\text{OI}|^{\perp\text{HFOIV}}$). The orthogonalization is done for each stock over the full sample by regressing one variable on the other and a constant and then taking the residuals as the orthogonalized values. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. Spreads are winsorized at 0.05% and 99.95% each year. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Standard errors are double-clustered by date and stock.

	Median value across years							
	Small stocks				Large stocks			
β_τ	-0.33	-0.19	-0.21	-0.15	-0.28	0.00	-0.12	0.04
	(-33.60)	(-35.85)	(-34.54)	(-26.99)	(-19.71)	(0.54)	(-12.10)	(3.57)
β_σ	0.46	0.43	0.44	0.43	0.46	0.37	0.41	0.35
	(45.05)	(39.17)	(41.92)	(38.05)	(30.95)	(20.63)	(24.16)	(18.98)
β_{HFOIV}	0.20				0.29			
	(16.34)				(19.47)			
$\beta_{\|\text{OI}\|}$		0.03				0.03		
		(10.92)				(6.95)		
$\beta_{\text{HFOIV}^{\perp\|\text{OI}\|}}$			0.10				0.20	
			(11.79)				(14.99)	
$\beta_{\|\text{OI}\|^{\perp\text{HFOIV}}}$				-0.01				-0.01
				(-5.06)				(-8.31)
$R^2(\%)$	31.46	28.29	29.68	27.93	26.19	17.44	22.24	16.74

Table 5. Order imbalance volatility, turnover, and calendar effects.
Log high-frequency order imbalance volatility (HFOIV) and log turnover are regressed on a set of explanatory variables and a set of fixed effects, separately for small stocks and large stocks. The explanatory variables include an indicator for the third Friday of each month, an indicator for the third Friday of end-of-quarter months (3rd Friday*EoQ), beginning-of-month (BoM) and end-of-month (EoM) indicators, an indicator for Russell reconstitution dates, indicators for the day before, of, and after an earnings announcement (EA-1, EA, and EA+1), and previous day price and market capitalization. The regression includes stock fixed effects, day-of-week fixed effects, calendar month fixed effect, and year fixed effects. The sample consists of NYSE, Amex, and NASDAQ common stocks from 2002 to 2017. Standard errors are double-clustered by date and stock and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level.

	Small stocks		Large stocks	
	log HFOIV	log turnover	log HFOIV	log turnover
log turnover	0.687***		0.911***	
	(0.003)		(0.006)	
3rd Friday	0.162***	0.106***	0.389***	0.069***
	(0.021)	(0.021)	(0.029)	(0.018)
3rd Friday*EoQ	0.340***	0.405***	0.530***	0.199***
	(0.048)	(0.046)	(0.058)	(0.027)
BoM	-0.045***	0.134***	-0.024***	0.083***
	(0.006)	(0.012)	(0.006)	(0.011)
EoM	-0.021***	0.091***	-0.018**	0.003
	(0.006)	(0.012)	(0.007)	(0.013)
Russell	-0.098	0.663***	-0.056	0.067
	(0.063)	(0.099)	(0.068)	(0.044)
EA-1	-0.004	0.083***	-0.046***	0.239***
	(0.003)	(0.006)	(0.003)	(0.006)
EA	0.003	0.674***	-0.074***	0.809***
	(0.004)	(0.012)	(0.006)	(0.011)
EA+1	-0.028***	0.402***	-0.056***	0.428***
	(0.003)	(0.008)	(0.005)	(0.006)
log price	0.376***	-0.024	0.014	0.037
	(0.020)	(0.048)	(0.011)	(0.032)
log mkt. cap.	-0.730***	0.670***	-0.095***	-0.356***
	(0.019)	(0.046)	(0.013)	(0.029)
Calendar month/day FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
R^2	60.79%	7.74%	60.31%	25.34%
Num. obs.	2,048,959		2,148,513	

Table 6. Effective spread, calendar effects, and order imbalance volatility.
The log effective spread (ES%) is regressed on log order imbalance volatility (HFOIV) or log absolute daily imbalance (|OI%|), a set of explanatory variables, and a set of control variables and fixed effects, separately for small stocks and large stocks. The explanatory variables include an indicator for the third Friday of each month, an indicator for the third Friday of end-of-quarter months (3rd Friday*EoQ), beginning-of-month (BoM) and end-of-month (EoM) indicators, an indicator for Russell reconstitution dates, indicators for the day before, of, and after an earnings announcement (EA-1, EA, and EA+1). Control variables include log turnover, log volatility, and previous day log price and log market capitalization. The regression includes stock fixed effects, day-of-week fixed effects, calendar month fixed effect, and year fixed effects. Spreads are winsorized at 0.05% and 99.95% each year. The sample consists of NYSE, Amex, and NASDAQ common stocks from 2002 to 2017. Standard errors are double-clustered by date and stock and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level.

	Small stocks			Large stocks		
	log ES%	log ES%	log ES%	log ES%	log ES%	log ES%
log HFOIV		0.236***			0.309***	
		(0.003)			(0.004)	
log |OI%|			0.045***			0.030***
			(0.001)			(0.001)
3rd Friday	0.166***	0.129***	0.161***	0.183***	0.066***	0.174***
	(0.019)	(0.015)	(0.018)	(0.014)	(0.010)	(0.013)
3rd Friday*EoQ	0.252***	0.173***	0.238***	0.286***	0.128***	0.271***
	(0.041)	(0.031)	(0.039)	(0.030)	(0.022)	(0.029)
BoM	0.006	0.015***	0.007	0.012**	0.015***	0.012**
	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)
EoM	0.005	0.010*	0.004	-0.012**	-0.006	-0.013**
	(0.006)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)
Russell	0.012	0.030	0.004	-0.055	-0.039**	-0.053
	(0.035)	(0.030)	(0.033)	(0.034)	(0.016)	(0.033)
EA-1	0.038***	0.036***	0.039***	-0.004	0.003	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
EA	0.149***	0.136***	0.150***	0.115***	0.107***	0.115***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
EA+1	0.035***	0.037***	0.037***	-0.004	0.006*	-0.004
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Calendar month/day FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	40.42%	44.55%	41.19%	52.71%	58.26%	53.11%
Num. obs.		2,048,959			2,148,513	

Table 7. Effective spread regressed on common and idiosyncratic components of volume, volatility, and order imbalance volatility.

This table reports median estimate, median t-statistic (in parentheses), and median adjusted R-squared across years. The following panel regression with stock fixed effects is estimated each year for small and large stocks: $\log s_{i,t} = \alpha_i + \beta_{\tau,C}\tau^C_{i,t} + \beta_{\tau,I}\tau^I_{i,t} + \beta_{\sigma,C}\sigma^C_{i,t} + \beta_{\sigma,I}\sigma^I_{i,t} + \beta_{\text{HFOIV},C}\text{HFOIV}^C_{i,t} + \beta_{\text{HFOIV},I}\text{HFOIV}^I_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is daily intraday turnover, $\sigma_{i,t}$ is realized volatility estimated using five-minute returns over the current day, and $\text{HFOIV}_{i,t}$ is high-frequency order imbalance volatility computed as the standard deviation of five-minute share imbalance scaled by total shares outstanding These variables are decomposed into common (C) and idiosyncratic (I) components as described in the text. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. Spreads are winsorized at 0.05% and 99.95% each year. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Standard errors are double-clustered by date and stock.

	Median value across years			
	Small stocks		Large stocks	
$\beta_{\tau,C}$	-0.08	-0.20	0.12	-0.15
	(-2.29)	(-6.32)	(2.67)	(-4.84)
$\beta_{\tau,I}$	-0.16	-0.30	0.01	-0.27
	(-30.85)	(-33.64)	(0.98)	(-21.48)
$\beta_{\sigma,C}$	0.00	0.00	0.01	0.03
	(0.63)	(1.07)	(3.52)	(6.78)
$\beta_{\sigma,I}$	0.42	0.45	0.33	0.40
	(34.84)	(39.50)	(21.74)	(27.50)
$\beta_{\text{HFOIV},C}$		0.04		0.03
		(8.39)		(5.20)
$\beta_{\text{HFOIV},I}$		0.12		0.23
		(12.98)		(22.23)
$R^2(\%)$	28.06	31.39	16.62	24.39

Table 8. High-frequency order imbalance volatility, turnover, and stock returns.
Every week, portfolios are formed by sequentially sorting stocks using NYSE breakpoints. The table reports portfolios' four-factor value-weighted alpha (in percent). Panel (a): sort on turnover then on HFOIV. Panel (b): sort on HFOIV then on turnover. Turnover is the average daily turnover over the previous month. HFOIV is an exponentially-weighted moving average of prior high-frequency order imbalance volatility with a half-life of one day. To be included in a portfolio, a stock must have a price greater than \$5 on the formation date. The sample consists of NYSE, Amex, and NASDAQ common stocks over 2002-2017 (797 weekly observations). t-statistics are reported in parentheses and computed using Newey-West standard errors with one lag. *, **, and *** denote significance at the 10%, 5%, and 1% level.

(a) Turnover then HFOIV: $\alpha_{FF4}^{\mathrm{VW}}$ (%)

	Low HFOIV	2	3	4	High HFOIV	H-L
Low turn.	0.00	0.01	0.03	0.02	0.12***	0.12***
	(0.15)	(0.21)	(0.86)	(0.76)	(3.93)	(2.85)
2	-0.00	0.06*	0.04	0.06*	0.10***	0.10***
	(-0.09)	(1.84)	(1.26)	(1.68)	(3.24)	(2.64)
3	-0.03	0.02	0.07**	0.04	0.09***	0.13***
	(-1.20)	(0.73)	(2.09)	(1.47)	(3.10)	(3.06)
4	-0.10***	-0.01	0.03	0.01	0.13***	0.23***
	(-3.03)	(-0.29)	(0.77)	(0.27)	(3.74)	(4.79)
High turn.	-0.02	-0.10*	0.04	-0.08	0.04	0.05
	(-0.34)	(-1.85)	(0.71)	(-1.58)	(0.72)	(0.86)

(b) HFOIV then turnover: $\alpha_{FF4}^{\mathrm{VW}}$ (%)

	Low turn.	2	3	4	High turn.	H-L
Low HFOIV	0.03	-0.00	-0.01	0.04	-0.08***	-0.11**
	(0.82)	(-0.01)	(-0.41)	(1.30)	(-2.64)	(-2.15)
2	0.02	0.03	0.04	-0.03	-0.07*	-0.09
	(0.54)	(0.77)	(1.29)	(-0.95)	(-1.95)	(-1.58)
3	0.07**	0.07**	0.06*	0.04	-0.03	-0.10
	(1.98)	(2.04)	(1.96)	(1.07)	(-0.72)	(-1.63)
4	0.09***	0.06*	0.03	-0.04	-0.07	-0.16**
	(2.76)	(1.93)	(0.99)	(-0.96)	(-1.27)	(-2.44)
High HFOIV	0.13***	0.15***	0.06*	0.03	-0.04	-0.16**
	(4.17)	(4.26)	(1.72)	(0.61)	(-0.56)	(-2.16)

Table 9. Value-weighted Fama-MacBeth regressions of weekly returns (in percent) on liquidity characteristics.

Order imbalance volatility (HFOIV_{t-1}) is an exponentially-weighted moving average (ewma) of prior high-frequency order imbalance volatility with a half-life of one day. Turnover is the average daily turnover over the previous month. ME_{t-1} is the market capitalization at the end of the previous week. ILLIQ_{t-1} is the illiquidity coefficient at the end of the previous week computed using the past 250 trading days with a minimum of 100 observations. Realized volatility (RVol_{t-1}) is an ewma of prior daily realized volatilities with a half-life of one day. Effective spread (ES_{t-1}) is an ewma of prior daily effective spreads with a half-life of one day. Price impact (lambda_{t-1}) is an ewma of prior Kyle's lambda with a half-life of one day. Absolute order imbalance ($|\text{OI}|_{t-1}$) is an ewma of prior daily absolute shares order imbalances (as a fraction of shares outstanding) with a half-life of one day. Depth (Depth/VOL_{t-1}) is an ewma of prior daily share depth over daily share volume with a half-life of one day. PIN is the absolute daily trade imbalance over the total number of trades (Aktas et al. (2007)). The standard deviation of share order imbalance divided by share volume ($\sigma(\text{OI/VOL})_{t-1}$) is computed at the end of the previous week using the past 22 trading days with a minimum of 11 observations. All explanatory variables (except the lagged return) are in logs. All explanatory variables are winsorized at 0.5% and 99.5%. The sample consists of NYSE, Amex, and NASDAQ common stocks 2002-2017 (797 weeks) with a price greater than $5 at the end of the previous week. $\bar{N}$ is the average number of stocks at each date. t-statistics are shown in parentheses and based on Newey-West standard errors. *, **, and *** denote significance at the 10%, 5%, and 1% level.

	(1)	(2)	(3)	(4)
HFOIV_{t-1}	0.064** (2.30)	0.079*** (3.42)	0.077*** (3.17)	0.078*** (3.15)
turn_{t-1}		-0.033 (-0.79)	-0.037 (-0.75)	-0.007 (-0.15)
ME_{t-1}			-0.005 (-0.424)	-0.010 (-0.26)
r_{t-1}			-1.691*** (-4.21)	-1.622*** (-4.06)
ILLIQ_{t-1}			0.008 (0.23)	0.005 (0.15)
RVol_{t-1}			-0.051 (-0.63)	-0.006 (-0.07)
ES_{t-1}			0.007 (0.23)	-0.018 (-0.56)
Depth/VOL_{t-1}			-0.032 (-1.43)	-0.035 (-1.58)
lambda_{t-1}			-0.106 (-0.79)	-0.185 (-1.43)
$\|\text{OI}\|_{t-1}$				-0.030* (-1.68)
PIN_{t-1}				0.03* (1.81)
$\sigma(\text{OI/VOL})_{t-1}$				0.091** (2.21)
$\bar{N}$	2,628	2,628	2,424	2,395
$\bar{R}^2$	0.020	0.038	0.110	0.119

Internet Appendix to "Liquidity, Volume, and Order Imbalance Volatility"

This appendix provides additional results to supplement the main text.

Appendix IA.A. Vector autoregressions

As a robustness check of the results in Section 3.3, we estimate a reduced-form vector autoregression (VAR) of spread, volume, and volatility using ordinary least squares, where the number of lags is chosen based on the Akaike information criterion. We then perform a Cholesky decomposition to orthogonalize the error terms and obtain a recursive VAR. The Cholesky decomposition is sensitive to the ordering of the variables. We report results with the following ordering: volume, volatility, and spread. The results are not substantially affected if we switch volume and volatility in the ordering.

For simplicity, we focus on large stocks in the last year of the sample (2017) and require stocks to be traded over the whole year. The results are consistent for other years. The VAR is estimated separately for each stock. Since we are interested in comparing the results across stocks, all the variables are normalized. First, we perform Granger causality tests. Both volatility and volume tend to Granger-cause spreads for the median stock. Spreads tend not to Granger-cause volatility and volume: for volume (volatility), we cannot reject the null of no Granger-causality for more than 76% (80%) of the stocks at a 10% level of statistical significance. Volume Granger-causes volatility for around 73% of the stocks, but volatility Granger-causes volume for only around 21% of the stocks (at a 10% level of statistical significance).

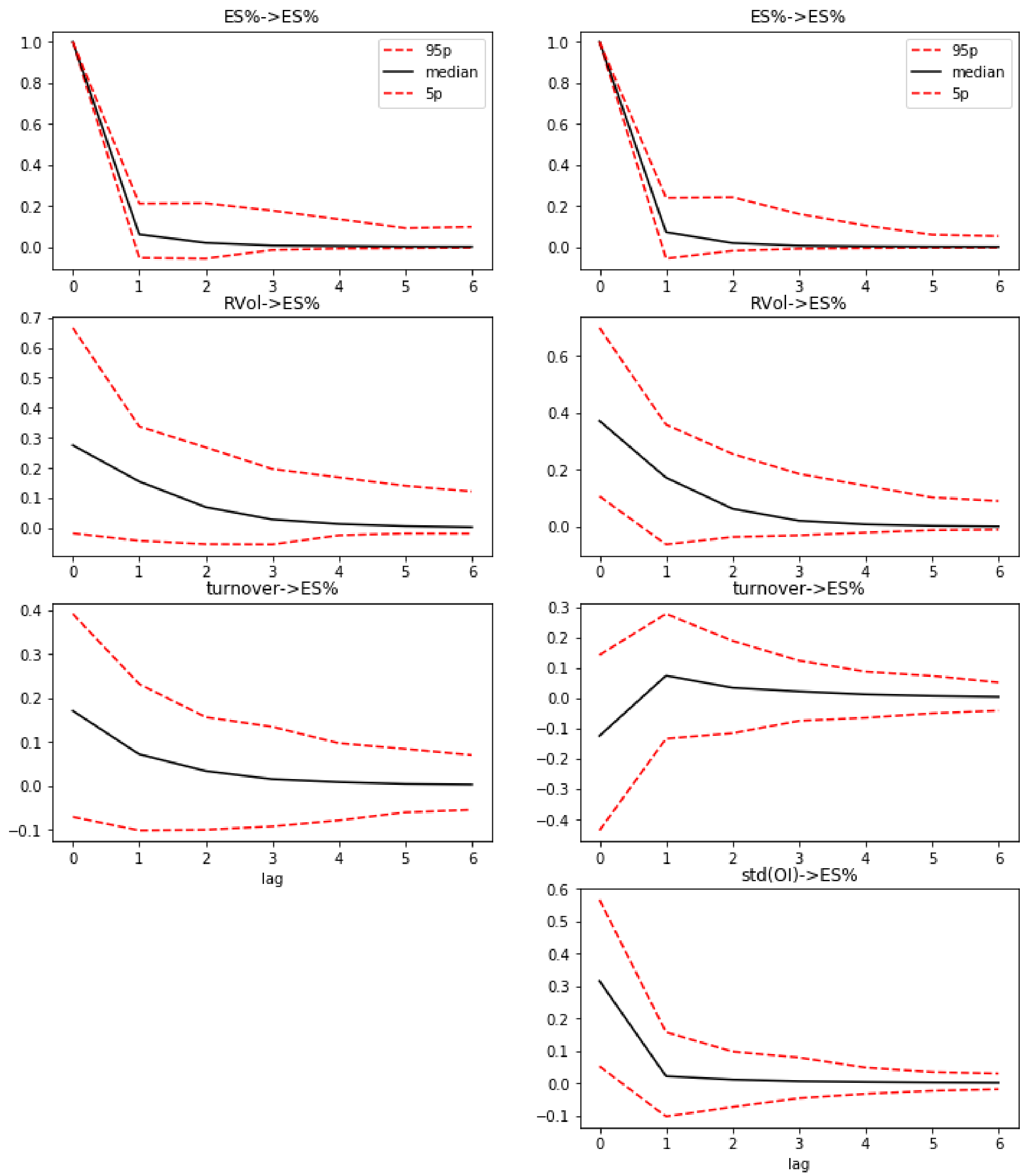
Next, we compute impulse responses to a one standard-deviation shock for each variable. Figure IA.1 reports the cross-sectional median and 5$^{\text{th}}$ and 95$^{\text{th}}$ percentiles impulse responses. The plots in the left column report the results with the baseline specification (the plots in the right columns are discussed later). The results confirm the evidence from the panel regressions. The contemporaneous response of spreads to a volume shock is mostly positive across stocks. Spreads remain higher after one day for the majority of stocks. As expected, a volatility shock causes a large contemporaneous increase in spreads.

Next, we include order imbalance volatility in the VAR. The right plots in Figure IA.1 show the impulse responses. When order imbalance volatility is included in the VAR, a volume shock lowers spreads for most stocks. This contrasts with the impulse response of a volume shock in the baseline model (left plots). An order imbalance volatility shock increases spreads consistently across stocks.

Appendix IA.B. Tick size

In this appendix, we discuss the role of the tick size for our results. The binding tick size can matter among large stocks in our sample period. Table IA.1 reports the distribution of average quoted

Figure IA.1. Vector autoregressions of spread, volume, and volatility.
For each stock, a VAR is estimated of (log) effective spread (ES%), (log) turnover, and (log) realized volatility (RVol), where the number of lags is chosen based on the Akaike information criterion and all the variables are normalized. The reduced-form VAR is estimated using ordinary least squares and then a Cholesky decomposition is performed to orthogonalize the error terms with the following ordering: volume, volatility, and spread. The figure reports the cross-sectional median and 5$^{\text{th}}$ and 95$^{\text{th}}$ percentiles impulse response to a one standard-deviation shock for each variable. The sample consists of stocks in the top size quintile among NYSE, Amex, and NASDAQ common stocks in 2017 that are traded over the whole year. The left column plots report the baseline specification. The right column plots report results with order imbalance volatility added to the VAR (ordered first).



spread across large stocks for each year in our sample. For example, in 2010 the 25$^{\text{th}}$ percentile large stock average quoted spread is 1.04 cent. Assuming that the quoted spread is either 1 or 2 cents would imply that the tick size binds 96% of the time for this stock. However, for the median large-stock in our sample, the average quoted spread is close to two ticks and therefore the binding tick size seems much less constraining.

Table IA.1. Quoted spread for large stocks. This table reports the distribution of average daily dollar quoted spread across large stocks for each year. Large stocks are stocks in the top quintile of market capitalization at the beginning of each month. Daily quoted spread is reported in dollar and computed by taking the time-weighted average of intraday quoted spread.

	Quoted spread ($)			
Year	25p	Median	75p	Mean
2002	0.0288	0.0370	0.0467	0.0413
2003	0.0191	0.0239	0.0299	0.0302
2004	0.0172	0.0227	0.0303	0.0335
2005	0.0160	0.0212	0.0292	0.0345
2006	0.0154	0.0217	0.0296	0.0348
2007	0.0138	0.0201	0.0288	0.0335
2008	0.0134	0.0206	0.0330	0.0358
2009	0.0108	0.0142	0.0202	0.0239
2010	0.0104	0.0128	0.0200	0.0572
2011	0.0105	0.0140	0.0259	0.0291
2012	0.0106	0.0147	0.0251	0.0294
2013	0.0112	0.0182	0.0354	0.0401
2014	0.0120	0.0204	0.0414	0.0447
2015	0.0129	0.0229	0.0479	0.0616
2016	0.0123	0.0217	0.0496	0.0631
2017	0.0128	0.0268	0.0569	0.0670

To evaluate the effect of the tick size on the volume-spread elasticity, we sort each month large stocks into quintiles based on their average quoted spread in the previous month. We then regress each month spread on volume within each quoted spread quintile (with and without controlling for volatility). Descriptive statistics of the monthly regression results are reported in Table IA.2. The positive volume-spread relation is stronger among large stocks for which the tick size binds more often (low quoted spread stocks). In univariate regressions, the median coefficient (t-statistic) is 0.22 (5.14) for these stocks versus a median coefficient (t-statistic) of 0.13 (3.20) for high quoted spread stocks. Even when controlling for volatility (Panel (b)), the spread-volume elasticity is in general positive for large stocks with high quoted spread (i.e., a positive median monthly elasticity), and positive and statistically significant for about 20% of months in this restricted sample.

It is important to point out that, in a world with no additional risk factor (i.e., only "good" volume), we would not expect to observe a positive volume-spread relation for some stocks (after controlling for volatility). The tick size only makes the "bad" volume more apparent by imposing a lower bound on the spread.

Finally, the relation between order imbalance volatility and spread is robust to focusing on large stocks with a price greater than \$80, \$100, or \$120 (Table IA.7), for which the tick size is unlikely to bind.

Table IA.2. Volume elasticity of spread among large stocks with low and high quoted spreads. This table reports descriptive statistics for the monthly volume coefficients (β_τ) and their associated t-statistics over the sample period. Large stocks are sorted each month into quintiles based on their average quoted spread in the previous month. For large stocks in the bottom and top quoted spread quintiles, the following panel regression is estimated each month: $\Delta s_{i,t} = \alpha_i + \beta_\tau \Delta\tau_{i,t} + \beta_\sigma \Delta\sigma_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\Delta x_t \equiv \log(\frac{x_t}{x_{t-1}})$, and $\tau_{i,t}$ is the daily intraday turnover and $\sigma_{i,t}$ is the realized volatility estimated using five-minute returns over the current day. In Panel (a), $\Delta\sigma_{i,t}$ is not included in the regression. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. Large stocks are the top quintile of stocks at the beginning of each month based on average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock.

(a) Univariate (without volatility)

	Low quoted spread		High quoted spread	
	Turnover	t-statistic	Turnover	t-statistic
min	0.0680	2.4508	-0.0482	-1.4862
25%	0.1617	4.0008	0.0819	1.8456
50%	0.2155	5.1416	0.1263	3.1991
75%	0.2702	6.8110	0.1763	4.8460
max	0.5496	12.1386	0.3660	9.5966
Obs.	191	191	191	191

(b) Multivariate (with volatility)

	Low quoted spread		High quoted spread	
	Turnover	t-statistic	Turnover	t-statistic
min	-0.0482	-0.7369	-0.2527	-10.2102
25%	0.1220	2.7753	-0.0610	-1.4674
50%	0.1596	3.6422	0.0133	0.3185
75%	0.2208	4.7263	0.05887	1.6397
max	0.5162	9.5564	0.2593	4.5323
Obs.	191	191	191	191

Appendix IA.C. A simple reduced-form model

We consider a simple reduced-form model to gain intuition. Split the day into $k = 1, \ldots, K$ periods. We assume that spread is directly related to absolute order imbalance (OI) in each period:

$$s_{t,k} = a_k + b_k|\mathrm{OI}_{t,k}| + \epsilon_{k,t}, \tag{IA.C1}$$

for intraday period k in day t. (IA.C1) with $b > 0$ is motivated by theory as absolute order imbalance proxies for shocks to the inventory of liquidity providers. It can also be shown in the data by regressing intraday spread on intraday order imbalance. For simplicity, we assume that $b_k = b \quad \forall k$, and we ignore additional explanatory variables as they would not change the main intuition.

When we consider variables at the daily level, we have

$$s_t = \frac{1}{K}\sum_k s_{t,k}, \tag{IA.C2}$$

$$\mathrm{OI}_t = \sum_k \mathrm{OI}_{t,k}. \tag{IA.C3}$$

In the empirical implementation, the daily effective spread is a weighted average of intraday spreads, where the weights are proportional to volume. However, this would only complicate the intuition here. At the daily level, we consider two regressions:

$$s_t = a + b|\mathrm{OI}_t| + \epsilon, \tag{IA.C4}$$

$$s_t = a' + b'\sum_k |\mathrm{OI}_{k,t}| + \epsilon'. \tag{IA.C5}$$

Note that $\sum_k |\mathrm{OI}_{k,t}|$ in (IA.C5) is closely related to HFOIV. When are (IA.C4) and (IA.C5) equivalent?

1. $K = 1$ (i.e., there is only one trading period).

2. $\mathrm{OI}_{k,t}$ all have the same sign (i.e., order imbalances are highly persistent).

In other cases, the explanatory power of (IA.C5) will be greater than that of (IA.C4). In (IA.C4), the explanatory variable is a noisy proxy of the explanatory variable in (IA.C5). (The explanatory power of (IA.C5) is the same as that of (IA.C1) when $|\mathrm{OI}_{t,k}|$ is not autocorrelated.)

Even if order imbalances are positively autocorrelated over the day, order imbalance volatility should still outperform absolute order imbalance as long as the correlation is not perfect. In the data, intraday order imbalances are positively autocorrelated, but there is substantial noise. In the cross-section, the simple model laid out above can shed light on why order imbalance volatility outperforms absolute order imbalance more strongly for some stocks than for others. Thinly-traded stocks and stocks with highly persistent order imbalances should experience less of an improvement than other stocks.

Appendix IA.D. Balanced volume and absolute order imbalance

Volume can be decomposed into balanced volume (BV) and absolute order imbalance (OI):[1]

$$V = \mathrm{BV} + |\mathrm{OI}|. \tag{IA.D1}$$

A key point of our paper is that one million shares bought in the morning and sold in the afternoon is not the same for the inventory risk of liquidity providers as one share bought and one sold every second throughout the day. In both cases, (IA.D1) yields similar results. However, our measure of high-frequency order imbalance volatility (HFOIV) differs and is higher in the first case. Furthermore, this decomposition is only valid for |OI| but our main variable of interest is HFOIV.

The above caveats in mind, we examine whether this decomposition can explain our results. We compute balanced volume from (IA.D1), and then estimate the following regression:

$$\log s_{i,t} = \alpha_i + \beta_{\mathrm{BV}} \log \mathrm{BV}_{i,t} + \beta_\sigma \log \sigma_{i,t} + \beta_{|\mathrm{OI}|} \log |\mathrm{OI}|_{i,t} + \text{controls} + \epsilon_{i,t}, \tag{IA.D2}$$

where BV is normalized by shares outstanding (i.e., balanced turnover) since we are using turnover as a measure of volume. As for all of our specifications in the paper, we also consider a specification with changes in the variables. The results for each year are reported in Table IA.13 (levels) and Table IA.14 (changes) below.

The specification in levels shows that |OI| has an inconsistent and weak effect for small stocks. The coefficient tends to be negative. For large stocks, balanced volume has an inconsistent behavior and switches to the expected negative sign in 2008. The specification in changes gives similar results. Here, balanced volume has a positive relation with spread for large stocks for the majority of the sample. In sum, the decomposition produces inconsistent results.

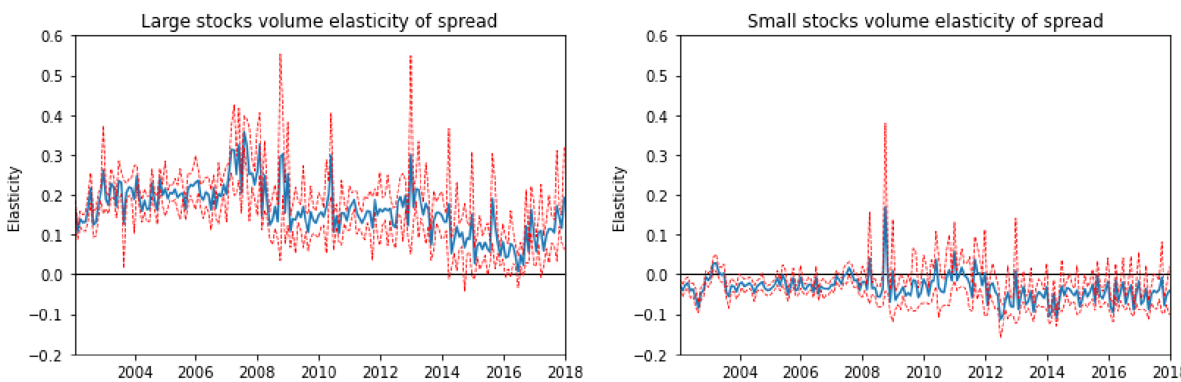
In Table IA.15 (levels) and Table IA.16 (changes), we replace |OI| with HFOIV. As can be seen, our measure produces consistent relations over the sample period. First, HFOIV is always positively associated with spread for small stocks. Second, balanced volume is alway negatively associated with spread for large stocks. If we regress spread only on volatility and balanced volume, we run again into a puzzling positive relation between spread and volume for large stocks. Table IA.13 and Table IA.14 show that absolute order imbalance does not resolve the puzzle, in contrast to HFOIV.

Appendix IA.E. Additional figures

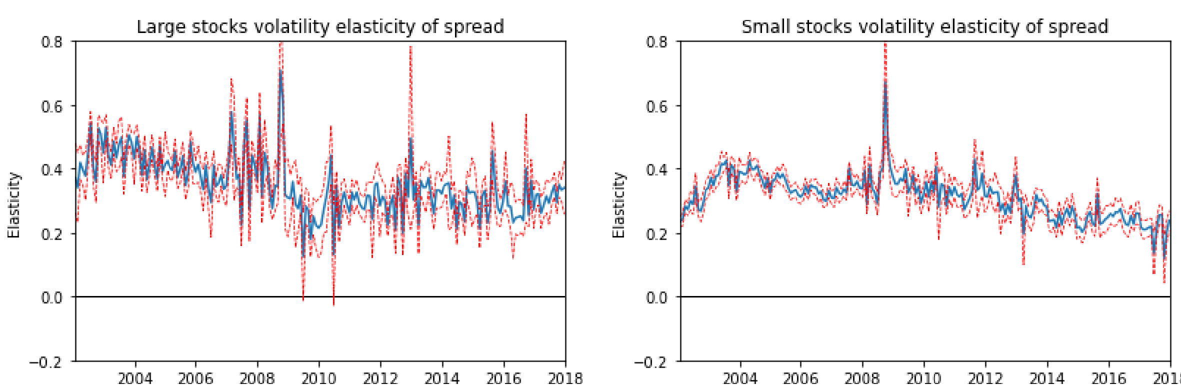
[1] We thank an anonymous referee for suggesting this decomposition.

Figure IA.2. Effective spread regressed on volume and volatility across size quintiles (univariate regression).
We estimate each month for stocks in a given size quintile panel regressions of spread on volume (Panel (a)) and spread on volatility (Panel (b)). Spread is the daily effective spread, volume is the daily intraday turnover, and volatility is the realized volatility computed using five-minute intraday midquote returns. The regressions include stock fixed effects and control for (log) market capitalization, (log) price, and day-of-the-week indicators. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each month. Standard errors are double-clustered by date and stock.

(a) Volume elasticity of spread



(b) Volatility elasticity of spread



Appendix IA.F. Additional tables

Table IA.3. Effective spread regressed on turnover and realized volatility.
The table reports estimates from the following panel regression: $\log s_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover and $\sigma_{i,t}$ is the realized volatility estimated using five-minute returns over the current day. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in a given size quintile. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading.. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

Year	Small stocks		Large stocks	
	β_τ	β_σ	β_τ	β_σ
2002	-0.15*** (-38.88)	0.39*** (52.50)	0.03*** (3.15)	0.43*** (14.67)
2003	-0.14*** (-32.27)	0.48*** (65.57)	0.07*** (5.78)	0.43*** (41.55)
2004	-0.15*** (-41.65)	0.48*** (70.99)	0.07*** (10.01)	0.37*** (34.65)
2005	-0.15*** (-37.92)	0.44*** (73.51)	0.08*** (10.37)	0.35*** (30.41)
2006	-0.15*** (-41.01)	0.42*** (70.33)	0.08*** (10.22)	0.30*** (28.99)
2007	-0.13*** (-35.03)	0.42*** (62.11)	0.11*** (8.65)	0.35*** (18.19)
2008	-0.15*** (-12.65)	0.49*** (35.45)	0.02* (1.95)	0.43*** (16.11)
2009	-0.16*** (-17.33)	0.43*** (36.05)	0.04*** (3.49)	0.23*** (13.08)
2010	-0.14*** (-14.72)	0.44*** (37.95)	0.04*** (3.83)	0.27*** (11.85)
2011	-0.15*** (-17.00)	0.42*** (38.33)	0.03*** (3.42)	0.29*** (22.58)
2012	-0.20*** (-20.74)	0.44*** (35.93)	0.05*** (3.40)	0.28*** (16.74)
2013	-0.19*** (-25.37)	0.40*** (25.46)	0.03** (2.19)	0.31*** (17.35)
2014	-0.18*** (-27.21)	0.35*** (33.05)	-0.05*** (-2.70)	0.36*** (22.36)
2015	-0.17*** (-27.70)	0.35*** (33.22)	-0.09*** (-8.18)	0.40*** (17.44)
2016	-0.18*** (-23.20)	0.35*** (34.33)	-0.10*** (-7.62)	0.38*** (19.84)
2017	-0.16*** (-20.70)	0.30*** (16.03)	-0.08*** (-5.09)	0.41*** (23.97)
$\bar{R}^2$(%)	26.67		19.94	

Table IA.4. Effective spread regressed on turnover and realized volatility (changes).
The table reports estimates from the following panel regression: $\Delta s_{i,t} = \alpha_i + \beta_\tau \Delta\tau_{i,t} + \beta_\sigma \Delta\sigma_{i,t} +$ controls$+\epsilon_{i,t}$ for stock i on day t, where $\Delta x_t \equiv \log(\frac{x_t}{x_{t-1}})$, and $\tau_{i,t}$ is the daily intraday turnover and $\sigma_{i,t}$ is the realized volatility estimated using five-minute returns over the current day. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in a given size quintile. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R^2}$ denotes the average adjusted R-squared across years.

	Small stocks		Large stocks	
Year	β_τ	β_σ	β_τ	β_σ
2002	-0.09***(-26.11)	0.28***(37.52)	0.08***(6.37)	0.36***(10.24)
2003	-0.09***(-20.44)	0.37***(47.84)	0.12***(7.76)	0.40***(35.28)
2004	-0.09***(-24.12)	0.37***(47.61)	0.13***(16.60)	0.35***(35.38)
2005	-0.09***(-21.60)	0.35***(52.67)	0.17***(15.78)	0.31***(28.19)
2006	-0.08***(-21.72)	0.33***(48.44)	0.17***(16.74)	0.26***(26.89)
2007	-0.08***(-18.92)	0.34***(48.13)	0.25***(11.70)	0.27***(15.44)
2008	-0.07***(-5.97)	0.36***(25.58)	0.13***(8.27)	0.35***(13.76)
2009	-0.08***(-7.39)	0.35***(31.03)	0.14***(7.56)	0.19***(9.47)
2010	-0.05***(-4.37)	0.32***(29.84)	0.15***(9.25)	0.22***(9.52)
2011	-0.06***(-5.66)	0.32***(28.21)	0.12***(8.89)	0.24***(18.40)
2012	-0.10***(-8.95)	0.30***(30.20)	0.16***(5.98)	0.20***(11.31)
2013	-0.10***(-14.64)	0.28***(25.83)	0.12***(7.33)	0.25***(16.80)
2014	-0.09***(-13.93)	0.24***(23.82)	0.06**(2.47)	0.29***(17.48)
2015	-0.09***(-13.30)	0.23***(22.68)	0.00(0.08)	0.33***(12.63)
2016	-0.08***(-10.07)	0.24***(23.15)	-0.01(-0.88)	0.35***(21.25)
2017	-0.08***(-8.47)	0.19***(9.56)	0.02(0.92)	0.33***(19.25)
$\bar{R^2}$(%)	9.21		8.43	

Table IA.5. Effective spread regressed on turnover, realized volatility, and order imbalance volatility.
The table reports estimates from the following panel regression: $\log s_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \beta_{\text{HFOIV}} \log \text{HFOIV}_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover, $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns, and $\text{HFOIV}_{i,t}$ is the volatility of order imbalance computed using five-minute order imbalances over the trading day. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in the bottom and top size quintiles. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, , and at least 100 days of prior trading.. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

	Small stocks			Large stocks		
Year	β_τ	β_σ	β_{HFOIV}	β_τ	β_σ	β_{HFOIV}
2002	-0.24*** (-42.17)	0.42*** (52.83)	0.13*** (20.67)	-0.26*** (-12.65)	0.51*** (14.46)	0.30*** (14.81)
2003	-0.25*** (-32.97)	0.51*** (64.46)	0.14*** (22.50)	-0.25*** (-13.62)	0.51*** (53.51)	0.29*** (18.63)
2004	-0.29*** (-51.03)	0.51*** (70.86)	0.18*** (31.31)	-0.24*** (-15.42)	0.46*** (38.85)	0.28*** (18.77)
2005	-0.29*** (-51.30)	0.48*** (74.19)	0.18*** (22.64)	-0.26*** (-17.81)	0.45*** (40.45)	0.30*** (20.07)
2006	-0.30*** (-55.55)	0.46*** (73.23)	0.19*** (29.30)	-0.26*** (-23.35)	0.41*** (49.95)	0.29*** (27.18)
2007	-0.29*** (-46.81)	0.46*** (64.36)	0.20*** (26.26)	-0.28*** (-16.40)	0.48*** (30.44)	0.33*** (19.37)
2008	-0.39*** (-26.97)	0.54*** (37.93)	0.29*** (11.19)	-0.39*** (-18.21)	0.53*** (24.41)	0.37*** (18.14)
2009	-0.37*** (-34.23)	0.48*** (43.30)	0.25*** (15.78)	-0.32*** (-19.09)	0.36*** (31.45)	0.33*** (19.57)
2010	-0.34*** (-31.63)	0.48*** (47.98)	0.25*** (14.35)	-0.29*** (-21.91)	0.37*** (21.75)	0.30*** (21.49)
2011	-0.34*** (-36.30)	0.46*** (46.81)	0.24*** (16.91)	-0.26*** (-26.05)	0.38*** (34.49)	0.26*** (26.79)
2012	-0.35*** (-26.52)	0.47*** (39.93)	0.20*** (11.41)	-0.28*** (-15.09)	0.38*** (25.24)	0.27*** (13.57)
2013	-0.33*** (-35.05)	0.43*** (25.85)	0.17*** (20.06)	-0.30*** (-28.11)	0.40*** (26.17)	0.28*** (26.55)
2014	-0.34*** (-31.57)	0.38*** (34.97)	0.20*** (13.49)	-0.42*** (-20.32)	0.47*** (36.32)	0.32*** (13.83)
2015	-0.33*** (-29.14)	0.37*** (34.20)	0.20*** (13.25)	-0.42*** (-31.97)	0.49*** (26.17)	0.29*** (23.16)
2016	-0.36*** (-28.08)	0.38*** (38.43)	0.23*** (12.54)	-0.43*** (-27.76)	0.47*** (26.82)	0.30*** (20.90)
2017	-0.33*** (-23.34)	0.31*** (16.02)	0.20*** (10.88)	-0.41*** (-24.92)	0.50*** (45.84)	0.29*** (14.36)
$\bar{R}^2$(%)		30.72			30.53	

Table IA.6. Effective spread regressed on turnover, realized volatility, and order imbalance volatility (changes).
The table reports estimates from the following regression: $\Delta s_{i,t} = \alpha_i + \beta_\tau \Delta\tau_{i,t} + \beta_\sigma \Delta\sigma_{i,t} + \beta_{\text{HFOIV}} \Delta\text{HFOIV}_{i,t} + \text{controls} + \epsilon_{i,t}$, for stock i on day t where $\Delta x_t \equiv \log(\frac{x_t}{x_{t-1}})$, $\tau_{i,t}$ is the daily intraday turnover, $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns, and $\text{HFOIV}_{i,t}$ is the volatility of order imbalance computed using five-minute order imbalances over the trading day. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in the bottom and top size quintiles. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

	Small stocks			Large stocks		
Year	β_τ	β_σ	β_{HFOIV}	β_τ	β_σ	β_{HFOIV}
2002	-0.17***(-32.16)	0.30***(37.64)	0.09***(16.38)	-0.21***(-10.75)	0.42***(10.10)	0.29***(16.18)
2003	-0.17***(-24.00)	0.39***(46.61)	0.10***(17.21)	-0.21***(-11.33)	0.46***(42.69)	0.28***(19.72)
2004	-0.21***(-37.12)	0.40***(47.94)	0.14***(25.52)	-0.19***(-13.14)	0.43***(40.40)	0.27***(20.37)
2005	-0.21***(-37.74)	0.38***(53.61)	0.14***(19.06)	-0.21***(-17.16)	0.40***(42.78)	0.29***(23.69)
2006	-0.21***(-38.28)	0.36***(50.87)	0.14***(21.82)	-0.20***(-19.72)	0.36***(46.67)	0.28***(29.89)
2007	-0.23***(-35.41)	0.38***(50.38)	0.17***(22.06)	-0.19***(-10.47)	0.38***(27.74)	0.32***(22.37)
2008	-0.30***(-20.70)	0.40***(28.04)	0.25***(11.42)	-0.33***(-18.41)	0.45***(21.03)	0.35***(22.00)
2009	-0.28***(-29.86)	0.39***(37.17)	0.22***(16.31)	-0.29***(-19.70)	0.30***(23.54)	0.33***(24.95)
2010	-0.27***(-26.01)	0.36***(39.28)	0.23***(16.13)	-0.25***(-22.40)	0.32***(18.36)	0.30***(26.91)
2011	-0.26***(-28.50)	0.36***(36.01)	0.22***(17.28)	-0.21***(-22.83)	0.32***(32.36)	0.24***(29.42)
2012	-0.26***(-22.53)	0.32***(35.14)	0.17***(10.86)	-0.22***(-18.59)	0.30***(22.29)	0.28***(16.11)
2013	-0.23***(-26.90)	0.30***(26.19)	0.14***(18.83)	-0.25***(-26.83)	0.34***(29.91)	0.27***(30.59)
2014	-0.25***(-25.51)	0.26***(25.27)	0.17***(13.39)	-0.38***(-25.78)	0.38***(36.34)	0.33***(17.99)
2015	-0.25***(-24.05)	0.25***(24.07)	0.18***(13.40)	-0.38***(-31.46)	0.42***(17.77)	0.29***(26.27)
2016	-0.26***(-22.52)	0.25***(25.86)	0.19***(11.53)	-0.39***(-28.40)	0.42***(32.79)	0.30***(22.76)
2017	-0.24***(-19.56)	0.20***(9.57)	0.17***(11.45)	-0.36***(-29.14)	0.42***(35.30)	0.29***(17.93)
$\bar{R}^2$(%)		12.86			19.99	

Table IA.7. Effective spread regressed on turnover, realized volatility, and order imbalance volatility for large stocks with price filter. Levels: $\log s_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \beta_{\text{HFOIV}} \log \text{HFOIV}_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover and $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns and $\text{HFOIV}_{i,t}$ is the volatility of order imbalance computed using five-minute order imbalances over the trading day. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in the top size quintile. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$X and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

	Price filter					
	$p \geq \$80$		$p \geq \$100$		$p \geq \$120$	
Year	β_τ	β_{HFOIV}	β_τ	β_{HFOIV}	β_τ	β_{HFOIV}
2002	-0.27*** (-5.20)	0.26*** (6.80)	-0.28*** (-3.28)	0.23*** (4.28)	-0.41*** (-7.69)	0.29*** (16.25)
2003	-0.30*** (-5.19)	0.33*** (9.26)	-0.27*** (-2.63)	0.32*** (4.07)	-0.44*** (-7.21)	0.35** (2.44)
2004	-0.36*** (-11.50)	0.32*** (16.98)	-0.38*** (-11.84)	0.32*** (9.41)	-0.39*** (-10.38)	0.24*** (6.40)
2005	-0.33*** (-12.17)	0.30*** (16.22)	-0.39*** (-12.63)	0.30*** (9.71)	-0.42*** (-23.95)	0.26*** (5.62)
2006	-0.32*** (-16.20)	0.29*** (19.26)	-0.35*** (-12.27)	0.27*** (11.09)	-0.35*** (-11.38)	0.25*** (6.42)
2007	-0.31*** (-14.11)	0.32*** (15.96)	-0.32*** (-13.43)	0.31*** (15.28)	-0.36*** (-13.89)	0.31*** (12.49)
2008	-0.43*** (-12.99)	0.37*** (11.47)	-0.38*** (-10.83)	0.33*** (9.46)	-0.36*** (-8.52)	0.31*** (6.46)
2009	-0.38*** (-14.62)	0.30*** (16.24)	-0.38*** (-13.35)	0.30*** (15.61)	-0.37*** (-11.86)	0.30*** (13.94)
2010	-0.34*** (-17.03)	0.29*** (16.71)	-0.33*** (-12.52)	0.28*** (12.30)	-0.34*** (-9.89)	0.27*** (9.56)
2011	-0.35*** (-23.02)	0.26*** (16.03)	-0.33*** (-17.87)	0.23*** (11.56)	-0.30*** (-11.66)	0.22*** (9.23)
2012	-0.38*** (-15.93)	0.32*** (10.72)	-0.38*** (-14.65)	0.31*** (10.45)	-0.38*** (-10.27)	0.30*** (7.59)
2013	-0.38*** (-25.17)	0.29*** (18.01)	-0.40*** (-21.94)	0.29*** (14.19)	-0.38*** (-18.75)	0.26*** (13.19)
2014	-0.49*** (-20.62)	0.34*** (11.51)	-0.50*** (-18.75)	0.32*** (9.82)	-0.49*** (-18.09)	0.30*** (8.46)
2015	-0.47*** (-30.09)	0.28*** (16.02)	-0.46*** (-28.57)	0.25*** (13.69)	-0.46*** (-25.17)	0.25*** (12.35)
2016	-0.49*** (-28.83)	0.30*** (15.74)	-0.48*** (-28.16)	0.28*** (14.72)	-0.46*** (-24.33)	0.27*** (12.61)
2017	-0.45*** (-24.48)	0.29*** (11.62)	-0.45*** (-22.87)	0.28*** (11.07)	-0.45*** (-21.35)	0.28*** (9.93)
$\bar{R}^2$(%)	33.31		35.78		37.96	
$\bar{\text{obs}}$	22,149		12,859		8,358	

Table IA.8. Effective spread regressed on turnover, realized volatility, and absolute order imbalance.
The table reports estimates from the following panel regression: $\log s_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \beta_{|\mathrm{OI}|} \log |\mathrm{OI}|_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover, $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns, and $|\mathrm{OI}|_{i,t}$ is the absolute daily order imbalance. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in the bottom and top size quintiles. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

	Small stocks			Large stocks		
Year	β_τ	β_σ	$\beta_{\lvert\mathrm{OI}\rvert}$	β_τ	β_σ	$\beta_{\lvert\mathrm{OI}\rvert}$
2002	-0.18*** (-42.85)	0.40*** (53.32)	0.03*** (18.66)	0.03*** (2.73)	0.43*** (14.68)	0.01** (2.20)
2003	-0.16*** (-34.20)	0.48*** (67.25)	0.02*** (14.66)	0.06*** (5.08)	0.43*** (41.90)	0.01*** (2.90)
2004	-0.18*** (-43.22)	0.48*** (72.25)	0.03*** (19.82)	0.06*** (9.43)	0.37*** (34.62)	0.00*** (2.96)
2005	-0.18*** (-51.22)	0.45*** (75.31)	0.03*** (15.05)	0.06*** (9.27)	0.35*** (31.45)	0.02*** (6.06)
2006	-0.19*** (-51.20)	0.43*** (72.44)	0.03*** (20.07)	0.05*** (8.11)	0.30*** (30.96)	0.02*** (8.29)
2007	-0.17*** (-45.74)	0.43*** (64.16)	0.03*** (18.05)	0.07*** (7.69)	0.36*** (20.38)	0.04*** (6.89)
2008	-0.21*** (-33.71)	0.51*** (35.96)	0.05*** (7.03)	-0.02** (-2.41)	0.44*** (16.91)	0.04*** (7.01)
2009	-0.20*** (-28.93)	0.44*** (38.44)	0.04*** (8.81)	0.00 (0.35)	0.25*** (15.55)	0.03*** (6.36)
2010	-0.17*** (-24.94)	0.45*** (39.90)	0.03*** (8.29)	0.01 (0.72)	0.28*** (13.21)	0.03*** (7.93)
2011	-0.19*** (-30.12)	0.43*** (40.95)	0.04*** (9.37)	-0.00 (-0.35)	0.30*** (25.54)	0.03*** (9.30)
2012	-0.24*** (-34.08)	0.45*** (36.91)	0.04*** (8.73)	0.01 (0.74)	0.30*** (18.15)	0.03*** (6.73)
2013	-0.22*** (-33.43)	0.41*** (25.62)	0.03*** (12.23)	-0.02 (-1.52)	0.32*** (18.92)	0.04*** (11.86)
2014	-0.21*** (-39.09)	0.36*** (33.91)	0.03*** (10.50)	-0.09*** (-7.77)	0.37*** (24.99)	0.04*** (6.49)
2015	-0.21*** (-38.31)	0.35*** (33.87)	0.04*** (11.34)	-0.13*** (-14.33)	0.41*** (18.65)	0.03*** (9.32)
2016	-0.22*** (-37.50)	0.36*** (35.79)	0.04*** (9.76)	-0.14*** (-12.39)	0.39*** (20.89)	0.03*** (8.97)
2017	-0.20*** (-28.48)	0.30*** (16.05)	0.04*** (9.98)	-0.13*** (-10.90)	0.42*** (27.14)	0.04*** (7.24)
$\bar{R}^2$(%)		27.33			20.71	

Table IA.9. Effective spread regressed on turnover, realized volatility, and absolute order imbalance (changes).
The table reports estimates from the following regression: $\Delta s_{i,t} = \alpha_i + \beta_\tau \Delta\tau_{i,t} + \beta_\sigma \Delta\sigma_{i,t} + \beta_{|\mathrm{OI}|}\Delta|\mathrm{OI}_{i,t}| + \text{controls} + \epsilon_{i,t}$, for stock i on day t where $\Delta x_t \equiv \log(\frac{x_t}{x_{t-1}})$, $\tau_{i,t}$ is the daily intraday turnover, $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns, and $|\mathrm{OI}_{i,t}|$ is the absolute daily order imbalance. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in the bottom and top size quintiles. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

	Small stocks			Large stocks		
Year	β_τ	β_σ	$\beta_{\lvert\mathrm{OI}\rvert}$	β_τ	β_σ	$\beta_{\lvert\mathrm{OI}\rvert}$
2002	-0.11***(-30.17)	0.29***(38.62)	0.02***(13.90)	0.07***(5.81)	0.36***(10.24)	0.01***(4.65)
2003	-0.10***(-22.84)	0.38***(49.68)	0.01***(10.17)	0.12***(7.32)	0.40***(35.47)	0.00***(2.86)
2004	-0.11***(-29.36)	0.38***(49.12)	0.02***(15.19)	0.13***(16.66)	0.35***(35.16)	0.00**(2.54)
2005	-0.11***(-29.81)	0.36***(54.11)	0.02***(11.61)	0.14***(16.18)	0.32***(29.02)	0.02***(7.13)
2006	-0.11***(-28.30)	0.34***(50.40)	0.02***(13.80)	0.14***(16.57)	0.27***(28.53)	0.02***(8.90)
2007	-0.11***(-27.49)	0.35***(49.99)	0.03***(14.56)	0.21***(12.05)	0.28***(17.11)	0.03***(9.42)
2008	-0.11***(-16.28)	0.37***(26.57)	0.04***(7.13)	0.09***(6.68)	0.36***(14.65)	0.03***(8.44)
2009	-0.11***(-13.51)	0.36***(32.85)	0.03***(8.99)	0.10***(6.71)	0.21***(10.85)	0.03***(8.16)
2010	-0.08***(-8.12)	0.33***(31.60)	0.02***(8.27)	0.11***(8.48)	0.23***(10.68)	0.03***(9.75)
2011	-0.09***(-11.08)	0.32***(30.32)	0.03***(9.11)	0.08***(8.00)	0.25***(20.88)	0.02***(10.32)
2012	-0.13***(-15.85)	0.30***(31.65)	0.03***(7.74)	0.12***(5.48)	0.21***(12.60)	0.03***(8.05)
2013	-0.12***(-20.71)	0.28***(26.13)	0.02***(11.73)	0.08***(5.68)	0.26***(18.52)	0.03***(13.73)
2014	-0.12***(-22.02)	0.24***(24.44)	0.02***(10.23)	0.01(0.66)	0.30***(19.61)	0.03***(8.49)
2015	-0.11***(-21.71)	0.24***(23.35)	0.03***(11.63)	-0.04**(-2.52)	0.34***(13.41)	0.03***(10.87)
2016	-0.12***(-19.40)	0.24***(24.00)	0.03***(10.07)	-0.05***(-3.86)	0.35***(22.46)	0.03***(9.81)
2017	-0.11***(-12.89)	0.19***(9.56)	0.03***(10.17)	-0.02(-1.31)	0.34***(21.16)	0.03***(9.18)
$\bar{R}^2$(%)		9.72			9.19	

Table IA.10. Effective spread regressed on turnover, realized volatility, and residual order imbalance volatility.

The table reports estimates from the following panel regression: $\log s_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \beta_{\mathrm{HFOIV}^{\perp|\mathrm{OI}|}} \log \mathrm{HFOIV}_{i,t}^{\perp|\mathrm{OI}|} + \mathrm{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover, and $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns. For each stock, log order imbalance volatility (computed using five-minute order imbalances over the trading day) is regressed on log absolute daily order imbalance and a constant. $\log \mathrm{HFOIV}_{i,t}^{\perp|\mathrm{OI}|}$ is the residual obtained from this regression. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in the bottom and top size quintiles. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than $5 and lower than $1,000, a market capitalization greater than $100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

	Small stocks			Large stocks		
Year	β_τ	β_σ	$\beta_{\mathrm{HFOIV}^{\perp\|\mathrm{OI}\|}}$	β_τ	β_σ	$\beta_{\mathrm{HFOIV}^{\perp\|\mathrm{OI}\|}}$
2002	-0.17*** (-41.34)	0.40*** (52.57)	0.06*** (13.28)	-0.10*** (-8.70)	0.47*** (14.56)	0.19*** (15.11)
2003	-0.17*** (-34.64)	0.49*** (66.07)	0.07*** (18.16)	-0.06*** (-5.32)	0.47*** (48.89)	0.17*** (18.58)
2004	-0.19*** (-49.89)	0.49*** (72.17)	0.10*** (24.28)	-0.07*** (-8.68)	0.41*** (37.38)	0.18*** (16.82)
2005	-0.19*** (-55.62)	0.46*** (75.72)	0.10*** (18.56)	-0.08*** (-10.26)	0.40*** (36.88)	0.19*** (16.55)
2006	-0.20*** (-52.82)	0.44*** (72.31)	0.11*** (21.56)	-0.09*** (-12.24)	0.35*** (41.09)	0.21*** (20.71)
2007	-0.18*** (-44.51)	0.44*** (62.70)	0.11*** (20.34)	-0.10*** (-8.01)	0.41*** (25.16)	0.25*** (14.88)
2008	-0.22*** (-32.53)	0.51*** (37.07)	0.16*** (8.50)	-0.18*** (-11.96)	0.47*** (19.71)	0.26*** (12.24)
2009	-0.22*** (-28.43)	0.44*** (39.27)	0.14*** (10.15)	-0.13*** (-11.81)	0.29*** (22.96)	0.23*** (12.41)
2010	-0.21*** (-33.61)	0.45*** (41.79)	0.15*** (11.07)	-0.12*** (-13.44)	0.32*** (16.44)	0.21*** (14.43)
2011	-0.21*** (-34.06)	0.43*** (42.06)	0.15*** (12.51)	-0.11*** (-15.63)	0.33*** (31.00)	0.18*** (16.52)
2012	-0.25*** (-28.76)	0.45*** (37.56)	0.11*** (8.56)	-0.11*** (-11.77)	0.33*** (21.90)	0.19*** (10.40)
2013	-0.24*** (-34.45)	0.42*** (43.23)	0.10*** (15.61)	-0.14*** (-16.58)	0.35*** (22.18)	0.20*** (18.32)
2014	-0.23*** (-38.72)	0.36*** (33.73)	0.10*** (10.71)	-0.23*** (-21.72)	0.41*** (31.99)	0.22*** (10.69)
2015	-0.22*** (-37.85)	0.35*** (33.84)	0.10*** (10.45)	-0.24*** (-25.32)	0.44*** (20.84)	0.19*** (15.68)
2016	-0.22*** (-34.15)	0.35*** (35.37)	0.11*** (8.75)	-0.25*** (-21.78)	0.42*** (23.15)	0.20*** (14.56)
2017	-0.21*** (-28.45)	0.30*** (15.46)	0.09*** (7.61)	-0.23*** (-26.10)	0.45*** (35.29)	0.19*** (10.31)
$\bar{R}^2$(%)		28.35			25.82	

Table IA.11. Effective spread regressed on turnover, realized volatility, and residual absolute order imbalance.
The table reports estimates from the following panel regression: $\log s_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \beta_{|\mathrm{OI}|^{\perp \mathrm{HFOIV}}} \log |\mathrm{OI}|^{\perp \mathrm{HFOIV}}_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover, and $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns. For each stock, log absolute daily order imbalance is regressed on log order imbalance volatility (computed using five-minute order imbalances over the trading day) and a constant. $\log |\mathrm{OI}|^{\perp \mathrm{HFOIV}}_{i,t}$ is the residual obtained from this regression. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in the bottom and top size quintiles. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

	Small stocks			Large stocks		
Year	β_τ	β_σ	$\beta_{\lvert\mathrm{OI}\rvert^{\perp \mathrm{HFOIV}}}$	β_τ	β_σ	$\beta_{\lvert\mathrm{OI}\rvert^{\perp \mathrm{HFOIV}}}$
2002	-0.15*** (-37.89)	0.39*** (52.64)	0.00** (2.38)	0.04*** (3.66)	0.43*** (14.52)	-0.02*** (-18.24)
2003	-0.14*** (-32.53)	0.48*** (66.39)	-0.01*** (-4.39)	0.07*** (6.08)	0.43*** (41.55)	-0.02*** (-13.47)
2004	-0.15*** (-40.38)	0.48*** (72.09)	-0.01*** (-6.58)	0.07*** (10.48)	0.37*** (34.83)	-0.02*** (-14.65)
2005	-0.15*** (-37.23)	0.44*** (75.50)	-0.01*** (-5.72)	0.08*** (10.72)	0.35*** (30.46)	-0.01*** (-12.98)
2006	-0.15*** (-40.11)	0.43*** (71.25)	-0.01*** (-7.63)	0.08*** (10.57)	0.30*** (29.01)	-0.01*** (-12.70)
2007	-0.13*** (-34.21)	0.43*** (62.46)	-0.01*** (-6.65)	0.12*** (8.87)	0.35*** (18.12)	-0.01*** (-9.75)
2008	-0.15*** (-12.29)	0.50*** (36.06)	-0.01*** (-7.74)	0.02** (2.06)	0.43*** (16.03)	-0.01*** (-8.95)
2009	-0.16*** (-17.11)	0.43*** (36.47)	-0.01*** (-5.87)	0.04*** (3.62)	0.23*** (13.04)	-0.01*** (-8.28)
2010	-0.14*** (-14.20)	0.44*** (37.74)	-0.01*** (-7.92)	0.04*** (3.91)	0.27*** (11.77)	-0.01*** (-8.33)
2011	-0.14*** (-16.57)	0.42*** (38.35)	-0.01*** (-6.21)	0.03*** (3.53)	0.29*** (22.52)	-0.01*** (-6.56)
2012	-0.20*** (-20.45)	0.44*** (35.97)	-0.00** (-2.31)	0.05*** (3.45)	0.28*** (16.68)	-0.00*** (-4.64)
2013	-0.19*** (-26.58)	0.42*** (41.72)	-0.01*** (-3.60)	0.03** (2.24)	0.31*** (17.32)	-0.00*** (-5.32)
2014	-0.18*** (-27.40)	0.36*** (33.11)	-0.00 (-1.35)	-0.04*** (-2.67)	0.36*** (22.36)	-0.01*** (-4.65)
2015	-0.17*** (-27.71)	0.35*** (33.27)	0.00 (1.54)	-0.09*** (-8.14)	0.40*** (17.40)	-0.00*** (-4.77)
2016	-0.18*** (-23.75)	0.35*** (34.25)	0.01*** (3.42)	-0.10*** (-7.58)	0.38*** (19.84)	-0.00*** (-3.72)
2017	-0.17*** (-20.57)	0.30*** (15.45)	0.00*** (3.65)	-0.08*** (-5.11)	0.41*** (23.98)	-0.00 (-1.47)
$\bar{R}^2$(%)		26.86			20.03	

Table IA.12. Order imbalance volatility computed at different frequencies. This table reports the median R-squared of estimating (15) with order imbalance volatility computed at different frequencies. For instance, "σ(OI) 30mn" is order imbalance volatility computed using 30-minute intervals over the day.

	Median R^2 across years (%)			
	Level regression		Change regression	
	Small	Large	Small	Large
Without OI	27.70	16.63	9.25	7.99
|OI|	28.29	17.44	9.72	9.19
σ(OI) 65mn	30.40	20.96	11.81	13.26
σ(OI) 30mn	30.89	23.18	12.36	15.47
σ(OI) 15mn	31.31	24.78	12.91	17.20
σ(OI) 10mn	31.41	25.45	12.63	18.95
σ(OI) 5mn	31.46	26.19	13.15	19.02

Table IA.13. Effective spread regressed on balanced turnover, realized volatility, and absolute order imbalance.
The table reports estimates from the following panel regression: $\log s_{i,t} = \alpha_i + \beta_{\mathrm{BV}} \log \mathrm{BV}_{i,t} + \beta_\sigma \log \sigma_{i,t} + \beta_{|\mathrm{OI}|} \log |\mathrm{OI}|_{i,t} + \mathrm{controls} + \epsilon_{i,t}$ for stock i on day t, where $\mathrm{BV}_{i,t}$ is the daily balanced turnover (volume minus absolute order imbalance, divided by shares outstanding), $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns, and $|\mathrm{OI}|_{i,t}$ is the absolute daily order imbalance. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in the bottom and top size quintiles. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

Year	Small stocks			Large stocks		
	β_{BV}	β_σ	$\beta_{\lvert\mathrm{OI}\rvert}$	β_{BV}	β_σ	$\beta_{\lvert\mathrm{OI}\rvert}$
2002	-0.12*** (-46.09)	0.43*** (58.88)	-0.02*** (-13.29)	0.01* (1.73)	0.44*** (14.68)	0.01*** (3.06)
2003	-0.12*** (-35.40)	0.50*** (69.38)	-0.02*** (-13.24)	0.05*** (4.23)	0.43*** (42.30)	0.01*** (5.99)
2004	-0.13*** (-45.33)	0.51*** (84.85)	-0.02*** (-12.81)	0.05*** (7.66)	0.37*** (34.35)	0.01*** (6.55)
2005	-0.14*** (-55.01)	0.49*** (89.73)	-0.01*** (-6.56)	0.04*** (6.20)	0.36*** (32.03)	0.02*** (7.42)
2006	-0.14*** (-53.11)	0.46*** (81.33)	-0.01*** (-7.82)	0.03*** (4.69)	0.32*** (32.90)	0.03*** (9.27)
2007	-0.13*** (-45.77)	0.46*** (74.48)	-0.00** (-2.31)	0.04*** (4.36)	0.38*** (22.13)	0.04*** (7.32)
2008	-0.17*** (-38.89)	0.55*** (40.72)	0.01 (1.24)	-0.04*** (-4.54)	0.45*** (17.04)	0.04*** (6.91)
2009	-0.18*** (-35.67)	0.46*** (42.00)	0.01 (1.55)	-0.02*** (-2.68)	0.27*** (17.90)	0.03*** (6.25)
2010	-0.15*** (-33.78)	0.46*** (42.74)	0.00 (1.01)	-0.02** (-2.56)	0.30*** (14.60)	0.03*** (7.72)
2011	-0.16*** (-37.88)	0.45*** (44.25)	0.01* (1.72)	-0.02*** (-3.65)	0.31*** (27.45)	0.03*** (8.80)
2012	-0.20*** (-39.57)	0.46*** (38.45)	-0.00 (-1.03)	-0.02*** (-2.98)	0.32*** (18.28)	0.04*** (6.34)
2013	-0.18*** (-33.70)	0.42*** (25.17)	-0.01*** (-3.16)	-0.04*** (-5.43)	0.34*** (20.60)	0.04*** (10.45)
2014	-0.16*** (-33.06)	0.37*** (35.40)	-0.01** (-2.40)	-0.12*** (-14.32)	0.40*** (29.39)	0.03*** (5.52)
2015	-0.16*** (-32.71)	0.37*** (35.30)	-0.00 (-0.65)	-0.15*** (-18.55)	0.43*** (20.15)	0.03*** (7.51)
2016	-0.18*** (-34.51)	0.37*** (38.39)	0.01 (1.11)	-0.16*** (-15.37)	0.41*** (22.29)	0.03*** (7.34)
2017	-0.16*** (-27.48)	0.31*** (15.82)	0.00 (0.46)	-0.15*** (-19.13)	0.45*** (34.68)	0.03*** (5.76)
$\bar{R}^2$(%)		28.67			20.83	

Table IA.14. Effective spread regressed on balanced turnover, realized volatility, and absolute order imbalance (changes).
The table reports estimates from the following regression: $\Delta s_{i,t} = \alpha_i + \beta_{\mathrm{BV}} \Delta \mathrm{BV}_{i,t} + \beta_\sigma \Delta \sigma_{i,t} + \beta_{|\mathrm{OI}|} \Delta |\mathrm{OI}_{i,t}| + \text{controls} + \epsilon_{i,t}$, for stock i on day t where $\Delta x_t \equiv \log(\frac{x_t}{x_{t-1}})$, $\mathrm{BV}_{i,t}$ is the daily balanced turnover (volume minus absolute order imbalance, divided by shares outstanding), $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns, and $|\mathrm{OI}_{i,t}|$ is the absolute daily order imbalance. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in the bottom and top size quintiles. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

Year	Small stocks			Large stocks		
	β_{BV}	β_σ	$\beta_{\lvert\mathrm{OI}\rvert}$	β_{BV}	β_σ	$\beta_{\lvert\mathrm{OI}\rvert}$
2002	-0.08***(-33.83)	0.32***(44.51)	-0.01***(-7.71)	0.04***(4.45)	0.37***(10.26)	0.02***(6.26)
2003	-0.08***(-24.34)	0.40***(51.83)	-0.01***(-7.83)	0.09***(6.45)	0.40***(36.41)	0.01***(7.36)
2004	-0.08***(-32.72)	0.42***(66.67)	-0.01***(-6.20)	0.10***(13.70)	0.36***(35.21)	0.01***(8.78)
2005	-0.08***(-33.32)	0.41***(71.13)	-0.01***(-3.66)	0.10***(13.41)	0.33***(30.81)	0.03***(10.36)
2006	-0.09***(-32.36)	0.38***(66.72)	-0.00***(-2.83)	0.09***(13.67)	0.29***(32.03)	0.03***(11.84)
2007	-0.09***(-29.50)	0.39***(60.90)	0.00(1.06)	0.14***(10.64)	0.31***(20.24)	0.05***(10.11)
2008	-0.10***(-23.36)	0.43***(33.00)	0.02**(2.53)	0.04***(3.58)	0.38***(15.42)	0.04***(8.81)
2009	-0.11***(-22.90)	0.38***(39.94)	0.01***(2.72)	0.04***(4.10)	0.24***(13.15)	0.03***(8.27)
2010	-0.08***(-16.02)	0.35***(36.15)	0.01**(2.57)	0.06***(6.29)	0.26***(12.51)	0.03***(9.86)
2011	-0.10***(-20.05)	0.35***(35.70)	0.01***(3.11)	0.04***(5.90)	0.27***(24.10)	0.03***(10.20)
2012	-0.12***(-26.84)	0.32***(34.90)	0.00(0.67)	0.06***(4.51)	0.25***(13.85)	0.04***(7.49)
2013	-0.11***(-24.40)	0.30***(26.43)	0.00(0.09)	0.02***(2.76)	0.29***(22.20)	0.04***(12.34)
2014	-0.10***(-23.71)	0.26***(27.29)	0.00(0.88)	-0.05***(-5.04)	0.34***(25.78)	0.04***(7.33)
2015	-0.10***(-24.53)	0.26***(26.40)	0.01**(2.35)	-0.08***(-8.16)	0.37***(15.46)	0.03***(8.83)
2016	-0.11***(-25.70)	0.26***(27.17)	0.01***(3.29)	-0.10***(-9.93)	0.39***(25.75)	0.03***(8.28)
2017	-0.10***(-16.03)	0.20***(9.32)	0.01***(2.85)	-0.09***(-9.31)	0.39***(28.43)	0.03***(7.52)
$\bar{R}^2$(%)		10.99			8.92	

Table IA.15. Effective spread regressed on balanced turnover, realized volatility, and order imbalance volatility.
The table reports estimates from the following panel regression: $\log s_{i,t} = \alpha_i + \beta_{\text{BV}} \log \text{BV}_{i,t} + \beta_\sigma \log \sigma_{i,t} + \beta_{\text{HFOIV}} \log \text{HFOIV}_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\text{BV}_{i,t}$ is the daily balanced turnover (volume minus absolute order imbalance, divided by shares outstanding), $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns, and HFOIV is high-frequency order imbalance volatility computed as the standard deviation of five-minute share imbalance scaled by total shares outstanding. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in the bottom and top size quintiles. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

Year	Small stocks			Large stocks		
	β_{BV}	β_σ	β_{HFOIV}	β_{BV}	β_σ	β_{HFOIV}
2002	-0.14*** (-41.46)	0.43*** (56.30)	0.02*** (4.71)	-0.18*** (-10.84)	0.49*** (14.66)	0.24*** (13.26)
2003	-0.14*** (-34.05)	0.50*** (67.44)	0.03*** (7.71)	-0.16*** (-10.31)	0.49*** (50.94)	0.22*** (17.09)
2004	-0.17*** (-46.77)	0.52*** (81.84)	0.06*** (14.34)	-0.16*** (-13.74)	0.43*** (40.14)	0.22*** (17.91)
2005	-0.17*** (-46.22)	0.49*** (86.33)	0.07*** (10.13)	-0.20*** (-15.06)	0.43*** (40.13)	0.25*** (18.61)
2006	-0.18*** (-51.04)	0.47*** (81.11)	0.07*** (13.31)	-0.21*** (-20.91)	0.40*** (49.28)	0.25*** (25.67)
2007	-0.18*** (-39.42)	0.47*** (73.94)	0.08*** (12.95)	-0.24*** (-15.24)	0.48*** (30.50)	0.30*** (18.99)
2008	-0.25*** (-22.48)	0.56*** (42.23)	0.14*** (6.43)	-0.36*** (-17.82)	0.53*** (24.40)	0.34*** (18.09)
2009	-0.26*** (-29.49)	0.48*** (45.39)	0.14*** (9.90)	-0.30*** (-18.36)	0.36*** (31.34)	0.31*** (19.95)
2010	-0.24*** (-29.04)	0.48*** (47.43)	0.15*** (10.01)	-0.28*** (-21.02)	0.38*** (22.29)	0.28*** (21.40)
2011	-0.24*** (-35.28)	0.47*** (48.59)	0.15*** (11.90)	-0.25*** (-25.39)	0.38*** (35.01)	0.24*** (27.03)
2012	-0.26*** (-29.59)	0.47*** (40.36)	0.11*** (7.47)	-0.27*** (-13.40)	0.39*** (23.87)	0.26*** (13.35)
2013	-0.22*** (-32.79)	0.43*** (24.93)	0.07*** (10.01)	-0.28*** (-26.82)	0.41*** (27.91)	0.26*** (25.55)
2014	-0.20*** (-24.93)	0.38*** (35.10)	0.06*** (5.27)	-0.40*** (-18.73)	0.48*** (35.31)	0.29*** (13.35)
2015	-0.20*** (-24.49)	0.37*** (34.33)	0.07*** (6.04)	-0.39*** (-30.39)	0.49*** (26.60)	0.26*** (22.16)
2016	-0.23*** (-23.89)	0.38*** (39.02)	0.10*** (6.40)	-0.40*** (-27.05)	0.47*** (27.10)	0.27*** (20.32)
2017	-0.20*** (-20.65)	0.31*** (15.72)	0.08*** (5.12)	-0.39*** (-22.99)	0.51*** (46.65)	0.26*** (14.09)
$\bar{R}^2$(%)		29.93			30.49	

Table IA.16. Effective spread regressed on balanced turnover, realized volatility, and order imbalance volatility (changes).
The table reports estimates from the following regression: $\Delta s_{i,t} = \alpha_i + \beta_{\text{BV}}\Delta\text{BV}_{i,t} + \beta_\sigma \Delta\sigma_{i,t} + \beta_{\text{HFOIV}_{i,t}}\Delta\text{HFOIV}_{i,t} + \text{controls} + \epsilon_{i,t}$, for stock i on day t where $\Delta x_t \equiv \log(\frac{x_t}{x_{t-1}})$, $\text{BV}_{i,t}$ is the daily balanced turnover (volume minus absolute order imbalance, divided by shares outstanding), $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns, and HFOIV is high-frequency order imbalance volatility computed as the standard deviation of five-minute share imbalance scaled by total shares outstanding. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in the bottom and top size quintiles. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

Year	Small stocks			Large stocks		
	β_{BV}	β_σ	β_{HFOIV}	β_{BV}	β_σ	β_{HFOIV}
2002	-0.09***(-33.74)	0.31***(43.03)	0.02***(5.26)	-0.14***(-9.49)	0.41***(10.29)	0.24***(15.27)
2003	-0.09***(-24.43)	0.39***(50.55)	0.03***(6.53)	-0.12***(-7.54)	0.44***(41.79)	0.22***(18.46)
2004	-0.11***(-36.41)	0.42***(65.36)	0.06***(13.11)	-0.11***(-10.11)	0.40***(40.04)	0.22***(19.89)
2005	-0.12***(-32.69)	0.40***(68.84)	0.06***(9.54)	-0.15***(-13.59)	0.39***(40.59)	0.26***(22.60)
2006	-0.12***(-36.90)	0.38***(65.24)	0.06***(11.70)	-0.16***(-17.04)	0.35***(45.37)	0.25***(28.84)
2007	-0.13***(-29.78)	0.39***(59.98)	0.08***(12.05)	-0.18***(-11.02)	0.38***(28.19)	0.30***(22.79)
2008	-0.18***(-17.92)	0.43***(34.32)	0.14***(7.64)	-0.30***(-17.01)	0.45***(20.82)	0.32***(21.99)
2009	-0.18***(-24.98)	0.39***(41.32)	0.13***(10.81)	-0.28***(-18.73)	0.31***(23.62)	0.31***(25.41)
2010	-0.17***(-25.37)	0.35***(40.20)	0.15***(11.91)	-0.24***(-20.91)	0.32***(18.94)	0.28***(26.62)
2011	-0.17***(-27.52)	0.35***(38.62)	0.14***(12.47)	-0.20***(-22.14)	0.33***(33.22)	0.23***(29.16)
2012	-0.17***(-21.82)	0.32***(37.27)	0.10***(7.43)	-0.22***(-14.96)	0.31***(20.98)	0.26***(15.48)
2013	-0.15***(-27.07)	0.30***(26.26)	0.07***(11.43)	-0.24***(-25.85)	0.35***(31.30)	0.26***(28.94)
2014	-0.13***(-21.51)	0.25***(27.15)	0.07***(7.14)	-0.36***(-22.54)	0.39***(36.86)	0.30***(17.24)
2015	-0.14***(-22.91)	0.26***(25.75)	0.08***(7.87)	-0.35***(-30.04)	0.42***(18.20)	0.26***(24.90)
2016	-0.15***(-21.39)	0.26***(27.23)	0.09***(7.22)	-0.37***(-26.86)	0.42***(33.17)	0.27***(22.04)
2017	-0.14***(-18.45)	0.20***(9.30)	0.09***(6.95)	-0.35***(-25.69)	0.43***(36.69)	0.27***(17.44)
$\bar{R}^2$(%)		12.67			20.04	

Table IA.17. Effective spread regressed on common and idiosyncratic components of volume, volatility, and order imbalance volatility (change regression).
This table reports median estimate, median t-statistic (in parentheses), and median adjusted R-squared across years. The following panel regression with stock fixed effects is estimated each year for small and large stocks: $\Delta \log s_{i,t} = \alpha_i + \beta_{\Delta\tau,C}\Delta\tau^C_{i,t} + \beta_{\Delta\tau,I}\Delta\tau^I_{i,t} + \beta_{\Delta\sigma,C}\Delta\sigma^C_{i,t} + \beta_{\Delta\sigma,I}\Delta\sigma^I_{i,t} + \beta_{\Delta\text{HFOIV},C}\Delta\text{HFOIV}^C_{i,t} + \beta_{\Delta\text{HFOIV},I}\Delta\text{HFOIV}^I_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\Delta x_t \equiv \log(\frac{x_t}{x_{t-1}})$. $\tau_{i,t}$ is daily intraday turnover, $\sigma_{i,t}$ is realized volatility estimated using five-minute returns over the current day, and $\text{HFOIV}_{i,t}$ is high-frequency order imbalance volatility computed as the standard deviation of five-minute share imbalance scaled by total shares outstanding These variables are decomposed into common (C) and idiosyncratic (I) components as described in the text. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock.

	Median value across years			
	Small stocks		Large stocks	
$\beta_{\Delta\tau,C}$	0.01	-0.07	0.27	-0.03
	(0.31)	(-2.58)	(3.95)	(-0.57)
$\beta_{\Delta\tau,I}$	-0.09	-0.20	0.10	-0.21
	(-15.98)	(-24.59)	(7.60)	(-14.37)
$\beta_{\Delta\sigma,C}$	0.01	0.01	0.01	0.02
	(3.66)	(3.85)	(3.69)	(5.52)
$\beta_{\Delta\sigma,I}$	0.31	0.33	0.27	0.33
	(29.27)	(32.34)	(17.59)	(22.66)
$\beta_{\Delta\text{HFOIV},C}$		0.04		0.03
		(9.38)		(6.90)
$\beta_{\Delta\text{HFOIV},I}$		0.09		0.23
		(11.03)		(24.31)
$R^2(\%)$	9.76	13.14	8.49	15.70

Table IA.18. Effective spread regressed on turnover and volatility in the cross-section.
The table reports estimates from the following panel regression: $\log s_{i,t} = \alpha_t + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover and $\sigma_{i,t}$ is the realized volatility estimated using five-minute returns over the current day. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in a given size quintile. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than $5 and lower than $1,000, a market capitalization greater than $100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

	Small stocks		Large stocks	
Year	β_τ	β_σ	β_τ	β_σ
2002	-0.23*** (-36.52)	0.47*** (47.56)	-0.08*** (-4.20)	0.51*** (12.84)
2003	-0.22*** (-40.92)	0.57*** (65.57)	-0.07*** (-3.35)	0.54*** (14.52)
2004	-0.25*** (-50.89)	0.57*** (63.02)	-0.12*** (-4.06)	0.56*** (12.38)
2005	-0.27*** (-53.30)	0.55*** (61.62)	-0.10*** (-2.86)	0.51*** (12.84)
2006	-0.29*** (-53.03)	0.52*** (54.37)	-0.11*** (-3.03)	0.46*** (12.55)
2007	-0.29*** (-52.00)	0.50*** (47.11)	-0.08** (-2.07)	0.43*** (15.04)
2008	-0.38*** (-47.01)	0.53*** (36.07)	-0.12*** (-2.81)	0.54*** (14.37)
2009	-0.35*** (-42.53)	0.52*** (35.15)	-0.03 (-1.32)	0.41*** (8.75)
2010	-0.30*** (-37.13)	0.51*** (34.47)	-0.01 (-0.38)	0.40*** (11.27)
2011	-0.33*** (-40.37)	0.53*** (35.78)	-0.00 (-0.14)	0.40*** (13.06)
2012	-0.36*** (-41.30)	0.53*** (35.48)	-0.09* (-1.68)	0.53*** (8.11)
2013	-0.33*** (-42.69)	0.49*** (31.09)	-0.09 (-1.64)	0.52*** (8.05)
2014	-0.33*** (-42.72)	0.47*** (36.73)	-0.13*** (-5.15)	0.59*** (15.30)
2015	-0.32*** (-43.70)	0.44*** (33.28)	-0.12*** (-5.67)	0.57*** (18.02)
2016	-0.34*** (-44.60)	0.45*** (35.28)	-0.13*** (-4.79)	0.54*** (14.53)
2017	-0.30*** (-32.61)	0.39*** (16.63)	-0.14*** (-5.71)	0.60*** (17.16)
$\bar{R}^2$(%)	52.52		36.22	

Table IA.19. Effective spread regressed on turnover, realized volatility, and order imbalance volatility in the cross-section.
The table reports estimates from the following panel regression: $\log s_{i,t} = \alpha_t + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \beta_{\text{HFOIV}} \log \text{HFOIV}_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover, $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns, and $\text{HFOIV}_{i,t}$ is the volatility of order imbalance computed using five-minute order imbalances over the trading day. Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in the bottom and top size quintiles. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

	Small stocks			Large stocks		
Year	β_τ	β_σ	β_{HFOIV}	β_τ	β_σ	β_{HFOIV}
2002	-0.37*** (-39.47)	0.51*** (50.69)	0.21*** (24.57)	-0.36*** (-12.26)	0.59*** (13.50)	0.34*** (17.97)
2003	-0.36*** (-44.40)	0.61*** (68.65)	0.21*** (28.42)	-0.37*** (-10.57)	0.64*** (17.53)	0.34*** (15.12)
2004	-0.42*** (-63.27)	0.62*** (71.06)	0.28*** (40.68)	-0.44*** (-10.14)	0.66*** (15.89)	0.37*** (13.16)
2005	-0.44*** (-64.56)	0.59*** (71.29)	0.29*** (39.79)	-0.45*** (-8.92)	0.62*** (16.40)	0.39*** (12.72)
2006	-0.47*** (-72.55)	0.57*** (68.59)	0.32*** (49.65)	-0.45*** (-9.74)	0.58*** (17.06)	0.39*** (14.36)
2007	-0.48*** (-76.30)	0.56*** (64.34)	0.34*** (53.62)	-0.43*** (-10.59)	0.56*** (22.97)	0.39*** (17.47)
2008	-0.58*** (-77.14)	0.63*** (55.53)	0.39*** (43.44)	-0.47*** (-11.16)	0.63*** (21.74)	0.40*** (18.13)
2009	-0.52*** (-50.88)	0.61*** (41.50)	0.30*** (37.15)	-0.39*** (-10.67)	0.52*** (12.62)	0.39*** (17.40)
2010	-0.48*** (-39.97)	0.54*** (42.58)	0.29*** (31.09)	-0.36*** (-10.58)	0.49*** (14.61)	0.38*** (17.09)
2011	-0.51*** (-56.01)	0.56*** (45.28)	0.31*** (40.03)	-0.35*** (-13.11)	0.48*** (15.92)	0.37*** (24.42)
2012	-0.52*** (-47.84)	0.55*** (43.25)	0.27*** (30.18)	-0.42*** (-8.38)	0.59*** (12.23)	0.37*** (19.05)
2013	-0.47*** (-48.99)	0.51*** (32.06)	0.24*** (31.61)	-0.45*** (-8.33)	0.58*** (11.87)	0.39*** (15.88)
2014	-0.51*** (-47.68)	0.51*** (41.92)	0.29*** (28.50)	-0.53*** (-17.41)	0.66*** (20.47)	0.42*** (22.88)
2015	-0.51*** (-51.56)	0.48*** (38.57)	0.31*** (31.59)	-0.52*** (-18.57)	0.65*** (22.65)	0.41*** (25.21)
2016	-0.53*** (-50.47)	0.48*** (41.03)	0.32*** (30.88)	-0.54*** (-15.96)	0.63*** (20.14)	0.43*** (22.15)
2017	-0.50*** (-33.37)	0.42*** (17.06)	0.30*** (21.63)	-0.50*** (-15.91)	0.68*** (22.13)	0.38*** (22.36)
$\bar{R}^2$(%)		57.82			46.41	

Table IA.20. High-frequency order imbalance volatility, turnover, and stock returns (raw returns). Every week, portfolios are formed by sequentially sorting stocks using NYSE breakpoints. The table reports portfolios' excess returns. Panel (a): sort on turnover then on high-frequency order imbalance volatility (HFOIV). Panel (b): sort on high-frequency order imbalance volatility then on turnover. Turnover is the average daily turnover over the previous month. Order imbalance volatility is an exponentially-weighted moving average of prior order imbalance with a half-life of one day. To be included in a portfolio, a stock must have a price greater than $5 on the formation date. The sample consists of NYSE, Amex, and NASDAQ common stocks over 2002-2017 (797 weekly observations). t-statistics are reported in parentheses and computed using Newey-West standard errors. *, **, and *** denote significance at the 10%, 5%, and 1% level.

(a) $\bar{r}^{\mathrm{VW}}$ (turnover then order imbalance volatility)

	Low HFOIV	2	3	4	High HFOIV	H-L
Low turn.	0.12**	0.16**	0.20**	0.20**	0.29***	0.17***
	(1.99)	(2.09)	(2.42)	(2.41)	(3.59)	(2.90)
2	0.16**	0.21***	0.20***	0.24***	0.27***	0.12***
	(2.06)	(2.75)	(2.59)	(2.73)	(3.33)	(2.64)
3	0.15*	0.20**	0.25***	0.24***	0.27***	0.12***
	(1.71)	(2.35)	(2.91)	(2.63)	(3.23)	(2.64)
4	0.11	0.19**	0.23**	0.22**	0.32***	0.22***
	(1.05)	(1.97)	(2.37)	(2.15)	(3.37)	(4.31)
High turn.	0.21*	0.16	0.28**	0.17	0.26**	0.05
	(1.71)	(1.23)	(2.27)	(1.34)	(2.24)	(0.75)

(b) $\bar{r}^{\mathrm{VW}}$ (order imbalance volatility then turnover)

	Low turn.	2	3	4	High turn.	H-L
Low HFOIV	0.17**	0.12*	0.14*	0.20**	0.12	-0.06
	(2.27)	(1.90)	(1.89)	(2.50)	(1.20)	(-1.09)
2	0.20**	0.19**	0.21**	0.16*	0.15	-0.05
	(2.36)	(2.36)	(2.53)	(1.76)	(1.38)	(-0.79)
3	0.25***	0.25***	0.25***	0.24**	0.20	-0.05
	(2.86)	(2.90)	(2.84)	(2.51)	(1.62)	(-0.72)
4	0.26***	0.25***	0.24**	0.18	0.18	-0.08
	(3.16)	(2.78)	(2.38)	(1.63)	(1.46)	(-1.08)
High HFOIV	0.29***	0.34***	0.27***	0.26**	0.23	-0.06
	(3.71)	(3.60)	(2.73)	(2.25)	(1.60)	(-0.70)

Table IA.21. Price impact (lambda) regressed on turnover, realized volatility, and order imbalance volatility.
The following regression is estimated: $\Delta\lambda_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_\sigma \Delta\sigma_{i,t} + \beta_{\text{HFOIV}} \Delta\text{HFOIV}_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover and $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns, $\text{HFOIV}_{i,t}$ is the volatility of order imbalance computed using five-minute order imbalances over the trading day, and $\Delta x_t = \log(\frac{x_t}{x_{t-1}})$. $\lambda_{i,t}$ is obtained from the regression $r_{i,t,k} = \delta_{i,t} + \lambda_{i,t}\sqrt{|\text{OI}^\$_{i,t,k}|}\text{sign}(\text{OI}^\$_{i,t,k}) + e_{i,t}$, where $r_{i,t,k}$ is the five-minute midquote return in interval k, and $\text{OI}^\$_{i,t,k}$ is the dollar order imbalance. Negative estimates of $\lambda_{i,t}$ are excluded. Controls are (log) market capitalization, (log) price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in the bottom and top size quintile. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Price impacts are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

Year	Small stocks β_τ	Small stocks β_σ	Small stocks β_{HFOIV}	Large stocks β_τ	Large stocks β_σ	Large stocks β_{HFOIV}
2002	-0.51***(-48.63)	1.32***(121.11)	-0.06***(-5.35)	-0.39***(-20.69)	1.19***(27.62)	-0.32***(-33.39)
2003	-0.42***(-35.71)	1.26***(85.49)	-0.09***(-8.36)	-0.30***(-18.73)	1.25***(77.85)	-0.36***(-39.50)
2004	-0.40***(-35.91)	1.26***(91.99)	-0.07***(-5.82)	-0.33***(-23.69)	1.29***(96.91)	-0.35***(-42.17)
2005	-0.43***(-38.11)	1.28***(105.17)	-0.08***(-8.00)	-0.36***(-28.69)	1.26***(94.22)	-0.37***(-50.40)
2006	-0.43***(-38.20)	1.25***(108.90)	-0.07***(-6.78)	-0.36***(-24.11)	1.21***(88.64)	-0.40***(-38.68)
2007	-0.38***(-29.62)	1.23***(100.53)	-0.12***(-10.49)	-0.33***(-15.81)	1.14***(40.36)	-0.39***(-32.37)
2008	-0.35***(-24.89)	1.13***(75.88)	-0.14***(-11.34)	-0.40***(-20.88)	1.25***(48.59)	-0.41***(-34.01)
2009	-0.27***(-15.97)	1.15***(65.98)	-0.15***(-12.06)	-0.23***(-12.73)	1.28***(55.74)	-0.42***(-39.73)
2010	-0.23***(-14.04)	1.16***(70.78)	-0.12***(-9.68)	-0.16***(-8.43)	1.32***(44.57)	-0.37***(-34.11)
2011	-0.31***(-15.77)	1.20***(58.38)	-0.08***(-5.94)	-0.20***(-8.81)	1.33***(45.39)	-0.31***(-21.88)
2012	-0.29***(-11.02)	1.14***(53.35)	-0.10***(-6.18)	-0.21***(-9.91)	1.39***(51.26)	-0.30***(-22.91)
2013	-0.32***(-16.59)	1.14***(56.43)	-0.10***(-7.41)	-0.31***(-11.62)	1.43***(46.11)	-0.27***(-17.73)
2014	-0.24***(-19.59)	1.13***(82.63)	-0.19***(-17.13)	-0.26***(-16.54)	1.24***(67.66)	-0.44***(-62.37)
2015	-0.22***(-16.48)	1.10***(76.69)	-0.20***(-17.96)	-0.25***(-19.64)	1.21***(54.71)	-0.44***(-61.86)
2016	-0.18***(-14.25)	0.96***(65.79)	-0.21***(-18.98)	-0.27***(-19.27)	1.22***(72.38)	-0.41***(-48.07)
2017	-0.20***(-6.76)	1.00***(12.18)	-0.19***(-14.92)	-0.32***(-23.35)	1.22***(85.61)	-0.42***(-56.08)
$\bar{R}^2$(%)		19.43			19.45	

Table IA.22. Price impact (ILLIQ) regressed on turnover, realized volatility, and order imbalance volatility.
The following regression is estimated: $\Delta \text{ILLIQ}_{i,t} = \alpha_i + \beta_\tau \Delta\tau_{i,t} + \beta_\sigma \Delta\sigma_{i,t} + \beta_{\sigma(\text{OI})} \Delta \text{HFOIV}_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover, $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns, $\text{HFOIV}_{i,t}$ is the volatility of order imbalance computed using five-minute order imbalances over the trading day, $\text{ILLIQ}_{i,t} = \frac{1}{\#\text{traded intervals}} \sum_{k \epsilon \{j|\text{DVOL}_{i,t,j}>0\}} \frac{|r_{itk}|}{\text{DollarVolume}_{i,t,k}}$, and $\Delta x_t = \log(\frac{x_t}{x_{t-1}})$. Controls are (log) market capitalization, (log) price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in the bottom and top size quintile. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than $5 and lower than $1,000, a market capitalization greater than $100 million, and at least 100 days of prior trading. Price impacts are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R}^2$ denotes the average adjusted R-squared across years.

Year	Small stocks			Large stocks		
	β_τ	β_σ	β_{HFOIV}	β_τ	β_σ	β_{HFOIV}
2002	-0.51***(-72.61)	0.97***(117.23)	0.14***(23.24)	-0.87***(-35.73)	0.77***(17.31)	0.16***(12.76)
2003	-0.48***(-38.45)	0.85***(69.95)	0.13***(16.23)	-1.08***(-46.12)	0.77***(65.26)	0.25***(24.95)
2004	-0.48***(-51.89)	0.87***(109.96)	0.11***(16.66)	-1.06***(-91.94)	0.74***(88.74)	0.22***(40.56)
2005	-0.50***(-61.78)	0.86***(101.93)	0.12***(20.68)	-1.04***(-89.08)	0.74***(116.90)	0.21***(44.16)
2006	-0.51***(-61.71)	0.91***(104.09)	0.12***(19.63)	-1.01***(-96.18)	0.77***(94.44)	0.18***(40.67)
2007	-0.57***(-61.52)	0.91***(76.46)	0.14***(20.64)	-0.99***(-94.69)	0.82***(76.85)	0.14***(33.03)
2008	-0.55***(-56.60)	0.89***(69.82)	0.12***(16.81)	-0.98***(-71.41)	0.79***(60.78)	0.09***(19.16)
2009	-0.57***(-44.37)	0.82***(61.43)	0.11***(14.70)	-0.98***(-102.64)	0.76***(65.12)	0.09***(17.18)
2010	-0.63***(-49.21)	0.90***(69.94)	0.12***(14.95)	-0.95***(-69.76)	0.74***(29.18)	0.09***(20.55)
2011	-0.57***(-46.26)	0.87***(86.37)	0.10***(14.49)	-0.97***(-98.20)	0.77***(59.61)	0.10***(21.33)
2012	-0.54***(-30.93)	0.91***(69.58)	0.08***(7.27)	-0.95***(-92.17)	0.69***(67.00)	0.10***(21.87)
2013	-0.64***(-43.48)	0.86***(37.61)	0.13***(13.23)	-0.97***(-60.90)	0.69***(52.34)	0.12***(26.34)
2014	-0.83***(-39.87)	0.95***(68.27)	0.20***(15.68)	-1.00***(-152.46)	0.71***(83.02)	0.14***(37.16)
2015	-0.81***(-38.99)	0.96***(59.94)	0.22***(16.26)	-0.99***(-159.76)	0.71***(82.41)	0.13***(34.68)
2016	-0.80***(-40.54)	0.98***(70.58)	0.23***(12.78)	-0.98***(-130.38)	0.70***(75.81)	0.12***(33.87)
2017	-0.71***(-23.47)	0.87***(11.14)	0.16***(11.05)	-0.95***(-160.66)	0.64***(101.90)	0.13***(40.77)
$\bar{R}^2$(%)		29.14			58.75	

Table IA.23. Depth regressed on turnover, realized volatility, and order imbalance volatility.
The following regression is estimated: $\Delta \text{Depth}_{i,t} = \alpha_i + \beta_\tau \Delta\tau_{i,t} + \beta_\sigma \Delta\sigma_{i,t} + \beta_{\text{HFOIV}} \Delta \text{HFOIV}_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where Depth is the average of the time-weighted share depth at the best bid and best ask (as a fraction of shares outstanding), $\tau_{i,t}$ is the daily intraday turnover and $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns, $\text{HFOIV}_{i,t}$ is the volatility of order imbalance computed using five-minute order imbalances over the trading day, and $\Delta x_t = \log(\frac{x_t}{x_{t-1}})$. Controls are (log) market capitalization, (log) price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a year-by-year basis for stocks in the bottom and top size quintile. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Depths and effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. $\bar{R^2}$ denotes the average adjusted R-squared across years.

	Small stocks			Large stocks		
Year	β_τ	β_σ	β_{HFOIV}	β_τ	β_σ	β_{HFOIV}
2002	0.11***(22.52)	-0.10***(-21.58)	0.01*(1.73)	0.29***(19.54)	-0.17***(-10.69)	-0.02***(-3.30)
2003	0.12***(21.85)	-0.13***(-27.07)	-0.01***(-3.72)	0.42***(19.98)	-0.27***(-33.19)	-0.07***(-8.52)
2004	0.11***(24.35)	-0.13***(-28.34)	-0.01***(-3.32)	0.45***(40.45)	-0.28***(-40.29)	-0.08***(-18.41)
2005	0.09***(22.46)	-0.12***(-26.23)	-0.01**(-2.44)	0.39***(39.92)	-0.27***(-41.68)	-0.07***(-20.10)
2006	0.12***(27.64)	-0.13***(-25.76)	-0.01***(-4.76)	0.32***(42.15)	-0.30***(-38.75)	-0.05***(-15.42)
2007	0.12***(28.09)	-0.14***(-28.70)	-0.00(-0.19)	0.28***(38.36)	-0.32***(-29.99)	-0.03***(-7.69)
2008	0.11***(23.17)	-0.12***(-21.25)	0.01***(2.97)	0.27***(26.13)	-0.33***(-22.10)	-0.02***(-3.61)
2009	0.12***(20.47)	-0.15***(-21.07)	0.01**(2.14)	0.25***(32.91)	-0.33***(-28.89)	-0.02***(-3.73)
2010	0.14***(25.20)	-0.19***(-25.97)	0.00(0.40)	0.28***(25.61)	-0.37***(-20.08)	-0.02***(-3.17)
2011	0.11***(18.32)	-0.16***(-26.11)	0.01**(2.20)	0.25***(32.80)	-0.34***(-26.93)	-0.01(-1.56)
2012	0.11***(17.50)	-0.14***(-20.80)	0.01***(2.63)	0.24***(39.13)	-0.23***(-26.37)	-0.02***(-5.51)
2013	0.13***(23.27)	-0.13***(-19.63)	-0.00(-0.24)	0.25***(35.82)	-0.23***(-23.03)	-0.03***(-10.48)
2014	0.10***(17.13)	-0.11***(-18.39)	0.03***(6.52)	0.24***(36.38)	-0.21***(-19.57)	-0.02***(-4.18)
2015	0.08***(12.75)	-0.08***(-13.57)	0.03***(8.12)	0.19***(33.13)	-0.16***(-15.31)	-0.01***(-5.47)
2016	0.06***(9.14)	-0.08***(-14.77)	0.04***(9.32)	0.19***(38.14)	-0.16***(-18.53)	-0.01***(-5.92)
2017	0.06***(9.31)	-0.07***(-9.25)	0.03***(7.25)	0.19***(39.22)	-0.13***(-18.31)	-0.02***(-9.16)
$\bar{R^2}$(%)		4.13			16.23	

Table IA.24. Price impact and realized spread regressed on turnover, realized volatility, and order imbalance volatility for large stocks in the time series.
$x_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \beta_{\text{HFOIV}} \log \text{HFOIV}_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover and $\sigma_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns and $\text{HFOIV}_{i,t}$ is the volatility of order imbalance computed using five-minute order imbalances over the trading day. $x_{i,t}$ denotes the price impact or realized spread (in basis points) obtained by decomposing the effective spread using the midquote five minutes after a trade. Both measures are in percent and computed by dollar-weighting over all trades in a day. Controls are (log) market capitalization, (log) price, and day-of-the-week and month-of-the-year indicators. The regression includes stock fixed effects and is estimated on a month-by-month basis for stocks in the top size quintile. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days. The sample consists of NYSE, Amex, and NASDAQ common stocks in 2017. To be included in a given month, a stock is required to have at the beginning of the month a price greater than $5 and lower than $1,000, a market capitalization greater than $100 million, and at least 100 days of prior trading. Price impact and realized spread are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock, and t-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level.

	(a) Price impact			(b) Realized spread		
Month	β_τ	β_σ	β_{HFOIV}	β_τ	β_σ	β_{HFOIV}
1	-1.03** (-2.51)	1.85*** (9.00)	0.31 (0.95)	-1.26*** (-9.18)	0.38** (2.16)	1.47*** (11.51)
2	-0.74*** (-3.91)	1.60*** (10.22)	0.19 (1.26)	-1.19*** (-7.06)	0.40** (2.29)	1.37*** (10.98)
3	-1.29*** (-3.57)	1.20*** (4.77)	1.11* (1.78)	-1.31*** (-7.86)	0.22 (0.94)	1.69*** (7.45)
4	-0.73*** (-4.03)	1.45*** (16.55)	0.14 (0.91)	-0.92*** (-9.73)	0.14 (0.69)	1.49*** (8.78)
5	-0.68*** (-3.45)	1.51*** (10.02)	0.15 (0.73)	-1.08*** (-8.62)	0.28** (2.36)	1.30*** (10.26)
6	-1.53*** (-2.72)	1.35*** (4.72)	1.30 (1.49)	-1.33*** (-4.45)	0.31 (1.53)	1.79*** (8.16)
7	-0.77*** (-3.89)	1.53*** (11.51)	0.13 (0.73)	-1.06*** (-7.68)	0.28* (1.91)	1.48*** (9.57)
8	-0.84*** (-3.86)	1.77*** (17.53)	0.22 (1.09)	-1.05*** (-8.98)	0.50*** (3.33)	1.29*** (10.42)
9	-1.24*** (-3.28)	1.31*** (7.14)	1.06 (1.59)	-1.19*** (-6.84)	0.06 (0.51)	1.82*** (6.74)
10	-0.71*** (-4.76)	1.64*** (12.93)	0.13 (0.77)	-0.92*** (-7.99)	0.53*** (3.88)	1.15*** (11.65)
11	0.17 (0.19)	1.35* (1.81)	-0.78** (-2.56)	-1.84** (-2.05)	0.87 (1.25)	2.26*** (6.00)
12	-1.11 (-1.20)	0.75 (0.82)	0.42 (0.91)	-2.21* (-1.90)	1.26 (1.22)	2.74** (2.49)

Table IA.25. Comparison of volatility measures.

The table reports median adjusted R-squared across years from the following panel regression with stock fixed effects: $\log s_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \beta_{\sigma(\mathrm{OI})} \log \sigma(\mathrm{OI})_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $s_{i,t}$ is the effective spread of stock i on date t, $\tau_{i,t}$ is the daily intraday turnover, and $\sigma(\mathrm{OI})_{i,t}$ is the volatility of order imbalance computed using five-minute order imbalances over the trading day. Volatility, $\sigma_{i,t}$, is measured as either the absolute daily return ($|r|$), the average absolute return over the previous week including the current day ($\overline{|r|}$), or the realized volatility computed using five-minute intraday midquote returns (RV_5). Controls are (log) lagged market capitalization, (log) lagged price, and day-of-the-week and month-of-the-year indicators. The regression is also estimated with daily changes in spread, volume, and volatility, where the change in a variable is defined by $\Delta x_t \equiv \log(\frac{x_t}{x_{t-1}})$. The regression is estimated for stocks in a given size quintile on a year-by-year basis. The sample consists of NYSE, Amex, and NASDAQ common stocks. To be included in a given month, a stock is required to have at the beginning of the month a price greater than \$5 and lower than \$1,000, a market capitalization greater than \$100 million, and at least 100 days of prior trading. Effective spreads are winsorized at 0.05% and 99.95% each year. Standard errors are double-clustered by date and stock.

	R^2(%) (median)					
	Small stocks			Large stocks		
	$\|r\|$	$\overline{\|r\|}$	RV_5	$\|r\|$	$\overline{\|r\|}$	RV_5
Level regressions	13.55%	14.88%	31.46%	19.26%	19.42%	26.19%
Change regressions	3.79%	3.60%	13.15%	13.67%	13.68%	19.02%