Accepted Manuscript

Distilling Liquidity Costs from Limit Order Books Forthcoming in Journal of Banking and Finance

Diego Amaya, Jean-Yves Filbien, Cédric Okou, Alexandre F. Roch

PII: S0378-4266(18)30135-3
DOI: 10.1016/j.jbankfin.2018.06.009
Reference: JBF 5369

To appear in: *Journal of Banking and Finance*

Received date: 21 September 2016
Revised date: 18 December 2017
Accepted date: 15 June 2018

Please cite this article as: Diego Amaya, Jean-Yves Filbien, Cédric Okou, Alexandre F. Roch, Distilling Liquidity Costs from Limit Order Books Forthcoming in Journal of Banking and Finance, *Journal of Banking and Finance* (2018), doi: 10.1016/j.jbankfin.2018.06.009

Distilling Liquidity Costs from Limit Order Books Forthcoming in Journal of Banking and Finance

Diego Amaya[a], Jean-Yves Filbien[b], Cédric Okou[c], Alexandre F. Roch[d]

[a] *Wilfrid Laurier University, 75 University Avenue West, Waterloo, CANADA, N2L 3C5*
[b] *University of Lille, 104 Avenue du Peuple Belge, 59800 Lille (France)*
[c] *University of Quebec in Montreal, Finance Department, 315, rue Sainte-Catherine Est, Montréal, CANADA, H2X 3X2*
[d] *Corresponding author, University of Quebec in Montreal, Finance Department, 315, rue Sainte-Catherine Est, Montréal, CANADA, H2X 3X2, roch.alexandre_f@uqam.ca. This author's research was supported by the Institute of Financial Mathematics of Montreal and the Natural Sciences and Engineering Research Council of Canada.*

Abstract

This paper proposes a method to compute ex-ante trading costs at the intraday level from limit order books. Using nearly 500 of the largest traded companies in the NYSE ArcaBook, we show that these costs have nontrivial intraday dynamics, are negatively related to volume and positively related to volatility. When ex-ante trading costs are incorporated into price impact specifications, the results show that this measure provides relevant information about price changes of the market at a high frequency level. Our evidence suggest that ex-ante trading costs constitute a new source of information for the study of intraday liquidity.

Keywords: Intraday liquidity, limit order books, price impact, price formation
JEL Classification: G12; G14

1. Introduction

Electronic limit order markets have emerged as important venues for trading, offering a real-time glimpse of existing supply and demand in the equity market. During the period of 2010 to 2014, the volume traded in NYSE ArcaBook and BATS accounted for 12% and 8% of the total U.S. equity market, respectively.[1] Electronic limit order books contain very rich and complex sources of information about liquidity provision and price formation at the high frequency level. In these markets, ex-ante commitments to offer liquidity are made by investors who submit limit orders, specifying both the price and the quantity to buy or sell. On the other hand, liquidity is demanded by investors who place market orders.

☆We thank the seminar participants at the IFM2 2014 Conference, EFMA 2016 meetings, HEC Montreal, Brock University, and Wilfrid Laurier University.

[1] Data from http://www.batstrading.com/market_data/venue/market/tapea/

The goal of this paper is to construct and apply an intraday measure of ex-ante trading costs based on limit orders. We show the pertinence of this measure by assessing the joint effect of execution costs and order flows on stock price changes of actively traded companies. Our empirical studies rely on a data set consisting of two years (January 2011 to December 2012) of intraday observations for nearly 500 of the largest traded companies in the NYSE ArcaBook.

Recent studies such as Hautsch and Huang (2012), Brogaard et al. (2015), and Fleming et al. (2017), among others, have shown that limit order activity contributes to the price formation process. The common approach employed in these studies is the standard vector autoregression specification introduced in Hasbrouck (1991), which uses tick-by-tick data to jointly model ask and bid quote dynamics in a cointegrated system. The evidence provided in these frameworks relates directly to the market mechanism in which trades are executed and shows how different characteristics of the trading process deviate prices from their efficient values in the short-run. Different from these studies, we center our analyses on frequencies beyond the tick level to capture the impact that trade demand exerts on the price process. To this end, we synthesize market supply/demand schedules into estimates of ex-ante trading costs over intraday periods of time. This synthetic liquidity measure corresponds to the marginal cost of buying or selling one unit of stock, which combined with the order imbalance, provides an estimate of the price change generated by investors' trade interest.

Empirically, we employ a panel model in which one-minute price changes are related to contemporaneous order imbalances and to a new variable that measures ex-ante costs of trading demand. This variable, which we call the implied liquidity cost of the imbalance, is defined as the dollar value of the trade imbalance times the ex-ante execution costs obtained from the stock's limit order book. We find that the implied cost of the imbalance has a contemporaneous, positive effect on price changes. Moreover, this new specification improves the explanatory power of order imbalances as determinants of price changes, illustrating the role of order limit activity on price formation beyond tick-levels. Next, we extend our study and examine the effect of buyer-initiated transactions and seller-initiated ones separately. To this end, we synthesize demand (offer) from the bid (ask) side of the limit order book and construct ex-ante trading costs of buying (selling) orders. Our results indicate that there can be periods in which prices change asymmetrically to buy versus sell trade initiation, with the latter generating almost twice as much the impact than the former.

Several tests are implemented to confirm the positive relation of implied costs of the imbalance and price changes. We verify that this relation exists over different periods of the day and that it is present in firms of different sizes. When we look at the explanatory power of our specification for price changes formed at longer horizons, we find that although model variables are still significant at two, three, five, and ten-minute intervals, their explanatory power markedly decreases with the horizon. This is consistent with the view that pressure from order imbalances is inversely related to the horizon of observations (Chordia et al., 2005). The robustness of our results comes from our panel model that consists of 482 firms present in the S&P500 index and uses observations sampled at a one-minute frequency (390 one-minute intervals per day) over 496 trading days. Moreover, the fact that we aggregate ex-ante trading costs to one-minute intervals allows us to have a measure that is robust to unusual activity happening at tick-level (Cumming et al., 2011).

Our work is part of the emerging literature that uses information from limit order book markets to study market activity. Hasbrouck and Saar (2002) construct order-driven measures such as average depth in the book, proportion of limit orders filled, and duration between execution times, to analyze the effect of volatility on limit order books. Bessembinder et al. (2009) reconstruct the limit order book and compute data estimates of price aggressiveness and order size to study the benefits of order exposure. Hasbrouck and Saar (2013) use order-level data about submissions, cancellations, and executions of orders to measure low-latency activity and to investigate how this variable affects market quality. Ainsworth and Lee (2014) compute market depth from best bid and ask prices to test for the hypothesis that waiting costs impact order choice around a trading deadline. Cao et al. (2009) and Brogaard et al. (2015) propose several summary statistics beyond the best bid and ask to capture features that are amenable for empirical analysis of future price movements at tick level.

This study complements and extends previous works in different dimensions. First, we provide a conceptual framework of the role that ex-ante trading costs have on price changes at the high frequency level. This measure of liquidity contrasts with the standard bid-ask spread since ex-ante trading costs are related to the total average execution price of large orders, while spreads capture the cost of a small market order. We illustrate this difference by looking at the empirical performance of specifications that relate contemporaneous returns and different liquidity measures. We find that the variability explained by spread-based measures is lower than the one associated with execution costs. When we look at the type of information included in the liquidity measure, we observe that measures relying exclusively on best quotes are not always significant in our analyses. This last point shows that gauging the sensitivity of price to depth becomes an important empirical feature for a liquidity measure in order to explain price changes.

Second, as shown by the coefficient of determination associated with the computation of our measure, the ex-ante trading cost subsumes most of the information contained in a snapshot of the limit order book. This characteristic and the fact that it relates to price changes beyond tick-levels make this variable amenable for empirical research. Additionally, this measure can be of particular appeal to investors using strategies that slice and dice large orders into smaller lots and requiring real-time guidance of execution costs.

Third, we provide evidence that beyond tick-level data, information from limit order books is useful to understand price formation in the overall market. This last point is particularly important since fragmentation of the equity market makes traders search for liquidity across venues, introducing frictions from trading in multiple markets. The information extracted from the visible part of the market can be used as a liquidity benchmark to help reduce trading costs incurred in this environment.

We focus our analysis on ex-ante liquidity costs computed from limit order books for two reasons. First, submitting limit or market orders is a dynamic forward-looking decision process in which investors assess the types of risks they will face when placing a specific order. Risks such as non-execution (Handa and Schwartz, 1996) or adverse selection (Glosten, 1994, Copeland and Galai, 1983), to name a few, require investors to submit orders that reflect their expectations about future prices conditional on the current state of the limit order book. To the extent that investors include such expectations in their order submissions, order books provide ex-ante information about general liquidity in the market. Second, algorithmic trading has changed the basic unit of market information, passing from a structure in which trades were the central unit to a new one where orders convey the relevant information (O'Hara, 2015). As such, the development of intraday liquidity measures that rely on order data becomes of paramount importance in the study of microstructure effects on asset prices.

This paper is organized as follows. In the next section, we describe the conceptual framework. Section 3 outlines our methodology. In Section 4, we describe our sample selection and data. Section 5 gives results on how trading liquidity costs affect returns. Section 6 studies alternative liquidity measures and Section 7 presents a set of robustness tests. Section 8 concludes.

2. Conceptual Framework of Liquidity Costs

Limit order books are composed of orders to buy and sell an asset for specific prices and quantities. The price and submission time of a limit order generates a priority in the order book when a market order arrives at time t: market orders are first matched with limit orders at the best bid, b_t, or ask price, a_t, according to time priority. If the size of a market order is bigger than the number of shares available at the best price, the remaining part of the order is matched to the next-best price according again to time priority until completion.

To represent an order book, we consider a dynamic version of Obizhaeva and Wang (2013) with a simple block-shape density, denoted as π_t, representing the number of shares available at every price outside the bid-ask spread. The assumption that the density is constant across the limit order book but time-varying is supported by the empirical evidence provided in Section 4.

A constant price density allows us to picture an order book on a tick-by-tick basis in which supplied and demanded quantities are the same for any given price. In this setting, π_t represents the depth of the market, i.e., the size of the order flow required to shift prices by one dollar. Define x as the signed size of an incoming market order: a positive value of x is a buy order, a negative value of x represents a sell order. In the spirit of Obizhaeva and Wang (2013), the best ask price at time t in the order book immediately after incoming market orders of size x can be written as

$$a_{t+} = a_t + \frac{1}{\pi_t}x, \tag{1}$$

whereas the best bid price at time t in the order book is represented as

$$b_{t+} = b_t + \frac{1}{\pi_t}x. \tag{2}$$

Therefore, if an investor buys x shares at time t she will pay a total dollar outlay of $a_t x + \lambda_t x^2$,[2] where the liquidity parameter λ is defined as

$$\lambda_t = \frac{1}{2\pi_t}. \tag{3}$$

The price per share for a buy transaction of size x, denoted $S_t(x)$, is therefore

$$S_t(x) = p_t + \frac{1}{2}\delta_t + \lambda_t x, \tag{4}$$

in which $p_t = \frac{a_t+b_t}{2}$ represents the midprice and $\delta_t = a_t - b_t$ is the bid-ask spread.[3] Several observations can be made at this point.

First, notice that the price per share given in Equation 4 is composed of the efficient price of the asset (proxied by the midprice) plus two terms that incorporate liquidity costs. The first term reflects the proportional cost associated with the quoted bid-ask spread at time t. This is the cost that traders face when they decide to submit their orders at the current best bid or ask values and it is independent of the order size. The second term captures the cost associated with the execution of the order. It is important to notice that this ex-ante trading execution cost is not necessarily the final cost paid by the investor, as incoming orders could hit non-displayed limit orders. In this case, this term will represent an upper bound of this cost if hidden orders are present in the order book.

Second, the above discussion shows that the principal cost associated with a large transaction depends on the value of λ_t. In this sense, this variable can be regarded as the *marginal cost per share* associated with a transaction. Moreover, it is a measure of illiquidity since it is influenced by the depth of the market: the higher the value of λ_t, the larger the price impact caused by a trade of x shares. When depth is large (small), the coefficient of illiquidity λ_t is low (high) and the average price for a trade is close (far) to the midprice p_t.

Third, the structure proposed for the price density of the order book implies that the price per share is a linear function of the transaction size, so that big orders will negatively impact the final price paid by investors. This negative relation is consistent with how trading strategies are executed in electronic markets, as these strategies slice and dice orders into smaller lots to minimize this effect (O'Hara et al., 2014).

Fourth, the trading execution costs extracted from the order book constitute a new source of information for the deployment of investment strategies through the execution of trading algorithms. Whereas bid-ask spread costs are useful to compute the return of a trading strategy, the trading execution cost captured by λ_t provides a closer look at the final impact of the transaction. This last piece of information becomes more important as trading in electronic exchanges is now mostly performed by trading algorithms that strategically submit orders to the market.

[2]This result follows from Eq. 13 in Obizhaeva and Wang (2013). Similarly, if an investor sells x shares at time t, she will receive a total dollar outlay of $b_t x - \lambda_t x^2$

[3]In a similar way, the price per share when selling x units of stock is $S_t(x) = p_t - \frac{1}{2}\delta_t - \lambda_t x$.

We end this section by noticing that the marginal liquidity cost measured by λ_t gives the additional premium per share that must be paid for a specific transaction. This implies however that λ_t cannot be directly compared across assets as it depends on the value of a share. Given that our objective is to conduct empirical analysis across companies, we need to express the marginal liquidity cost in percentage terms. This can be done by rewriting Equations 1 and 2 in terms of the midprice as

$$\frac{p_{t+} - p_t}{p_t} = 2M_t q_t, \tag{5}$$

where $q_t = p_t x_t$ is the (signed) dollar-value of the trade x_t, and

$$M_t = \frac{\lambda_t}{p_t^2} \tag{6}$$

is the normalized marginal liquidity cost per dollar. Observe that M_t has the desirable property that if there is a split in the stock, its value remains unchanged. Indeed, Equation 5, as opposed to Equation 4, is invariant in terms of the relative value of one unit of the stock. Thus, the variable M_t only depends on the total dollar value of the trade, which is a more useful quantity to compare from one asset to another.

3. Empirical Strategy

The first part of this section presents the empirical approach to estimate the ex-ante trading cost λ_t from snapshots of the order book. It also presents a case study for a specific company to illustrate this methodology. The second part lays out the empirical framework to assess the information content of limit order books beyond the tick level frequency.

3.1. Measuring Marginal Costs from Limit Order Books

The variable λ_t cannot be directly measured from order books since it assumes a continuum of prices. To circumvent this issue, we employ the discrete set of offer and bid prices available in the order book to infer this variable. This procedure is similar to the way Dierker et al. (2016) compute the elasticity of demand and supply schedules from limit-order data to obtain and estimate of this quantity directly.

Specifically, suppose that at a given time t, the best N offer prices are given by $p_1 < ... < p_N$ and the best N bid prices by $p_{-1} > ... > p_{-N}$. Denote the quantity offered at price p_i by x_i, the total quantity available up to the i-th best ask price by X_i, and the average price per share by $\hat{S}_t(X_i)$. Namely,

$$X_i = \sum_{j=1}^{i} x_j, \tag{7}$$

$$\hat{S}_t(X_i) = \frac{\sum_{j=1}^{i} p_j x_j}{X_i}. \tag{8}$$

In a similar way for the bid side, we calculate

$$X_{-i} = \sum_{j=1}^{i} x_{-j}, \tag{9}$$

$$\hat{S}_t(X_{-i}) = \frac{\sum_{j=1}^{i} p_{-j}x_{-j}}{X_{-i}}. \tag{10}$$

Equation 4 states that the average price per share is linearly proportional to the total number of shares composing the transaction. This suggests that we can use the pairs $(\hat{S}_t(X_n), X_n)$ to infer the marginal liquidity cost per share λ_t from the current state of the order book with the following linear regression model

$$\hat{S}_t(X_n) = \hat{\alpha}_t + \hat{\lambda}_t X_n + \epsilon_t, \quad n = -N, ... -1, 1, ..., N. \tag{11}$$

The coefficient $\hat{\lambda}_t$ represents the estimate of the marginal liquidity cost per share from an order limit book.

To obtain an estimate of the marginal liquidity cost over a specified time interval τ, we estimate $\hat{\lambda}_t$ every time there is an update of the order book during this interval. From this set of estimates and the times between updates, we take the time-weighted average of all values of $\hat{\lambda}_{t_j}$ within that interval as an estimate of ex-ante liquidity cost for the interval

$$\tilde{\lambda}_\tau = \sum_{j=1}^{n} \frac{(t_j - t_{j-1})\hat{\lambda}_{t_j}}{\tau}, \tag{12}$$

with n being the total number of order-book updates in the interval τ. Taking time-weighted averages is convenient in this context since it allows us to interpret this average as the prevailing cost over the time interval.

3.1.1. A Case Study: Abbott Laboratories

To illustrate the proposed method, we present a case study for the firm Abbott Laboratories (NYSE ticker: ABT).

Panel A in Figure 1 shows the order book for ABT on January 3, 2011 at 10:00 AM ET, using two different representations. The left figure shows the price p_j and the number of shares available x_j for the 10 best limit orders in the book. The best bid offer consists of 200 shares for a price of \$47.98 per share, whereas the best ask offer is for 600 shares selling for \$47.99 per share, implying a mid-quote price of 47.985. As we walk down (up) the book, we observe different order sizes with prices decreasing (increasing) by a magnitude of 1 cent. A different picture emerges when aggregated orders X_i and average prices per share $\hat{S}_t(X_i)$ are plotted, as depicted in the right figure of Panel A. Contrary to the left figure, we observe a strong linear relation between aggregated orders and average prices. Estimating the model in Equation 11 for this particular order book gives an R-squared (R^2) of 98.29% with $\hat{\lambda}_t = 1.15 \times 10^{-5}$. If liquidity costs were completely absent (if all limit orders were at the midprice), an investor willing to trade $10,000$ shares with a market order would have to pay \$479,850. With liquidity cost, the previous estimates and Equation 4 imply that the investor will have to pay .5 cents per share due to the bid-ask spread, plus 11.5 cents per share due to trading costs, for a total liquidity cost of $\$1,200$. If the investor decides to trade in lots of $1,000$ shares, she would have to pay for each lot the same bid-ask spread cost of .5 cents per share and a trading cost of 1.15 cents per share for the first lot. If liquidity providers are able to quickly re-adjust their inventory and place limit orders for the same prices and quantities soon after a trade (so that best ask price remains the same), then the total liquidity cost of trading 10 lots of $1,000$ shares will be 10 times \$16.5 (\$165). On the other hand, if no new limit orders to sell enter after a trade, the new best ask price after the first 1000-share lot is at \$48.013 ($= \$47.99 + 2 \times 1000 \times \hat{\lambda}_t$) from Equation 1. The bid-ask spread cost per share after the first lot transaction is therefore 2.8 cents, whereas the trading cost remains at 1.15 cents per share, for a total of \$39.50 for the second lot. Overall, trading the 10 lots will generate a total cost of \$1,200. These two different outcomes show that the advantage of splitting orders in smaller sizes resides in the ability of the liquidity providers to balance their inventory and post new limit orders at competitive prices after large trades. This outcome also means that beyond the tick level, trading costs are important determinants of price changes.

Panel B in Figure 1 shows how the marginal liquidity costs fluctuate across the trading day of January 3, 2011. The left figure shows the activity at the opening (9:30 AM to 9:45 AM), while the one on the right depicts the evolution across the rest of the day (9:45 AM to 4:00 PM). A total of 16,907 order-book updates were recorded during the day, producing the same number of estimates for $\hat{\lambda}_t$. We observe large spikes at the opening, with values going beyond 10^{-4}, that tend to stabilize as the day goes along. Large marginal liquidity costs during the opening are consistent with hefty demand for liquidity during this period as investors rally to position themselves in the market. Once overnight information is impounded at the opening, liquidity demand decreases. This variability is different from the one shown by the bid-ask spread, which tends to fluctuate around the minimum tick size across the day for highly traded stocks.

3.2. Model Specification

Equation 5 gives the immediate impact on midprices due to an incoming market order. On larger time horizons, the interactions between different agents and their need for liquidity create a more complex relation between order imbalances and returns. To analyze the relation between trading activity and stock price movements beyond tick-by-tick levels, we hypothesize that the relation between price changes and orders over the time interval $(t-1, t]$ satisfies:

$$R_t = \Lambda_t Q_t + \epsilon_t, \tag{13}$$

where R_t is the return, Q_t is the signed dollar volume over the interval $(t-1, t]$ and ϵ_t is an error term, whereas Λ_t represents the price impact coefficient over this time interval.

The model described in Equation 13 requires the specification of Λ_t. We hypothesize that this impact is related to the ex-ante trading costs embedded in the order book and assume that it can be written as a constant term plus the normalized marginal cost per dollar over the interval,[4] that is,

$$\Lambda_t = \beta + \gamma M_t. \tag{14}$$

In the previous equation, the constant β captures the average impact on prices and the coefficient γ measures the effect induced by fluctuations in supply/demand at time t, as quantified by the ex-ante trading costs M_t.

Combining Equations 14 with 13, the impact on returns is expressed as:

$$R_t = \beta Q_t + \gamma M_t Q_t + \epsilon_t. \tag{15}$$

The first term in the right-hand side of Equation 15 states that stock price changes are proportional to the signed volume (order flow) over an interval of time, as proposed in Chordia and Subrahmanyam (2004). The second one captures price changes coming from the combined effect of order flows and price impacts associated with trading demand. This specification extends the idea that prices move only because of buying and selling pressures —the order flow —to one in which the lack of liquidity in the market, as measured by the marginal liquidity costs, also contributes to these changes.

3.2.1. Panel Regressions

In order to test this theoretical framework, we employ a panel of intraday observations. Let us start by defining $\$OIB_{i,t}$ as the dollar order imbalance over the interval of time $(t-1, t]$ for asset i and $R_{i,t}$ as the return over this interval. To measure the component related to trading costs (second term in Equation 15), we define the implied liquidity cost at time t, $ILC_{i,t}$, as the normalized liquidity cost times the dollar order imbalance, i.e., $ILC_{i,t} = M_{i,t} \times \$OIB_{i,t}$. As explained in Section 2, the normalized liquidity cost variable $M_{i,t}$ is obtained by dividing the ex-ante liquidity cost $\tilde{\lambda}_t$ of Equation 12 by p_t^2.

Since our objective is to measure the distinct impact of trading demand beyond that of order flow, we start by regressing the implied cost of the imbalance on the dollar order imbalance, that is:

$$ILC_{i,t} = \delta \times \$OIB_{i,t} + ILC_{i,t}^{\perp \$OIB}, \tag{16}$$

where $ILC_{i,t}^{\perp \$OIB}$ is the orthogonal component of $ILC_{i,t}$ to $\$OIB_{i,t}$ and δ denotes the coefficient of the orthogonal projection on the order imbalance.

Then, we employ the two orthogonal variables $ILC_{i,t}^{\perp \$OIB}$ and $\$OIB_{i,t}$ to explain the contemporaneous asset return over the intra-day interval t by running the panel regression

$$R_{i,t} = \beta \times \$OIB_{i,t} + \gamma \times ILC_{i,t}^{\perp \$OIB} + \epsilon_{i,t}, \tag{17}$$

where $\epsilon_{i,t}$ is an i.i.d shock with zero mean.

The coefficient β in the above regression captures the overall market-averaged impact of the order imbalance on price changes. As argued in Chordia and Subrahmanyam (2004),

[4] In Section 6 we study different specifications that contain other ex-ante measures of liquidity costs obtained from the limit order book.

a positive coefficient reveals the presence of price pressures in the stock market due to inventory effects. The key coefficient of interest in our study, γ, measures the complementary impact associated with trading demand.[5] Under the null hypothesis, imbalances can be perfectly rebalanced by liquidity providers so that trading demand has no effect on prices. If this hypothesis is rejected, a positive and significant γ would indicate that liquidity providers schedule their trading positions as a function of variations in trading demand: high liquidity cost require a slow re-adjustment of the imbalance to minimize the total trading cost, producing a temporal price pressure effect on subsequent prices. This result would also indicate that limit order books provide useful information about trading costs that affect market prices.

4. Data and Descriptive Statistics

4.1. Sample Selection

We analyze a sample of stocks representative of the overall movement in the U.S. stock market. This sample is composed of stocks trading in the NYSE ArcaBook that belong to the S&P500 index in the period from January 1, 2011 to December 31, 2012. We conduct our analysis in parallel over these two years since they were marked by different economic events. The year 2011 was characterized by extreme economic uncertainty, powered by fears of recession in the U.S., the credit rating downgrade of the U.S. debt, and a mounting debt crisis in the Eurozone. All these effects generated important trading activity in the market. Contrary to the year 2011, the year 2012 showed signs of recovery, particulary in the U.S., with the S&P500 increasing by 13% during that year, stimulated in part by the Fed's decision to support the economy with spending programs and promises to keep low rates through 2014.

Given that our objective is to analyze the interactions between liquidity and price formation, we conduct our study with intraday data at a one-minute horizon. We choose this frequency as we expect it to be high enough to capture the effects of trading activity on liquidity, but sufficiently low so that prices impound information related to the trading process and are not distorted by the microstructure of the exchange in which the asset is trading.[6] Eliminating holidays and days with an early close, the final sample includes 482 stocks with 249 trading days in 2011 and 247 trading days in 2012. This provides a rich panel of 46,507,006 firm-period observations in 2011 and 46,266,376 in 2012.

Information about the limit order book comes from the TAQ NYSE ArcaBook historical database obtained from NYSE Market Data. It provides all timestamped messages disseminated through the NYSE ArcaBook, including all limit orders entered, removed and modified before, during and after trading hours. Each order contains a timestamp (millisecond), a price, a quantity and a buy or sell ID. It also has a unique identifier that allows

[5] A similar investigation of price pressures beyond the size of the imbalance was considered by Chordia et al. (2008), in which the authors construct a liquidity dummy variable based on effective spread to highlight the difference in impact of imbalances between days of low liquidity and days of high liquidity.

[6] In Section 5, we also employ two, three, five, and ten-minute intervals to analyze the strength of the results.

us to follow an order from its creation through to its modifications and deletion. An order can be modified by its initiator or modified due to a partial fill. An order is deleted when it is canceled by its initiator or when it is matched to a market order. To reconstruct the order book, we follow each order from its creation by recording its limit price and associated quantity. At each point in time, we sort buy and sell orders according to prices and add up quantities of limit orders with equal prices.

We extract intraday data from the NYSE Trade and Quote database with millisecond timestamps (DTAQ). This database contains trades and quotes originating from all trading platforms in the US market. We then apply the following filters to the data:

1. a trade is included in any given period if its field *CORR* in the TAQ database has a value of 0 or 1, and the field *COND* a value of either blank, @, or F;
2. a trade is excluded if its price is greater than 150% or less than 50% of the previous trade price;
3. a quote is excluded if its bid price is higher than the ask price;
4. data for ticker symbols ALTR, AMZN, AAPL, QGEN, and CECO is excluded on November 7, 2011, since NASDAQ reported market crossings in the bid and ask prices for these companies; and
5. only trades and quotes occurring during regular market trading hours of 9:30 AM to 4:00 PM are considered.

Finally, monthly market capitalization information about these companies is obtained from the Center for Research in Security Prices (CRSP). We define three market size categories: small size for firms with market capitalizations, calculated on January of each year, that are equal to or less than the market capitalization of the first tercile of companies in the sample for a given year; large size for firms with market capitalizations higher than or equal to the third tercile; and mid-cap size for all others.

4.2. Variable Definitions

We work with three categories of variables: returns, trading activity variables, and liquidity costs. Midpoints are obtained from The National Best Bid and Offer (NBBO), from which returns (*returns*) are computed over intraday time intervals; this definition of returns avoids serial dependence induced by bid-ask bounces. Variables related to trading activity are extracted from the DTAQ database and include a) volume, as measured by the total number of trades (*Volume (Trades)*) and total dollars traded (*Volume (Dollars)*), b) order imbalances, given by the number of buyer-initiated trades less the number of seller-initiated trades (*#OIB*) and the dollars paid by buyer-initiators less the dollars received by seller-initiators (*$OIB*), and c) the magnitude of the imbalance, computed from the absolute value of both measures of order imbalance. To determine the initiator of a trade, we apply the Lee and Ready (1991) procedure. Note that the previous variables are constructed from trade and quote information from all exchange venues.

Liquidity cost variables are computed from the ArcaBook limit order book. For each company in our sample, we compute the marginal liquidity cost (Equations 11 and 12) over a one-minute interval using the 10 best bids and asks prices in the order book. To assess the quality of the model fitted in Equation 12, we record the R-squared of the model at a single point of time, and time-average it over one-minute intervals. Figure 2 shows three percentiles associated with the one-minute time-averaged R-squared that result from the computation of the marginal liquidity cost (Equation 11) across the whole sample. This figure shows that a linear model provides a very good description of order-book snapshots. The cross-sectional average and the 25^{th} percentile are above 90% across the day, whereas the first percentile lies above of 80% for most of the day. Overall, the quality of the approximation offered by the reduced-form representation of the limit order book is remarkably high, taken into account the complexity underlying the interactions that drive limit order markets.

4.3. Descriptive Statistics

Table 1 presents various descriptive statistics. The left column presents results for the year 2011 and the right column for 2012. The statistics are calculated from one-minute intervals for all firms in the sample (Panel A) and according to market capitalization (Panel B to Panel D). Returns at this frequency are characterized by an average value close to zero and a standard-deviation that is slightly less than 10 basis points for 2011 and eight basis points for 2012. The average number of trades is 31 and 23 (390,000 and 340,000 in dollar trades) for 2011 and 2012, respectively. We find that, on average, order imbalances (in trades and in dollars) are close to zero, which shows that liquidity providers are able to accommodate overall imbalances. Nonetheless, standard deviations and percentiles of order imbalances show the existence of large imbalances. These imbalances are substantial, as can be seen by the average absolute value of order imbalances (expressed both in the number of trades and in dollars) representing about 25% of the average volume traded over a one-minute interval. Volume and order-imbalance measures reflect a slight decline in trading activity from 2011 to 2012.

Regarding trading liquidity costs, the mean of the normalized liquidity cost is 1.12×10^{-8}. In economic terms, this value means that if an investor wants to trade one million dollars, the trading cost associated with the trade would be of \$11,200, that is, 1.12% of the total value of the transaction.[7] With this interpretation in mind, we observe that this liquidity cost increased on average from 2011 to 2012.

Table 1 is also informative about the size effect on trading activity and liquidity costs. We observe that trading activity, measured from volume order imbalance, decreases with the size of the firm, and that this difference persists over the two sample periods. Conditioning on market size reveals that trading liquidity costs vary with this variable, with small capitalization firms presenting a larger liquidity cost on average. To put it into perspective, an investor wanting to trade one million dollars of a small firm in 2011 will have to pay on average 137 basis points more than one trading the same dollar amount on a large firm. We investigate the potential effect of size differences later in Section 7.

[7]Similarly, for a trade of \$100,000, the liquidity cost would be \$112 or 0.112% of this trade. These costs are obtained from multiplying the normalized liquidity cost by the squared of the order size in dollars.

4.4. Characteristics of Liquidity Costs and Trading Activity

We now turn our attention to characteristics about trading liquidity costs and trading activity. In Figure 3, we provide the daily time-series pattern of the cross-sectional mean of normalized liquidity costs for 2011 and 2012. First, observe that these costs vary across time with sporadic spikes throughout the sample period. For instance, the downgrade of U.S. credit rating by Standard and Poor's on August 5, 2011 had a noticeable impact on the dynamics of this variable. Consistent with statistics provided in Table 1 for this variable, we observe an increasing trend over the 2012 period, especially by the end of the year, possibly fueled by uncertainty about the U.S. fiscal cliff and its impact on the economy.

In Figure 4, we illustrate the trading activity at a one-minute horizon across the day for 2011 (Panel A) and 2012 (Panel B) by market capitalization. This figure shows the cross-sectional average of dollars traded for each of the 390 one-minute intervals averaged across days. Consistent with empirical evidence (Jain and Joh, 1988; Foster and Viswanathan, 1993), we observe a U-shaped pattern for the intraday volume for both years, with high activity at the beginning and at the end of the day. The large trading activity around the opening could also be attributed to macroeconomic announcements frequently released before market opening and around 10:00 AM.

Figure 5 depicts the cross-sectional mean of the normalized liquidity cost for each of the 390 one-minute intervals, averaged across days, for 2011 (Panel A) and 2012 (Panel B) by market capitalization. In this case a different pattern emerges: liquidity costs remarkably decline from the opening to the close of trading, generating an L-shaped curve whose level is inversely related to the firm's market capitalization. This large decline at the opening could be associated with the level of trading activity observed over this period, which rapidly dries up liquidity. In contrast to what is observed at the opening, liquidity is less affected by the trading activity happening near the close. This suggests that liquidity costs computed from limit order books follow different dynamics to those associated with trading activity.[8] Whether this difference is important from a price formation perspective constitutes the objective of the analysis in the next section of the paper.

5. Interactions between Trading Activity and Liquidity Costs

The preponderance of either buyers or sellers in the market can transitorily alter returns as liquidity providers try to adapt to a specific pattern of trades. In this section we present two different views of this interaction. The first one consists of contemporaneous regressions of returns on order imbalances and trading liquidity costs associated with these imbalances. This specification allows us to understand the components that affect price changes and their relative contribution in the formation of prices. The second specification relies on predictive regressions, which shed light on how price pressures caused by lagged order imbalances and trading liquidity costs affect future prices at the intraday level. Finally, we look at

[8]This pattern is also different from the U-shaped pattern found in bid-ask spreads by Brock and Kleidon (1992) and McInisih and Wood (1992). This difference provides preliminary evidence of the complementary information found in limit order books, compared to that contained in best quoted prices.

determinants of trading liquidity costs, in particular, at interactions of the variable M_t and several market variables.

5.1. Contemporaneous Regressions

Table 2 focuses on three distinct model specifications to understand one-minute returns. The first specification regresses stock returns on contemporaneous dollar order imbalances ($\$OIB$). Consistent with the evidence documented in Chordia and Subrahmanyam (2004), we observe a positive relationship between stock returns and order imbalances that is significant at the 1% level for both 2011 and 2012. The economic significance of this coefficient is also evident: an imbalance shock of one standard deviation impacts the return by an amount of 2.15 basis points in 2011,[9] whereas in 2012 the impact is 1.34 basis points. This impact represents 23.4% of the return standard deviation in 2011 and 18.1% in 2012. Regarding the coefficient of determination resulting from this regression, we observe that order imbalances explain 5.486% of the return variability in 2011 and that this value decreases to 3.287% in 2012.

The second specification looks at the relation between contemporaneous returns and the implied liquidity costs (ILC) associated with the imbalance. We find a positive link between this measure and the contemporaneous return that is significant at the 1% level. The economic significance of this variable is also evident, with an impact of 3.24 (2.36) basis points on one-minute returns that represents 35.2% (25.6%) of the return standard deviation for 2011 (2012). It is also clear that ILC is able to explain a larger variability of the contemporaneous returns, with an adjusted R-squared of 12.45% in 2011 and 10.19% in 2012.

Having seen the individual effect of order imbalances and implied liquidity costs on contemporaneous returns, we now proceed to assess the joint contribution of $\$OIB$ and ILC on the determination of stock returns. To this end, we use Equations 16 and 17 to obtain orthogonalized components that isolate liquidity cost effects from those of the order imbalance. We observe that the coefficients of both components remain positive and significant at the 1% level, with a modest improvement in terms of adjusted R-squared relative to the second specification. Note also that the accuracy of the estimated $\$OIB$ coefficient is substantially enhanced, with the corresponding standard error cut by half after the inclusion of the implied liquidity variable in the model.

To measure the extent to which the goodness-of-fit provided by the third specification is stable across periods, we look at the performance of this specification on a monthly basis relative to the one in which only order imbalances explain contemporaneous returns. Accordingly, we estimate both models for every month in our sample and obtain the coefficients of determination for each model. Figure 6 presents the coefficient associated with these models for both years, as well as the ratio between the adjusted R-squared of Model 3 and the adjusted R-squared of Model 1. We observe that the adjusted R-squared of Model 3 is almost always more than two times the coefficient of the first specification, and that

[9] This value can be obtained by multiplying the coefficient of $\$OIB$ in specification 1 of Table 2 and the standard deviation of $\$OIB$ from Table 1

this ratio is relatively more important in 2012 as compared to the one observed in 2011. These results suggest that the proposed implied liquidity cost contains relevant information beyond that provided by the $\$OIB$ variable.

The results obtained for 2012, as compared to those of 2011, seem to indicate a decline in the explanatory power of the model. However, notice from Table 1 that 2011 was characterized by larger imbalances than 2012. This points towards a possible explanation for the decline: the performance of the model is related to periods experiencing large imbalances. We explore this explanation by working with intervals of time that experienced large order imbalances, as measured by absolute values. Panels B, C, and D in Table 3 show results for intervals experiencing the largest order imbalances. We form these subsamples by computing for each interval the average absolute order imbalance ($|\$OIB|$) and then selecting intervals with values greater than a specific percentile (75^{th}, 90^{th}, and 95^{th}) computed over all days in a given year. We observe that the goodness-of-fit of the model increases for intervals with extreme order imbalances.

Table 3 also supports the empirical evidence that large order imbalances have marginally less contemporaneous impact on returns than small order imbalances. Indeed, in all three models coefficients decrease as the magnitude of $\$OIB$ increases, giving rise to a concave price impact function in terms of order imbalance size. This finding is also consistent with the notion that investors divide larger traders into smaller ones that are spread out in time to take advantage of the ability of liquidity providers to absorb inventory imbalances, even at the one-minute horizon. Indeed, as shown in Section 3.1.1, the marginal price impact of a large trade would be the same regardless of how it is divided up if liquidity providers were not able to absorb inventory imbalances.

5.2. Results According to Trade Side

The evidence presented so far indicates that seller- (or buyer-) initiated trades provide information about excessive investor interest in a stock. In a market experiencing a large imbalance, the capacity of liquidity providers to accommodate subsequent orders without disturbing prices depends on the depth of the market. For instance, low depth and a large imbalance imply that liquidity providers will struggle to readjust their inventory without incurring significant costs. We now examine whether price changes could response differently to the side of a trade.

Since our measure of liquidity costs intrinsically assumes that marginal price impacts are on average equal for sell and buy sides, we disentangle these costs by running the regression in Equation (11) separately for bid and ask limit orders. To conduct this analysis we split our sample by trade initiation, so intervals with buyer- (seller-) initiated order imbalances are matched with ex-ante trading costs computed from ask (bid) prices.

Table 4 provides results by discriminating on the trade side of the imbalance. We observe that the impact of the order imbalance is approximately the same across both years. On the contrary, for the year 2011, we observe an asymmetric response of price changes to the implied illiquidity costs of the imbalance. For this year, the magnitude of coefficient estimates associated with ILC for sell-initiated imbalances is almost twice the value observed for buy-initiated imbalances. In unreported results, we observe that the standard deviation of the imbalance does not change significantly between buy and sell initiated trade intervals, suggesting that this statistical difference is also economically significant. For 2012, the coefficient of ILC is larger for sell-initiated imbalances, but the difference is no longer significant. As argued in the previous section, 2011 was characterized by larger imbalances than 2012, which could explain why prices in 2012 appear to respond similarly to buyer or seller initiated imbalances.

5.3. Determinants of Trading Liquidity Costs

Now that we have established the explanatory power of the variable ILC, we study the determinants of the dynamics of the liquidity cost variable M_t sampled at a one-minute frequency level. To this end, we look at interactions of market variables with ex-ante liquidity costs.

Table 5 presents regressions of M_t on market variables such as order imbalance, dollar volume, and volatility. Volatility is measured by the range, calculated as the difference between the maximum and minimum trade prices over a given one-minute interval divided by the average price during that interval. We provide separate contemporaneous and predictive (lagged) regressions for 2011 and 2012. However, these two types of regressions do not present appreciable differences given the high persistence of volume and volatility. The regressions provide evidence that volume contributes positively and volatility contributes negatively to liquidity in a significant way (recall that M is a measure of illiquidity). On the other hand, there is no evidence that order imbalance has an impact on illiquidity. The positive effect of volume on liquidity is well documented in the literature (Demsetz, 1968), and it can be explained by the fact that high volumes imply high interest on the asset, more competitive quotes, and lower trading costs. Likewise, the negative impact of volatility on liquidity is also an empirical regularity (Amihud and Mendelson, 1980; Pastor and Stambaugh, 2003) that can be justified by the assumption that as volatility increases, inventory risk increases, which influences the quoting process and the risk premia required by liquidity providers and market makers.

6. Alternative Liquidity Cost Measures

Thus far, we have focused the analysis on marginal costs associated with the ex-ante liquidity in the limit order book. We have seen that the effect of this cost on contemporaneous returns is positive and that it improves the explanation power of order imbalances. We now assess how these results compare to other liquidity cost measures.

To this end, we augment the specification of Equation 14 by introducing other ex-ante trading cost variables. We guide our choice by looking at the components of the price per share discussed in Section 2. On a tick-by-tick basis, the price per share is a function of the asset's efficient price plus two terms related to liquidity costs: the spread and the marginal cost per share. Thus, to capture price pressure effects of spreads over longer intervals of time, we add to Equation 14 a term related to this component:

$$\Lambda_t = \beta + \gamma M_t + \eta D_t, \tag{18}$$

where D_t represents a measure of spread at time t. The associated coefficient η captures the effect induced by this quantity. Combining Equation 18 with 13, we obtain a new specification for the impact of trading activity on returns as:

$$R_t = \beta Q_t + \gamma M_t Q_t + \eta D_t Q_t + \epsilon_t. \tag{19}$$

This specification extends the one in Equation 15 by adding a new term that depends proportional liquidity costs, as measured by spreads.

We consider three measures of the spread based on different levels of information. The first one corresponds to the *proportional quoted spread* ($QSPR$), defined as the difference between the ask and bid price divided by the current midpoint price, all from the NBBO. This measure aggregates information only from best quote prices, leaving aside any information about quantity and limit orders behind best prices. To include the latter in the calculation of spreads, we follow Cao et al. (2009) and define the *weighted average midquote* (WP^1) and the *weighted price beyond best quotes* ($WP^{2,10}$) as

$$\begin{aligned} WP_1 &= \frac{x_1 p_1 + x_{-1} p_{-1}}{x_1 + x_{-1}}, \\ WP_{2,10} &= \frac{\sum_{i=2}^{10} x_i p_i + x_{-i} p_{-i}}{\sum_{i=2}^{10} x_i + x_{-i}}, \end{aligned}$$

in which x_i (x_{-i}) denotes the total quantity offered (demanded) at price p_i (p_{-i}) in the order book, as in Section 3.1. These two measures generalize the calculation of the midquote by weighing the best bids and offers with the associated quantities available in the order book. Based on these prices, we then define two weighted spread measures in terms of WP_1 and $WP_{2,10}$ as

$$\begin{aligned} WSPR_1 &= \frac{|WP_1 - MID|}{MID}, \\ WSPR_2 &= \frac{|WP_{2,10} - MID|}{MID}, \end{aligned}$$

in which MID is the midquote, as measured by the NBBO. To obtain the value of one of these spread variables over a given interval of time, we compute time-weighted averages of observations over the interval. The proportional costs measured by these two variables are the same as those of $QSPR$ when prices and quantities are the same across the order book. When imbalances are present, these costs will be different, as $WSPR_1$ captures liquidity aspects from best quotes only and $WSPR_2$ does so for quotes beyond the best.

We adapt the panel regression of Section 3.2 to include information from spread variables. Specifically, we use Equation 19 to define the following panel model of asset returns:

$$R_{i,t} = \alpha + \beta \times \$OIB_{i,t} + \gamma \times ILC_{i,t}^{\perp \$OIB} + \eta \times DC_{i,t}^{\perp \$OIB} + \epsilon_{i,t}, \tag{20}$$

This regression adds to the base model a new component that measures the trading costs of the imbalance based on the spread: $DC_{i,t} = D_{i,t} \times \$OIB_{i,t}$, where D_{it} represents any of the three spread measures presented. To measure the impact of this variable beyond that of order flows, we employ the orthogonal component that comes from the projection of this quantity on the order imbalance.

Table 6 contains regression results for 4 specifications of Equation 20 using 2011 and 2012 sample periods separately. Panel A reports estimates based on $QSPR$, Panel B on $WSPR_1$, and Panel C on $WSPR_2$. Specification 2 in the panels only includes $DC_{i,t}$ and shows that implied liquidity costs based on different definitions of spreads are positive and significantly related to contemporaneous price changes over one-minute intervals. The economic significance of these variables varies with the sample period. For 2011, the impact of spread cost based on $QSPR$, $WSPR_1$, and $WSPR_2$ are 3.37, 2.62, and 3.49 bps, respectively. For 2012, the impacts are higher for all variables, with 5.96 for $QSPR$, 8.29 for $WSPR_1$, and 10.01 for $WSPR_2$.

A noticeable difference with results for ILC (Specification 2 in Table 2) is the R-squared of the regressions. From Table 2, the variability explained by ILC alone is always higher than those obtained from spread-based measures. One reason for this difference could be the fact that ILC employs a measure that responds to changes in the shape of the book, thus capturing the dynamics of imbalances between offer and supply that generate price pressures. Spread measures like $QSPR$ and $WSPR_1$ are less likely to capture these pressures since information at best prices does not allow to gauge the sensitivity of price to depth. It is interesting to observe that the variability explained by the spread measure $WSPR_2$ is similar in magnitude to the other spread-based measures, even though it employs information from the order book beyond the best bid and offer. This last observation illustrates the empirical difference of working with marginal costs as opposed to proportions (spreads), since the former costs relate directly to the elasticity of supply/demand curves, thus better reflecting high limit order activity.

We assess the joint contribution of $DC_{i,t}$ and $\$OIB$ in specification 3. In this model, the orthogonal components that isolate liquidity costs from order imbalance effects are employed as regressors. We observe that coefficients for $QSPR$ and $WSPR_1$ remain positive but are no longer significant for 2011, contrary to coefficients for $WSPR_2$, which remain significant and positive in both years. Thus, our results indicate that there exists important information beyond that found at top of the book.

Finally, specification 4 in Table 2 investigates spread effects when ILC is added to the model. The coefficient of ILC remains positive and significant for both years, while coefficients for $QSPR$ and $WSPR_1$ behave the same as those in specification 3. Regarding the coefficient of $WSPR_2$ in Panel C, we observe that its sign changes in 2011 and it is no longer significant in 2012. In unreported results, we find that there in an important level of correlation between ILC and DC based on $WSPR_2$, pointing to a multicolinearity in these variables. This level of commonality is not surprising since both liquidity costs use the same limit order information beyond best quotes.

7. Further Analysis and Robustness Tests

In this section, we illustrate the interaction of price changes, order imbalances, and ex-ante liquidity costs under different specifications. We run our base regressions at lower frequency intervals, look at lagged effects, and perform several robustness analyses across dimensions such as firm size and period of the day. Throughout these analyses, we employ the framework described in Section 3.2.

7.1. Lower Frequencies and the Lagged Effect of Order Imbalances

We now focus on the evaluation of the impact of order imbalances and liquidity costs on returns aggregated on longer time horizons. In Table 7 we present estimation results for returns computed over intervals of one, two, three, five, and ten minutes for the three model specifications previously considered. Observe first that the marginal effect of the regressors weakens at lower frequencies from an economic perspective (smaller magnitude of estimated coefficients) as well as from a statistical standpoint (smaller t-test statistics). The decreasing coefficients suggest that the pressure induced by imbalances is less strong when aggregated over larger intervals of time. This is the effect of the trade scheduling by liquidity providers who take some time to re-adjust the inventory as a function of the liquidity costs and arbitrage away any serial dependence remaining after prices adjust to their new equilibrium values.

Over larger intervals of time, liquidity providers have time to re-balance their inventory so that the impact of large order imbalances is less perceptible and the model coefficients are less significant. Nonetheless, liquidity effects —related to both $\$OIB$ and ILC —remain significant up to ten minutes and are sizeable over the entire sample period, even though they are less apparent in 2012 that in 2011. Notice also that as the time scale increases, there is a reduction across the board in the goodness-of-fit of all models. For instance, one-minute regressions yield adjusted R-squared values that are about 10 to 20-fold higher than their counterparts from five-minute and ten-minute regressions, respectively.

The loss in model explanatory power induced by the time aggregation of returns seems robust across various specifications and sample periods, and goes in line with the empirical evidence in Chordia et al. (2005). The above interpretation also implies that liquidity providers took less time to re-adjust their inventory in 2012 than it was in 2011, i.e., the time required to converge back to equilibrium prices is shorter. Indeed, although both series of coefficients decrease as the length of the time interval increases, the coefficients of 2012 are systematically lower by approximately a ratio of 2:1 than those of 2011. Furthermore, a comparison of panels of Table 7 leads us to determine that a given dollar order imbalance has a similar impact on returns over a two-minute interval in 2011 than over a one-minute interval in 2012, providing evidence that the liquidity providers were faster in 2012 to absorb the impact of larger orders. Note also that the model coefficients are systematically lower in 2012 for all specifications and robustness tests associated to Equation 17.

Above we showed that order imbalances and liquidity costs significantly explain contemporaneous stock returns. Nonetheless, Chordia and Subrahmanyam (2004) show that the lagged effect of order imbalances on stock returns is also significant. In Figure 7, we see that order imbalances as well as their associated liquidity costs are autocorrelated at least up to lags of ten minutes, suggesting that some traders are spreading out their large orders over time to minimize the price impact of their trades and thus take advantage of the market's ability to revert back to equilibrium prices. This autocorrelation pattern has repercussions on the price formation mechanism since imbalances and their rebalancing costs become partly predictable by liquidity providers, so that a fraction of liquidity premia is already included in the price impact of time $t-1$. All this implies that, conditional on past values, one should observe smaller contemporaneous price impacts.

Table 8 presents panel return regressions that include contemporaneous imbalances and 10 lags of order imbalances. This table provides evidence that liquidity providers take into account the autocorrelation of order imbalances to manage their inventory risk since the regression coefficients for lagged imbalances are negative. Note, however, that the significance of the lagged variables decreases in 2012, supporting the idea that the liquidity providers took less time to re-adjust their inventory after large orders in 2012 than 2011.

7.2. Results According to Market Capitalization

Additional estimation of the model for different firm sizes allows us to better understand the role of order imbalances and their ex-ante rebalancing costs, as well as to address some possible empirical concerns. We build at the beginning of each year three bins of firms (small, medium, large) using the first and third market capitalization terciles as cut-off points. Table 9 displays the estimation results of the alternative model specifications for each market capitalization group.

Observe first that the corresponding slope coefficients remain both statistically and economically significant for all three market-capitalization groups. This suggests that the proposed implied liquidity measure is a major driver of returns across market capitalization levels.

We can gain further intuition about the differences between price pressures coming from order imbalances and their implied rebalancing costs, by looking at the variability of coefficient estimates among different groups of firms. Small firms (panel B) have larger $\$OIB$ coefficients compared to those of medium (panel C) or large firms (panel D). For instance, in the third column of 2011, this coefficient declines from 37.777×10^{-10} for the smallest firms to 7.997×10^{-10} for the largest ones. On the contrary, coefficient estimates for ILC vary much less across firm sizes. As argued before, large order imbalances have marginally less contemporaneous impact on returns than small order imbalances, which could explain why the price pressure for large firms—they have larger absolute order imbalances—is smaller than that for smaller firms. On the contrary, price pressure coming from liquidity is less affected by the size of the firm, as the marginal impact is more closely related to prevailing market conditions.

7.3. Results Related to Time of the Day

In this subsection, we analyze some of the results previously discussed for different trading periods. Jain (1988) and Foster and Viswanathan (1993) document intraday patterns of volume and liquidity costs that can be related to different trading dynamics during the trading day. In addition, many scheduled macroeconomic news are released during the first trading hour. For instance, the Consumer Confidence Index and Retail Sales are announced at 10:00 AM and 8:30 AM, respectively. To assess whether these patterns have an effect on the relation between returns and order imbalances, we split the trading day between the first hour and the rest of the trading day and report these results in Table 10.

We observe that our results are consistent across trading periods. This suggests that our measure of trading liquidity cost embodies relevant information that is not contained in the standard $\$OIB$ measure. The relation between this measure and returns is present and robust during different trading periods across the day.

8. Conclusion

In this paper, we provide a method to compute ex-ante liquidity costs from limit order books. We show that these costs are important determinants of price changes at intraday frequencies. These costs have non-trivial dynamics, they are negatively related to volume activity and positively to volatility. Overall, our work provides strong support for the idea that information about liquidity in the visible part of the limit order book is useful to characterize liquidity impacts.

By investigating different periods of the trading day, firm sizes, and sides of the imbalance, we demonstrate that this mechanism is robust and provides new information beyond the order imbalance alone. Moreover, the economic significance in panel regression reveals that both order imbalances and their rebalancing costs have considerable impacts on intraday returns.

The intraday panel model that we exploit for identification allows us to further explore the impact of order imbalances and liquidity costs on returns aggregated over different intervals. Our results suggest that the pressure induced by imbalances is less strong when aggregated over larger intervals of time, consistent with the view that liquidity providers take some time to re-adjust inventory as a function of the prevailing liquidity costs. We find that liquidity effects caused by imbalances remain significant up to ten minutes. However, these impacts are far less substantial than those observed for one-minute intervals.

The proposed ex-ante liquidity measure can help solve some of the complexities introduced by market fragmentation. Since there is no central market, traders need to search for liquidity across markets, thus incurring search costs. As a by-product, this study provides a liquidity benchmark for traders in the market that could help reduce search frictions originating from the information asymmetry associated with multi-venue trading.

References

References

Ainsworth, A., Lee, A., 2014. Waiting Costs and Limit Order Book Liquidity: Evidence from the Ex-dividend Deadline in Australia. Journal of Financial Markets 20, 101–28.

Amihud, Y., Mendelson, H., 1980. Dealership Market: Market-Making with Inventory. Journal of Financial Economics 8, 31–53.

Bessembinder, H., Panayides, M., Venkataraman, K., 2009. Hidden Liquidity: An Analysis of Order Exposure Strategies in Electronic Stock Markets. Journal of Financial Economics 94 (3), 361–383.

Brock, W., Kleidon, A. W., 1992. Periodic Market Closure and Trading Volume: a Model of Intraday Bids and Asks. Journal of Economic Dynamics and Control 16 (3), 451–489.

Brogaard, J., Hendershott, T., Riordan, R., 2015. Price Discovery without Trading: Evidence from Limit Orders. Working paper.

Cao, C., Hansch, O., Wang, X., 2009. The Information Content of an Open Limit-Order Book. The Journal of Futures Markets 29, 16–41.

Chordia, T., Roll, R., Subrahmanyam, A., 2005. Evidence on the Speed of Convergence to Market Efficiency. Journal of Financial Economics 76, 271–292.

Chordia, T., Roll, R., Subrahmanyam, A., 2008. Liquidity and Market Efficiency. Journal of Financial Economics 87, 249–268.

Chordia, T., Subrahmanyam, A., 2004. Order Imbalance and Individual Stock Returns: Theory and Evidence. Journal of Financial Economics 72, 485–518.

Copeland, T., Galai, D., 1983. Information Effects on the Bid–Ask Spread. Journal of Finance 38, 1457–1469.

Cumming, D., Johan, S., Li, D., 2011. Exchange Trading Rules and Stock Market Liquidity. Journal of Financial Economics 99, 651–671.

Demsetz, H., 1968. The Cost of Transacting. Quarterly Journal of Economics 82, 33–53.

Dierker, M., Kim, J., Lee, J., Morck, R., 2016. Investor's Interacting Demand and Supply Curves for Common Stocks. Review of Finance 20 (4), 1517–1547.

Fleming, M., Mizrach, B., Nguyen, G., 2017. The Microstructure of a U.S. Treasury ECN: The BrokerTec Platform. Journal of Financial Markets, Forthcoming.

Foster, F., Viswanathan, S., 1993. Variations in Trading Volume, Return Volatility, and Trading Costs: Evidence on Recent Price Formation Models. Journal of Finance 48, 187–211.

Glosten, L., 1994. Is the Electronic Open Limit Order Book Inevitable? Journal of Finance 49 (4), 1127–1161.

Handa, P., Schwartz, R., 1996. Limit Order Trading. Journal of Finance 51 (5), 1835–1861.

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

Hasbrouck, J., Saar, G., 2002. Limit Orders and Volatility in a Hybrid Market: The Island ECN. Working paper.

Hasbrouck, J., Saar, G., 2013. Low-latency Trading. Journal of Financial Markets 16, 646–679.

Hautsch, N., Huang, R., 2012. The Market Impact of a Limit Order. Journal of Economic Dynamics and Control 36 (4), 501–522.

Jain, P., 1988. Response of Hourly Stock Prices and Trading Volume to Economic News. Journal of Business 61, 219–231.

Jain, P., Joh, G., 1988. The Dependance between Hourly Prices and Trading Volume. Journal of Financial and Quantative Analysis 23 (3), 269–283.

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

McInisih, T., Wood, R., 1992. An Analysis of Intraday Patterns in Bid–Ask Spreads for NYSE Stocks. Journal of Finance 47, 753–764.

Obizhaeva, A., Wang, J., 2013. Optimal Trading Strategy and Supply/Demand Dynamics. Journal of Financial Markets 16 (1), 1–32.

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

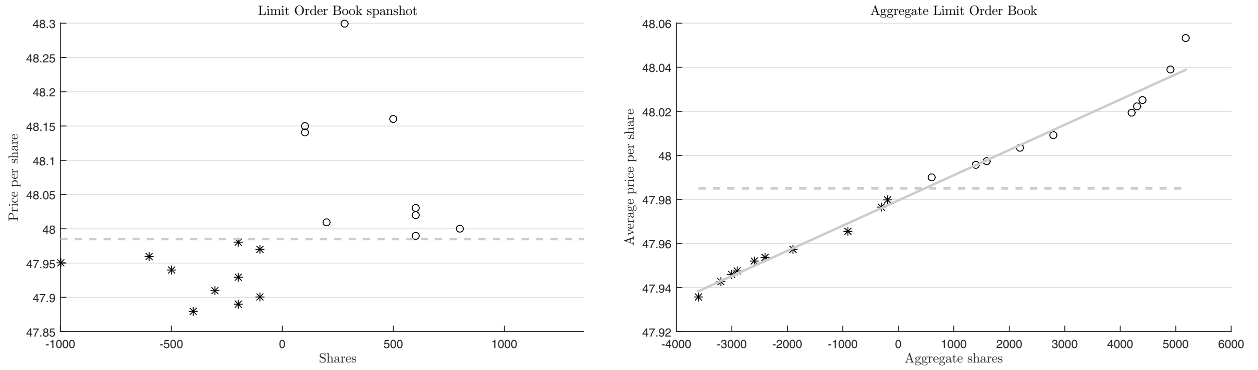
O'Hara, M., Yao, C., Ye, M., 2014. What's Not There: Odd Lots and Market Data. Journal of Finance 69 (5), 2199–2236.

Pastor, L., Stambaugh, R., 2003. Liquidity Risk and Expected Stock Returns. Journal of Political Economy 111 (3), 642–685.

Figure 1: Limit order book and marginal liquidity cost per share for Abbot Laboratories

Snapshots of the limit order book for ABT at 10:00 AM are presented in Panel A. In this panel, the left figure shows limit buy orders (stars), limit ask orders (circles), and the mid-quote (horizontal dotted line). The right figure shows the limit order book for aggregate shares and the fitted linear model of Equation 11. The marginal liquidity cost per share at opening and across the day is presented in Panel B.

Panel A: Snapshots of the limit order book



Panel B: Marginal liquidity cost per share

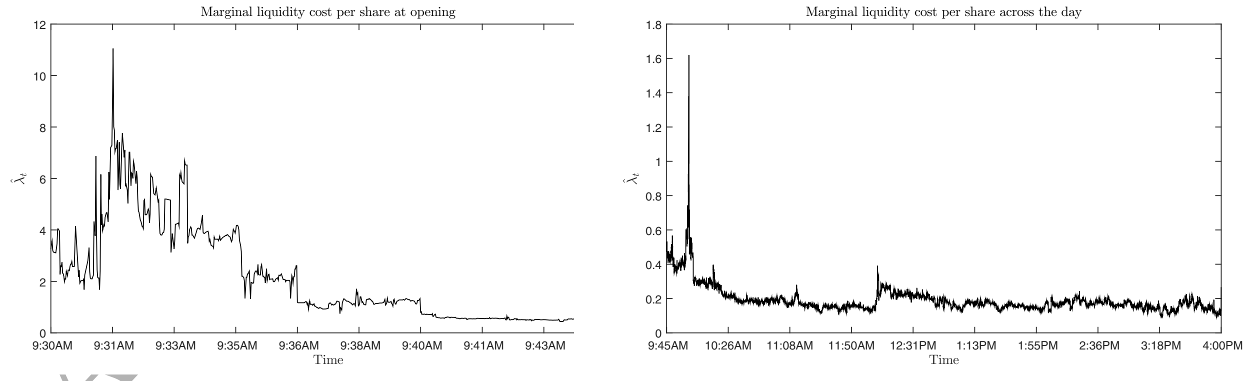


Figure 2: Distribution of R-squared for limit order book regressions (Equation 11) per minute of trading day

This figure displays percentiles P1 and P25, and the mean associated with the one-minute time-averaged R^2 of the regressions.

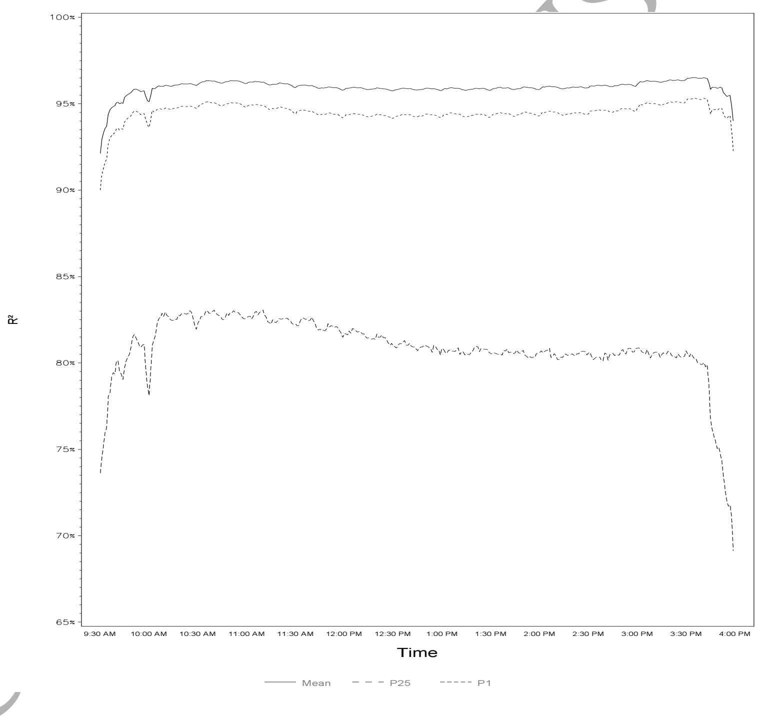


Figure 3: Times series of normalized cost of liquidity

This figure displays the daily mean of normalized cost of liquidity M_t, defined in Equation 6, according to the entire sample, small caps, middle caps, and large caps. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively. The first one-minute interval of the trading day is excluded. To simplify the presentation, the values are multiplied by 10^8.

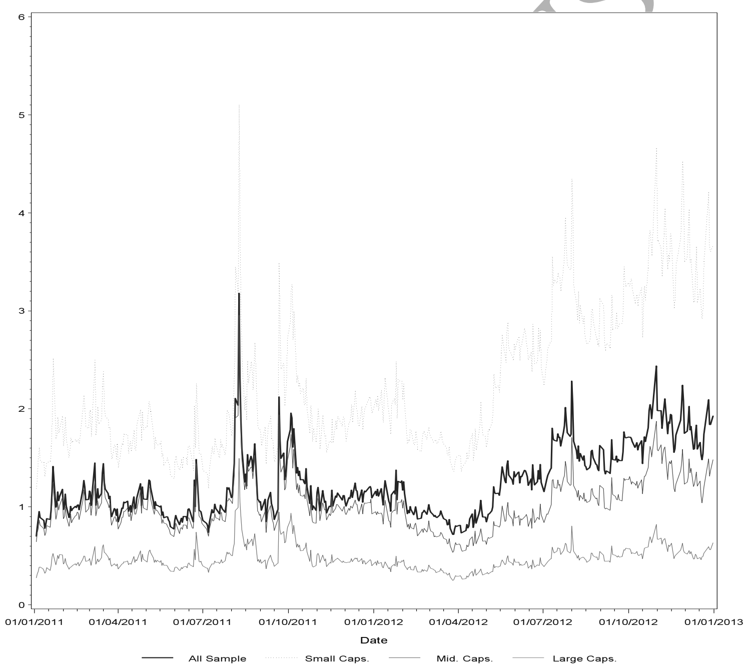
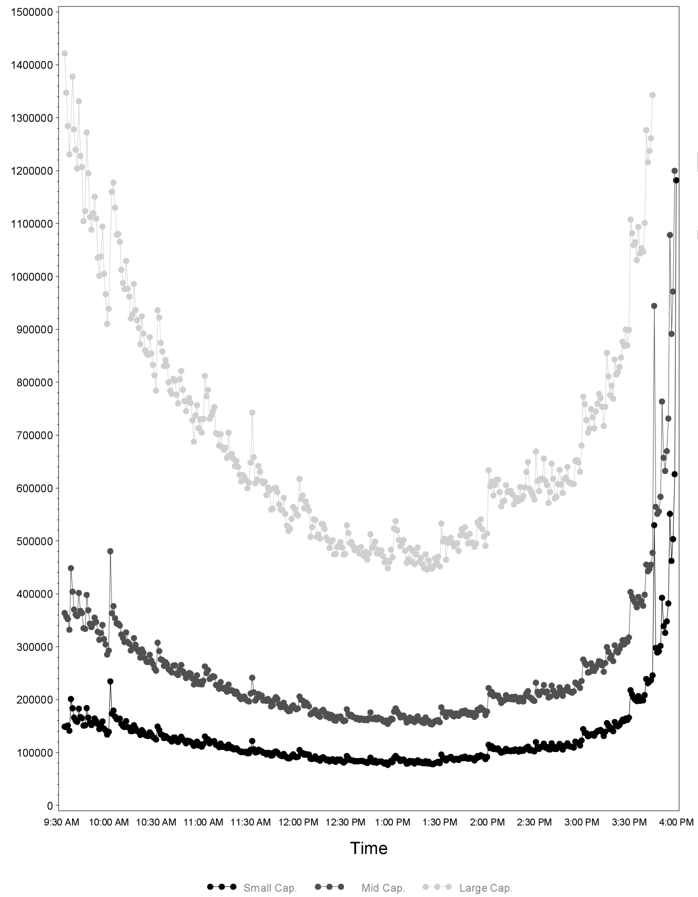


Figure 4: Trading activity across a trading day for 2011 (a) and 2012 (b)

These figures display the cross-sectional mean of the number of dollars traded across a trading day. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively. The first one-minute interval of the trading day is excluded. To simplify the presentation, the values are multiplied by 10^8.

(a) 2011



(b) 2012

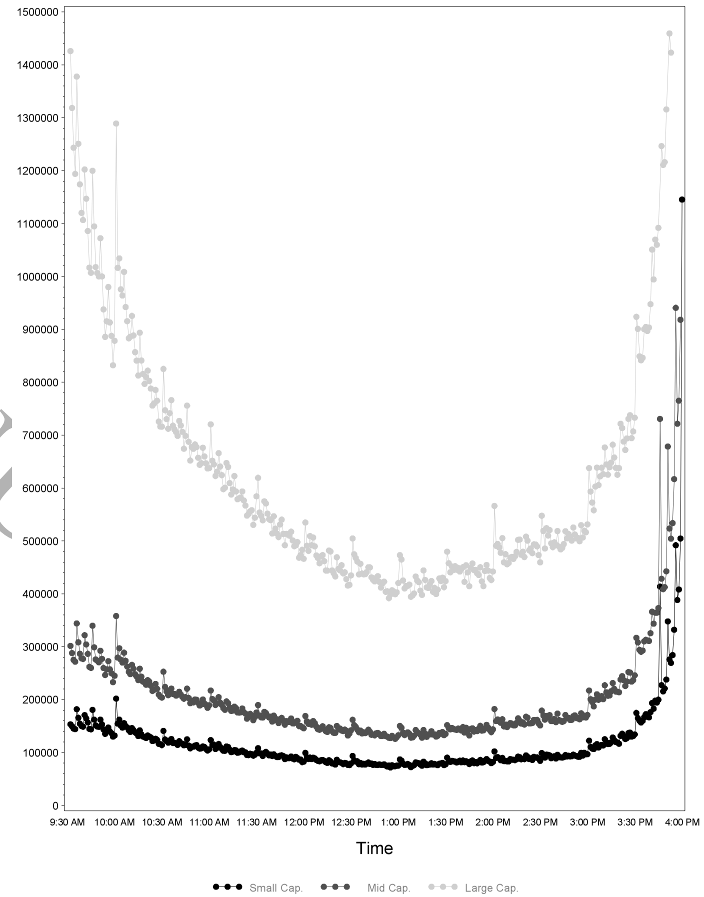
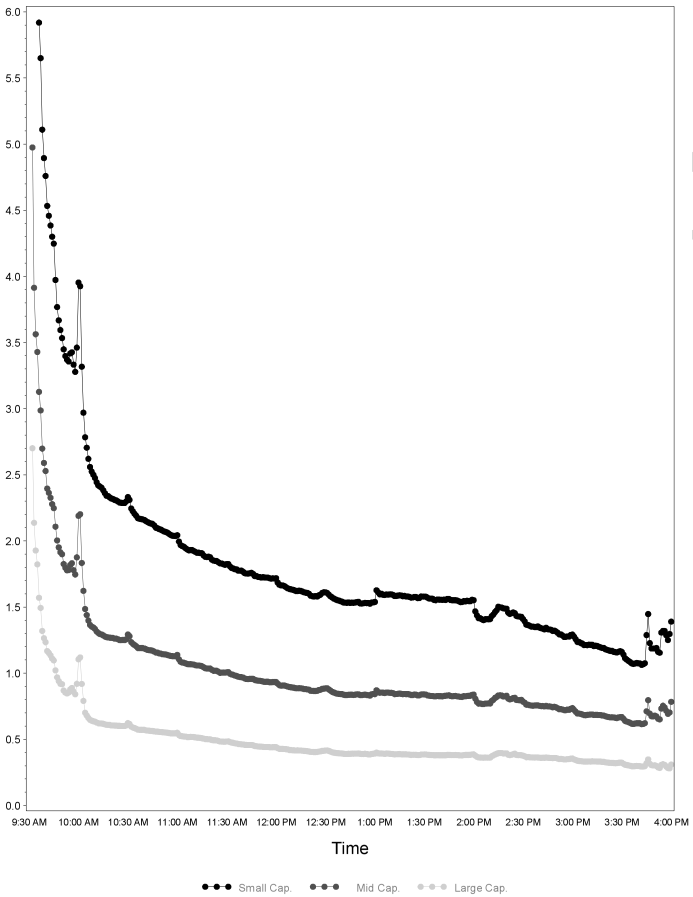


Figure 5: Normalized cost of liquidity across a trading day for 2011 (a) and 2012 (b)

These figures display the cross-sectional mean of normalized liquidity costs, as defined in Equation 6, across a trading day. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively. The first one-minute interval of the trading day is excluded. To simplify the presentation, the values are multiplied by 10^8.

(a) 2011



(b) 2012

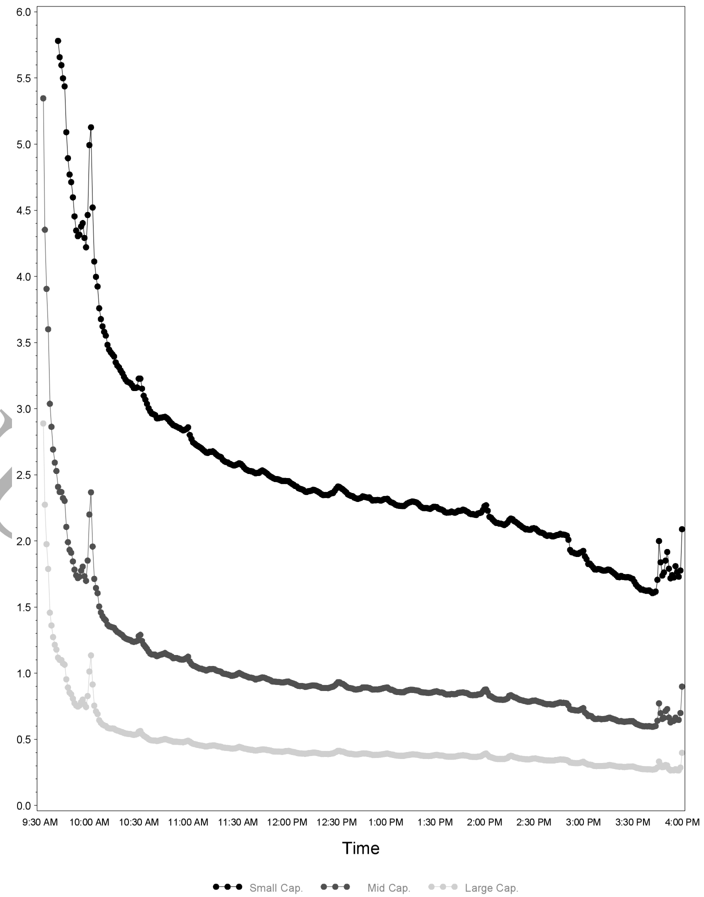
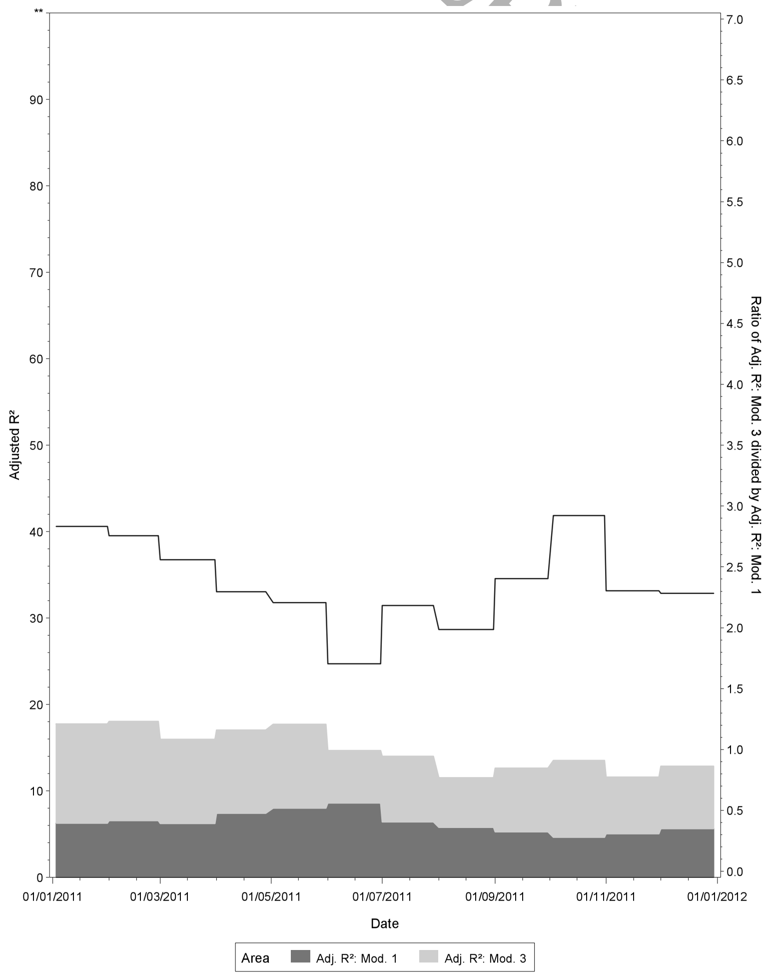


Figure 6: Monthly adjusted R-squared (R^2) for 2011 (a) and 2012 (b)

These figures display the adjusted R^2 obtained from the monthly regression derived from Equation 17 over the one-minute interval sample. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively. The first one-minute interval of the trading day is excluded. To simplify the presentation, the values for adjusted R^2 are expressed in percent. The continuous line represents the ratio of adjusted R^2 from Model 3 to that of Model 1 (values on the right-hand Y axis). A ratio above one indicates that Model 3 yields a better fit than Model 1.

(a) 2011



(b) 2012

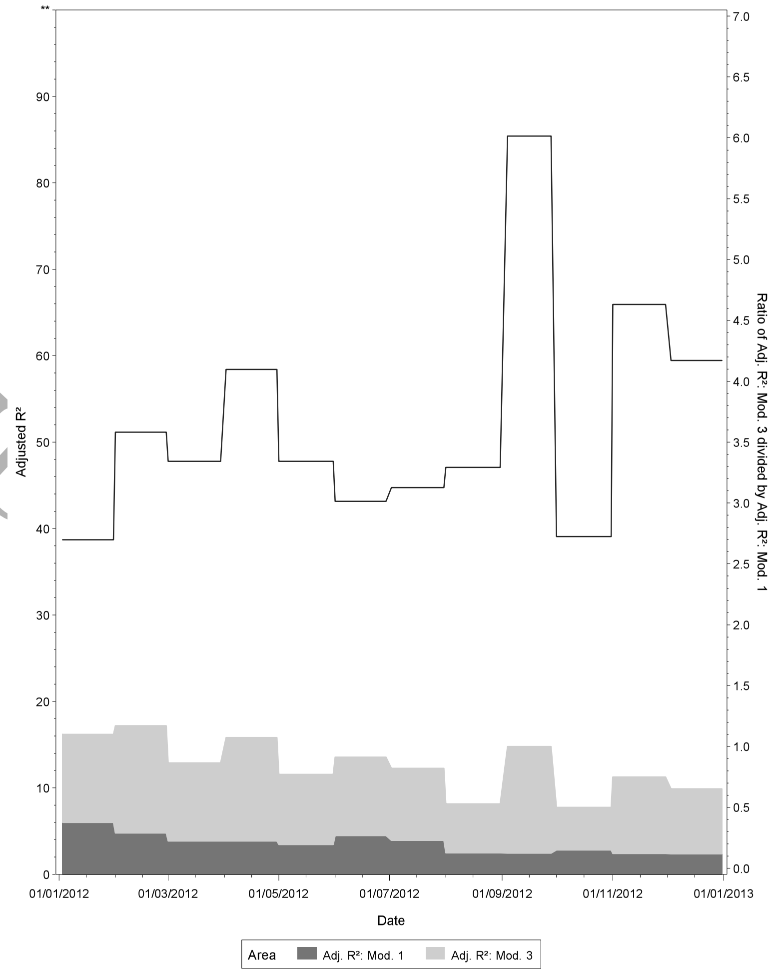
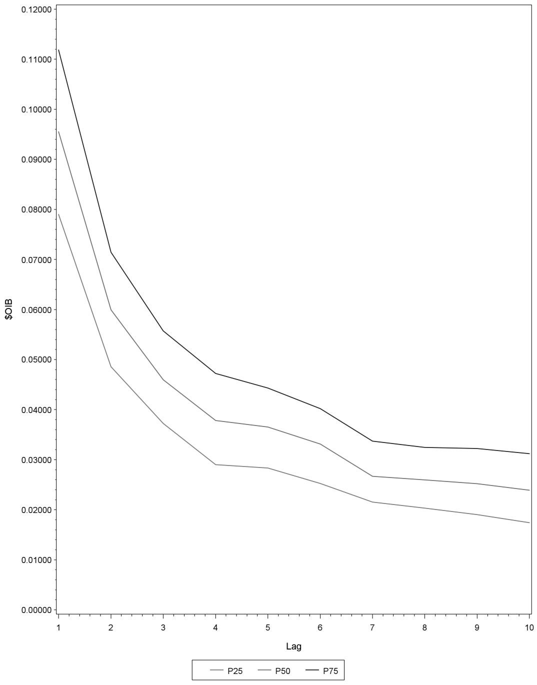


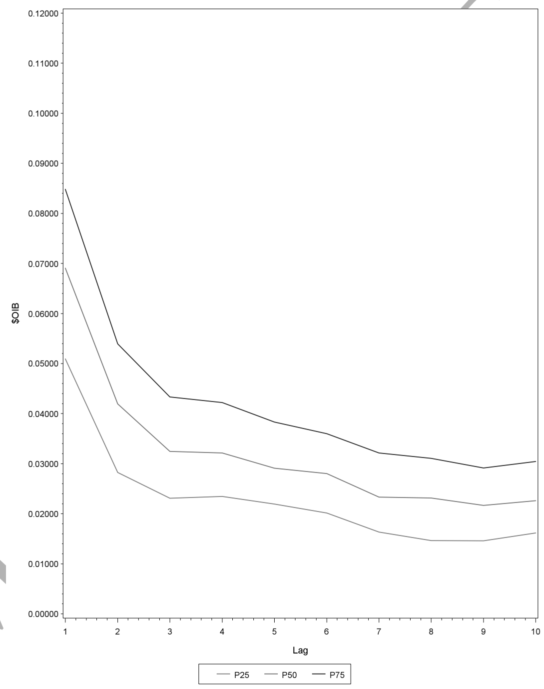
Figure 7: Autocorrelation for 2011 (a and c) and 2012 (b and d)

These figures display the distribution of autocorrelation coefficients for $\$OIB$ and $ILC^{\perp \$OIB}$ over the one-minute interval sample. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively. The first one-minute interval of the trading day is excluded.

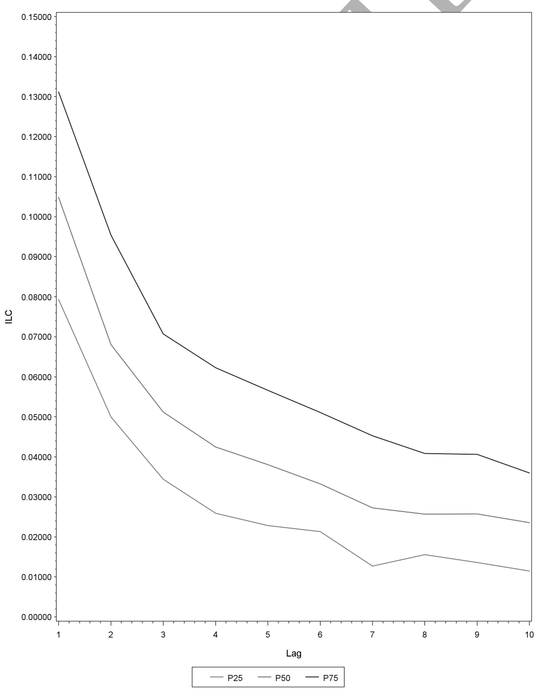
(a) 2011



(b) 2012



(c) 2011



(d) 2012

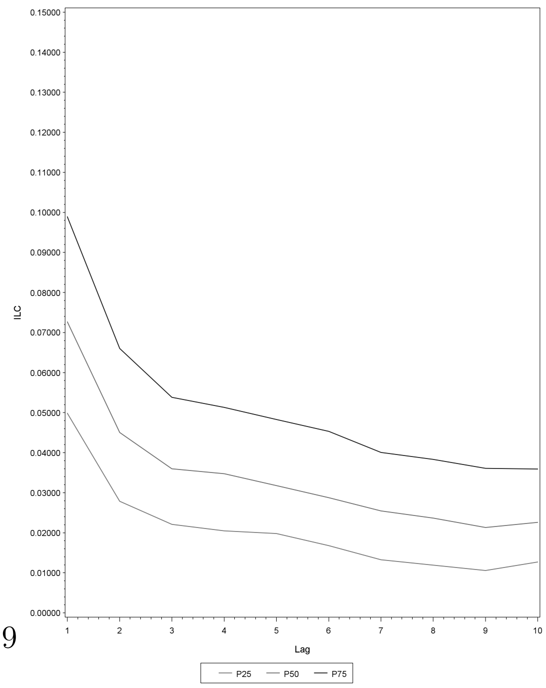


Table 1: Descriptive Statistics

This table presents the descriptive statistics for the one-minute panel sample. The first one-minute interval of the trading day is excluded. To simplify the presentation, the values for returns in the table are expressed in percent, Volume (Dollars) is expressed in thousands, \$OIB is expressed in thousands, $|\$OIB|$ is expressed in thousands, normalized cost of liquidity (M) is multiplied by 10^8, and ILC is multiplied by 10^5. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively.

Variable	2011				2012			
	Mean	St-Dev.	P1	P99	Mean	St-Dev.	P1	P99
Panel A: All Sample								
Return	0.000	0.092	-0.266	0.267	0.000	0.074	-0.209	0.209
Volume (Trades)	31.087	43.975	1.000	201.000	23.150	34.716	1.000	150.000
Volume (Dollars)	387.304	1058.990	2.720	3792.881	335.902	1293.606	1.909	3227.651
#OIB	-0.064	15.085	-41.000	41.000	-0.026	13.440	-36.000	36.000
$\|\#OIB\|$	8.142	12.699	0.000	55.000	6.859	11.558	0.000	48.000
\$OIB	-1.451	355.373	-760.883	744.374	-0.434	446.553	-691.606	686.391
$\|\$OIB\|$	114.606	340.677	0.002	1152.381	108.697	442.693	0.001	1071.427
M	1.119	1.455	0.056	6.466	1.366	2.434	0.055	10.344
ILC	-0.790	206.071	-457.222	449.964	-0.125	237.027	-418.051	420.074
Panel B: Small Cap.								
Return	0.000	0.099	-0.291	0.292	0.000	0.090	-0.262	0.261
Volume (Trades)	19.156	24.805	1.000	115.000	16.525	24.415	1.000	109.000
Volume (Dollars)	130.543	238.739	1.614	1028.414	117.313	261.114	0.920	1064.396
#OIB	-0.002	10.314	-29.000	29.000	0.013	11.033	-29.000	30.000
$\|\#OIB\|$	6.010	8.382	0.000	38.000	5.603	9.504	0.000	40.000
\$OIB	0.029	104.927	-264.700	267.302	0.163	111.539	-267.043	272.737
$\|\$OIB\|$	46.474	96.552	0.001	380.848	44.131	106.667	0.001	400.998
M	1.856	2.045	0.239	9.591	2.626	3.751	0.310	16.874
ILC	0.174	204.839	-441.339	444.492	0.317	304.504	-509.492	517.636
Panel C: Mid. Cap.								
Return	0.000	0.092	-0.266	0.266	0.000	0.069	-0.193	0.193
Volume (Trades)	26.611	32.997	1.000	155.000	19.372	27.243	1.000	116.000
Volume (Dollars)	258.613	453.316	3.999	1933.578	209.511	360.223	3.495	1570.073
#OIB	-0.049	12.473	-35.000	35.000	0.004	11.088	-29.000	29.000
$\|\#OIB\|$	7.235	10.160	0.000	46.000	5.865	9.410	0.000	38.000
\$OIB	-0.512	185.609	-465.048	466.591	0.215	156.767	-405.121	411.867
$\|\$OIB\|$	82.265	168.463	0.002	667.972	73.274	142.178	0.001	576.398
M	1.015	0.987	0.147	4.541	1.023	0.987	0.187	4.557
ILC	-0.662	219.208	-473.151	468.020	-0.044	214.471	-400.425	403.789
Panel D: Large Cap.								
Return	0.000	0.083	-0.239	0.240	0.000	0.059	-0.168	0.168
Volume (Trades)	47.127	60.358	2.000	279.000	33.150	45.643	1.000	202.000
Volume (Dollars)	764.963	1687.671	10.801	6646.177	667.606	2132.172	6.849	6105.836
#OIB	-0.141	20.423	-55.000	55.000	-0.091	17.128	-47.000	46.000
$\|\#OIB\|$	11.117	17.133	0.000	73.000	9.029	14.555	0.000	63.000
\$OIB	-3.890	579.087	-1354.620	1324.810	-1.688	750.354	-1254.655	1247.787
$\|\$OIB\|$	213.004	540.525	0.007	1990.314	204.789	727.910	0.003	1932.561
M	0.485	0.504	0.037	2.460	0.456	0.521	0.038	2.373
ILC	-1.885	193.086	-456.900	436.851	-0.650	173.256	-336.693	331.314

Table 2: Explaining the returns: OLS regression

This table presents the results of ordinary least square regressions that explain returns. The dependent variable is returns. Double-clustered t-statistics are used. The first one-minute interval of the trading day is excluded. The first number is the estimated regression coefficient. The second number is the t-statistic from the panel regressions. The third number is the standard-deviation. To simplify the presentation, the coefficients for $OIB in the table are multiplied by 10^{10}, the coefficients of ILC are multiplied by 10^2, and adjusted R-squared are expressed in percent. For Model 3 only, ILC stands for the variable $ILC^{\perp \$OIB}$. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively.

Variable	2011			2012		
	(1)	(2)	(3)	(1)	(2)	(3)
$OIB	6.049***	.	9.274***	2.999**	.	4.828***
	4.208	.	11.800	2.036	.	5.624
	1.437	.	0.786	1.473	.	0.858
ILC	.	15.717***	13.795***	.	9.946***	9.070***
	.	13.525	11.739	.	7.768	6.947
	.	1.162	1.175	.	1.280	1.306
Adj. R^2	5.486	12.450	13.230	3.287	10.190	11.040

Table 3: Explaining the returns by extreme values of absolute $OIB: OLS regression

This table presents the results of ordinary least square regressions that explain returns for different percentiles of the variable $\$OIB$. The dependent variable is returns. Double-clustered t-statistics are used. The first one-minute interval of the trading day is excluded. The first number is the estimated regression coefficient. The second number is the t-statistic from the panel regressions. The third number is the standard-deviation. To simplify the presentation, the coefficients for $OIB in the table are multiplied by 10^{10}, the coefficients for ILC are multiplied by 10^2, and adjusted R-squared are expressed in percent. For Model 3 only, ILC stands for the variable $ILC^{\perp \$OIB}$. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively.

Variable	2011 (1)	2011 (2)	2011 (3)	2012 (1)	2012 (2)	2012 (3)
Panel A: All						
$OIB	6.049***	.	9.274***	2.999**	.	4.828***
	4.208	.	11.800	2.036	.	5.624
	1.437	.	0.786	1.473	.	0.858
ILC	.	15.717***	13.795***	.	9.946***	9.070***
	.	13.525	11.739	.	7.768	6.947
	.	1.162	1.175	.	1.280	1.306
Adj. R^2	5.486	12.450	13.230	3.287	10.190	11.040
Panel B: Largest $\|\$OIB\|$ *(≥ P75)*						
$OIB	5.187***	.	7.894***	2.628**	.	4.151***
	4.279	.	12.017	2.075	.	5.298
	1.212	.	0.657	1.266	.	0.783
ILC	.	13.100***	10.988***	.	8.146***	7.239***
	.	16.232	13.104	.	7.539	6.632
	.	0.807	0.839	.	1.081	1.091
Adj. R^2	10.570	19.550	21.600	5.789	14.050	15.950
Panel C: Largest $\|\$OIB\|$ *(≥ P90)*						
$OIB	4.210***	.	6.480***	2.142**	.	3.401***
	4.339	.	12.120	2.140	.	5.195
	0.970	.	0.535	1.001	.	0.655
ILC	.	10.886***	8.792***	.	6.444***	5.602***
	.	16.599	12.626	.	7.015	6.202
	.	0.656	0.696	.	0.919	0.903
Adj. R^2	13.540	21.600	24.850	7.192	14.280	17.140
Panel D: Largest $\|\$OIB\|$ *(≥ P95)*						
$OIB	3.514***	.	5.436***	1.780**	.	2.851***
	4.371	.	12.071	2.215	.	5.183
	0.804	.	0.450	0.804	.	0.550
ILC	.	9.452***	7.440***	.	5.275***	4.515***
	.	16.036	11.937	.	6.436	5.800
	.	0.589	0.623	.	0.820	0.778
Adj. R^2	15.420	22.380	26.620	8.127	13.880	17.560

Table 4: Explaining returns by trade side: OLS regression

This table presents the results of ordinary least square regressions that explain returns according to trade sign. We classify trades according to the Lee and Ready (1991) procedure. Panel A uses all intervals classified as buy initiated, while Panel B does so for sell initiated. The variable ILC in Panel A is based on ex-ante trading costs computed from the ask side of the limit order book, while ILC in Panel B is based on the bid side. The dependent variable is returns. Double-clustered t-statistics are used. The first one-minute interval of the trading day is excluded. The first number is the estimated regression coefficient. The second number is the t-statistic from the panel regressions. The third number is the standard-deviation. To simplify the presentation, the coefficients for \$OIB in the table are multiplied by 10^{10}, those for ILC are multiplied by 10^2, and adjusted R-squared are expressed in percent. For Model 3 only, ILC stands for the variable $ILC^{\perp \$OIB}$. The symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively.

Panel A: Buy side

Variable	2011			2012		
	(1)	(2)	(3)	(1)	(2)	(3)
\$OIB	6.004***	.	7.610***	3.044**	.	5.006***
	3.971	.	5.985	2.019	.	4.125
	1.512	.	1.271	1.508	.	1.214
ILC	.	7.774***	5.629***	.	7.682***	6.271***
	.	3.500	3.130	.	4.000	3.628
	.	2.221	1.798	.	1.920	1.728
Adj. R^2	5.407	5.475	8.011	3.314	7.077	7.674

Panel B: Sell side

Variable	2011			2012		
	(1)	(2)	(3)	(1)	(2)	(3)
\$OIB	6.094***	.	8.053***	2.954**	.	3.419***
	4.461	.	8.063	2.041	.	3.029
	1.366	.	0.999	1.447	.	1.129
ILC	.	14.668***	10.625***	.	9.648***	6.610***
	.	9.778	7.798	.	6.407	5.412
	.	1.500	1.363	.	1.506	1.221
Adj. R^2	5.592	9.942	9.788	3.288	8.398	6.757

Table 5: Determinants of Normalized Liquidity Costs

This table presents the results of ordinary least square regressions that explain normalized liquidity costs. The dependent variable is the normalized liquidity cost. Double-clustered t-statistics are used. The first one-minute interval of the trading day is excluded. To simplify the presentation, the coefficients for Intercept in the table are multiplied by 10^8, those for $\$OIB_{t-1}$ are multiplied by 10^{16}, those for Volume $(\text{Dollars})_{t-1}$ are multiplied by 10^{16}, those for Volat_{t-1} are multiplied by 10^6, and adjusted R-squared are expressed in percent. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Intraday fixed effects are controlled with dummy variables constructed over half-hour intervals. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively.

Parameter	2011				2012			
estimates	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Intercept	0.616^{***}	10.98	0.613^{***}	11.13	0.622^{***}	7.15	0.620^{***}	7.44
$\$OIB_t$	0.428	1.45	.		-0.524	-0.73		
$\$OIB_{t-1}$	.		0.480	1.24	.		0.280	0.53
Volume $(\text{Dollars})_t$	-23.588^{***}	-3.18	.		-20.839^{*}	-1.85		
Volume $(\text{Dollars})_{t-1}$	.		-24.955^{***}	-3.06	.		-21.870^{*}	-1.78
Volat_t	2.988^{***}	7.05	.		5.181^{***}	4.20		
Volat_{t-1}	.		2.965^{***}	6.91	.		5.165^{***}	4.23
Adj. R^2	15.51		15.57		8.60		8.64	

Table 6: Alternative Liquidity Measures and Price Changes

This table presents the results of ordinary least square regressions using alternative liquidity measures, with returns as the dependent variable. Panel A employs proportional quoted spreads ($QSPR$) to compute the liquidity cost of the imbalance (DC), Panel B employs the weighted average midquote spread ($WSPR_1$), and Panel C the weighted average beyond best quotes ($WSPR_2$). Definitions of these variables are provided in Section 6. The first one-minute interval of the trading day is excluded. Each cell in the table presents the estimated regression coefficient, followed by the t-statistic using double-clustered errors, and the standard-deviation. The coefficients for $\$OIB$, ILC, and DC are multiplied by 10^{10}, 10^2, and 10^2, respectively. The adjusted R-squared are expressed in percent. For Model 3 and 4, DC stands for the variable $DC^{\perp \$OIB}$, and for Model 3 ILC for the variable $ILC^{\perp \$OIB}$. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively.

Panel A: Proportional Quoted Spread $QSPR$

Variable	2011				2012			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
\$OIB	6.049***	.	6.148***	9.189***	2.999**	.	4.249***	5.349***
	4.208	.	4.313	11.581	2.036	.	3.852	7.649
	1.437	.	1.425	0.793	1.473	.	1.103	0.699
DC	.	6.271*	0.974	0.496	.	9.490***	8.657***	4.410***
	.	1.698	0.306	0.367	.	4.225	3.088	3.120
	.	3.693	3.179	1.350	.	2.246	2.804	1.413
ILC	.	.	.	13.290***	.	.	.	8.459***
	.	.	.	9.904	.	.	.	6.685
	.	.	.	1.342	.	.	.	1.265
Adj. R^2	5.486	2.845	5.525	12.930	3.287	5.261	5.295	11.530

Panel B: Spread from weighted average midquote $WSPR_1$

Variable	2011				2012			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
\$OIB	6.049***	.	5.946***	9.248***	2.999**	.	3.664***	5.051***
	4.208	.	4.312	11.806	2.036	.	3.098	6.362
	1.437	.	1.379	0.783	1.473	.	1.183	0.794
DC	.	29.092**	7.552	2.006	.	44.731***	31.865***	11.338**
	.	2.356	0.743	0.537	.	4.138	3.038	2.270
	.	12.350	10.160	3.733	.	10.810	10.489	4.996
ILC	.	.	.	13.740***	.	.	.	8.735***
	.	.	.	11.700	.	.	.	6.354
	.	.	.	1.174	.	.	.	1.375
Adj. R^2	5.486	2.002	5.667	13.230	3.287	3.926	4.915	11.330

Panel C: Spread from weighted average beyond best quotes $WSPR_2$

Variable	2011				2012			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
\$OIB	6.049***	.	6.583***	9.020***	2.999**	.	3.988***	5.008***
	4.208	.	5.076	11.655	2.036	.	3.386	5.853
	1.437	.	1.297	0.774	1.473	.	1.178	0.856
DC	.	2.931***	1.453***	−1.061***	.	6.086***	4.571***	0.777
	.	4.515	3.169	−6.479	.	5.202	3.950	0.876
	.	0.649	0.458	0.164	.	1.170	1.157	0.887
ILC	.	.	.	14.618***	.	.	.	8.831***
	.	.	.	12.160	.	.	.	6.190
	.	.	.	1.202	.	.	.	1.427
Adj. R^2	5.486	1.846	5.970	13.410	3.287	4.003	5.315	11.200

Table 7: Explaining the returns by time horizon: OLS regression

This table presents the results of ordinary least square regressions that explain returns at different time horizons. The dependent variable is returns. Double-clustered t-statistics are used. We exclude the first time interval for each time horizon. The first number is the estimated regression coefficient. The second number is the t-statistic from panel regressions. The third number is the standard-deviation. To simplify the presentation, the coefficients for \$OIB in the table are multiplied by 10^{10}, the coefficients for ILC are multiplied by 10^2, and adjusted R-squared are expressed in percent. For Model 3 only, ILC stands for the variable $ILC^{\perp \$OIB}$. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively.

Variable	2011			2012		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A: 1-minute interval						
\$OIB	6.049***	.	9.274***	2.999**	.	4.828***
	4.208	.	11.800	2.036	.	5.624
	1.437	.	0.786	1.473	.	0.858
ILC	.	15.717***	13.795***	.	9.946***	9.070***
	.	13.525	11.739	.	7.768	6.947
	.	1.162	1.175	.	1.280	1.306
Adj. R^2	5.486	12.450	13.230	3.287	10.190	11.040
Panel B: 2-minute interval						
\$OIB	3.117***	.	4.291***	1.494**	.	2.383***
	4.243	.	10.915	2.175	.	5.895
	0.735	.	0.393	0.687	.	0.404
ILC	.	8.265***	7.193***	.	4.671***	4.245***
	.	12.958	11.209	.	7.979	7.325
	.	0.638	0.642	.	0.585	0.580
Adj. R^2	1.659	3.608	3.857	0.901	2.671	2.939
Panel C: 3-minute interval						
\$OIB	2.264***	.	2.909***	1.033**	.	1.475***
	4.273	.	10.233	2.184	.	4.727
	0.530	.	0.284	0.473	.	0.312
ILC	.	5.946***	5.130***	.	2.801***	2.485***
	.	12.425	10.743	.	4.505	4.079
	.	0.479	0.478	.	0.622	0.609
Adj. R^2	0.934	1.952	2.104	0.460	1.067	1.239
Panel D: 5-minute interval						
\$OIB	1.720***	.	2.055***	0.759**	.	1.025***
	4.372	.	9.614	2.153	.	4.698
	0.393	.	0.214	0.353	.	0.218
ILC	.	4.425***	3.778***	.	2.187***	1.970***
	.	11.825	10.232	.	7.199	6.455
	.	0.374	0.369	.	0.304	0.305
Adj. R^2	0.590	1.176	1.280	0.272	0.729	0.822
Panel E: 10-minute interval						
\$OIB	1.246***	.	1.310***	0.538**	.	0.598***
	4.478	.	9.119	2.140	.	3.259
	0.278	.	0.144	0.251	.	0.183
ILC	.	2.973***	2.455***	.	1.168**	1.010*
	.	9.222	7.654	.	2.128	1.858
	.	0.322	0.321	.	0.549	0.543
Adj. R^2	0.353	0.611	0.692	0.148	0.261	0.331

Table 8: Lagged-Effect of Imbalance

This table presents the results of ordinary least square regressions that explain returns with contemporaneous and lagged order imbalances and illiquidity costs. The dependent variable is returns. Double-clustered t-statistics are used. The first one-minute interval of the trading day is excluded. To simplify the presentation, the coefficients for \$OIB in the table are multiplied by 10^{10}, the coefficients for ILC are multiplied by 10^2, and adjusted R-squared are expressed in percent. In this table, ILC stands for the variable $ILC^{\perp \$OIB}$. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively.

2011						2012					
Variable	Coeff.	t-stat	Variable	Coeff.	t-stat	Variable	Coeff.	t-stat	Variable	Coeff.	t-stat
\$OIB$_t$	6.10***	4.42	ILC_t	6.29***	3.41	\$OIB$_t$	2.96**	2.06	ILC_t	5.45***	2.91
\$OIB$_{t-1}$	−0.54***	-6.58	ILC_{t-1}	−0.52***	-4.05	\$OIB$_{t-1}$	−0.23***	-3.08	ILC_{t-1}	−0.34	-1.54
\$OIB$_{t-2}$	−0.33***	-5.27	ILC_{t-2}	−0.34***	-3.07	\$OIB$_{t-2}$	−0.13*	-1.86	ILC_{t-2}	−0.45***	-2.64
\$OIB$_{t-3}$	−0.25***	-4.63	ILC_{t-3}	−0.26***	-3.10	\$OIB$_{t-3}$	−0.10*	-1.81	ILC_{t-3}	−0.26*	-1.78
\$OIB$_{t-4}$	−0.20***	-4.01	ILC_{t-4}	−0.25***	-2.81	\$OIB$_{t-4}$	−0.09	-1.64	ILC_{t-4}	−0.29**	-2.06
\$OIB$_{t-5}$	−0.12***	-3.89	ILC_{t-5}	−0.11*	-1.72	\$OIB$_{t-5}$	−0.05*	-1.92	ILC_{t-5}	0.03	0.21
\$OIB$_{t-6}$	−0.12***	-4.24	ILC_{t-6}	−0.14**	-2.45	\$OIB$_{t-6}$	−0.05**	-2.05	ILC_{t-6}	−0.07	-0.73
\$OIB$_{t-7}$	−0.15***	-4.34	ILC_{t-7}	−0.22***	-3.14	\$OIB$_{t-7}$	−0.04	-1.41	ILC_{t-7}	−0.18*	-1.96
\$OIB$_{t-8}$	−0.11***	-3.77	ILC_{t-8}	−0.09*	-1.78	\$OIB$_{t-8}$	−0.06*	-1.87	ILC_{t-8}	−0.18*	-1.95
\$OIB$_{t-9}$	−0.12***	-3.46	ILC_{t-9}	−0.12**	-2.29	\$OIB$_{t-9}$	−0.06*	-1.92	ILC_{t-9}	−0.21**	-2.28
\$OIB$_{t-10}$	−0.12***	-4.36	ILC_{t-10}	−0.08	-1.64	\$OIB$_{t-10}$	−0.04*	-1.80	ILC_{t-10}	−0.13*	-1.93
Adj. R^2	5.85		Adj. R^2	1.74		Adj. R^2	3.57		Adj. R^2	3.04	

Table 9: Explaining returns by firm size: OLS regression

This table presents the results of ordinary least square regressions that explain returns according to market capitalization. We define three size categories: small size for firms with market capitalizations, calculated on January of each year, that are equal to or less than the market capitalization of the first tercile of companies in the sample for a given year; large size for firms with market capitalizations higher than or equal to the third tercile; and mid-cap size for all others. The dependent variable is returns. Double-clustered t-statistics are used. The first one-minute interval of the trading day is excluded. The first number is the estimated regression coefficient. The second number is the t-statistic from the panel regressions. The third number is the standard-deviation. To simplify the presentation, the coefficients for \$OIB in the table are multiplied by 10^{10}, those for ILC are multiplied by 10^2, and adjusted R-squared are expressed in percent. For Model 3 only, ILC stands for the variable $ILC^{\perp \$OIB}$. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively.

Variable	2011			2012		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A: All						
\$OIB	6.049***	.	9.274***	2.999**	.	4.828***
	4.208	.	11.800	2.036	.	5.624
	1.437	.	0.786	1.473	.	0.858
ILC	.	15.717***	13.795***	.	9.946***	9.070***
	.	13.525	11.739	.	7.768	6.947
	.	1.162	1.175	.	1.280	1.306
Adj. R^2	5.486	12.450	13.230	3.287	10.190	11.040
Panel B: Small-Cap firms sample						
\$OIB	31.216***	.	37.777***	26.834***	.	31.350***
	8.677	.	11.492	8.279	.	9.121
	3.597	.	3.287	3.241	.	3.437
ILC	.	16.798***	11.013***	.	9.145***	5.412***
	.	12.858	12.783	.	4.906	3.534
	.	1.306	0.861	.	1.864	1.531
Adj. R^2	10.870	12.000	13.610	11.040	9.553	13.400
Panel C: Mid-Cap firms sample						
\$OIB	16.984***	.	25.614***	16.230***	.	25.270***
	7.054	.	10.927	10.452	.	9.426
	2.408	.	2.344	1.553	.	2.681
ILC	.	14.161***	8.330***	.	10.326***	4.599***
	.	9.344	6.566	.	8.421	4.853
	.	1.515	1.269	.	1.226	0.948
Adj. R^2	11.740	11.380	14.010	13.690	10.370	14.890
Panel D: Large-Cap firms sample						
\$OIB	4.077***	.	7.997***	1.888**	.	4.248***
	4.140	.	14.438	2.167	.	7.358
	0.985	.	0.554	0.871	.	0.577
ILC	.	16.538***	13.887***	.	11.830***	10.187***
	.	15.025	12.173	.	7.624	6.382
	.	1.101	1.141	.	1.552	1.596
Adj. R^2	8.042	14.710	15.870	5.684	11.890	13.070

Table 10: Explaining returns by time of the day: OLS regression

This table presents the results of ordinary least square regressions that explain returns according to different periods of the trading day. The dependent variable is returns. Double-clustered t-statistics are used. The first one-minute interval of the trading day is excluded. The first number is the estimated regression coefficient. The second number is the t-statistic from the panel regressions. The third number is the standard-deviation. To simplify the presentation, the coefficients for \$OIB in the table are multiplied by 10^{10}, those for ILC are multiplied by 10^2, and adjusted R-squared are expressed in percent. For Model 3 only, ILC stands for the variable $ILC^{\perp \$OIB}$. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively.

Variable	2011			2012		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A: All						
\$OIB	6.049***	.	9.274***	2.999**	.	4.828***
	4.208	.	11.800	2.036	.	5.624
	1.437	.	0.786	1.473	.	0.858
ILC	.	15.717***	13.795***	.	9.946***	9.070***
	.	13.525	11.739	.	7.768	6.947
	.	1.162	1.175	.	1.280	1.306
Adj. R^2	5.486	12.450	13.230	3.287	10.190	11.040
Panel B: Before 10:30						
\$OIB	7.278***	.	9.748***	3.895*	.	5.338***
	3.305	.	12.488	1.897	.	5.385
	2.202	.	0.781	2.053	.	0.991
ILC	.	15.078***	13.788***	.	9.672***	8.943***
	.	17.095	14.388	.	10.952	9.631
	.	0.882	0.958	.	0.883	0.929
Adj. R^2	4.938	13.930	14.360	2.922	10.490	11.130
Panel C: After 10:30						
\$OIB	5.660***	.	5.660***	2.727**	.	2.727***
	4.524	.	7.708	2.071	.	3.269
	1.251	.	0.734	1.317	.	0.834
ILC	.	16.284***	13.836***	.	10.201***	9.146***
	.	7.943	6.824	.	4.251	3.850
	.	2.050	2.028	.	2.400	2.376
Adj. R^2	5.948	11.550	12.520	3.736	9.963	10.990

Table 11: Explaining returns by trade side: OLS regression

This table presents the results of ordinary least square regressions that explain returns according to trade sign. We use trades following the Lee and Ready (1991) procedure to measure trade sign. Our sample consists of 482 stocks over 249 and 247 trading days for 2011 and 2012, respectively. The dependent variable is returns. Double-clustered t-statistics are used. The first one-minute interval of the trading day is excluded. The first number is the estimated regression coefficient. The second number is the t-statistic from the panel regressions. The third number is the standard-deviation. To simplify the presentation, the coefficients for \$OIB in the table are multiplied by 10^{10}, those for ILC are multiplied by 10^2, and adjusted R-squared are expressed in percent. For Model 3 only, ILC stands for the variable $ILC^{\perp \$OIB}$. The symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Variable	2011			2012		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A: All						
\$OIB	6.049***	.	9.274***	2.999**	.	4.828***
	4.208	.	11.800	2.036	.	5.624
	1.437	.	0.786	1.473	.	0.858
ILC	.	15.717***	13.795***	.	9.946***	9.070***
	.	13.525	11.739	.	7.768	6.947
	.	1.162	1.175	.	1.280	1.306
Adj. R^2	5.486	12.450	13.230	3.287	10.190	11.040
Panel B: Buy						
\$OIB	6.094***	.	8.482***	2.954**	.	3.591***
	4.461	.	9.367	2.041	.	3.257
	1.366	.	0.906	1.447	.	1.103
ILC	.	15.838***	12.151***	.	9.196***	6.540***
	.	10.975	8.807	.	4.582	3.972
	.	1.443	1.380	.	2.007	1.646
Adj. R^2	5.592	12.520	11.170	3.288	9.293	7.396
Panel C: Sell						
\$OIB	6.004***	.	9.593***	3.044**	.	5.445***
	3.971	.	11.635	2.019	.	5.961
	1.512	.	0.824	1.508	.	0.913
ILC	.	15.598***	12.734***	.	10.773***	8.906***
	.	16.702	13.297	.	10.034	8.444
	.	0.934	0.958	.	1.074	1.055
Adj. R^2	5.407	12.440	11.900	3.314	11.270	10.000