Liquidity, Volume, and 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

December 23, 2019

Abstract

We examine the relation between liquidity, volume, and volatility using a comprehensive sample of U.S. stocks in the post-decimalization period. For large stocks, effective spread and volume are positively related in the time series even after controlling for volatility, contrary to most theoretical predictions. This relation is mostly driven by the systematic component of volume. In contrast, for small stocks the evidence matches the predictions of standard adverse selection models. In line with a continuous-time inventory model, we show that the volatility of order imbalances can reconcile our puzzling finding with standard intuition. Order imbalance volatility is strongly associated with spreads both in the time series and cross-section. Evidence from alternative liquidity measures (price impact and depth), spread decomposition, and intraday patterns support our interpretation of order imbalance volatility as a measure of inventory risk. Furthermore, order imbalance volatility is priced in the cross-section of weekly returns.

1 Introduction

This paper examines the relation between stock liquidity, trading volume, and price volatility.

Microstructure models based on Kyle (1985) suggest that higher volume should be associated with lower trading costs, as volume is mainly driven by uninformed trading, which re-

*We thank our discussants: David Cimon, Itay Goldstein, Andrei Kirilenko, Dmitry Livdan, and Bart Yueshen. We also thank Pierluigi Balduzzi and conference and seminar participants at 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.

duces adverse selection risk.[1] There is considerable empirical evidence that trading volume is positively related to the stochastic stock price volatility.[2] However, the relation between price volatility and trading costs is theoretically ambiguous. Microstructure models of spreads based on adverse selection predict that price volatility and trading costs should be negatively related if price volatility is mostly driven by shocks to uninformed volume (Admati and Pfleiderer (1988), Collin-Dufresne and Fos (2016a)), but positively related if price volatility is mostly driven by shocks to information (Foster and Viswanathan (1990), Collin-Dufresne and Fos (2016b)). Inventory-based models on the other hand suggest that price volatility and trading costs should be positively related (Stoll (1978b)).

Thus, the relation between trading costs, volatility, and volume is ultimately an empirical question. Early 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.[3]

In this paper, we take a systematic look at the relation between trading costs as measured by daily effective spreads, volume as measured by daily turnover, and volatility measured using both daily absolute return and high-frequency realized volatility. Our sample covers U.S. stocks from 2002 to 2017. We focus on the time-series relation but find that most of our results hold in the cross-section as well. We find that daily effective spreads are negatively related to volume and positively related to volatility both in the cross-section and in the time series. This is consistent with the intuition from Kyle-type adverse selection models and in line with the literature cited above. However, when we sort stocks into quintiles based on market capitalization, we observe a different pattern for large stocks. Specifically, for large stocks effective spreads are 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.

[1] Though higher volume can be associated with higher trading costs if the increase in trading volume reflects an increase in the likelihood of informed trading (Easley and O'Hara (1992)).

[2] See, e.g., Clark (1973); Tauchen and Pitts (1983); Epps and Epps (1976); Gallant, Rossi, and Tauchen (1992); Andersen (1996).

[3] See also Foster and Viswanathan (1993) for an early empirical examination of variations in volume, volatility, and trading costs in a sample of stocks in 1988.

In the cross-section of large stocks, we find that spreads are mostly unrelated to volume, even after controlling for volatility.

The results suggest that factors other than adverse selection play an important role in driving effective spreads of large stocks. To gain more intuition, we decompose volume and volatility into common and idiosyncratic components.[4] Intuition suggests that adverse selection risk should be mostly driven by the idiosyncratic component of volatility. Similarly, the common component of volume is less likely to be driven by firm-specific information events. For small stocks, we find that 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, where idiosyncratic volume is mostly driven by noise trading and idiosyncratic volatility is a proxy for the amount of private information in the market.

For large stocks, we find that trading costs are also positively related to idiosyncratic volatility. However, they are positively related to both idiosyncratic and common components of volume. Moreover, it is the common component of volume that is economically more significant. Since it is unlikely that the common component of volume proxies the likelihood of an information event, which could have explained the positive volume-spread relation as shown by Easley and O'Hara (1992), these findings further suggest that trading costs for large stocks are likely driven by other factors than solely adverse selection.[5]

To better understand the findings, we develop a theoretical model of a risk-averse liquidity provider who faces stochastic arrival of buyers and sellers. In this model, a bid-ask spread arises as a result of inventory costs incurred by the liquidity provider as long as she waits for offsetting order flow, as in Grossman and Miller (1988). In the model, one can compute the equilibrium bid-ask spread charged by the liquidity provider in response to stochastically arriving buy and sell orders. The model shows that when buy and sell orders arrive at the same intensity, then increasing the

[4] In a study of commonality in liquidity, Chordia, Roll, and Subrahmanyam (2000) show that industry and market trading volumes affect individual stocks' spreads. They do not control for volatility in their time-series tests, however.

[5] Hendershott and Menkveld (2014) find economically large price pressures in a sample of NYSE stocks from 1994 to 2004. In a cross-sectional study, Bollen, Smith, and Whaley (2004) find that the adverse selection component of the spread is small. This contrasts with prior work such as Glosten and Harris (1988). We do not mean to imply that adverse selection is unimportant for large stocks. Intraday patterns in spread and volatility strongly suggest that asymmetric information is an important driver of spreads (see Section 5.2). More generally, the fact that trades have a permanent price impact supports the role of asymmetric information in the trading process (Hasbrouck (1991)).

intensity of both buyers and sellers both increases trading volume and reduces the inventory risk of the liquidity provider, who will more easily find an offsetting trade, thus generating a negative volume-spread relation. However, if buy and sell order intensities are asymmetric then changes in these intensities that leave average volume constant but increase the volatility of order imbalance lead to a higher equilibrium bid-ask spread. The intuition is that controlling for an average rate of trading, the higher the volatility of the order imbalance the greater the inventory risk faced by the liquidity provider. The model's insights apply equally well to the time-series than to the cross-sectional evidence we document.[6] From a microstructure point of view, it is natural to distinguish between volume and order imbalance (e.g., Chordia, Roll, and Subrahmanyam (2002)).

In the data, we find that when we introduce a measure of the volatility of high-frequency order imbalances over the trading day, we can reconcile the behavior of small and large stocks. Consistent with the model, once we control for order imbalance volatility, the relation between turnover and effective spread becomes strongly negative. Furthermore, order imbalance volatility increases the explanatory power of both levels and changes regressions by more than 10 percentage points on average. Interestingly, controlling for the volatility of order imbalances makes sensitivities of trading costs to volatility and volume similar in magnitude for large and small firms. 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. When we do not control for the volatility of order imbalances, large firms' volume elasticity of spread is in fact positive, which is opposite from the prediction of the invariance theory. We find similar results for the cross-sectional relation between spread, volume, and volatility. Order imbalance volatility substantially increases the fit of the regression across stocks in both small and large size quintiles.[7]

We consider how order imbalance volatility relates to alternative liquidity measures. Order imbalance volatility is positively related to an Amihud-type measure computed using intraday

[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.

[7] In a empirical study of market liquidity at the daily frequency, Chordia et al. (2002) find that absolute aggregate imbalance is negatively associated with spreads even when controlling for contemporaneous volume and absolute return. Our results are consistent with their findings. An important difference is that, in line with a simple inventory model, we focus on the volatility of order imbalance. Furthermore, 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.

data.[8] Furthermore, order imbalance volatility is negatively related to total depth at the best prices. Hence, more volatile imbalances are associated with lower liquidity as measured by spread, depth, and Amihud's measure. Moreover, in line with our interpretation based on inventory risk, a standard effective spread decomposition shows that order imbalance volatility is mostly associated with realized spread (liquidity provision) rather than price impact (adverse selection).

To complement our daily empirical results, we examine the intraday relation between volume, volatility, and spread. In line with the daily results, the intraday relation between volume and spread is generally positive and significant. Interestingly, spreads tend to be the most sensitive to volume around the close. We also document that absolute order imbalances tend to be highest in the last thirty minutes of trading, which has to the best of our knowledge not been documented before. This increase at the end of the trading day, a time when inventory considerations likely dominate, supports our interpretation based on inventory risk.

Finally, we show that order imbalance volatility predicts the cross-section of weekly returns.[9] This predictability holds for value-weighted returns even after controlling for many other liquidity variables. This evidence supports the idea that inventory risk is priced and is of interest since many high-frequency liquidity measures do not appear to be priced (Lou and Shu (2017)).

Our results hold when we use realized volatility computed from five-minute intraday midquote returns as our measure of volatility. Even though realized volatility improves substantially the explanatory power of the spread regressions, volume remains in general positively and significantly associated with spread in the time-series. We examine why realized volatility has a significantly higher explanatory power for the spread compared to cruder measures such as the daily absolute return. We show that realized volatility is linked to the profit of an intraday reversal strategy, which is naturally linked to the spread.

This paper is organized as follows. Section 2 reviews the determinants of spreads. Section 3 explores empirically the relation between spread, volume, and volatility. Section 4 develops a continuous-time inventory model to explain our findings. Section 5 examines order imbalance volatility. Section 6 concludes.

[8] By construction order imbalance volatility should be negatively related to price impact measures that employ order flow as a measure of trading activity. We discuss this point in Section 5.1.

[9] Chordia, Hu, Subrahmanyam, and Tong (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. Our measure differs from theirs along several dimensions and is not subsumed by it (see Section 5.3).

2 Determinants of Spreads

In this section, we discuss the determinants of spreads with a special focus on the role of volume and volatility. A more comprehensive review of market microstructure theories can be found in the survey of Biais, Glosten, and Spatt (2005).

2.1 Adverse Selection

Consider the classic continuous-time model of informed trading (Kyle (1985)). Informed trader and noise traders send order flow to a risk-neutral market maker. Denote noise trading volatility by σ_{noise} and the market maker's prior uncertainty about the informed's signal by σ_{informed}. The latter reflects how much information is going to be eventually revealed in the price. In this model, it can be shown that:

$$\begin{aligned}
&\text{price impact} = \frac{\sigma_{\text{informed}}}{\sigma_{\text{noise}}}, \\
&\text{price volatility} = \sigma_{\text{informed}}, \text{ and} \\
&\text{volume} = \sigma_{\text{noise}}.
\end{aligned}$$

Noise trading drives volume and informed trading drives volatility. All else equal, an increase in noise trading volatility results in a higher volume and a lower price impact. This is intuitive as noise trading reduces the market maker's adverse selection problem. For price impact to be positively associated with volume, $\frac{\partial\sigma_{\text{informed}}/\sigma_{\text{informed}}}{\partial\sigma_{\text{noise}}/\sigma_{\text{noise}}} > 1$. This condition is difficult to satisfy. To illustrate, in the Internet Appendix we solve 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.[10] 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. In all cases, 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

[10] The Internet Appendix is available at http://www.vincentbogousslavsky.com/liquidity_appendix.pdf

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 move price impact adversely.

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 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[11] (e.g, Admati and Pfleiderer (1988), Collin-Dufresne and Fos (2016a)), but can generate a negative relation between volume (or market depth) and volatility, if the higher volume is associated with a lower rate of news arrival[12] (Foster and Viswanathan (1990), Collin-Dufresne and Fos (2016b)). Lastly, we note that price impact can be positively linked to volume and volatility if there is a direct positive link between volume and information. For example, if the increase in trading volume reflects an increase in the likelihood of informed trading (Easley and O'Hara (1992)) or if the increase in trading volume comes with an increase in the rate of news arrival (Collin-Dufresne and Fos (2016b)) then volume, volatility and spreads may all be positively related.

A final remark on the definition of volume and order imbalance in these adverse selection models. As the net order flow comes from three groups of traders (informed, noise, and market maker), volume is naturally defined (e.g., Admati and Pfleiderer (1988)) as as one-half of the sum of the absolute value of each. In a continuous time model, where it is optimal for the informed to trade in an absolutely continuous fashion, the expected volume (per unit time) is proportional to the noise trading volatility (as the volume due to noise trading dwarfs the volume due to informed traders), which is also equal to the volatility of the cumulative net order flow submitted to the market maker.[13]

Instead, below we will propose an inventory model where it is natural to distinguish the two

[11] This is because a higher noise trading volatility increases the average volume and leads to more aggressive informed trading which increases price volatility.

[12] A higher volume pushes the insider to trade more aggressively, but a lower rate of news arrival reduces her incentives to trade aggressively. The latter effect can dominate and lead to a decrease in price volatility.

[13] Volume is defined as $VOL = \frac{1}{2}(|dX_t^i|+|dX_t^u|+|dX_t^i+dX_t^u|)$. Now, in continuous time, it is optimal for the insider to trade in absolutely continuous fashion, that is $dX_t^i = \mu_i dt$, whereas $dX_t^u = \sigma_u dZ_t$ for some Brownian motion Z_t. It follows that (dropping terms of order higher than dt and recalling that if $dX \sim N(0, \sigma^2)$ then $E[|dX|] = \sqrt{2/\pi}\sigma$) squared expected volume (per unit time) is proportional to the variance of noise trading, that is $E[VOL]^2 = 2/\pi\sigma_u^2 dt$. Total cumulative order flow is $Y_t = X_t^u + X_t^i$ and again $Var[dY_t] = \sigma_u^2 dt$.

notions and indeed, they have different implications for spreads.

2.2 Inventory Risk

Liquidity providers face inventory risk. This inventory risk is lower when it is easier for them to find an offsetting trade. Hence, as long as volume is not one-sided, a higher volume is generally associated with improved liquidity in inventory models.

In contrast, risk-averse liquidity providers require a compensation to absorb one-sided supply shocks (Grossman and Miller (1988)). Consider a model along the lines of Campbell, Grossman, and Wang (1993) in which liquidity providers with exponential utility and risk aversion γ absorb every period liquidity shocks with volatility σ_{noise}. In this model, it can be shown that:

$$\begin{aligned}
&\text{price impact} \propto \gamma\sigma_{\text{ret}}^2,\\
&\text{volatility} \equiv \sigma_{\text{ret}}^2, \text{ and}\\
&\text{volume} \propto \sigma_{\text{noise}}^2.
\end{aligned}$$

Since noise trading moves prices, then $\frac{\partial}{\partial\sigma_{\text{noise}}^2}\sigma_{\text{ret}}^2 > 0$. As a result, volume and price impact are positively related. This positive relation depends crucially on holding fixed the number of liquidity providers. In fact, it can be shown that if entry of liquidity providers (at a fixed cost) is allowed then price impact decreases with noise trading volatility. The intuition being that a higher volatility of noise trading increases the profits of incumbent liquidity providers, which attracts additional liquidity providers and ultimately lowers price impact.[14] As in the case of adverse selection, there is a direct link between return volatility and price impact.

2.3 Competition

If liquidity provision is not perfectly competitive, then imperfect competition can affect spreads in adverse selection and inventory models. But even absent inventory concerns and adverse selection, the lack of competition can matter for liquidity. For instance, in the model of Foucault, Kadan, and Kandel (2005) there is no inventory risk and no asymmetric information. An increase in the total order

[14] In equilibrium, the fraction of noise trading volatility that each market maker has to absorb remains constant. The result is that the price impact per unit of noise trading volatility shock goes down while the product of price impact and noise trading volatility remains constant. The model is detailed in the Internet Appendix.

arrival rate lowers spread since traders wait on average a smaller amount of time before their limit orders are executed, in line with the intuition of Demsetz (1968). An increase in the fraction of impatient traders can, however, increase both volume and spreads due to the strategic behavior of patient traders and imperfect competition.

3 An Empirical Exploration of the Relation between Spreads, Volume, and Volatility

We examine the time-series relation between spreads, volume, and volatility. Our focus is on the post-decimalization period. We first 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. We compute daily and intraday 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 have a single valid TAQ trade in a given month are excluded. The liquidity measures (described below) are computed over the regular trading day (9:30am to 4:00pm). Days with early closures are excluded from the analysis (the NYSE closes at 1pm on the day before Independence day, the day after Thanksgiving, and Christmas Eve).

[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. Monthly TAQ is used before 2014 and Daily TAQ is used since 2014.

3.2 Variables and Descriptive Statistics

We use the percentage effective spread as our primary measure of liquidity. The percentage effective spread on a trade t 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. Importantly, our results hold if we use as dependent variable the *dollar* effective spread, computed by dollar-weighting or share-weighting $2|P_{i,t} - M_{i,t}|$ over the day. This shows that price effects do not drive our results.[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 effective spreads. Our results are unchanged if we use instead total turnover obtained from CRSP. We use the average absolute return over the past five trading days (including the current day) as a measure of volatility (Chordia et al. (2001)). Alternative measures of volatility are discussed in Section 3.6.

Figure 1 plots the daily cross-sectional median of each measure 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. Turnover increases until the crisis then drops and remains relatively stable. Volatility does not show any marked trend over the sample period.

In the analysis that follows we split stocks into groups based on market capitalization. We show that it is important to consider separately large stocks from 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. We require a stock to have a minimum of 100 observations. The results are similar when we sort stocks based on dollar volume instead of market capitalization.

On average each quintile contains 540 stocks, with a minimum of 456 and a maximum of 634. The median market capitalization of a stock in the lowest (highest) size quintile is \$0.17 (\$7.09)

[16]Effective spreads may be a biased measure of transaction costs due to the binding tick size (Hagströmer (2019)). Our main results are robust to focusing on large stocks with a price above \$100 (for which Hagströmer (2019) does not find evidence of bias).

billion in 2002 and grows to \$0.23 (\$17.81) billion in 2017. The median daily dollar volume of a stock in the lowest (highest) size quintile is \$0.28 (\$41.34) million in 2002 and grows to \$0.61 (\$111.68) million in 2017. Table 1 reports descriptive statistics for our main variables of interest—percentage effective spread, turnover, and volatility—for stocks in the bottom, middle, and top size quintiles in even years (to save space). The values confirm the evidence in Figure 1.

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.

Table 2 reports cross-sectional averages of the individual stocks' time-series correlations for the different variables.[17] As expected, spread and volatility are positively correlated for both small and large stocks. More surprising, spread is positively correlated with turnover for large stocks. We show below that this relation is not explained by volatility.

3.3 Spread, Volume, and Volatility in the Time Series

Building on the theories in Section 2, our goal is to evaluate how liquidity varies with volume and volatility for a large cross-section of stocks. We examine what drives spreads by estimating the following panel regressions:

$$\log s_{i,t} = \alpha_i + \beta_\tau^u \log \tau_{i,t} + \text{controls} + \epsilon_{i,t}, \quad \text{and} \tag{1}$$

$$\log s_{i,t} = \alpha_i + \beta_\sigma^u \log \sigma_{i,t} + \text{controls} + \epsilon_{i,t}, \tag{2}$$

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). We also control for market capitalization and price (in logs). The results are, however, similar if we do not include these controls. Furthermore, as mentioned above the results hold if we use dollar

[17]Correlations between log variables are reported since log transformed variables are used in the analysis. The correlations between raw variables are not substantially different. Additional measures of volatility and order imbalance are introduced later in the analysis.

effective spreads instead of percent effective spreads.

Even though we focus on the post-decimalization period, Figure 1 shows that spread and turnover still exhibit trends over parts of the sample period. To avoid as much as possible issues associated 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 the variables stationary but should alleviate the issue and be informative about what happens in the data over time. Furthermore, it is not clear in Figure 1 that spread and turnover exhibit any trend over the second part of the sample. Second, we estimate as robustness checks regressions with percentage changes in the variables and vector autoregressions (discussed below).

Figure 2 reports the results of estimating (1) and (2) on a month-by-month basis separately for stocks in the bottom and top size quintiles as explained above. Results for the other size quintiles lie in-between these two extremes. Surprisingly, Panel (a) shows that a higher volume is associated with a higher spread among the quintile of large stocks. The reverse holds among small stocks. Mid-cap stocks lie in-between (not reported). In line with theory, a higher volatility is associated with a higher spread (Panel (b)).

Spread, volume, and volatility are endogenous quantities. It is well-known that volatility and volume are strongly associated. Panel (c) of Figure 2 confirms the strong positive association between volume and volatility. We next estimate the following multivariate regression:

$$\log s_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \text{controls} + \epsilon_{i,t}. \tag{3}$$

Equation (3) can also 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.

Figure 3 reports the month-by-month estimated elasticities across size quintiles. In the Internet Appendix, we report similar results when estimating (3) year-by-year. Spread is positively related to turnover for large stocks (Figure 2) and controlling for volatility does not explain this positive relation (Figure 3). Economically, a one (within) standard deviation increase in turnover from its mean level leads to a roughly 10% increase in spread for large stocks. For small stocks, the spread decreases by more than 15%. The average monthly adjusted R^2 is 7.0% for large stocks and 3.7% for small stocks and peaks during the Financial crisis. 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 (reported in the Internet Appendix). But this relation breaks down and turns positive once we focus on large stocks.[18]

As a first robustness check, we use percentage changes in the variables like Chordia et al. (2000). 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:

$$\Delta s_{i,t} = \alpha_i + \beta_\tau \Delta\tau_{i,t} + \beta_\sigma \Delta\sigma_{i,t} + \text{controls} + u_{i,t}, \tag{4}$$

where $\Delta x_t \equiv \log(\frac{x_t}{x_{t-1}})$, and the controls are daily market capitalization, daily price, and day-of-the-week and month-of-the-year indicators.[19]

The previous results are robust when we estimate equation (4). Among large stocks changes in spreads are positively associated with changes in volume and among small stocks this relation is in general negative (reported in the Internet Appendix). As another robustness check, we estimate for each stock a time-series regression of spread on turnover and volatility each year and then examine the distribution of volume and volatility betas across stocks in a given size quintile (we use the size quintile allocation at the beginning of each year). The results are reported in the Internet Appendix and confirm the panel evidence. For large stocks, the median sensitivity of spread to turnover is positive while it is negative for small stocks. Over 2002 to 2012, more than two-thirds

[18]Most of the evidence for a 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. We estimate (3) with day fixed effects to study cross-sectional variation with the same controls as before except for the calendar indicators. The cross-sectional relation between volume and spread is strongly negative for stocks in the bottom and middle size quintiles. Among large stocks, the relation is mostly insignificant. Among stocks in the top market capitalization decile, the cross-sectional relation between turnover and spread is often positive and statistically significant. These results are reported in the Internet Appendix and suggest that the cross-sectional relation between spread, volume, and volatility in the post-decimalization period may be more complicated than previously thought.

[19]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.

of the volume betas are positive and statistically significant at the level of 10%. In recent years, this proportion is closer to 40%, but only few large stocks have negative and significant turnover betas.

Reverse causality is a concern in (3). Our specification builds on microstructure theories that suggest that volume and volatility are likely to have exogenous drivers, while spreads are mostly endogenous. Furthermore, for large stocks reverse causality cannot explain the empirical result since it seems implausible that an increase in spread could cause an increase in volume. To provide another perspective on the previous findings, we estimate vector autoregressions (VAR) of spread, volume, and volatility. More precisely, we estimate a reduced-form VAR 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. At the daily level, market microstructure theory suggests that volatility and volume drive spreads rather than the opposite. Chordia, Sarkar, and Subrahmanyam (2005) make a similar point when estimating VAR of aggregate stock and bond liquidity.

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. Since we are interested in comparing the results across stocks, all the variables are normalized. The VAR is estimated separately for each stock. 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. Interestingly, 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 4 reports the cross-sectional median and 5^{th} and 95^{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 spread to a turnover 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 spread.

3.4 Decomposing Volume and Volatility

To gain more intuition about the relation between spread, volume, and volatility, we decompose volume and volatility into common and idiosyncratic components. We are interested in understanding whether the common and idiosyncratic components of volume and volatility affect liquidity in different ways to shed light on the theories of Section 2.

We expect asymmetric information to play a more important role for liquidity via the idiosyncratic component than the common component of volume and volatility. It is unlikely that the common component of turnover reflects the likelihood of an information event in a specific stock. Thus if we find a positive relation between the common component of volume and spreads then it seems difficult to ascribe this to an adverse selection theory of spreads. Instead, idiosyncratic volume could be driven by firm-specific information events that trigger more informed trading and thus could cause a positive relation with spreads as shown in Easley and O'Hara (1992). Alternatively, if idiosyncratic volume was mostly driven by noise trading, then we would expect a negative relation with spreads as in Kyle (1985).

Similarly, we would 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.[20]

The role of idiosyncratic and systematic volume and volatility shocks in inventory theories is more difficult to evaluate. The existence of actively-traded basket securities should make systematic volume shocks easier to hedge than idiosyncratic volume shocks for individual liquidity providers. Further, if liquidity provider do not hold well-diversified portfolios, perhaps because they specialize in making markets on a limited number of securities, then idiosyncratic risk should be the primary driver of inventory cost.[21] At the same time, a systematic volume shocks consumes liquidity

[20] That said, adverse selection could affect common volume and volatility components if there were asymmetric information at the industry or market level, or if agents have different abilities to interpret public signals, which might generate information asymmetry related to systematic risk shocks (Kim and Verrecchia (1994)).

[21] Moreover, even if a liquidity provider holds a diversified portfolio of securities, idiosyncratic risk still limits her ability to arbitrage away any short-term price deviation (Pontiff (2006)).

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 discuss the impact of volume shocks on spreads in inventory theories in Section 4 using a dynamic inventory model.

We decompose turnover into common and idiosyncratic components as follows. For each stock i, we regress daily (log) turnover on a common turnover measure:

$$\log \tau_{i,t} = a_i + b_i \tau_{m,t} + \tau_{i,t}^I, \tag{5}$$

where the common turnover, $\tau_{m,t}$, 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, $\tau_{i,t}^I$, and the common component of turnover by the fitted value (i.e., $\tau_{i,t}^C \equiv \log \tau_{i,t} - \tau_{i,t}^I$). For simplicity, we estimate (5) for each stock using the full sample of data. The results are almost identical when we estimate the components on a year-by-year basis. The results are also qualitatively similar if we decompose raw turnover instead of log turnover. We decompose volatility into common and idiosyncratic components similarly. For each stock i, we compute 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, $\sigma_{i,t}^C$ ($\sigma_{i,t}^I$), 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.

Using the decomposed measures, we estimate the following regression:

$$\log s_{i,t} = \alpha_i + \beta_{\tau,C} \tau_{i,t}^C + \beta_{\tau,I} \tau_{i,t}^I + \beta_{\sigma,C} \sigma_{i,t}^C + \beta_{\sigma,I} \sigma_{i,t}^I + \text{controls} + \epsilon_{i,t}. \tag{6}$$

Figure 5 reports the month-by-month estimated coefficients and associated t-statistics for the bottom and top size quintiles. The year-by-year estimation results are reported in the Internet Appendix. The common volume elasticity of spread is positive and significant for large stocks over the sample. The downward trend in the volume elasticity observed in Figure 3 is not reflected in the pattern of the common component. In contrast, for small stocks the common volume elasticity tends to be negative or insignificant except in the aftermath of the Financial Crisis where it is positive and significant. The idiosyncratic volume elasticity of spread is positive for large stocks and negative for small stocks. For large stocks, the elasticity tends to trend downwards over the sample period. With respect to volatility, the common component displays a very noisy pattern. It is in general positive but not statistically insignificant. The idiosyncratic volatility elasticity is positive and strongly significant for small stocks, consistent with asymmetric information theories. The idiosyncratic volatility elasticity tends to be positive but is not consistently significant over the sample period. The idiosyncratic component of volatility appears to be more important for small stocks than for large stocks.[22]

Overall, the results suggest that competition and inventory effects are important as drivers of spreads for large stocks since common and idiosyncratic volume elasticities are large and positive. For small stocks, the evidence supports adverse selection as the primary driver of spreads. Idiosyncratic volatility elasticity is large and positive while common volatility elasticity is in general insignificant. Furthermore, idiosyncratic volume elasticity is negative. The standard adverse selection intuition works well for small stocks if we interpret idiosyncratic volume as mostly driven by noise trading, but not that well for large stocks.[23] As discussed in Section 5.2, we do not mean to imply that adverse selection does not matter for large stocks, only that competition and inventory effects play an important role for daily liquidity fluctuations.

Results for the specification based on changes in the variables instead of levels as in (6) are reported in the Internet Appendix. The results are similar except that, for small stocks, changes in the common component of turnover tend to be positively associated with changes in spreads. Moreover, this elasticity spikes during the Financial crisis. This result suggests that limited market making capital can also be important for small stocks.

3.5 Bias in Effective Spread

Hagströmer (2019) points out that due to the minimum tick size, effective spreads of large stocks

[22] Evidence from individual stock time-series regression is in line with these results and reported in the Internet Appendix. Among large stocks, common turnover betas are overwhelmingly positive. On average more than 90% of the betas are positive and a large proportion is statistically significant. Idiosyncratic turnover betas also tend to be positive and significant. Consistent with Figure 5, the proportion of idiosyncratic turnover betas that are significantly negative increases notably after 2013. Among small stocks, common turnover betas are in general negative and more likely to be significantly negative than significantly positive over the first part of the sample. This relation switches after 2007. Idiosyncratic turnover betas are reliably negative and significant.

[23] We note that 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.

can be biased. For the bias to affect our results, it should be positively correlated with volume within a given stock. Hagströmer (2019) shows that the bias is negligible for stocks with a price above \$100. In the Internet Appendix, we show that the previous results are robust to focusing on large stocks with prices above \$100 and \$120. The results tend to be weaker in the first years of the sample, but we only have a small sample of stocks that pass the price filter in these years.

3.6 Measuring Volatility: Realized Volatility

In this section, we examine the sensitivity of the previous results to the choice of the volatility measure. The results are similar if we use the contemporaneous daily absolute return or the daily absolute intraday return instead of the average absolute return over the past five trading days including the current day. We focus on a more sophisticated and arguably more precise measure of volatility: realized volatility. We argue, however, that realized volatility is correlated with the profit of an intraday reversal strategy, which makes the interpretation of the volatility elasticities more complicated.

We compute five-minute realized volatility using intraday midquote returns.[24] We then estimate (3) and (4) with realized volatility as measure of volatility. Table 3 reports the results for large stocks. The results for small stocks are in line with the previous results and reported in the Internet Appendix. Realized volatility substantially improves the fit of the regression and lowers the importance of turnover. Turnover remains, however, positive and statistically significant except in the last years of the sample for both level and change specifications. The results from the month-by-month estimation are similar (not reported). The impact of turnover is lowered by the inclusion of five-minute realized volatility but remains positive and in general statistically significant. In particular, 150 out of 192 common turnover monthly elasticities are positive, among which 62 significantly so. We also estimate (6) using realized volatility and report the results in the Internet Appendix. The elasticity of common turnover remains large and significant in most years of the sample.[25]

To better understand the difference between realized volatility and intraday absolute return,

[24] To minimize the influence of noisy opening quotes (e.g., Bogousslavsky (2019)), we take the volume-weighted average price over the first five minutes of trading as our opening price.

[25] Realized volatility can be decomposed into components using the methodology described in Patton and Verardo (2012). Since the decomposition does not significantly affect the results, we focus on the estimation with raw realized volatility for simplicity.

we examine what drives the volume coefficient. Consider Equation (3) and subtract the deviation from the time-mean to remove the stock fixed effects. By Frisch-Waugh's theorem, the bi-variate regression of (demeaned) effective spread on the residuals of (demeaned) turnover and volatility relative to the control variables yields the same coefficients β_τ and β_σ as in (3). The turnover coefficient in this (bi-variate) regression is given by

$$\beta_\tau = \sqrt{\frac{\mathrm{Var}[s]}{\mathrm{Var}[\tau](1-\rho_{\tau,\sigma}^2)}}(\rho_{s,\tau} - \rho_{\tau,\sigma}\rho_{s,\sigma}), \tag{7}$$

where ρ indicates the correlation coefficient between two variables and the variables are transformed as explained above. Equation (7) shows that the measure of volatility (σ) used in (3) affects the turnover elasticity through $\rho_{\tau,\sigma}$ and $\rho_{s,\sigma}$. Empirically, we find that $\rho_{\tau,\mathrm{RVol}} \approx 2\rho_{\tau,|r|}$ and $\rho_{s,\mathrm{RVol}} \approx 3\rho_{s,|r|}$. From (7), we see that $\rho_{s,\sigma}$ unambiguously decreases β_τ. To interpret the empirical evidence it is therefore important to understand how realized volatility affects $\rho_{s,\sigma}$.

Let $r_{t,k}$ denote the return on day t in intraday interval k. The K-interval realized variance is defined by $\sum_{k=1}^{K} r_{t,k}^2$. Using log returns for simplicity, it follows that

$$\mathrm{RVol}(K)_t^2 = r_t^2 + \Pi(K)_t, \tag{8}$$

where $\Pi(K)_t = \sum_{k=2}^{K}(-2\sum_{j=1}^{k-1} r_{t,j})r_{t,k}$. Hence, Π_t is the daily return of an intraday reversal strategy with weight $-2\sum_{j=1}^{k-1} r_{t,j}$ in the asset and weight $1+2\sum_{j=1}^{k-1} r_{t,j}$ in the risk-free security.[26] Equation (8) shows that if $\rho_{s,\Pi} > 0$, then $\rho_{s,\mathrm{RVol}^2} > \rho_{s,r^2}$. In words, if the daily effective spread is positively correlated with the daily return of an intraday reversal strategy defined as above, then all else equal we expect the inclusion of realized volatility to lower the turnover elasticity of the spread.

If realized volatility is computed using transaction prices, then a positive correlation is mechanical from the bid-ask bounce. Even with midquotes, however, one should expect the spread to be positively correlated with the intraday reversal strategy. For instance, a large buy trade consumes all the displayed liquidity at the best ask. Since the best bid is unchanged, the midquote increases. If part of the price impact is temporary (i.e., there is reversal in intraday midquote returns), then

[26] Since intraday trades may not get the risk-free rate due to the end-of-day settlement on transactions, we set the risk-free rate to zero.

a positive correlation between intraday reversal profit and average intraday spread follows. For example, in Grossman and Miller (1988) there is a one-to-one relation between return autocorrelation and price impact.

In what follows, we use realized volatility instead of the average absolute return as our measure of volatility. The above caveat should be kept in mind.

4 Continuous-Time Model with Stochastic Arrival of Buyers and Sellers

In this section, we develop a simple inventory model to shed light on the above results.

We consider a long-lived arbitrageur with constant absolute risk-aversion utility $u(c,t) = -e^{-\beta t - \alpha c_t}$ who maximizes his expected utility of intertemporal consumption by by trading continuously a stock S_t that pays a continuous dividend δ_t and whose price depends on the realization of a continuous time Markov Chain which takes on M discrete values $N_t = \{1, 2, \ldots, M\}$. The arbitrageur can also invest in a constant risk-free rate we set to zero for simplicity. The price and dividend dynamics are:

$$dS_t + \delta_t dt = \mu_t dt + \sigma_t dZ_t + \sum_{i=1}^{M} \mathbf{1}_{\{N_{t^-}=i\}} \sum_{j \neq i} \eta_{ij}(dN_{ij}(t) - \lambda_{ij} dt) \quad (9)$$

$$d\delta_t = \kappa_\delta(\overline{\delta}(N_t) - \delta_t)dt + \sigma_\delta dZ(t) \quad (10)$$

$$dN_t = \sum_{i=1}^{M} \mathbf{1}_{\{N_{t^-}=i\}} \sum_{j \neq i} (j-i)(dN_{ij}(t) - \lambda_{ij} dt) \quad (11)$$

where $N_{ij}(t)$ are point processes with transition intensities λ_{ij}. The states characterize the expected long-term fundamental value of the asset $\overline{\delta}(N_t) := \sum_{i=1}^{M} \overline{\delta}_i \mathbf{1}_{\{N_t=i\}}$ and the *total supply* or inventory that the market maker must hold in equilibrium $\theta(N_t) := \sum_{i=1}^{M} \theta_i \mathbf{1}_{\{N_t=i\}}$. We will assume 'adverse selection' in the sense that $\overline{\delta}(N)$ is inversely related to $\theta(N)$, that is when the liquidity provider must absorb a larger supply of shares, the fundamental value is lower. With only two states, inventory and fundamental are perfectly negatively correlated. However, with more than two states, we can capture richer patterns of adverse selection risk in this model. We first present the model in full generality and then investigate a few special cases to gain some insights.

The risk-averse liquidity provider maximizes

$$\max_{c_t, n_t} \mathrm{E}[\int_0^\infty -e^{-\beta t - \alpha c_t}] \tag{12}$$

subject to

$$dW_t = (rW_t - c_t)dt + n_t(\mu_t - rS)dt + n_t\sigma_t dZ_t + n_t \sum_{i=0}^{M} \mathbf{1}_{\{N_{t^-}=i\}} \sum_{j\neq i} \eta_{ij}(dN_{ij}(t) - \lambda_{ij}dt). \tag{13}$$

The model is solved in Appendix A. We now investigate a few special cases which shed some light on the role of order imbalance and market liquidity.

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. That we consider the simple model with two states $M = 2$ and

$$\lambda_{12} = \lambda_{21} = \lambda \tag{14}$$

$$\theta_2 = -\theta_1 = \theta \tag{15}$$

$$\overline{\delta}_1 = \overline{\delta}_2 = 0 \tag{16}$$

Note that the change in price when switching from state 1 to state 2 is given by $\eta_{12} = s_2 - s_1$. (This is the case since the long-run mean is constant and equal to zero across states.) We can prove that there exists a unique symmetric solution characterized by $s_1 = -s_2$. In the appendix, we further show that

$$-\theta_1\alpha\sigma^2 > s_1 > \frac{-\theta_1\alpha r\sigma^2}{2\lambda + r} > 0.$$

We see that in equilibrium there is a bid-ask spread in the sense that to absorb the supply, there is a jump in the price the liquidity provider requires to absorb the order flow. We can define the "price impact" by the sensitivity of price to order flow in absolute value that is $PI = \frac{|\eta_{ij}|}{|\theta_j - \theta_i|}$.

In the symmetric case without adverse selection, we find:

$$PI = \frac{|s_1|}{|\theta_1|}. \tag{17}$$

And in particular, we have:

$$\alpha\sigma^2 > PI > \frac{\alpha r\sigma^2}{2\lambda + r}. \tag{18}$$

Clearly PI has an upper bound that is tight when volume is lowest and the frequency of trading is smallest (i.e., $\lambda = 0$). As we would expect PI is lowest when trading intensity (λ) increases. In fact, in this example the liquidity provider faces zero risk, when trading intensity is infinite and he can constantly intermediate between buyers and sellers.

However, this symmetric case, does not allow to talk about order imbalance. To understand the effect of order imbalance we solve the case where $\lambda_{12} \neq \lambda_{21}$, but otherwise keep the same assumption (of symmetric depth and no adverse selection). That is, we consider the simple model with two states $M = 2$ and

$$\theta_2 = -\theta_1 = \theta \tag{19}$$

$$\bar{\delta}_1 = \bar{\delta}_2 = 0. \tag{20}$$

In the appendix, we show that solving for the equilibrium amounts to solving two non-linear equations for s_1, s_2. We solve this system numerically. We can then investigate how the price impact

$$PI = \frac{|s_1 - s_2|}{|\theta_1 - \theta_2|} \tag{21}$$

changes with the expected volume and order imbalance.

The expected volume is

$$VOL = |\theta_1 - \theta_2| E[|dN_t|]/dt = |\theta_1 - \theta_2| \mathrm{E}[\lambda_1 N_t + \lambda_2(1 - N_t)] \tag{22}$$

$$= |\theta_1 - \theta_2| \left(\frac{\frac{1}{\lambda_2} + \frac{1}{\lambda_1}}{2} \right)^{-1} \tag{23}$$

Note that the expected volume is the harmonic average intensity times the average trade size. We define the order imbalance as the unconditional variance of the cumulative order flow:

$$OI = p_2\theta_2^2 + p_1\theta_1^2 - (p_2\theta_2 + p_1\theta_1)^2 = \frac{(\theta_1 - \theta_2)^2 \lambda_2 \lambda_1}{(\lambda_2 + \lambda_1)^2} \tag{24}$$

where $p_i = \frac{\lambda_{ji}}{\lambda_{ij}+\lambda_{ji}}$ is the unconditional probability that the Markov chain is in state i.

Note that (for a given θ_2, θ_1) the variance of the order imbalance is maximized for $\lambda_2 = \lambda_1$ (since $\frac{\lambda_2\lambda_1}{(\lambda_2+\lambda_1)^2} \le \frac{1}{4}$ and attains the upper bound for $\lambda_2 = \lambda_1$). Thus, suppose we hold VOL constant equal to $2c|\theta_1 - \theta_2|$ then this implies that $\frac{1}{\lambda_2} + \frac{1}{\lambda_1} = \frac{1}{c}$. In other words, the set of feasible intensities belong to the open interval (c, ∞). Then $OI = |\theta_1 - \theta_2|^2 c^2 (\frac{1}{c} - \frac{1}{\lambda_1})\frac{1}{\lambda_1}$ for $\frac{1}{\lambda_1} \in (0, \frac{1}{c})$. Note that $(\frac{1}{c} - \frac{1}{\lambda_1})\frac{1}{\lambda_1}$ is a concave fonction of $1/\lambda_1$ which achieves its maximum of $\frac{1}{4c^2}$ at $\frac{1}{\lambda_1} = \frac{1}{2c}$, where $OI = \frac{|\theta_1-\theta_2|^2}{4}$. So as we vary $\frac{1}{\lambda_1}$ from 0 to $1/(2c)$ holding VOL constant then OI ranges from 0 to $\frac{|\theta_1-\theta_2|^2}{4}$.

In Figure 6, we plot equilibrium prices and show some comparative statics where we vary trading intensities and risk-aversion. We find that the bid-ask spread $s_1 - s_2$ decreases when we increase volume symmetrically, i.e., when $\lambda_2 = \lambda_1 = \lambda$ increases. The bid-ask spread increases when risk-aversion α increases. The bid-ask spread increases when we increase OI, but holding VOL constant.

In Figure 7 we plot the bid-ask spread $(s_1 - s_2)$ surface as a function of expected volume (VOL) and the variance of order imbalance (OI), which we obtain by varying λ_{12} and λ_{21} and normalizing $|\theta_1 - \theta_2| = 1$. The graph clearly shows that increasing volume holding order imbalance constant decreases spreads, but instead increasing order imbalance holding volume constant increases spreads. We also see that the model can generate the empirical finding that increasing volume (starting form the origin and moving diagonally towards the 'north-east') can lead to an increase in spread, if one does not control for the change in order imbalance.

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. Thus volume does not have an unambiguous effect on spreads unless one controls for the variance of order imbalance. Clearly, both effects are also tied to the risk-bearing capacity of the liquidity provider. More risk-aversion increases the impact on spreads of a change in the variance of order imbalance.

5 Volatility of Order Imbalance

In the previous section, we propose a simple inventory model. In the model, a higher volume reduces liquidity provider's inventory risk since it makes it easier to offset trades. However, holding volume constant, a more volatile order flow increases the inventory risk of the market maker, leading to higher spreads. Hence, volume is negatively related to spread, and the volatility of order flow is positively related to spread. 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. In contrast, Kim and Stoll (2014) argue that order imbalance is not indicative of private information. In the simplest version of our model, there is no asymmetric information and order imbalance volatility is positively related to spreads after we control for volume.

We now show empirically that taking into account the volatility of order flow can explain the puzzling positive spread-volume sensitivity. We compute order imbalance (as a proportion of shares outstanding) over every five-minute interval of the trading day using the Lee and Ready (1991) algorithm. The daily volatility of order imbalance is the standard deviation of the five-minute imbalance, computed over the trading day.[27] We then update (3) to include the daily volatility of order imbalance $\sigma(\text{OI})_{i,t}$:

$$\log s_{i,t} = \alpha_t + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \beta_{\sigma(\text{OI})} \log \sigma(\text{OI})_{i,t} + \text{controls} + \epsilon_{i,t}. \tag{25}$$

Similarly, we include the change in the volatility of order imbalance in (4).

Table 4 reports the estimation results for the quintile of large stocks. First, the inclusion of order imbalance volatility dramatically improves the explanatory power of both the level and change regressions. Order imbalance volatility is strongly associated with effective spreads at the daily level. Second, the inclusion of order imbalance volatility makes the volume elasticity of spread negative and significant, consistent with the idea that a higher volume is beneficial for liquidity. The results for small stocks are similar and not reported. Order imbalance volatility makes the role of volume consistent across small and large stocks.[28] The right plots in Figure 4 confirm these

[27] The choice of the aggregation interval is dictated by practical considerations. We use a five-minute interval to be consistent with our estimate of realized volatility. When estimating volatility, a five-minute interval has been found to provide a good balance between achieving greater precision from a higher sampling frequency and introducing microstructure noise.

[28] Relative to the decomposition results in Section 3.4, we find that order imbalance volatility makes both common

results. When order imbalance volatility is included in the VAR, a volume shock lowers spread for most stocks. The contrast is striking relative to the impulse response of a volume shock in the baseline model (left plots). In contrast, an order imbalance volatility shock increases spread consistently across stocks.

Importantly, we find that the absolute value of the daily order imbalance does not explain the positive volume-spread sensitivity. The volume result is robust to including the absolute order flow in the regression for large stocks (reported in the Internet Appendix). Moreover, the explanatory power of the regression is not substantially increased. This result suggests that, in the era of algorithmic trading, daily imbalance measures are not enough to capture the dynamics of liquidity.

Interestingly, the inclusion of order imbalance volatility does not reduce the volatility elasticity of spread. This can be seen by comparing the elasticities in Tables 3 and 4. The higher volatility elasticity and the lower volume elasticity are much 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 (25) 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. Clearly, invariance of transaction costs does not explicitly incorporate order imbalance volatility. Nevertheless, we view this evidence as encouraging and suggesting interesting opportunities for future research.

5.1 Alternative Liquidity Measures

In this section, we examine how alternative daily liquidity measures relate to order imbalance volatility. First, we look at price impact. Second, we look at depth. In summary, we find that order imbalance volatility is negatively associated with alternative measures of liquidity, in line with our spread results.

We consider two standard measures of price impact. First, we estimate for each stock-day: $r_{itk} = \delta_{it} + \lambda_{it}\sqrt{|\mathrm{OI}_{itk}^{\$}|}\mathrm{sign}(\mathrm{OI}_{itk}^{\$}) + e_{it}$, where r_{itk} is the five-minute midquote return for stock i on day t in interval k, and $\mathrm{OI}_{itk}^{\$}$ is the dollar order imbalance. A similar measure is used in Hasbrouck (2009). Second, we compute a measure of price impact based on Amihud (2002). We compute illiquidity

and idiosyncratic volume components consistently negative and significantly so.

for each stock-day using intraday five-minute midquote returns and dollar volume: $\text{ILLIQ}_{it} = \frac{1}{\#\text{traded intervals}} \sum_{k \epsilon \{j | \text{DVOL}_{itj} > 0\}} \frac{|r_{itk}|}{\text{DVOL}_{itk}}$.

Table 5 reports the results of estimating (25) every year with the two price impact measures as dependent variables. We focus on stocks in the top size quintile but obtain similar results with stocks in other size groups.[29] The first measure, λ, is negatively related to volume, positively related to volatility, and *negatively* related to order imbalance volatility (Panel (a)). This is inconsistent with the spread results. In contrast, ILLIQ is *positively* related to order imbalance volatility (Panel (b)), in line with the spread results.

What explains this discrepancy? It is not surprising that λ is negatively related to order imbalance volatility. If we assume that order imbalance is symmetric and equally likely to be positive or negative, then $\lambda = \frac{\sigma_r}{\text{E}[|\text{OI}|]} \text{corr}[r, \sqrt{|\text{OI}^\$|}\text{sign}(\text{OI}^\$)]$ for a given stock. Hence, λ is positively (negatively) associated with return volatility (order imbalance volatility) by construction. Similarly, ILLIQ is positively (negatively) associated with return volatility (volume) by construction. However, it is interesting to see that order imbalance volatility is positively associated with price impact. 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. In the context of such models, the interpretation most consistent with our results is that volume proxies for noise trading volatility, as explained in Section 2.1.

Another important dimension 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) using TAQ data. Unfortunately, due to data limitation we only observe depth at the best bid and best ask. This is problematic as it can lead to mechanical changes in depth. For instance, traders may 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. To attenuate this issue, we control for spreads in the regression. The results are, however, not sensitive to including this additional control.

Table 6 reports the results of estimating (25) every year with depth as dependent variable and spread as an additional control. In the time-series, depth is positively associated with volume

[29]Since we use log variables, we exclude the small number of estimated λ_{it} that are negative.

and negatively associated with volatility. More importantly for our purpose, order imbalance volatility is negatively associated with depth. This relation is statistically significant in most years of the sample. In addition to the spread result, this result suggests that volatile imbalances are accompanied by a general decrease in liquidity.

Finally, we examine a standard decomposition of 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 some horizon following a trade. Realized spread is generally associated with liquidity provision and equals the signed difference between the trade price and the midquote some time after the trade. We follow prior work and compute realized spread and price impact using the quote midpoint five minutes after a trade. Both measures are in percent and computed by dollar-weighting over all trades in a day.

Our interpretation of order imbalance volatility as a measure of inventory risk suggests that it should be mostly associated with realized spread. In Table 7 we report estimates of month-by-month panel regressions of price impact and realized spread on turnover, realized volatility, and order imbalance volatility. We focus on large stocks in 2017. The results are similar for other size quintiles. In line with our interpretation, we find that order imbalance volatility is weakly associated with price impact and strongly associated with realized spread.

5.2 Evidence from Intraday Returns

To complement the previous results, this section presents evidence from intraday returns. An examination of intraday patterns is useful since the degree of informed trading and liquidity trading is likely not constant over the day. This follows naturally from at least two reasons. First, the informational advantage of trading on overnight information is likely short-lived, which constrains informed traders to trade shortly around the open. The analysis of Madhavan, Richardson, and Roomans (1997) supports the idea that information asymmetry declines over the trading day. Second, liquidity traders may cluster their trades to reduce adverse selection. The period before the close is a natural focal point since many market participants trade at that time for non-informational reasons such as indexing or hedging. Foster and Viswanathan (1990) and Admati and Pfleiderer (1988) formally show that these two channels can generate variation in volume, volatility, and trading costs. Hence, we expect the relation between liquidity, volume, and volatility to vary over the day and that an examination of intraday elasticities can help us shed light on the underlying drivers of liquidity.

We estimate the following panel regression with intraday variables:

$$\log s_{i,t,k} = \alpha_{i,k} + \sum_k 1_k \beta_{\tau,k} \log \tau_{i,t,k} + \sum_k 1_k \beta_{|r|,k} \log |r_{i,t,k}| + \text{controls} + \epsilon_{i,t,k}, \tag{26}$$

where $s_{i,t,k}$ is the percentage effective spread for stock i on day t over intraday interval k, 1_k is an indicator variable that equals one for intraday interval k, and $\alpha_{i,k}$ are interval-stock fixed effects. Control variables are the midquote at the end of each interval and daily market capitalization. Days with FOMC announcements are excluded when estimating (26). We split the trading day into five-minute intervals and focus on stocks in the top market capitalization quintile. These stocks are in general regularly traded over the day, which is important for our analysis. In addition, we exclude the first five minutes of trading since volume over this interval is disproportionately affected by the opening auction.

Figure 8 reports the turnover elasticities and t-statistics for a sample of quarters. In Section 3.3, we document that turnover is positively associated with spread even when controlling for the absolute contemporaneous return. Figure 8 shows that this result holds across the trading day over most of the sample (to save space we only report the results for a sample of quarters). Intraday turnover elasticities are in general positive and statistically significant. Interestingly, turnover elasticities tend to be highest towards the end of the trading day: spreads are more sensitive to volume around the close. Importantly, the intraday pattern in the turnover elasticity of spread is not mechanically driven by the intraday pattern in volatility and volume since both volatility and volume are U-shaped over the trading day. Similarly, while spreads tend to be lowest at the very end of the day in recent years, this is not the case in 2008, when spreads tend to increase at the end of the day.[30] Furthermore, we control for interval-stock fixed effects.

The turnover elasticity is low at the beginning of the day, when information asymmetry is likely to be high, and high at the end of the day, when inventory risk or market power are likely to be

[30] One may worry about a mechanical effect due to the minimum tick size. For instance, if spreads on large stocks are always compressed to the minimum tick size at the end of the day, then (dollar) spreads 'can only go up.' This could generate a mechanical positive relation between volatility, turnover, and spread. We cannot totally rule out this concern for the last years of the sample (when spreads are at the lowest at the end of the day). We observe, however, a similar increase in the volume elasticity at the end of the day in years when spreads tend to increase at the end of the day, which suggests that this mechanical effect does not explain the result.

high. This result suggests that information asymmetry is not the primary driver of the positive spread-volume relation. Rather, inventory or competition effect seem more important, in line with the simple model presented in Section 5.[31]

So far, in this section we have focused on *elasticities*. To put the results into perspective, it is interesting to briefly discuss intraday levels for our main set of variables. Figure 9 reports median effective spreads, absolute return, turnover, and absolute order imbalance across large stocks over the day in 2006 and 2016. Spread and volatility tend to be high around the open (which is consistent with adverse selection theories). In contrast, turnover and absolute order imbalance are mostly high around the close. Intraday patterns in volume and volatility are well-documented by prior work. To the best of our knowledge, the intraday pattern in absolute order imbalance has not been documented before. Absolute order imbalance increases sharply in the last hour of trading. Combined with the evidence that spreads are more sensitive to volume around the close, this result suggests that trading around the close may be more risky and costly than implied by spreads. This is of particular importance since trading volume around the close has grown massively in recent years. We leave a detailed investigation of this interesting pattern for future research.

5.3 Order Imbalance Volatility and the Cross-Section of Stock Returns

Order imbalance volatility also helps explain spreads in the cross-section. For example across large stocks the average fit of panel regression estimated each year increases from 37.15% to 47.16% (reported in the Internet Appendix). This suggests that our analysis of time-series effects extends to the cross-section.

Relatedly, an interesting question is whether order imbalance volatility is priced in the cross-section of stock returns. Intuitively, if order imbalance volatility represents a source of risk for liquidity providers, it should be associated with a risk premium. Importantly, our simple inventory model suggests that we should control for turnover. Since we are interested in "high-frequency" liquidity provision, we examine weekly returns over our sample period (2002-2017), which consists of 797 five-day return observations. Our main variable of interest is an exponentially-weighted

[31] The results in Figure 8 hold when we control for the absolute value of the contemporaneous order imbalance in (26) (not reported). Intraday absolute order imbalance is positively and significantly associated with spread. Moreover, it explains part of the positive turnover elasticity, which remains, however, mostly positive and significant. The evidence in Section 5 suggests that the volatility of order imbalance computed using higher frequency returns is a good candidate to explain away the puzzling sign of the turnover elasticity.

moving average of prior order imbalance volatility with a half-life of one day.

We first consider portfolio sorts. Table 8 reports value-weighted four-factor alpha of portfolios built from sequential sorts with NYSE breakpoints. Raw value-weighted returns give similar results and are reported in the Internet Appendix. In Panel (a), stocks are first sorted into quintiles based on prior-week average turnover and then within each turnover quintile on prior order imbalance volatility. The results support the idea that order imbalance volatility is priced in the cross-section of stock returns. Within all turnover quintiles but one, the long-short order imbalance volatility portfolio earns positive and statistically significant alpha. For example, among stocks with high turnover, the weekly (five-day) alpha is 0.13% with a t-statistic of 1.98. In Panel (b), the order of the sequential sort is reversed. Within each order imbalance volatility quintile, stocks with high turnover tend to earn lower alpha than stocks with low turnover. This is consistent with turnover reducing liquidity provider's risk conditional on order imbalance volatility. The alpha are, however, statistically significant only for the two top quintiles of order imbalance volatility.

Next, we use value-weighted Fama and MacBeth (1973) regressions, which allow us to control for many variables. Table 9 reports the results. Order imbalance volatility predicts higher weekly returns (first column). This relation is statistically significant and becomes stronger once we control for turnover (second column). In addition, order imbalance volatility remains a strong and statistically significant predictor of weekly returns even when we control for a host of other liquidity variables (third column). In particular, we control for turnover, market capitalization, past return, illiquidity (Amihud (2002)), realized volatility, effective spread, and monthly standard deviation of share order imbalance divided by share volume. This last variable is employed by Chordia et al. (2018), who show that it predicts future monthly returns. The denominator in their measure is the daily volume, whereas in ours it is the number of shares outstanding. They link their measure to adverse selection risk. In contrast, our measure is motivated by inventory risk. We note that their measure is positively associated with future returns, though it is not statistically significant. Hence, it appears that the two measures capture different aspects of liquidity.

Overall, the above results suggest that order imbalance volatility predicts future weekly returns in the cross-section even after controlling for well-known factors. We believe this evidence is of particular interest since many high-frequency liquidity measures do not appear to be priced (Lou and Shu (2017)).

6 Conclusion

In this paper, we provide new evidence about the time-series relation between daily liquidity, volume, and volatility. For large stocks, turnover tends to be positively associated with effective spread. This relation is not explained by volatility and is mostly driven by the common component of turnover. This evidence is difficult to explain with adverse selection theories and is more consistent with inventory risk theories. Hence, we argue that inventory risk plays an important role for daily fluctuations in spreads of large stocks. Adverse selection theories fit well the day-to-day variation in spread, turnover, and volatility of small stocks.

We develop a simple continuous-time inventory model to understand our finding. In the model, order imbalance volatility is an important driver of liquidity. Controlling for turnover, an increase in order imbalance volatility leads the market maker to widen the spread because of increased inventory risk.

In the data, order imbalance volatility computed from intraday data is strongly associated with effective spread and substantially improves the explanatory power of spread regressions. Consistent with the model, once we control for order imbalance volatility, the relation between turnover and spread becomes strongly negative. The results are similar for small and large stocks.

Order imbalance volatility is positively associated with other illiquidity measures (such as Amihud's measure and inverse depth). In line with our interpretation based on inventory risk, a standard effective spread decomposition shows that order imbalance volatility is mostly associated with realized spread (liquidity provision) rather than price impact (adverse selection). Furthermore, absolute order imbalance tends to spike at the end of the trading day, a time when inventory considerations likely dominate. Finally, we show that order imbalance volatility is priced in the cross-section of weekly returns. This predictability holds for value-weighted returns even after controlling for many other liquidity variables.

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. 2, 7, 27

Amihud, Y. (2002). Illiquidity and Stock Returns: Cross-Section and Time-Series Effects. *Journal of Financial Markets 5*(1), 31–56. 4, 25, 30

Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *Journal of Finance 51*(1), 169–204. 2

Barinov, A. (2014). Turnover: Liquidity or Uncertainty? *Management Science*. 13

Biais, B., L. Glosten, and C. Spatt (2005). Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications. *Journal of Financial Markets 8*(2), 217–264. 6

Bogousslavsky, V. (2019). The cross-section of intraday and overnight returns. 18

Bollen, N. P., T. Smith, and R. E. Whaley (2004). Modeling the bid/ask spread: Measuring the inventory-holding premium. *Journal of Financial Economics 72*(1), 97–141. 3

Campbell, J. Y., S. J. Grossman, and J. Wang (1993). Trading Volume and Serial Correlation in Stock Returns. *Quarterly Journal of Economics 108*(4), 905–939. 8

Chordia, T., J. Hu, A. Subrahmanyam, and Q. Tong (2018). Order Flow Volatility and Equity Costs of Capital. *Management Science*. 5, 24, 30

Chordia, T., R. Roll, and A. Subrahmanyam (2000). Commonality in Liquidity. *Journal of Financial Economics 56*(1), 3–28. 3, 13

Chordia, T., R. Roll, and A. Subrahmanyam (2001). Market Liquidity and Trading Activity. *Journal of Finance 56*(2), 501–530. 2, 10

Chordia, T., R. Roll, and A. Subrahmanyam (2002, jul). Order imbalance, liquidity, and market returns. *Journal of Financial Economics 65*(1), 111–130. 4

Chordia, T., A. Sarkar, and A. Subrahmanyam (2005). An empirical analysis of stock and bond market liquidity. *Review of Financial Studies 18*(1), 85–129. 14

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 135–155. 2

Collin-Dufresne, P. and V. Fos (2016a). Insider Trading, Stochastic Liquidity and Equilibrium Prices. *Econometrica 84*(4), 1441–1475. 2, 7

Collin-Dufresne, P. and V. Fos (2016b). Insider trading with a random horizon. *working paper SFI@EPFL*. 2, 7

Demsetz, H. (1968). The Cost of Transacting. *Quarterly Journal of Economics*, 33–53. 9

Easley, D. and M. O'Hara (1992). Time and the Process of Security Price Adjustment. *Journal of Finance 47*(2), 577–605. 2, 3, 7, 15, 17

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. 2

Fama, E. F. and J. D. MacBeth (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy 81*(3), 607–636. 30

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. 2, 7, 27

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. 2

Foucault, T., O. Kadan, and E. Kandel (2005). Limit Order Book as a Market for Liquidity. *Review of Financial Studies 18*(4), 1171–1217. 8

Gallant, R. A., P. E. Rossi, and G. Tauchen (1992). Stock Prices and Volume. *Review of Financial Studies 5*(2), 199—-242. 2, 13

Glosten, L. R. (1989). Insider trading, liquidity, and the role of the monopolist specialist. *Journal of Business*, 211–235. 6

Glosten, L. R. and L. E. Harris (1988). Estimating the Components of the Bid/Ask Spread. *Journal of Financial Economics 21*, 123–142. 3

Grossman, S. J. and M. H. Miller (1988). Liquidity and Market Structure. *Journal of Finance 43*(3), 617–633. 3, 8, 20

Hagströmer, B. (2019). Bias in the effective bid-ask spread. *Available at SSRN 2939579*. 10, 17, 18

Hasbrouck, J. (1991). Measuring the Information Content of Stock Trades. *The Journal of Finance 46*(1), 179–207. 3

Hasbrouck, J. (2009). Trading Costs and Returns for U . S . Equities: Estimating Effective Costs from Daily Data. *Journal of Finance 64*(3), 1445–1477. 25

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

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. 9

Johnson, T. (2008). Volume, Liquidity, and Liquidity Risk. *Journal of Financial Economics 87*(2), 388–417. 4

Kim, O. and R. E. Verrecchia (1994). Market Liquidity and Volume around Earnings Announcements. *Journal of Accounting and Economics 17*(1-2), 41–67. 15

Kim, S. T. and H. R. Stoll (2014). Are Trading Imbalances Indicative of Private Information? *Journal of Financial Markets 20*, 151–174. 24

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica 53*(6), 1315–1335. 1, 6, 15

Kyle, A. S. and A. A. Obizhaeva (2016). Market Microstructure Invariance: Empirical Hypotheses. *Econometrica 84*(4), 1345–1404. 4, 12, 25

Lee, C. and M. J. Ready (1991). Inferring Trade Direction from Intraday Data. *Journal of Finance 46*(2), 733–746. 24

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. 2

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. 12

Lou, X. and T. Shu (2017). Price Impact or Trading Volume: Why Is the Amihud (2002) Measure Priced? *Review of Financial Studies* (December). 5, 30

Madhavan, A., M. Richardson, and M. Roomans (1997). Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks. *Review of Financial Studies 10*(4), 1035–1064. 27

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. 18

Pontiff, J. (2006). Costly Arbitrage and the Myth of Idiosyncratic Risk. *Journal of Accounting and Economics 42*(1-2), 35–52. 15

Stoll, H. (2000). Friction. *Journal of Finance 55*(4), 1479–1514. 13

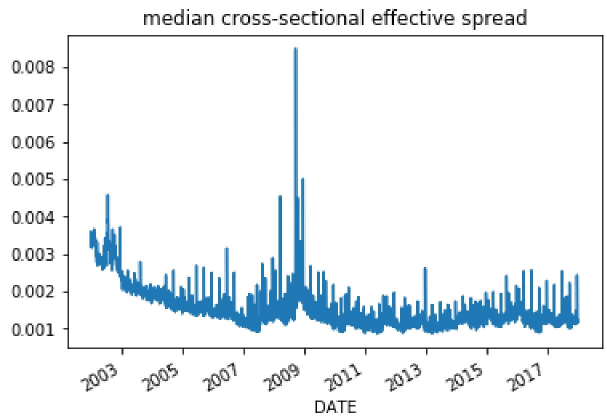
Stoll, H. R. (1978a). The pricing of security dealer services: An empirical study of nasdaq stocks. *Journal of Finance 33*(4), 1153–1172. 2

Stoll, H. R. (1978b). The supply of dealer services in securities markets. *Journal of Finance 33*(4), 1133–1151. 2

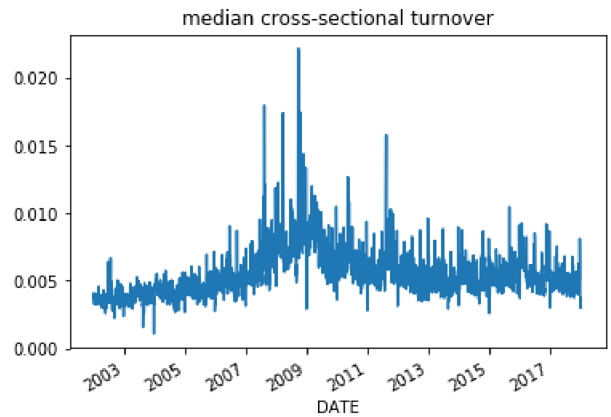
Tauchen, G. E. and M. Pitts (1983). The price variability-volume relationship on speculative markets. *Econometrica*, 485–505. 2

Figure 1. Spreads, volume, and volatility. 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

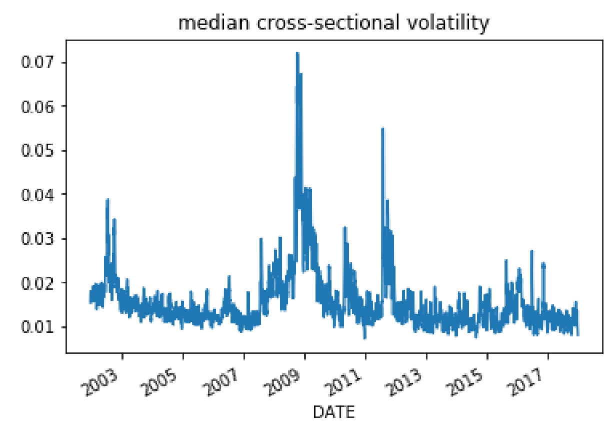
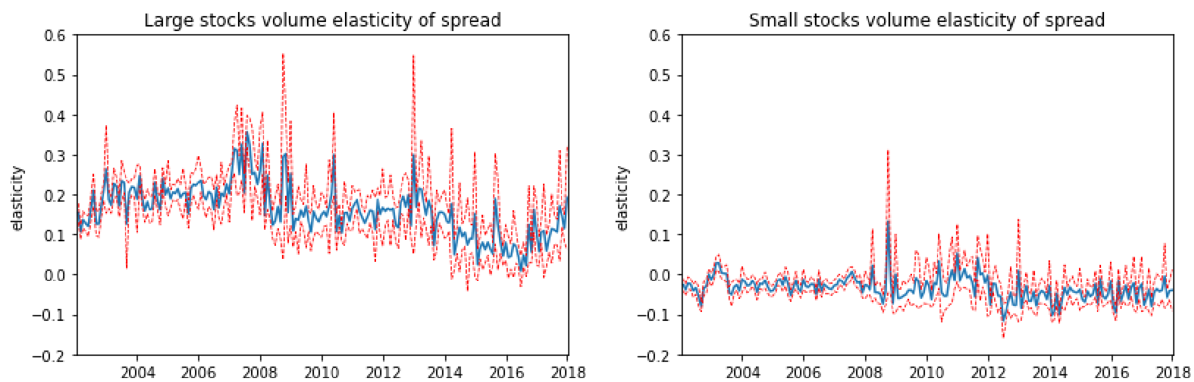
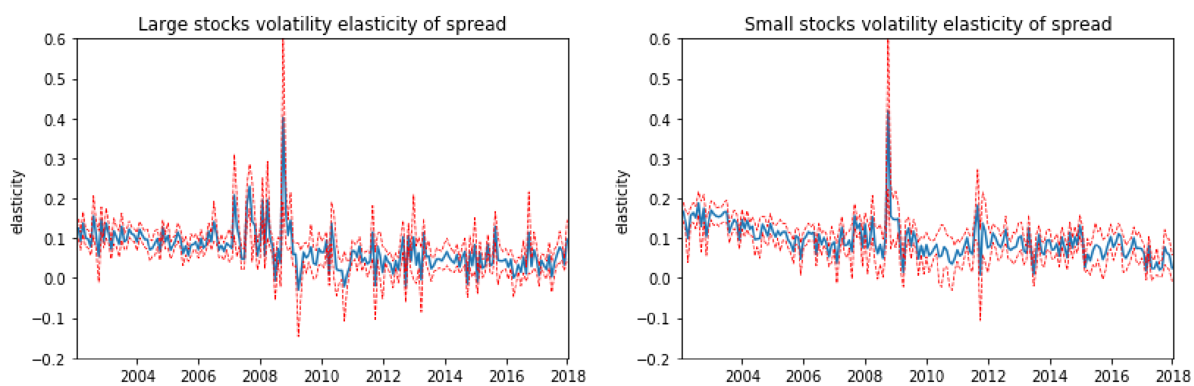


Figure 2. Univariate regressions of spread, volume, and volatility across size quintiles.
We estimate each month for stocks in a given size quintile panel regressions of spread on volume (Panel (a)), spread on volatility (Panel (b)), and volatility on volume (Panel (c)). Spread is the daily effective spread, volume is the daily intraday turnover, and volatility is the average absolute return over the past five trading days (including the current day). 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 (a minimum of 100 observations is required). 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 a market capitalization greater than $100 million. 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



(c) Volume elasticity of volatility

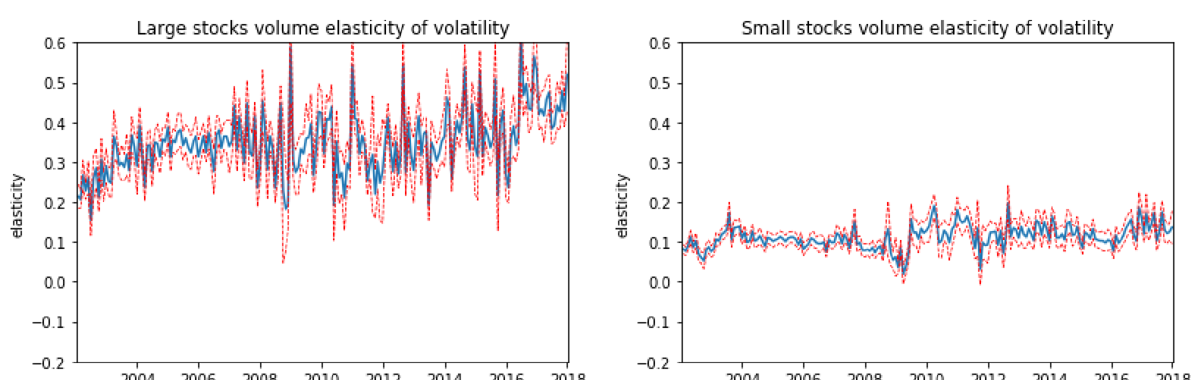
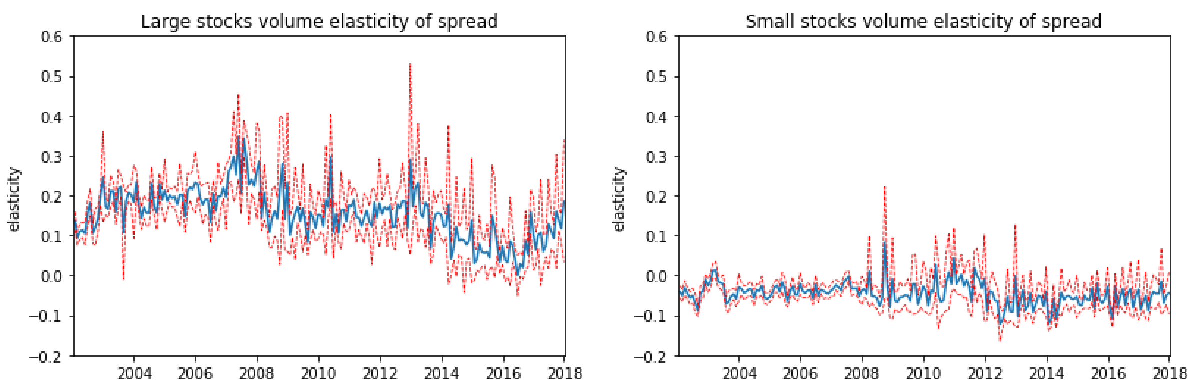


Figure 3. Effective spread regressed on turnover and average absolute return across size quintiles. Panel regression: $\log s_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_\sigma \log \sigma_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover and $\sigma_{i,t}$ is the average absolute return over the past five trading days (including the current day). 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 (a minimum of 100 observations is required). 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 a market capitalization greater than \$100 million. 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

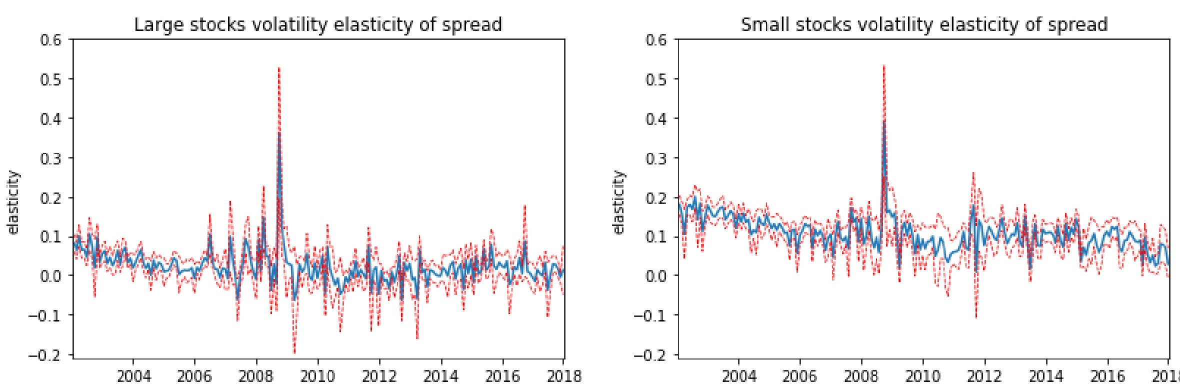


Figure 4. Vector autoregressions of spread, volume, and volatility. For each stock, a VAR is estimated of (log) effective spread (ES%), (log) turnover, and (log) realized volatility (RVol), where the number of lags is chosen based on the Akaike information criterion and all the variables are normalized. 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^{th} and 95^{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 the standard deviation of order imbalance added to the VAR (ordered first).

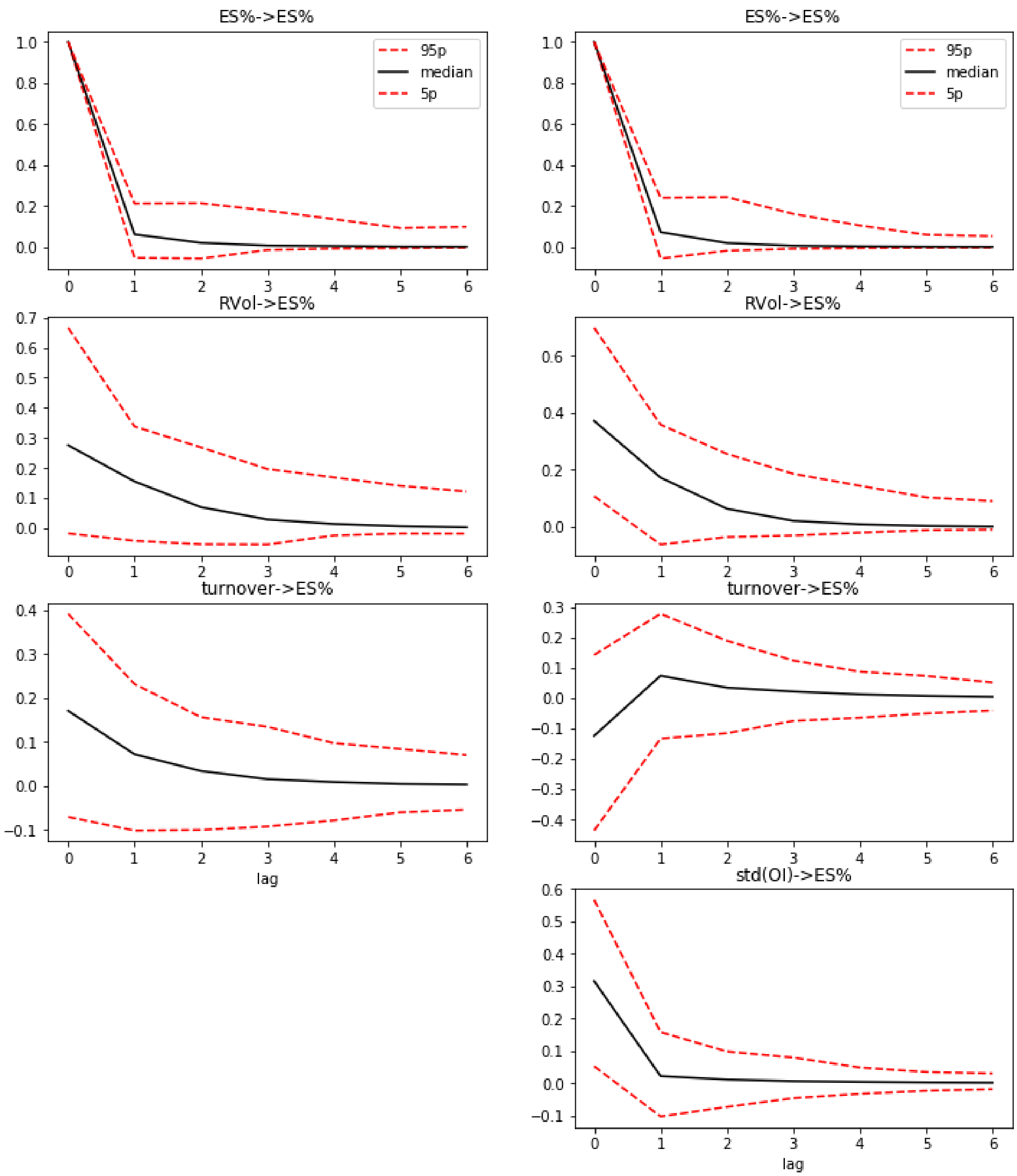
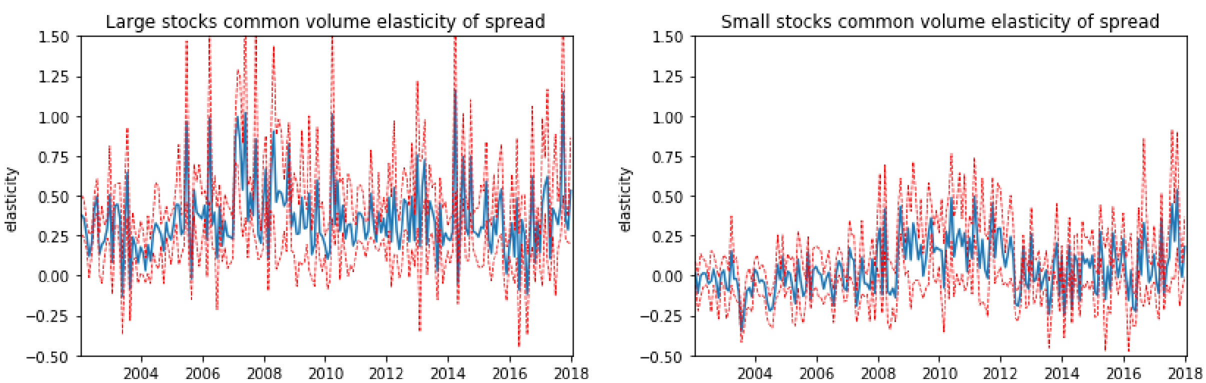
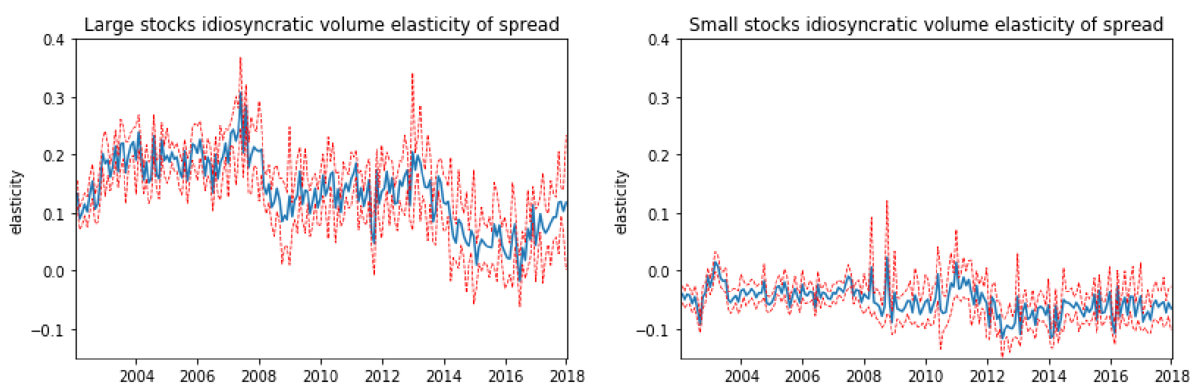


Figure 5. Effective spread regressed on common turnover, idiosyncratic turnover, common volatility, and idiosyncratic volatility across size quintiles.
Panel regression: $\log s_{i,t} = \alpha_i + \beta_{\tau,C}\tau^C_{i,t} + \beta_{\tau,I}\tau^I_{i,t} + \beta_{\sigma,C}\sigma^C_{i,t} + \beta_{\sigma,I}\sigma^I_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where the subscripts C and I denote common and idiosyncratic quantities computed as described in the text. 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 (a minimum of 100 observations is required). 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 a market capitalization greater than \$100 million. Effective spreads are winsorized at 0.05% and 99.95% each month. Standard errors are double-clustered by date and stock.

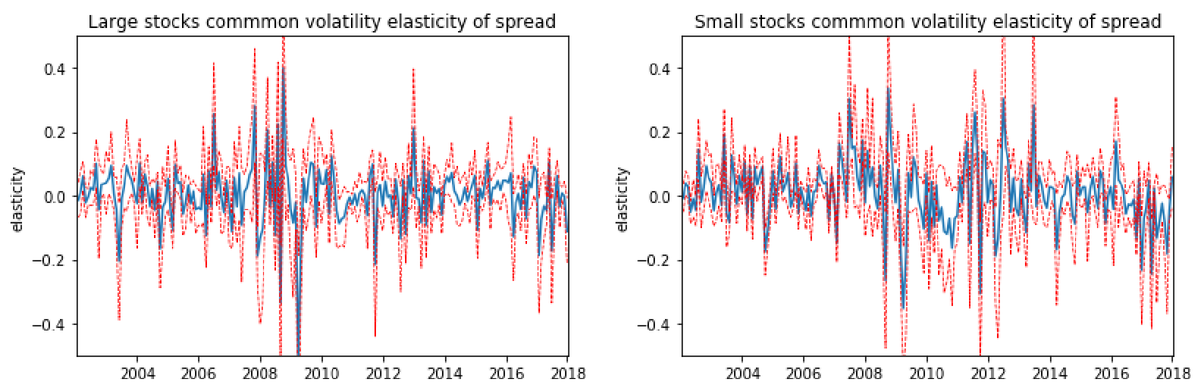
(a) Common turnover elasticity of spread



(b) Idiosyncratic turnover elasticity of spread



(c) Common volatility elasticity of spread



(d) Idiosyncratic volatility elasticity of spread

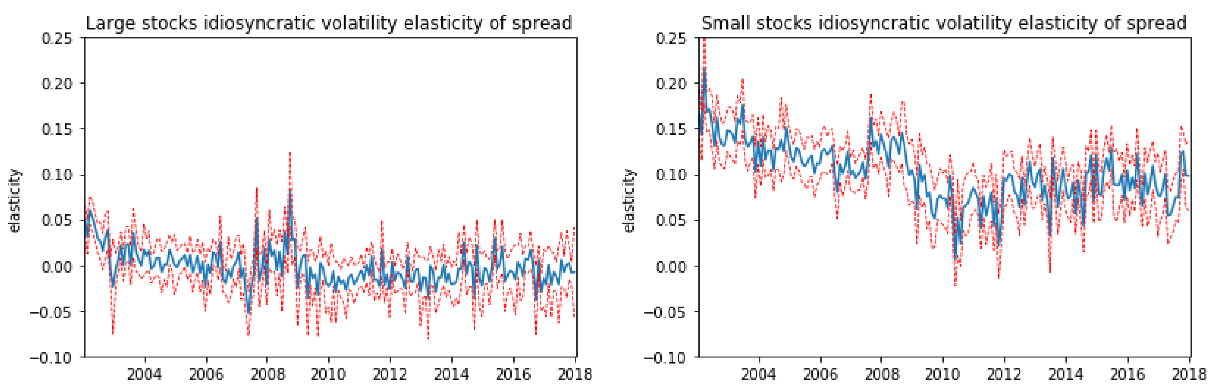
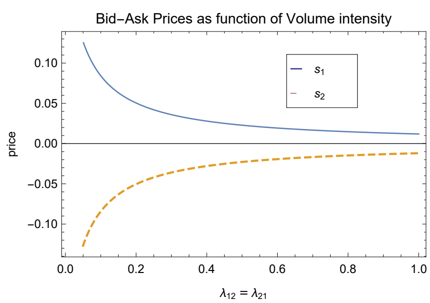
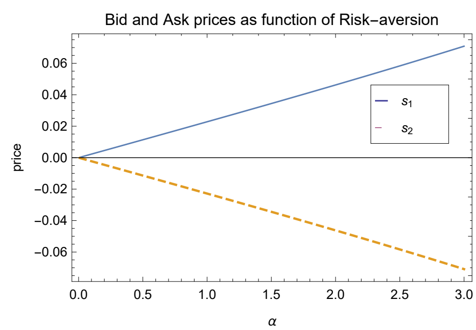


Figure 6. Bid-Ask spread comparative statics. Panel (a) shows bid and ask prices (s_2, s_1) as a function of the trading intensity in the symmetric model $\lambda_{12} = \lambda_{21}$. Panel (b) shows the bid and ask prices as a function of risk-aversion α in the symmetric case. Panel (c) shows the bid and ask prices as a function of the trading intensity λ_{12} in the asymmetric model where λ_{21} is set so that expected volume remains constant. Panel (d) shows the bid-ask spread from panel (c) (i.e., $s_1 - s_2$) as well as the order imbalance as a function of λ_{12} holding expected volume constant.

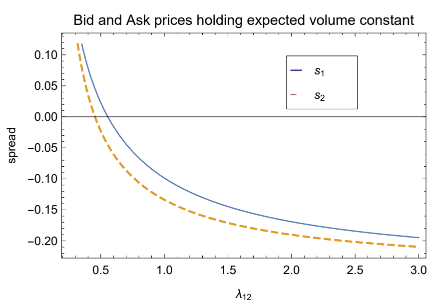
(a) Bid-Ask and Volume



(b) Bid-Ask and Risk-Aversion



(c) Bid-ask prices and order-imbalance



(d) Bid-ask spread and order-imbalance

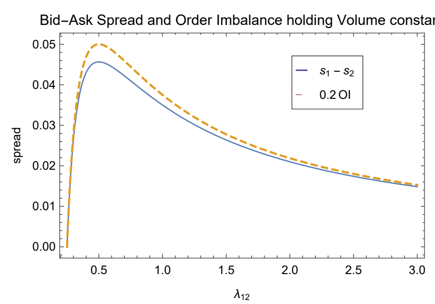


Figure 7. Bid-Ask spread as a function of Volume and Order Imbalance. This figure plots the bid-ask spreads $(s_1 - s_2)$ as a function of the expected volume VOL and the variance of order imbalance OI obtained by varying λ_{12} and λ_{21} between 0.1 and 2 (and normalizing $|\theta_2 - \theta_1| = 1$.

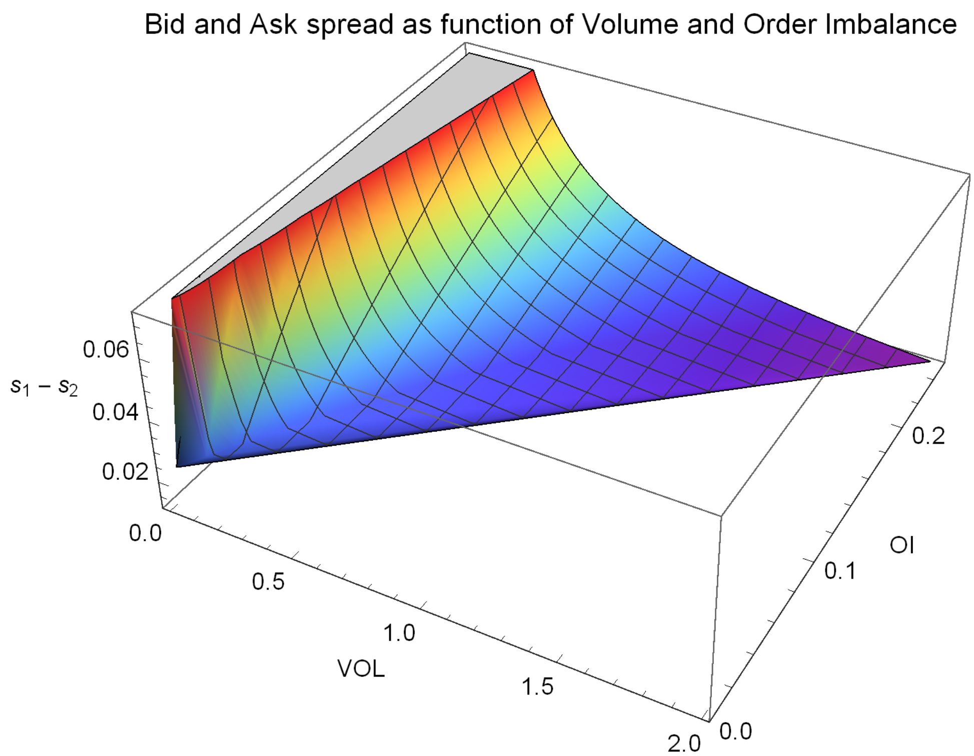
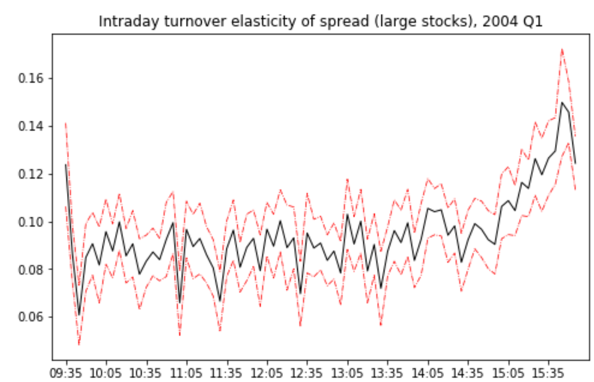
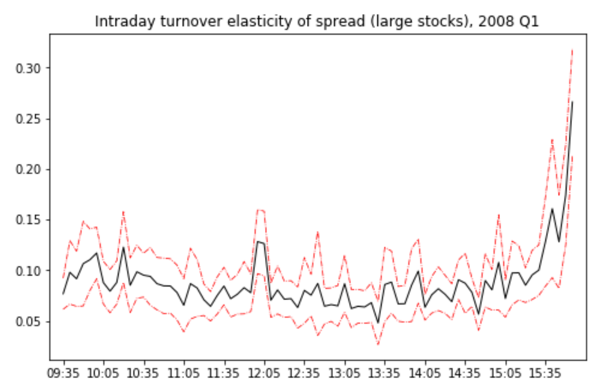


Figure 8. Intraday turnover elasticity of spread for large stocks.
Panel regression: $\log s_{i,t,k} = \alpha_{i,k} + \sum_k 1_k \beta_{\tau,k} \log \tau_{i,t,k} + + \sum_k 1_k \beta_{|r|,k} \log |r_{i,t,k}| + \text{controls} + \epsilon_{i,t,k}$ for stock i on day t in five-minute interval k, where 1_k is an indicator variable that equals one for intraday five-minute interval k and $\alpha_{i,k}$ are interval-stock fixed effects. The figure plots $\beta_{\tau,k}$. Control variables are the (log) midquote at the end of each interval and (log) daily market capitalization. The regression is estimated using the first quarter of each year for a sample of years for stocks in the top market capitalization quintile. The first five-minute interval of the trading day is excluded. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days (a minimum of 100 observations is required). The sample consists of NYSE, Amex, and NASDAQ common stocks. Days with FOMC announcements are excluded from the sample. 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 a market capitalization greater than \$100 million. Effective spreads and absolute returns are winsorized at 0.05% and 99.95%. Standard errors are double-clustered by date and stock.

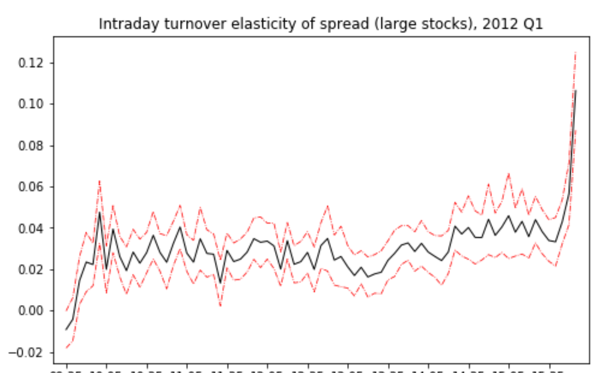
(a) 2004 Q1



(b) 2008 Q1



(c) 2012 Q1



(d) 2016 Q1

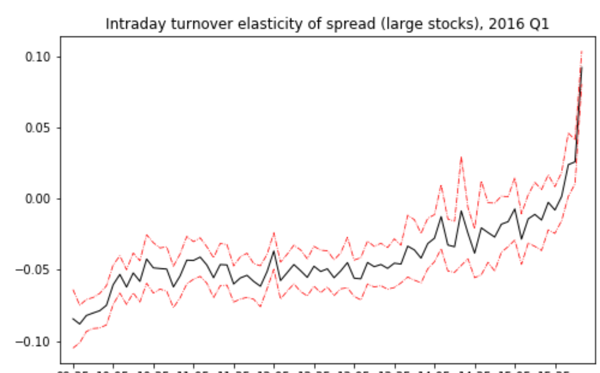
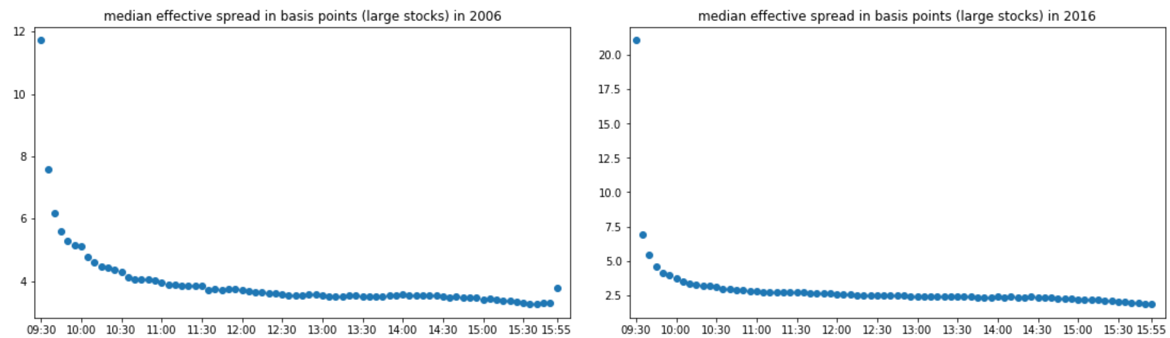
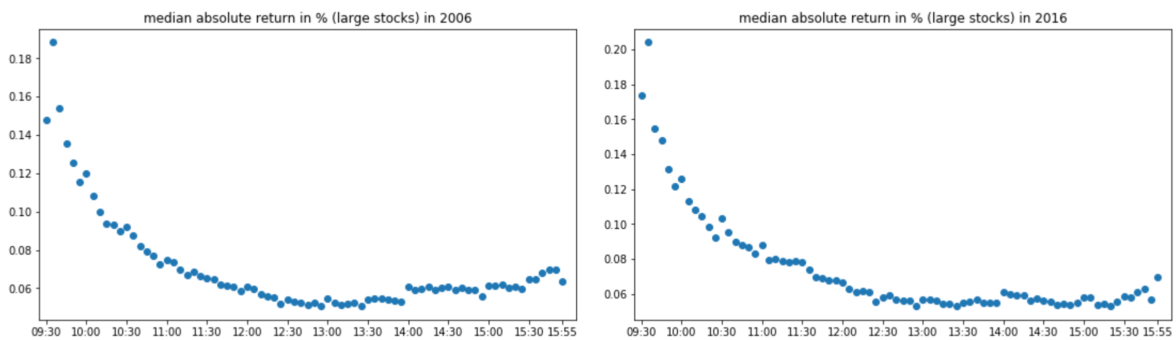


Figure 9. Intraday effective spreads, absolute return, turnover, and absolute order imbalance. This figure reports median values across stocks in the top market capitalization quintile. The sample consists of NYSE, Amex, and NASDAQ common stocks in 2006 and 2016.

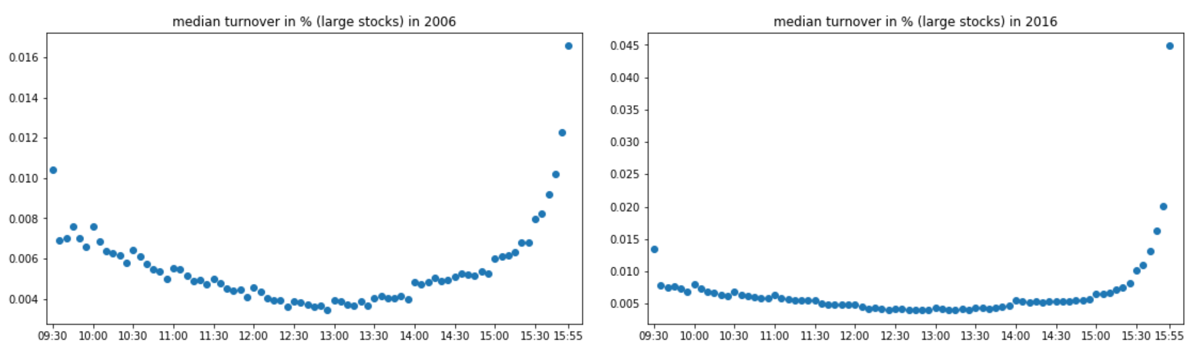
(a) Effective Spread



(b) Absolute Return



(c) Turnover



(d) Absolute Order Imbalance

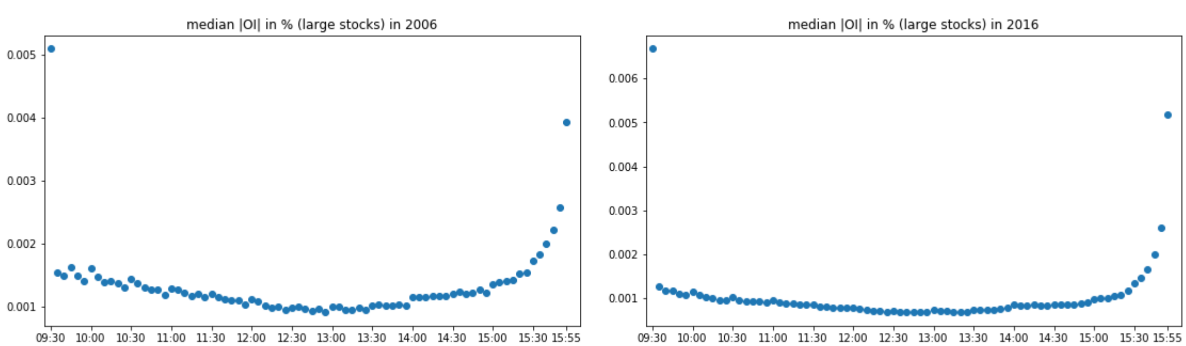


Table 1. Descriptive statistics for stocks sorted in size quintiles for a sample of years. The spread is the percent effective spread (reported in basis points), turnover is the intraday turnover, and volatility is the average absolute return over the past five trading days including the current day. 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 caps									
spread [bp]	mean	94.72	70.18	62.76	96.68	52.73	62.69	69.56	70.32
	median	73.87	51.33	41.75	50.35	35.64	40.85	46.39	44.66
	σ (within)	63.39	48.62	47.19	103.16	43.54	49.98	60.16	63.32
turnover [%]	mean	0.39	0.50	0.48	0.52	0.51	0.42	0.49	0.48
	median	0.17	0.19	0.19	0.27	0.29	0.23	0.23	0.25
	σ (within)	0.84	1.38	0.98	0.83	0.96	0.89	0.89	1.28
volatility [%]	mean	2.25	1.83	1.63	3.06	2.07	1.72	1.67	1.87
	median	1.87	1.53	1.35	2.44	1.83	1.50	1.39	1.51
	σ (within)	1.38	1.06	0.99	2.13	1.13	1.02	1.03	1.58
obs.		127,080	146,897	152,527	132,182	123,247	119,480	129,742	126,515
Mid caps (third quintile)									
spread [bp]	mean	38.26	21.57	15.98	21.39	13.51	12.84	16.28	15.09
	median	30.55	16.92	12.96	14.86	10.56	10.30	11.87	11.69
	σ (within)	24.87	15.19	11.04	33.40	11.07	9.81	19.21	14.52
turnover [%]	mean	0.77	0.86	0.96	1.31	1.09	0.94	0.95	1.01
	median	0.43	0.50	0.63	0.95	0.68	0.59	0.60	0.62
	σ (within)	1.03	1.49	1.20	1.23	1.52	1.23	1.13	1.38
volatility [%]	mean	2.48	1.71	1.59	3.07	1.82	1.56	1.53	1.77
	median	2.07	1.48	1.38	2.50	1.60	1.34	1.27	1.44
	σ (within)	1.37	0.88	0.89	1.88	0.97	0.86	0.99	1.07
obs.		129,496	150,570	157,936	137,397	124,983	121,106	131,846	129,165
Large caps									
spread [bp]	mean	16.15	8.27	6.67	8.29	5.00	4.65	4.61	4.77
	median	12.66	6.59	5.35	6.20	4.05	3.65	3.45	3.63
	σ (within)	13.59	5.95	4.84	10.23	4.09	3.04	4.57	4.31
turnover [%]	mean	0.73	0.67	0.76	1.42	1.12	0.90	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	2.14	1.17	1.13	2.70	1.35	1.16	1.03	1.23
	median	1.73	1.01	0.96	2.03	1.18	1.01	0.88	1.01
	σ (within)	1.28	0.57	0.57	1.99	0.66	0.58	0.56	0.72
obs.		130,092	151,157	158,327	137,730	125,443	121,479	132,074	129,411

Table 2. Correlations among (log) variables for stocks in the bottom and top size quintiles. s is the percent effective spread, τ is the intraday turnover, σ is the average absolute return over the past five trading days including the current day, $|r|$ is the absolute daily return, RVol is the realized volatility computed using five-minute midquote returns, $|\text{OI}|$ is the absolute daily order imbalance as a fraction of shares outstanding, and $\sigma(\text{OI})$ is the daily volatility of order flow imbalance computed using five-minute order imbalance within the day. All the variables are in logs. The table reports the cross-sectional averages of the individual stocks' time-series correlations. 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.

Large caps

	s	τ	σ	$\lvert r\rvert$	RVol	$\lvert\text{OI}\rvert$	$\sigma(\text{OI})$
s	1.00	0.15	0.34	0.22	0.51	0.15	0.30
τ	0.15	1.00	0.41	0.32	0.48	0.40	0.72
σ	0.34	0.41	1.00	0.50	0.61	0.14	0.22
$\lvert r\rvert$	0.22	0.32	0.50	1.00	0.41	0.13	0.19
RVol	0.51	0.48	0.61	0.41	1.00	0.14	0.26
$\lvert\text{OI}\rvert$	0.15	0.40	0.14	0.13	0.14	1.00	0.48
$\sigma(\text{OI})$	0.30	0.72	0.22	0.19	0.26	0.48	1.00

Small caps

	s	τ	σ	$\lvert r\rvert$	RVol	$\lvert\text{OI}\rvert$	$\sigma(\text{OI})$
s	1.00	-0.17	0.22	0.18	0.40	-0.06	-0.00
τ	-0.17	1.00	0.24	0.23	0.32	0.59	0.78
σ	0.22	0.24	1.00	0.49	0.47	0.10	0.12
$\lvert r\rvert$	0.18	0.23	0.49	1.00	0.41	0.13	0.14
RVol	0.40	0.32	0.47	0.41	1.00	0.12	0.17
$\lvert\text{OI}\rvert$	-0.06	0.59	0.10	0.13	0.12	1.00	0.60
$\sigma(\text{OI})$	-0.00	0.78	0.12	0.14	0.17	0.60	1.00

Table 3. Effective spread regressed on turnover and realized volatility for large stocks in the time series.
(a) Levels: $\log s_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_{\text{RVol}} \log \text{RVol}_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover and $\text{RVol}_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns. (b) Changes: $\Delta s_{i,t} = \alpha_i + \beta_\tau \Delta\tau_{i,t} + \beta_\sigma \Delta\text{RVol}_{i,t} + \text{controls} + \epsilon_{i,t}$, where $\Delta x_t \equiv \log(\frac{x_t}{x_{t-1}})$. 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 top size quintile. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days (a minimum of 100 observations is required). 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 a market capitalization greater than \$100 million. 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.

	(a) Levels		(b) Changes	
Year	β_τ	β_{RVol}	β_τ	β_{RVol}
2002	0.03** (2.56)	0.43*** (14.15)	0.07*** (5.64)	0.35*** (10.29)
2003	0.07*** (5.76)	0.43*** (42.35)	0.12*** (7.68)	0.40*** (35.97)
2004	0.07*** (9.83)	0.37*** (37.47)	0.13*** (16.38)	0.36*** (36.52)
2005	0.08*** (10.23)	0.35*** (30.73)	0.16*** (15.67)	0.32*** (28.65)
2006	0.08*** (9.88)	0.31*** (31.07)	0.17*** (16.54)	0.27*** (27.56)
2007	0.11*** (8.35)	0.36*** (18.90)	0.25*** (11.57)	0.28*** (15.78)
2008	0.01 (1.19)	0.45*** (17.51)	0.13*** (8.13)	0.35*** (14.06)
2009	0.04*** (3.08)	0.25*** (12.90)	0.13*** (7.47)	0.20*** (9.79)
2010	0.04*** (3.45)	0.28*** (12.19)	0.14*** (9.11)	0.22*** (9.84)
2011	0.02* (1.91)	0.31*** (19.18)	0.11*** (8.67)	0.24*** (18.80)
2012	0.05*** (3.32)	0.29*** (16.90)	0.16*** (6.04)	0.20*** (11.46)
2013	0.02* (1.89)	0.32*** (17.12)	0.12*** (7.13)	0.25*** (17.11)
2014	-0.05*** (-2.92)	0.36*** (22.64)	0.05** (2.34)	0.29*** (17.82)
2015	-0.10*** (-8.84)	0.43*** (18.66)	-0.00 (-0.10)	0.34*** (13.11)
2016	-0.10*** (-7.75)	0.39*** (18.81)	-0.02 (-1.10)	0.35*** (21.64)
2017	-0.09*** (-5.34)	0.41*** (24.53)	0.02 (0.77)	0.34*** (19.77)
$\bar{R}^2$(%)	20.38		8.48	

Table 4. Effective spread regressed on turnover, realized volatility, and order imbalance volatility for large stocks in the time series. (a) Levels: $\log s_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_{\text{RVol}} \log \text{RVol}_{i,t} + \beta_{\sigma(\text{OI})} \log \sigma(\text{OI})_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover and $\text{RVol}_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns and $\sigma(\text{OI})_{i,t}$ is the volatility of order imbalance computed using five-minute order imbalances over the trading day. (b) Changes: $\Delta s_{i,t} = \alpha_i + \beta_\tau \Delta\tau_{i,t} + \beta_\sigma \Delta \text{RVol}_{i,t} + \beta_{\sigma(\text{OI})} \Delta\sigma(\text{OI})_{i,t} + \text{controls} + \epsilon_{i,t}$, where $\Delta x_t \equiv \log(\frac{x_t}{x_{t-1}})$. 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 top size quintile. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days (a minimum of 100 observations is required). 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 a market capitalization greater than \$100 million. 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.

Year	(a) Levels β_τ	(a) Levels β_{RVol}	(a) Levels $\beta_{\sigma(\text{OI})}$	(b) Changes β_τ	(b) Changes β_{RVol}	(b) Changes $\beta_{\sigma(\text{OI})}$
2002	-0.26*** (-12.39)	0.51*** (13.85)	0.30*** (14.60)	-0.21*** (-10.73)	0.43*** (10.11)	0.29*** (16.17)
2003	-0.25*** (-13.74)	0.52*** (54.16)	0.29*** (18.90)	-0.21*** (-11.54)	0.47*** (43.08)	0.28*** (20.27)
2004	-0.25*** (-15.80)	0.47*** (42.07)	0.28*** (19.07)	-0.20*** (-13.70)	0.43*** (42.15)	0.27*** (21.25)
2005	-0.27*** (-18.06)	0.45*** (40.77)	0.30*** (20.26)	-0.21*** (-17.51)	0.41*** (43.20)	0.29*** (24.08)
2006	-0.27*** (-23.77)	0.41*** (54.66)	0.29*** (27.35)	-0.20*** (-20.06)	0.36*** (47.49)	0.28*** (30.21)
2007	-0.28*** (-16.59)	0.49*** (31.72)	0.33*** (19.30)	-0.20*** (-10.60)	0.39*** (28.40)	0.32*** (22.34)
2008	-0.40*** (-19.20)	0.55*** (27.57)	0.37*** (18.29)	-0.33*** (-18.80)	0.45*** (21.45)	0.35*** (22.28)
2009	-0.32*** (-18.99)	0.37*** (29.47)	0.33*** (19.57)	-0.29*** (-19.93)	0.31*** (23.94)	0.33*** (25.21)
2010	-0.29*** (-22.26)	0.38*** (22.91)	0.30*** (21.73)	-0.25*** (-22.64)	0.32*** (19.01)	0.30*** (27.09)
2011	-0.27*** (-26.55)	0.40*** (34.40)	0.26*** (26.88)	-0.21*** (-23.37)	0.33*** (33.00)	0.24*** (29.54)
2012	-0.28*** (-15.38)	0.38*** (25.66)	0.27*** (13.80)	-0.22*** (-19.04)	0.30*** (22.87)	0.27*** (16.78)
2013	-0.30*** (-28.21)	0.41*** (25.93)	0.28*** (26.71)	-0.25*** (-27.03)	0.35*** (30.28)	0.27*** (30.77)
2014	-0.42*** (-20.59)	0.48*** (36.59)	0.32*** (13.92)	-0.38*** (-26.25)	0.39*** (37.12)	0.33*** (18.12)
2015	-0.43*** (-33.39)	0.52*** (28.04)	0.29*** (24.11)	-0.38*** (-32.22)	0.42*** (18.38)	0.29*** (27.02)
2016	-0.44*** (-27.83)	0.48*** (24.90)	0.30*** (21.35)	-0.39*** (-28.85)	0.43*** (33.34)	0.30*** (23.11)
2017	-0.41*** (-25.57)	0.51*** (46.57)	0.29*** (14.64)	-0.36*** (-29.52)	0.42*** (36.04)	0.29*** (18.09)
$\bar{R}^2(\%)$		30.91			20.02	

Table 5. Price impact regressed on turnover, realized volatility, and order imbalance volatility for large stocks in the time series. The following regression is estimated: $\log \text{PriceImpact}_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_{\text{RVol}} \log \text{RVol}_{i,t} + \beta_{\sigma(\text{OI})} \log \sigma(\text{OI})_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover and $\text{RVol}_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns and $\sigma(\text{OI})_{i,t}$ is the volatility of order imbalance computed using five-minute order imbalances over the trading day. Panel (a): price impact equals λ_{it} obtained from the regression $r_{itk} = \delta_{it} + \lambda_{it}\sqrt{|\text{OI}^{\$}_{itk}|}\text{sign}(\text{OI}^{\$}_{itk}) + e_{it}$, where r_{itk} is the five-minute midquote return in interval k, and $\text{OI}^{\$}_{itk}$ is the dollar order imbalance. Panel (b): price impact equals $\text{ILLIQ}_{it} = \frac{1}{\#\text{traded intervals}} \sum_{k\epsilon\{j|\text{DVOL}_{itj}>0\}} \frac{|r_{itk}|}{\text{DVOL}_{itk}}$, where DVOL is the dollar volume. 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 top size quintile. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days (a minimum of 100 observations is required). 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 a market capitalization greater than $100 million. 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.

	(a) PriceImpact = λ_{it}			(b) PriceImpact = ILLIQ_{it}		
Year	β_τ	β_{RVol}	$\beta_{\sigma(\text{OI})}$	β_τ	β_{RVol}	$\beta_{\sigma(\text{OI})}$
2002	-0.27*** (-15.04)	1.15*** (36.50)	-0.39*** (-38.54)	-1.10*** (-54.89)	0.90*** (40.95)	0.24*** (18.72)
2003	-0.26*** (-17.42)	1.23*** (89.79)	-0.39*** (-45.85)	-1.21*** (-56.32)	0.88*** (81.51)	0.29*** (27.03)
2004	-0.28*** (-21.64)	1.23*** (100.38)	-0.39*** (-51.03)	-1.20*** (-100.27)	0.88*** (65.68)	0.27*** (37.15)
2005	-0.30*** (-26.32)	1.23*** (89.28)	-0.42*** (-54.70)	-1.17*** (-103.29)	0.90*** (44.62)	0.24*** (39.38)
2006	-0.28*** (-18.42)	1.15*** (69.91)	-0.44*** (-36.30)	-1.15*** (-98.54)	0.92*** (87.78)	0.21*** (35.00)
2007	-0.35*** (-17.33)	1.06*** (46.13)	-0.37*** (-26.42)	-1.12*** (-101.99)	0.98*** (72.96)	0.16*** (29.75)
2008	-0.30*** (-13.68)	1.14*** (38.51)	-0.44*** (-31.75)	-1.10*** (-82.55)	0.96*** (58.25)	0.10*** (18.51)
2009	-0.18*** (-11.00)	1.21*** (48.06)	-0.44*** (-38.19)	-1.07*** (-141.59)	0.92*** (41.61)	0.10*** (15.21)
2010	-0.12*** (-5.94)	1.19*** (31.61)	-0.38*** (-34.89)	-1.10*** (-72.77)	0.92*** (26.01)	0.11*** (21.17)
2011	-0.16*** (-7.21)	1.18*** (38.35)	-0.32*** (-22.37)	-1.12*** (-107.44)	0.96*** (45.96)	0.11*** (23.25)
2012	-0.20*** (-11.84)	1.31*** (63.60)	-0.30*** (-23.46)	-1.09*** (-101.16)	0.83*** (72.48)	0.12*** (19.40)
2013	-0.31*** (-15.55)	1.33*** (41.83)	-0.26*** (-20.12)	-1.14*** (-67.34)	0.89*** (30.81)	0.14*** (23.18)
2014	-0.24*** (-20.83)	1.22*** (88.63)	-0.48*** (-65.66)	-1.14*** (-148.76)	0.88*** (86.04)	0.15*** (36.93)
2015	-0.20*** (-16.35)	1.15*** (50.44)	-0.48*** (-57.84)	-1.15*** (-123.23)	0.89*** (66.42)	0.15*** (32.22)
2016	-0.22*** (-15.90)	1.17*** (64.77)	-0.44*** (-50.03)	-1.14*** (-108.79)	0.88*** (52.77)	0.14*** (32.52)
2017	-0.24*** (-21.30)	1.21*** (81.43)	-0.47*** (-55.41)	-1.11*** (-135.75)	0.79*** (67.02)	0.15*** (43.56)
$\bar{R}^2$(%)		28.79			77.05	

Table 6. Depth regressed on turnover, realized volatility, order imbalance volatility, and effective spread for large stocks in the time series.
The following regression is estimated: $\log \text{Depth}_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_{\text{RVol}} \log \text{RVol}_{i,t} + \beta_{\sigma(\text{OI})} \log \sigma(\text{OI})_{i,t} + \beta_s \log s_{i,t} \text{controls} + \epsilon_{i,t}$ 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 $\text{RVol}_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns, $\sigma(\text{OI})_{i,t}$ is the volatility of order imbalance computed using five-minute order imbalances over the trading day, and $s_{i,t}$ is the dollar-weighted percent effective spread. 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 top size quintile. At the beginning of each month, stocks are sorted by their average daily market capitalization over the past 250 trading days (a minimum of 100 observations is required). 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 a market capitalization greater than \$100 million. 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.

Year	β_τ	β_{RVol}	$\beta_{\sigma(\text{OI})}$	β_s
2002	0.35*** (20.58)	-0.22*** (-9.94)	-0.00 (-0.32)	-0.19*** (-15.34)
2003	0.43*** (22.47)	-0.31*** (-30.46)	-0.04*** (-5.18)	-0.09*** (-16.94)
2004	0.47*** (31.15)	-0.42*** (-14.13)	-0.04*** (-7.23)	-0.10*** (-15.55)
2005	0.46*** (30.23)	-0.44*** (-15.49)	-0.05*** (-10.80)	-0.07*** (-13.95)
2006	0.44*** (30.65)	-0.51*** (-18.67)	-0.06*** (-11.93)	-0.07*** (-12.71)
2007	0.41*** (25.41)	-0.56*** (-22.82)	-0.02*** (-4.58)	-0.04*** (-6.93)
2008	0.40*** (18.33)	-0.69*** (-17.47)	-0.01*** (-2.70)	0.02*** (2.78)
2009	0.38*** (23.72)	-0.66*** (-22.94)	-0.00 (-0.12)	-0.03*** (-3.54)
2010	0.39*** (16.04)	-0.66*** (-14.65)	-0.01 (-1.07)	-0.02** (-2.39)
2011	0.38*** (19.13)	-0.65*** (-17.34)	-0.02*** (-3.06)	0.03*** (4.03)
2012	0.35*** (29.22)	-0.40*** (-22.19)	-0.03*** (-6.01)	-0.00 (-0.22)
2013	0.40*** (18.79)	-0.48*** (-10.24)	-0.05*** (-9.43)	0.02** (2.47)
2014	0.31*** (34.06)	-0.39*** (-23.40)	-0.01 (-1.56)	-0.01*** (-2.92)
2015	0.30*** (21.97)	-0.34*** (-15.81)	-0.02*** (-4.05)	0.01*** (2.77)
2016	0.30*** (15.37)	-0.37*** (-11.26)	-0.02*** (-4.91)	0.03*** (4.61)
2017	0.28*** (26.71)	-0.27*** (-14.35)	-0.03*** (-10.16)	0.02*** (4.23)
$\bar{R}^2(\%)$		41.70		

Table 7. Price impact and realized spread regressed on turnover, realized volatility, and order imbalance volatility for large stocks in the time series.

$\log x_{i,t} = \alpha_i + \beta_\tau \log \tau_{i,t} + \beta_{\text{RVol}} \log \text{RVol}_{i,t} + \beta_{\sigma(\text{OI})} \log \sigma(\text{OI})_{i,t} + \text{controls} + \epsilon_{i,t}$ for stock i on day t, where $\tau_{i,t}$ is the daily intraday turnover and $\text{RVol}_{i,t}$ is the realized volatility computed using five-minute intraday midquote returns and $\sigma(\text{OI})_{i,t}$ 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 obtained by decomposing the effective spread using the midquote five minutes after a trade. 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 (a minimum of 100 observations is required). 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 and a market capitalization greater than \$100 million. 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_{RVol}	$\beta_{\sigma(\text{OI})}$	β_τ	β_{RVol}	$\beta_{\sigma(\text{OI})}$
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 8. 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. Panel (a): sort on turnover then on order imbalance volatility. Panel (b): sort on order imbalance volatility then on turnover. Turnover is the average turnover over previous five trading days. 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 with 3 lags. *, **, and *** denote significance at the 10%, 5%, and 1% level.

(a) α^{VW}_{FF4} (turnover then order imbalance volatility)

	low σ(OI)	2	3	4	high σ(OI)	H-L
low turn.	-0.02	0.02	0.02	0.02	0.08***	0.10**
	(-0.66)	(0.55)	(0.51)	(0.50)	(2.78)	(2.56)
2	-0.01	0.05*	-0.00	0.01	0.06*	0.06
	(-0.30)	(1.72)	(-0.05)	(0.39)	(1.66)	(1.56)
3	0.00	0.03	0.06**	0.09***	0.11***	0.11***
	(0.09)	(0.88)	(2.02)	(3.23)	(3.65)	(2.65)
4	-0.09***	0.00	0.01	-0.04	0.12***	0.20***
	(-2.91)	(0.13)	(0.24)	(-1.15)	(4.03)	(4.59)
high turn.	-0.05	-0.07	0.04	-0.05	0.08*	0.13**
	(-0.94)	(-1.28)	(0.68)	(-0.98)	(1.68)	(1.98)

(b) α^{VW}_{FF4} (order imbalance volatility then turnover)

	low turn	2	3	4	high turn	H-L
low σ(OI)	-0.02	0.01	-0.00	0.01	-0.05*	-0.03
	(-0.45)	(0.17)	(-0.14)	(0.42)	(-1.79)	(-0.60)
2	0.00	0.00	0.01	0.04	-0.06*	-0.06
	(0.01)	(0.04)	(0.43)	(1.26)	(-1.87)	(-1.18)
3	0.05	0.03	0.04*	0.04	0.00	-0.05
	(1.56)	(0.96)	(1.74)	(1.22)	(0.00)	(-0.89)
4	0.05	0.10***	0.03	-0.02	-0.09	-0.14**
	(1.49)	(3.22)	(1.19)	(-0.50)	(-1.58)	(-2.04)
high σ(OI)	0.10***	0.08**	0.05	0.09**	-0.05	-0.15**
	(3.55)	(2.41)	(1.53)	(2.37)	(-0.82)	(-2.10)

Table 9. Value-weighted Fama-MacBeth regressions of weekly returns (in percent) on liquidity characteristics. Order imbalance volatility ($\sigma(\text{OI})_{t-1}$) is an exponentially-weighted moving average of prior order imbalance volatility with a half-life of one day. Turnover (turn_{t-1}) is the average turnover over the previous five trading days. 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 exponentially-weighted moving average of prior daily realized volatilities with a half-life of one day. Effective spread (ES_{t-1}) is the average percentage effective spread over the previous five trading days. Monthly standard deviation of share order imbalance divided by share volume ($\sigma(\text{OI/VOL})^{\text{month}}_{t-1}$) 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 the 0.5% and 99.5% levels. 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 with 3 lags. *, **, and *** denote significance at the 10%, 5%, and 1% level.

dependent variable: r_t (weekly return in percent)			
	coeff. (t-stat)	coeff. (t-stat)	coeff. (t-stat)
$\sigma(\text{OI})_{t-1}$	0.064** (2.35)	0.086*** (3.02)	0.083*** (3.40)
turn_{t-1}		-0.037 (-1.00)	-0.026 (-0.67)
ME_{t-1}			-0.012 (-0.31)
r_{t-1}			-1.652*** (-3.91)
ILLIQ_{t-1}			-0.009 (-0.25)
RVol_{t-1}			-0.023 (-0.32)
ES_{t-1}			-0.023 (-0.63)
$\sigma(\text{OI/VOL})^{\text{month}}_{t-1}$			0.056 (1.42)
$\bar{N}$	2,628	2,628	2,591
$\bar{R}^2$	0.020	0.036	0.104

A Appendix: Model Details

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 $W, N = i$):

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

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 $N = i$).

$$J_W = \alpha e^{-\alpha c} \tag{27}$$

$$0 = J_W(\mu - \overline{\lambda\eta}_i - rS) + J_{WW} n\sigma^2 + \sum_{j \neq i} \lambda_{ij} \eta_{ij} J_W(W + n\eta_{ij}, j) \tag{28}$$

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:

$$c(W, N = i) = rW - b_i \tag{29}$$

The second FOC implies:

$$\mu - rS = \alpha r n \sigma^2 + \sum_{j \neq i} \lambda_{ij} \eta_{ij} (1 - e^{-\alpha(rn\eta_{ij} - b_j + b_i)}) \tag{30}$$

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

$$0 = -r + r\alpha(b + n(\mu - \overline{\lambda\eta}_i - rS)) - \frac{1}{2} r^2 \alpha^2 n^2 \sigma^2 + \beta - \sum_{j \neq i} \lambda_{ij} (e^{-\alpha(rn\eta - b_j + b_i)} - 1) \tag{31}$$

From equation (30) we can substitute $\mu - rS$ to obtain:

$$0 = -r + r\alpha b + \frac{1}{2} r^2 \alpha^2 n^2 \sigma^2 + \beta - \sum_{j \neq i} \lambda_{ij} (e^{-\alpha(rn\eta_{ij} - b_j + b_i)} (1 + r\alpha n \eta_{ij}) - 1) \tag{32}$$

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:

$$\mu - rS = \alpha r\theta_i\sigma^2 + \sum_{j\neq i}\lambda_{ij}\eta_{ij}(1 - e^{-\alpha(r\theta_i\eta_{ij}-b_j+b_i)}) \tag{33}$$

$$0 = -r + r\alpha b_i + \frac{1}{2}r^2\alpha^2\theta_i^2\sigma^2 + \beta - \sum_{j\neq i}\lambda_{ij}(e^{-\alpha(r\theta_i\eta_{ij}-b_j+b_i)}(1 + r\alpha\theta_i\eta_{ij}) - 1) \tag{34}$$

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:

$$V_\delta^i\kappa_\delta(\overline{\delta}_i - \delta) + \frac{1}{2}V_{\delta\delta}^i\sigma_\delta^2 + \sum_{j\neq i}\lambda_{ij}(V^j(\delta) - V^i(\delta)) = rV^i(\delta) - \delta \tag{35}$$

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:

$$\frac{\kappa_\delta\overline{\delta}_i}{r + \kappa_\delta} + \sum_{j\neq i}\lambda_{ij}(v_j - v_i) = rv_i \tag{36}$$

The solution obtains in terms of the transition matrix Λ (which has entry λ_{ij} and where $\lambda_{ii} = -\sum_{j\neq i}\lambda_{ij}$) 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:

$$S(\delta, N) = V(\delta, N) + s(N) \tag{37}$$

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 $N_t = i$):

$$\mu - rS \equiv \mathrm{E}_t[dS_t/dt + \delta - rS] = \mathrm{E}_t[ds(N_t)/dt - rs(N_t)] = \sum_{j\neq i} \lambda_{ij}(s_j - s_i) - rs_i \tag{38}$$

$$\sigma = \frac{\sigma_\delta}{r + \kappa_\delta} \tag{39}$$

$$\eta_{ij} = v_j - v_i + s_j - s_i \tag{40}$$

Substituting into our system of equilibrium conditions (41) and (42) 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 $\forall i, j \in \{1, M\}$:

$$0 = rs_i + \alpha r \theta_i \sigma^2 + \sum_{j\neq i} \lambda_{ij}\{v_j - v_i - \eta_{ij} e^{-\alpha(r\theta_i\eta_{ij} - b_j + b_i)}\} \tag{41}$$

$$0 = -r + r\alpha b_i + \frac{1}{2} r^2\alpha^2\theta_i^2\sigma^2 + \beta + \sum_{j\neq i} \lambda_{ij}\{1 - e^{-\alpha(r\theta_i\eta_{ij} - b_j + b_i)}(1 + r\alpha\theta_i\eta_{ij})\} \tag{42}$$

$$\sigma = \frac{\sigma_\delta}{r + \kappa_\delta} \tag{43}$$

$$\eta_{ij} = v_j - v_i + s_j - s_i \tag{44}$$

$$v = \frac{\kappa_\delta}{r + \kappa_\delta}(rI - \Lambda)^{-1}\overline{\delta} \tag{45}$$

The solution of this system characterizes the equilibrium.

A.1 Price impact without 'adverse selection' in a symmetric model

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. That we consider the simple model with two states $M = 2$ and

$$\lambda_{12} = \lambda_{21} = \lambda$$
$$\theta_2 = -\theta_1 = \theta$$
$$\overline{\delta}_1 = \overline{\delta}_2 = 0$$

Note that since the long-run mean is constant and equal to zero across states $v = 0$. Then we can prove that there exists a unique symmetric solution characterized by

$$s_1 = -s_2$$
$$b_1 = b_2$$

where s_1, b_1 solve the system of equations:

$$0 = rs_1 + \alpha r\theta_1\sigma^2 + 2\lambda s_1 e^{2\alpha r\theta_1 s_1} \tag{46}$$

$$0 = -r + r\alpha b_1 + \frac{1}{2} r^2\alpha^2\theta_1^2\sigma^2 + \beta + \lambda\{1 - e^{2\alpha r\theta_1 s_1}(1 - 2r\alpha\theta_1 s_1)\} \tag{47}$$

$$\sigma = \frac{\sigma_\delta}{r + \kappa_\delta} \tag{48}$$

Clearly, b_1 is uniquely solved in equation (47) given a solution for s_1. And note that (46) can be solved uniquely for s_1 as the intersection between the strictly positive decreasing curve $2\lambda e^{2s_1\alpha r\theta_1}$ and the curve $-\theta_1\alpha r\sigma^2/s_1 - r$. Further, since we assume $\theta_1 < 0$ it is clear that there is a unique positive solution $s_1 > 0$ which satisfies this equation. Further, since the optimal solution satisfies $0 < -\alpha r\theta_1\sigma^2/s_1 - r < 2\lambda$ we obtain that

$$-\theta_1\alpha\sigma^2 > s_1 > \frac{-\theta_1\alpha r\sigma^2}{2\lambda + r} > 0.$$

A.2 Price impact and order imbalance: Asymmetric model without adverse selection

To understand the effect of order imbalance we solve the case where $\lambda_{12} \neq \lambda_{21}$, but otherwise keep the same assumption (of symmetric depth and no adverse selection). That is, we consider the simple model with two states $M = 2$ and

$$\theta_2 = -\theta_1 = \theta \tag{49}$$

$$\overline{\delta}_1 = \overline{\delta}_2 = 0 \tag{50}$$

In this model the solution $s_i, b_i\, \forall i \in 1, 2$ satisfy the system of equations:

$$0 = rs_i + \alpha r\theta_i\sigma^2 + \lambda_{ij}(s_i - s_j)e^{-\alpha r\theta_i(s_j - s_i) + \alpha(b_j - b_i)} \tag{51}$$

$$0 = -r + r\alpha b_i + \frac{1}{2}r^2\alpha^2\theta_i^2\sigma^2 + \beta + \lambda_{ij}\{1 - e^{-\alpha r\theta_i(s_j - s_i) + \alpha(b_j - b_i)}(1 + r\alpha\theta_i(s_j - s_i))\} \tag{52}$$

$$\sigma = \frac{\sigma_\delta}{r + \kappa_\delta} \tag{53}$$

We solve this system numerically. It is easy to show that there is a unique solution for b_i in terms of s_1, s_2. Thus solving the system amounts to solving two non-linear equations for s_1, s_2.