Journal of Financial Markets 24 (2015) 25–48

Contents lists available at ScienceDirect

Journal of Financial Markets

journal homepage: www.elsevier.com/locate/finmar

Relative liquidity and future volatility ☆

Marcela Valenzuela [a], Ilknur Zer [b,*,1], Piotr Fryzlewicz [c], Thorsten Rheinländer [d]

[a] University of Chile, Department of Industrial Engineering, Republica 701, Santiago, Chile
[b] Federal Reserve Board, 20th Street and Constitution Avenue, N.W. Washington, DC 20551, USA
[c] London School of Economics, Department of Statistics, Houghton Street, London WC2A 2AE, UK
[d] Vienna University of Technology, Research Group Financial and Actuarial Mathematics (FAM), Wiedner Hauptstrasse 8-10/105-1, A-1040 Vienna, Austria

ARTICLE INFO

Article history:
Received 23 May 2014
Received in revised form
2 March 2015
Accepted 12 March 2015
Available online 23 March 2015

JEL classification:
G1
G20

Keywords:
Order-driven markets
Limit order book distribution
Volatility predictability
Liquidity

ABSTRACT

The main contribution of this paper is to identify the strong predictive power of the relative, rather than the absolute, volume of orders over volatility. To this end, we propose a new measure, relative liquidity, which accounts for how quoted depth is distributed in a limit order book and captures the level of consensus on a security's trading price. Higher liquidity provision farther away from the best quotes, relative to the rest of the book, is associated with a disagreement on the current price and followed by high volatility. The relationship is robust to the inclusion of several alternative measures.

Published by Elsevier B.V.

☆ We sincerely thank Vicente Cuñat, Jon Danielsson, Dobrislav Dobrev, Christian Julliard, Richard Payne, and an anonymous referee for their valuable comments. We would also like to thank to Coskun Gunduz, Recep Bildik, and Huseyin Eskici for providing us data and the support for understanding the market mechanisms. Valenzuela acknowledges the support of Fondecyt Project no. 11140541 and Instituto Milenio ICM IS130002.

* Corresponding author.
E-mail addresses: mvalenzuela@dii.uchile.cl (M. Valenzuela), ilknur.zerboudet@frb.gov (I. Zer), p.fryzlewicz@lse.ac.uk (P. Fryzlewicz), rheinlan@fam.tuwien.ac.at (T. Rheinländer).

[1] The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

http://dx.doi.org/10.1016/j.finmar.2015.03.001
1386-4181/Published by Elsevier B.V.

1. Introduction

In this paper we examine the link between two central concepts in financial markets: liquidity and volatility. Liquidity, the ease with which an asset can be traded without affecting the asset's price, is essential for well-functioning financial markets. Hence, understanding the effects of liquidity provision on market dynamics has gained an increased attention from regulators, market participants, and academics alike. On the other hand, information on volatility, variation in trade prices, is one of the main ingredients in assessing risk-return trade-off for portfolio valuation, derivatives pricing models, and it is important for the calibration of execution probability of limit orders. In this paper, we propose a new way of summarizing the distribution of liquidity in a limit order book and examine its informativeness on future volatility.

Our empirical investigation is motivated by the theoretical predictions of Goettler, Parlour, and Rajan (2005, 2009), in which the frequency of orders waiting to be executed at each quote reveals information of the disagreement on the true price. They predict that higher liquidity provision around the best quotes relative to the rest of the book is associated with a consensus on the current price, whereas the accumulation of orders at a quote farther away from the best prices signals to the market that current quotes are mispriced. We argue that, in the latter case price movements are more plausible, creating higher future volatility.

To examine the effects of the relative accumulation of orders on future volatility, we construct a measure, *relative liquidity* (*RLIQ*), which is the first principal component of an aggregate limit order book. To calculate *RLIQ*, we first obtain the empirical probability density function of a limit order book for a given stock. We then calculate the cross-sectional average of individual stock distributions to reach the aggregate distribution. In other words, we measure the proportions of orders waiting at each price level in the market. Finally, we employ a principal component analysis to summarize this information in as few interpretable quantities as possible.

As a summary measure, *RLIQ* has three ingredients: (1) it accounts for the *relative* distribution of orders waiting to be traded, which reveals information on the disagreement of the true price; (2) it includes the information contained beyond the best quotes; and (3) it weighs this information based on price distances. Depths at and farther away from the best quotes play different roles in traders' order choices. The weighting scheme is introduced to capture this variation on the informativeness of different price levels. Instead of imposing an exogenous weighting scheme, we use the loadings of the first principal component of the aggregate distribution as weights. Thus the principal component analysis enables us to avoid the subjective judgments regarding the relative importance of quotes.

We evaluate the predictive power of relative liquidity for both market and individual stock volatilities at an intraday level, with a particular interest in the former. It is challenging to estimate intraday volatility of the price process due to the microstructure noise inherent in the high-frequency data, such as informational effects, bid–ask bounces, or data recording errors. It has been recognized in the literature that the standard estimator—realized volatility—can be highly unreliable under the microstructure noise, especially if the sampling frequency is high. Zhang, Mykland, and Ait-Sahalia (2005) and Ait-Sahalia, Mykland, and Zhang (2011) address this specific problem and provide the volatility proxy that we use in this study: the two scales realized volatility estimator (TSRV).

The order and trade books of the largest 30 stocks from the Istanbul Stock Exchange (ISE) form our dataset. By matching these two books and removing the executed orders, we dynamically reconstruct the limit order book. That is, for a given time we have the best bid and ask prices, all of the orders waiting to be executed, their submitted prices, and their corresponding volumes. Since ISE is a fully centralized purely order-driven market and operates with a single trading platform, our data contains the entire order flow in the public domain, which is a major advantage compared to the main European and U.S. market exchanges.

We show that relative liquidity (*RLIQ*) is the strongest among standard liquidity and trading activity measures in explaining the in-sample variations in market volatility. On average, a one standard deviation increase in *RLIQ* decreases the 15-minutes-ahead volatility by 4.4 bps. Given that the mean value of volatility is 19 bps, this impact is economically significant. The results are

robust to the inclusion of alternative controls and jump-robust volatility measures. Out-of-sample forecasting tests provide evidence for the substantial forecasting power of relative liquidity. It predicts 15-minutes-ahead market volatility at the 5% level with an out-of-sample R^2 of 12.9%, where the forecasting power lasts up to 75 minutes ahead. *RLIQ* complements the bid–ask spread as it is related to the depth dimension of liquidity and by construction does not include the spread. Further analysis reveals that capturing relative liquidity along with the tightness dimension of liquidity delivers an out-of-sample R^2 of over 24%.

It can be argued that the documented relationship is "mechanical": when the limit order book is thin (i.e., when depth at the best quotes is low), any market order causes a price impact or a bid–ask bounce. However, unlike standard depth measures, relative liquidity focuses on the *relative* distribution of orders, rather than the *absolute* volume of orders waiting at each quote. Hence, although a thin book is associated with low depth, it is not necessarily associated with low *RLIQ* since the latter quantifies whether the orders are concentrated or spread over a book. The documented predictive power of *RLIQ* after controlling for standard depth measures suggests that the information content of *RLIQ* over volatility cannot be explained by absolute depth.

Finally, we show that the relationship between *RLIQ* and market volatility is not driven by variations in a particular stock or industry, but rather that it is shared by the majority of the stocks. We find a significant relationship between the individual stock level *RLIQ* and future volatility for 87% of the stocks in our sample.

This paper relates to the literature that attempts to measure the liquidity provision considering the whole book. Domowitz, Hansch, and Wang (2005) propose an illiquidity measure based on the supply and demand step functions and conclude that the liquidity commonality is priced in stock returns. Marshall (2006) measures liquidity by the weighted order value, which depends on the execution rate of orders waiting in each price band and their corresponding prices and volumes. The author documents a negative association between liquidity and monthly returns. In another related study, Naes and Skjeltorp (2006) introduce the slope of the book, which describes the average elasticity across all price levels with the corresponding volumes.

This paper is part of the market microstructure literature that examines the predictive power of liquidity on intraday volatility. In an early empirical work, Ahn, Bae, and Chan (2001) analyze the interactions between transitory volatility and order flow composition. They conclude that transitory volatility arises mainly from the scarcity of limit orders at the best quotes. Pascual and Veredas (2010) show that trade size and quoted depth both at the best and away from the quotes have predictive power for individual volatility. Duong and Kalev (2008) investigate the forecasting power of the Naes and Skjeltorp's (2006) definition of order book slope. By using data from the automated futures market, Coppejans, Domowitz, and Madhavan (2001) study the dynamic relationship between liquidity, return, and volatility in a vector autoregressive framework.

Finally, the paper is related to the few studies that use intraday data from the Istanbul Stock Exchange (ISE). Ekinci (2008) and Koksal (2012) provide descriptive analyses of the intraday liquidity patterns of the ISE by focusing on the behavior of spreads, depths, and trading volume. Valenzuela and Zer (2013) study how the market characteristics and information content of a limit order book affects the order choice of investors.

Our contribution to the literature is threefold. First, we provide a new variable that summarizes the information provided by a limit order book. Contrary to the aforementioned measures, which focus on the absolute volume of the orders waiting in a given book, *RLIQ* is based on the relative distribution of volume at a given time. Second, we show that relative liquidity contains information on future volatility that cannot be explained by the standard predictors of volatility. Finally, in contrast to these former studies, which examine the volatility–liquidity relationship at an individual stock level, we focus on the link between aggregate liquidity and future market volatility.

The rest of the paper is organized as follows. In the next section, we describe data and the trading structure in our market. In Section 3, we explain the construction of our measure in detail. In Section 4, we introduce the econometric methodology and variables included in the analysis. The in-sample and out-of-sample predictive results, comparison of relative liquidity with standard depth measures, and robustness checks are given in Section 5. Finally, we present concluding remarks in Section 6.

2. The market and data

Our dataset comprises order and trade books of the individual constituents of the Istanbul Stock Exchange ISE-30 Index for the period of June and July 2008. The index corresponds to almost 75% of the total trading volume of the ISE for the sample period. The ISE is a fully computerized, as well as a fully centralized purely order-driven stock exchange (i.e., the trading of the listed stocks has to be executed in the ISE via electronic order submissions without a market maker). Hence, our data fully capture the order flow.

The trading occurs between 09:30am and 5:00pm, with a lunch break. Similar to all other major exchanges, a trading day starts with a call market matching mechanism of 15 minutes to determine the opening price. In contrast to the opening session, during the continuous double auction, all of the orders submitted are either matched instantaneously based on the usual price and time priorities or booked until the corresponding match order arrives to the system. A submitted order is valid for a given session or for a day. All brokers have access to the full book. Prior to the submission of an order, they can see the quantity available at different prices; they are not limited to the best five or ten quotes.

The order book data consists of information regarding the orders submitted for a given stock, whereas the trade data includes the executed orders, both time-stamped at the accuracy of one second. The order and trade ID numbers generated by the exchange system allow us to identify the priority of orders submitted in the same second, to match orders in the order and trade books, and finally to track any order through submission to (possible) execution or modification. We use the order and trade books to reconstruct the limit order book dynamically for each stock and obtain relevant information, such as the bid and ask prices and corresponding volumes at a given time. The reconstruction methodology enables us to obtain snapshots of a limit order book at any given time. In particular, we have the same information that a trader observes: the volume of orders waiting to be executed for the entire price range. We use this information to calculate the relative frequency of orders waiting at every price level.

3. The limit order book distribution and relative liquidity

3.1. The limit order book distribution

The limit order book (LOB) distribution is obtained by employing the following steps, which are illustrated with an example in the Appendix:

1. For each security and each day, we sample the limit order books every 15 minutes, excluding the lunch break and the opening session. The first snapshot of the book contains the unexecuted orders submitted until 10:00am, whereas the last one contains all of the unexecuted orders submitted until 5:00pm. Hereafter, the time subscript τ indexes these trading intervals, with $\tau = 1, 2, \ldots, 21$. We repeat the empirical analysis with 30-minute sampling frequencies as a check of robustness. The results are presented in Section 5.6.
2. We calculate the (tick-adjusted) price distance, Δ, of each limit order relative to the best limit price in each snapshot. For each order i in the limit order book at τ, we define the price distances as

$$\Delta_{i,\tau}^{\text{buy}} = (p_\tau^B - p_i^{\text{buy}})/\text{tick},$$
$$\Delta_{i,\tau}^{\text{sell}} = (p_i^{\text{sell}} - p_\tau^A)/\text{tick},$$

   where p_τ^B (p_τ^A) is the best bid (ask) price at the end of interval τ and p_i^{buy} (p_i^{sell}) is the limit price of the ith order.
3. For each buy and sell sides of the book, day, and limit order book at τ, we get the limit order book probability density function for a given stock (*indPDF*) by calculating the percentage of total volume supplied/demanded at a given Δ for $\Delta = 0, 1, 2, \ldots, \Delta_c$, where Δ_c is the maximum price distance considered. Therefore, *indPDF* summarizes both the relative magnitude of the depth provision and its price location.

4. We calculate the equally weighted cross-sectional average of individual LOB probability density functions to obtain the aggregate LOB probability density function (*aggPDF*). That represents the proportion of orders waiting at each price level in the market. For a given trading interval τ and price distance Δ:

$$aggPDF_{\tau}^{\text{buy}}(\Delta) = \frac{1}{S}\sum_{s=1}^{S} indPDF_{s,\tau}^{\text{buy}}(\Delta), \tag{1}$$

where $indPDF_{s,\tau}^{\text{buy}}(\Delta)$ is the buy side limit order book probability density function of stock s and S is the total number of stocks. The measure for the sell side is calculated analogously. In order to consider the possible impact of bigger or more actively traded stocks, we also calculate the value-weighted and number-of-trades-weighted averages of the *indPDFs* to obtain the *aggPDF*. We reach qualitatively similar results, which are presented in Section 5.6.

In Fig. 1, we plot the *aggPDF* averaged across all trading intervals and days. It reveals that for both sides of the market, the frequency of orders submitted at the second best quote is the highest and the limit order book distribution is positively skewed. The liquidity provision is concentrated closer to the best quotes for the buy side compared to the sell side, which can be observed by comparing either the mean or the skewness of the distribution presented in Table 1. The mean of the distribution, for all of the time intervals, is higher for the sell side than the buy side. Wilcoxon rank sum test results show that the difference is statistically significant at the 5% level.

The cumulative frequency of orders waiting after 5 ticks away from the quotes is 28% for the sell side, whereas it is only 18% for the buy side. Moreover, the average variance of the sell side is 36% higher than the average variance of the buy side. Both observations indicate that the buy side is less dispersed compared to the sell side of the market. Indeed, the Kolmogorov-Smirnov test results indicate that we can reject the hypothesis that the buy-side and sell-side distributions are equal, at the 1% significance level.

About 90% of the submitted orders are waiting within the 10 best prices for both sides of the market. Hence, we consider the information contained in the book up to the 10th best quotes by setting $\Delta_c = 10$. However, we examine the robustness of our findings when Δ_c is equal to 20 and 30 (i.e., when we consider the whole book), in Section 5.6.

3.2. Summarizing the limit order book distribution: RLIQ

The shape of the limit order book distribution at time τ is given by the proportion of volume of orders waiting to be traded at different price distances Δ. There are several ways to summarize this information. We want our summary measure to weigh the information provided at different quotes based on price distances to capture the different levels of informativeness of the quotes. One, for example, could assume exogenously given weights or give equal weights to the frequency of orders waiting at each price distance Δ. We instead employ the principal component analysis (PCA), which can be used to extract the most important uncorrelated sources of variation in the LOB distribution. The advantage of this approach is that it assigns an objective weighting scheme, which aims to encode as much information about the LOB distribution in as few quantities (principal components) as possible.

The PCA applied on the ten price bins of the aggregate limit order book distribution function defined in (1) produces ten uncorrelated principal components. The first principal component is the leading eigenvector in the spectral decomposition of the covariance matrix of *aggPDF*, which explains the highest variation in the limit order book distribution. Relative liquidity summary measures, $RLIQ^{\text{buy}}$ and $RLIQ^{\text{sell}}$, are chosen to be the first principal components of *aggPDF* for both sides of the market. In Section 5.6, we discuss the sensitivity of the findings by (a) considering the first three and five principal components, (b) using the empirical frequencies of orders waiting at each price distance separately, and (c) employing LASSO (Tibshirani, 1996), a shrinkage and variable selection technique.

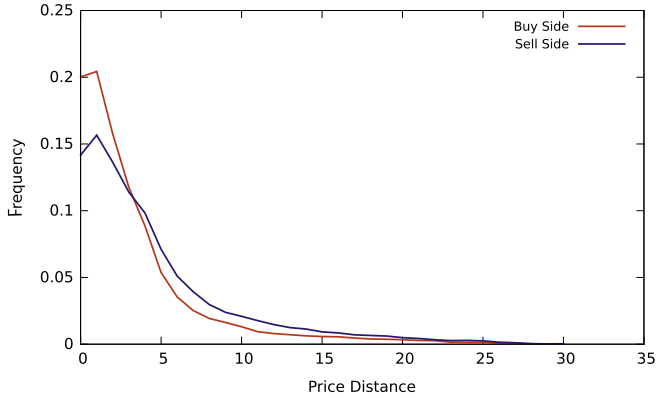

Fig. 1. Limit order book distribution. This figure plots the aggregated limit order book probability density function (*aggPDF*) for the period of June and July 2008, averaged across 21 15-minute trading intervals and 39 days considering the whole book.

Table 1
Summary statistics: the limit order book distribution.
This table presents statistics for the limit order book distribution for the buy and sell sides the market. The mean, variance, skewness, and the fraction of number of shares accumulated up to a given price distance Δ are reported. Column I shows the summary statistics of the limit order book distribution, which are obtained by averaging across intervals, days, and stocks. Columns II–V report the statistics for four limit order book distributions (averaged across stocks) at 10:00am (beginning of the day), 12:00pm (end of the morning session), 2:15pm (beginning of the afternoon session), and 5:00pm (end of the trading day), respectively.

		uncond. I	**10:00am** II	**12:00pm** III	**2:15pm** IV	**5:00pm** V
Buy side	Mean	3.43	3.64	3.32	3.41	3.42
	Variance	18.42	20.06	17.67	17.83	17.52
	Skewness	2.41	2.34	2.60	2.33	2.35
	Up to 1 Δ	0.40	0.38	0.40	0.41	0.41
	Up to 3 Δ	0.68	0.66	0.69	0.68	0.68
	Up to 5 Δ	0.82	0.81	0.84	0.82	0.82
	Up to 10 Δ	0.93	0.92	0.94	0.93	0.93
	Up to 20 Δ	0.99	0.99	0.99	0.99	0.99
	Up to 30 Δ	1.00	1.00	1.00	1.00	1.00
Sell side	Mean	4.63	4.68	4.64	4.56	4.73
	Variance	25.16	27.51	25.77	23.73	24.20
	Skewness	1.84	1.83	1.89	1.77	1.74
	Up to 1 Δ	0.30	0.31	0.29	0.30	0.28
	Up to 3 Δ	0.55	0.56	0.55	0.55	0.53
	Up to 5 Δ	0.72	0.73	0.72	0.72	0.70
	Up to 10 Δ	0.88	0.87	0.88	0.89	0.88
	Up to 20 Δ	0.98	0.98	0.98	0.98	0.98
	Up to 30 Δ	1.00	1.00	1.00	1.00	1.00

In Fig. 2, we plot the loadings of $RLIQ^{buy}$ and $RLIQ^{sell}$ that are used to weigh the information provided at different price levels. The signs of the loadings of the first principal component are chosen so that the sign corresponding to first price distance ($\Delta = 0$) is positive, both for the buy and sell sides of the market. The figure reveals that if the frequency of orders waiting around the best quotes increases, this increases *RLIQ* due to positive loadings assigned to the information provided by the top

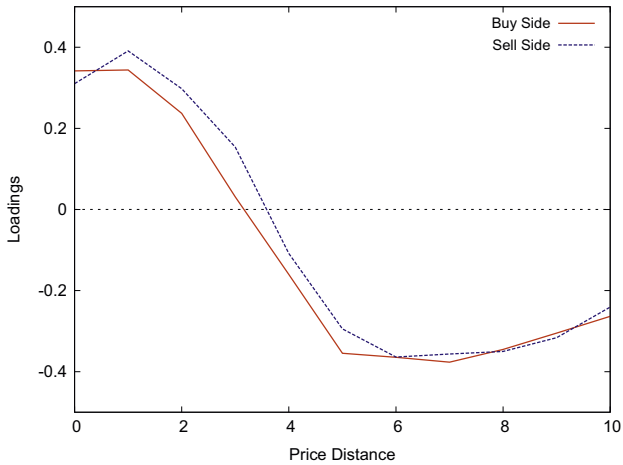

Fig. 2. Loadings of relative liquidity. This figure plots the loadings of the first principal component of the aggregate limit order book distribution function (*aggPDF*) defined in (1). The loadings are used to weigh the frequency of orders waiting at different price levels. The first 10 price distances are considered (i.e., $\Delta_c = 10$).

of the book. On the other hand, an increase in the proportion of orders waiting farther away from the best quotes translates into a decrease in *RLIQ*. This easy-to-interpret pattern in the loadings provides further justification for the use of this summary measure as a possible predictor of volatility.

4. Predictive analysis

4.1. Methodology

To evaluate the information content of a limit order book on future volatility, we rely on a standard predictive regression model of intraday volatility:

$$\sigma^M_{\tau+1} = a_0 + a_1\sigma^M_\tau + a_2 RLIQ^{\text{buy}}_\tau + a_3 RLIQ^{\text{sell}}_\tau + \sum_{j=1}^{20} b_j D_{j,\tau} + \text{controls} + \varepsilon_{\tau+1}, \tag{2}$$

where for a given interval τ, σ^M_τ is the mid-quote-volatility of the value-weighted index, and $RLIQ^{\text{buy}}_\tau$ and $RLIQ^{\text{sell}}_\tau$ are the proposed relative liquidity summary measures, calculated as the first principal component of the aggregate limit order book distribution for the buy and sell sides of the market, respectively. $D_{j,\tau}$ is the intraday dummy that equals to 1 if $j = \tau$. We include the lagged volatility, σ^M_τ, and intraday dummies in the set of explanatory variables to control the well-known systematic intraday patterns and clustering in volatility. Furthermore, we employ both the standard predictors of volatility and other liquidity measures as control variables, which are introduced in Section 4.3.

4.2. Measuring volatility: the two scales realized volatility estimator

A common practice to estimate return volatility is to use the sum of squared returns. However, it is not entirely appropriate to use the realized variance when using high-frequency data due to microstructure noise and jumps embedded in the data. Researchers have addressed this issue and proposed ways to improve the estimator. Ait-Sahalia, Mykland, and Zhang (2005), Bandi and Russell (2009), Ghysels and Sinko (2011), among others, focus on optimal sampling frequency to smooth the noise. Zhou (1996) and Hansen and Lunde (2006) consider a first-order autocorrelation to bias-correct the realized variance. Focusing on assets with large tick sizes, Delattre, Robert, and Rosenbaum (2013) propose a statistical methodology to estimate the efficient price of an asset through the order flow.

Barndorff-Nielsen and Shephard (2004), Andersen, Dobrev, and Schaumburg (2012), among others, propose volatility estimators that are robust to the presence of jumps in data.

In this paper, we employ the two scales realized volatility (TSRV) of Zhang, Mykland, Ait-Sahalia (2005) and Ait-Sahalia, Mykland, and Zhang (2011). In the presence of the microstructure noise, the TSRV estimator gives an unbiased and consistent estimate of volatility, under the assumptions that the log-price process follows a geometric Brownian motion and that the noise term is i.i.d. and independent of the true price process. The TSRV is defined as follows:

$$\sigma_\tau^M = \sqrt{\frac{1}{K}\sum_{k=1}^{K}[MQ, MQ]_\tau^{sparse,k} - \frac{1}{K}[MQ, MQ]_\tau^{(all)}}, \tag{3}$$

where $[MQ, MQ]_\tau^{(all)}$ is the realized variance of the mid-quote returns of the value-weighted index calculated in a trading interval τ with a sample size of T. To obtain $[MQ, MQ]^{sparse,k}$, we first divide the sample into K moving window subsamples [following Ait-Sahalia, Mykland, and Zhang, 2011, K is set to be 5 minutes] with a fixed length of N, where $N = T - K$. For example, the first subsample starts with the first and ends with the Nth observation, whereas the second subsample starts with the second and ends with $(N+1)$th observation. Then, we sample sparsely with a 30-second frequency. Thus $[MQ, MQ]^{sparse,k}$ is the realized variance estimator of the kth 30-second-sampled mid-quote returns.

Finally, note that although TSRV is a noise-robust estimator, it is not robust in the presence of jumps. In Section 5.6, we test the sensitivity of the results under alternative jump-robust volatility estimators.

4.3. Control variables

Our first set of covariates includes the variables that have been shown as predictors of volatility. First, consistent with Bollerslev and Domowitz (1993), Jones, Kaul, and Lipson (1994), and Foucault, Moinas, and Theissen (2007), the number of trades occurring in interval τ, NT, and the average trade size, AQ, are included to capture the trading activity. In a related study, Foucault, Moinas, and Theissen (2007) show that the bid–ask spread is informative of future individual stock volatility. Hence, we also include the relative spread, $relSPR_\tau$, which is calculated as the ratio of the bid–ask spread to the mid-quote prices for each interval. Finally, we consider the slope of a limit order book, $SLOPE$, as an explanatory variable following Naes and Skjeltorp (2006) and Duong and Kalev (2008). $SLOPE$ aggregates the price-quantity information in different quotes and measures the sensitivity of the quantity supplied in the book with respect to the prices.

Our second set of covariates includes other liquidity measures. We first consider standard depth measures. Depth, defined as the total volume available to be traded at the best bid or ask price, is one of the traditional measures of liquidity. We calculate $DEPTHi^{buy}$ ($DEPTHi^{sell}$) for $i = 1, 2, \ldots, 5$, which denotes the volume of orders waiting at the ith best bid (ask) to capture the volume available at and beyond the best quotes for the buy and sell sides of the market, respectively. Second, we employ the Amihud's (2002) illiquidity measure, AMR, which is calculated as the ratio of absolute stock return to the turnover. Another related illiquidity measure is the log quote slope, $\log QS$, which is introduced by Hasbrouck and Seppi (2001). A decrease in the $\log QS$ means that the slope of the best quotes is flatter and the market is more liquid. Finally, we consider the illiquidity measure proposed by Domowitz, Hansch, and Wang (2005), DHW, which measures the cost of buying and selling Q shares of the stock, simultaneously. We set Q as the median of the accumulated volume of orders waiting in the book for a given stock.

All of the control variables are calculated as the equal-weighted cross-sectional average of the individual stock measures. As a robustness check, we repeat the analysis by calculating the value-weighted and trade weighted averages of the explanatory variables to proxy the aggregate measures. The results presented in Section 5.6 reveal that the main findings are confirmed.

Compared to standard liquidity measures like spread, depth, and ratios based on both spread and depth, $RLIQ$ provides a more complete picture of the liquidity provision by considering the book beyond the best quotes. Moreover, instead of focusing on the volume of the orders waiting, $RLIQ$ is

based on the distribution of volume at a given time, hence, it reveals information of how quoted depths are spread or concentrated throughout a limit order book.

5. Empirical findings

5.1. One-period-ahead predictive regressions

We first examine the predictive power of relative liquidity, *RLIQ*, for the 15-minutes-ahead market volatility. To account for the intraday patterns, all of the specifications include 21 trading intervals as intraday dummies. To conserve space, we do not report the estimated coefficients of the dummy variables. To improve the ease of interpretation of the estimated coefficients, all of the explanatory variables are standardized to have mean zero and unit variance, and the dependent variable is presented in percentage terms.

Table 2 reveals that relative liquidity variables are significantly and negatively related to the one-period-ahead market volatility at the 5% level. The proportion of the variation in market volatility explained by our measures is about 22%. When the intraday dummies are not included in the specification, $RLIQ^{buy}$ and $RLIQ^{sell}$ alone explain around 16% of the variation in volatility. As Fig. 2 shows, the loadings of *RLIQ* are positive for the proportion of volume around the best quotes and turn negative for the orders waiting away from the best five quotes. Hence, the negative sign of the

Table 2
Predictability regressions.
The table provides the estimated coefficients of the regression model defined in (2). The dependent variable is the 15-minutes-ahead market volatility, $\sigma^M_{\tau+1}$, calculated as the mid-quote volatility of the value-weighted index via the TSRV estimator (multiplied by 100). $RLIQ^{buy}$ ($RLIQ^{sell}$) is the first principal component of the aggregate limit order book distribution for the buy (sell) side, as outlined in Section 3.2. All of the control variables are constructed as the cross-sectional average of the corresponding individual stock measures. *SLOPE* is the slope of the limit order book, *relSPR* is the relative spread, *NT* is the number of trades, and *AQ* is the average trade size. *AMR* is the Amihud (2002) illiquidity measure. The log*QS* is the log quote slope, introduced by Hasbrouck and Seppi (2001). Finally, *DHW* is the Domowitz, Hansch, and Wang (2005) illiquidity measure. All of the explanatory variables are standardized. *t*-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals and reported in parentheses. ***, ** and * indicate significance at the 0.1%, 1%, and 5% levels, respectively. For the sake of brevity, the estimated coefficients of the intraday dummies are omitted.

Dep. var.: $\sigma^M_{\tau+1}$	I	II	III	IV	V	VI
$RLIQ^{buy}_\tau$	−0.044***		−0.035***	−0.031***	−0.029***	−0.027***
	(−7.38)		(−7.47)	(−7.16)	(−6.29)	(−5.70)
$RLIQ^{sell}_\tau$	−0.028***		−0.022***	−0.017***	−0.010	−0.010
	(−4.25)		(−4.14)	(−3.45)	(−1.75)	(−1.73)
$\overline{SLOPE}_\tau$				0.009	0.016**	0.014*
				(1.45)	(2.90)	(2.38)
$\overline{relSPR}_\tau$				0.025***	0.013	0.012
				(4.89)	(1.77)	(1.54)
$\overline{NT}_\tau$				0.007		0.007
				(1.27)		(1.28)
$\overline{AQ}_\tau$				0.000		0.002
				(0.10)		(0.36)
$\overline{AMR}_\tau$					0.001	0.001
					(0.33)	(0.45)
$\overline{\log QS}_\tau$					0.024*	0.025*
					(2.13)	(2.16)
$\overline{DHW}_\tau$					0.001	0.001
					(0.17)	(0.20)
σ^M_τ		0.037***	0.023***	0.015***	0.015***	0.012*
		(6.28)	(5.04)	(3.32)	(3.39)	(2.51)
Adj. R^2(%)	21.95	17.25	25.15	28.48	29.45	29.53

coefficients indicates that an increase in liquidity beyond the best quotes relative to the top of the book is followed by a higher level of volatility in the next period. If the volume of orders waiting to be executed is accumulated more beyond the best prices, incoming investors may interpret this as mispricing of the current quotes. Hence, large price movements are more likely to happen, creating higher future volatility.

As expected, lagged volatility is highly and positively related to one-period-ahead volatility. However, the predictive power of $RLIQ^{buy}$ is higher than the lagged volatility. A one standard deviation increase in $RLIQ^{buy}$ decreases 15-minutes-ahead volatility by 4.4 bps, whereas a one standard deviation increase in volatility increases the next period volatility by 3.7 bps. Column III in Table 2 shows that when relative liquidity variables and lagged volatility are included in the specification, the adjusted R^2 increases to over 25%.

Columns IV and V confirm the robustness of the predictive power of *RLIQ* for the one-period-ahead market volatility when the standard predictors of volatility and alternative liquidity measures are considered. Not surprisingly, relative spread ($\overline{relSPR}$) is informative of future volatility at the 5% level of significance. Note that by construction, *RLIQ* is related to the depth dimension of liquidity and it can be thought as a complement of spread. Moreover, the slope of the book ($\overline{SLOPE}$), as well as the slope of the best quotes ($\overline{\log QS}$) are positively and significantly correlated with future volatility. On the other hand, the significance of $RLIQ^{sell}$ decreases when all of the control variables are included in the setting. This result is consistent with the literature documenting that buy orders are more information-driven than sell orders. The informed traders may exploit their informational advantage by submitting buy orders (e.g., Burdett and O'Hara, 1987; Griffiths et al., 2000; Duong and Kalev, 2008, among others).

Our results further extend the findings of Foucault, Moinas, and Theissen (2007) and Duong and Kalev (2008), who document that the relative spread and the slope of the book, respectively, have explanatory power for future individual stock volatility. We show that the cross-sectional average of both measures ($\overline{SLOPE}$ and $\overline{relSPR}$) have explanatory power for the market volatility as well. Moreover, we provide new empirical evidence that the measure of Hasbrouck and Seppi (2001), log quote slope, is significantly and positively related to the subsequent market volatility. Yet, the estimated (standardized) coefficients and *t*-statistics of $RLIQ^{buy}$ are always the highest among the alternative variables. They are robust and stable in all of the specifications examined. The adjusted R^2 increases from 25.15% to only 29.53% when all of the control variables are included in addition to *RLIQ* variables.

The documented results are not spurious since the augmented Dickey-Fuller and Phillips-Perron tests show that the null of a unit-root in volatility and in all of the explanatory variables can be rejected at the 5% level of significance. The estimated coefficient of the *non-standardized* lagged volatility – the AR(1) coefficient – equals to 0.34. We analyze the persistency of volatility and find that the estimated AR(1) coefficient increases to 0.42 and 0.60 for half-an-hour and daily sample periods, respectively. Persistency decreases with the sampling frequency; it is the lowest for the 15-minute volatility and the highest for the daily volatility. This result suggests that integrated volatility is estimated with an error, which is not surprising given that volatility is a latent variable, rather than an observable one. Although this problem is partially alleviated by subsampling and averaging in TSRV, we see that especially the 15-minute volatility is still subject to an estimation error. However, even in the presence of possibly high estimation error, we document a robust and strong dependency between relative liquidity and future volatility.

5.2. Relative liquidity vs. standard depth measures

Standard depth measures consider the volume of orders waiting at a given price. In a thin limit order book (when the volume at the best quotes is small) any "large" market order has a price impact and changes the quotes. Hence, it is not surprising to expect a link between the standard depth measures and future price movements. On the other hand, relative liquidity (*RLIQ*) extracts the relative accumulation of orders rather than the absolute volume of orders. Hence, a thin book is not necessarily associated with low *RLIQ*. *RLIQ* is lower when the orders are accumulated further away from the best quotes. However, as it is constructed from the distribution of volume at a given time, it may share common information with standard depth variables. Thus we next examine whether *RLIQ*

is still significant in explaining subsequent volatility under the presence of depth variables. To this end, we include the volume of orders at different prices along with the *RLIQ* measures in our analysis. Similarly, all of the specifications include the interval dummies and lagged volatility as control variables.

Table 3 shows that $\overline{DEPTH1}^{buy}$ and $\overline{DEPTH1}^{sell}$, the total volume of orders waiting at the best bid and ask prices, respectively, significantly explain future market volatility at the 5% level. A decrease in the volume of orders at the best quotes creates higher subsequent volatility. However, when relative liquidity measures are included in the specification, $\overline{DEPTH1}$ variables are no longer significant. The highest adjusted R^2 is only 26.11% even when the total depth up to the fifth quotes is included in the analysis. That is, by including 10 depth variables in addition to our measures, we only increase the adjusted R^2 by less than 1%.

We conclude that the relative concentration of depth provision, rather than the absolute volume, reveals more information about future volatility. *RLIQ* has a superior in-sample predictive power compared to the standard depth measures.

Table 3
RLIQ vs. standard depth measures.

This table shows the in-sample predictive power of *RLIQ* in comparison to the depth measures, which are defined as the quoted volume of orders waiting at a given threshold. The dependent variable is the 15-minutes-ahead market volatility, $\sigma^M_{\tau+1}$, calculated as the mid-quote volatility of the value-weighted index via the TSRV estimator (multiplied by 100). $RLIQ^{buy}$ ($RLIQ^{sell}$) is the first principal component of the aggregate limit order book distribution for the buy (sell) side, as outlined in Section 3.2. $\overline{DEPTHi}^{buy}$ ($\overline{DEPTHi}^{sell}$) is the total volume of buy (sell) orders waiting at price distance *i*. All of the explanatory variables are standardized. *t*-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals and reported in parentheses. ***, ** and * indicate significance at the 0.1%, 1%, and 5% levels, respectively. For the sake of brevity, the estimated coefficients of the intraday dummies are omitted.

Dep. var.: $\sigma^M_{\tau+1}$	I	II	III	IV	V
$RLIQ^{buy}_\tau$	−0.035***		−0.032***	−0.033***	−0.036***
	(−7.47)		(−7.02)	(−6.84)	(−7.04)
$RLIQ^{sell}_\tau$	−0.022***		−0.018**	−0.016*	−0.008
	(−4.14)		(−3.03)	(−2.54)	(−1.10)
$\overline{DEPTH1}^{buy}_\tau$		−0.012*	−0.003	−0.001	0.000
		(−2.26)	(−0.66)	(−0.21)	(0.00)
$\overline{DEPTH1}^{sell}_\tau$		−0.011*	−0.006	−0.003	−0.003
		(−2.50)	(−1.55)	(−0.51)	(−0.46)
$\overline{DEPTH2}^{buy}_\tau$				−0.007	−0.008
				(−1.11)	(−1.34)
$\overline{DEPTH2}^{sell}_\tau$				0.001	0.000
				(0.14)	(0.06)
$\overline{DEPTH3}^{buy}_\tau$				−0.001	0.010
				(−0.15)	(1.29)
$\overline{DEPTH3}^{sell}_\tau$				−0.001	−0.011
				(−0.10)	(−1.42)
$\overline{DEPTH4}^{buy}_\tau$					−0.017**
					(−2.67)
$\overline{DEPTH4}^{sell}_\tau$					0.012
					(1.38)
$\overline{DEPTH5}^{buy}_\tau$					−0.005
					(−0.84)
$\overline{DEPTH5}^{sell}_\tau$					0.002
					(0.23)
σ^M_τ	0.023***	0.032***	0.022***	0.022***	0.021***
	(5.04)	(6.31)	(4.96)	(5.00)	(4.73)
Adj. R^2(%)	25.15	19.94	25.32	25.08	26.11

5.3. Predicting further horizons

In this section, we examine the informativeness of the limit order book distribution at time τ on multiple-period-ahead volatility. Specifically, we run the baseline regression model (2), while we calculate the dependent variable as the mid-quote volatility of the index at time $\tau+h$, with $h=1,2,\ldots,11$ (i.e., up to 165 minutes ahead). The independent variables are calculated based on the limit order book information at trading interval τ. For example, $\tau+2$ refers to volatility in a 30-minutes-ahead trading interval.

Table 4 shows that the significance of the estimated coefficients, as well as the predictive power of relative liquidity measures are (almost) monotonically decreasing with the prediction horizon. $RLIQ^{\text{buy}}$ has a significant forecasting power with respect to market volatility up to 165-minutes-ahead. Moreover, the slope of the book, the relative spread, and the quote-slope significantly predict volatility at longer horizons. However, relative liquidity is both economically and statistically the strongest predictor.

5.4. Out-of-sample tests

In this section, we evaluate the out-of-sample forecasting ability of *RLIQ* compared to historical volatility. Specifically, for a subsample of observations up to a given time interval τ, we compare the h-period-ahead squared forecast errors with the squared difference between the realized value at $\tau+h$ and the sample mean value up to time τ. To do so, we split our data into two subsample periods: L_{train} is the training period and L_{test} is the testing period with $L_{\text{train}}+L_{\text{test}}=L$, the total number of time intervals. We then re-estimate the parameters of the model, in which we use the variable of interest, as the predictor. Recursive estimators of h-period-ahead forecasts are based on the sample starting from L_{train} up to $L-h$. For L_{train} equals to 350 and 400 observations, we calculate the following error terms:

$$\varepsilon_{1,\tau+h}=\sigma^{M}_{\tau+h}-\widehat{\sigma^{M}_{\tau+h}},$$
$$\varepsilon_{2,\tau+h}=\sigma^{M}_{\tau+h}-\overline{\sigma^{M}_{\tau}},$$

where $\sigma^{M}_{\tau+h}$ is the two scales realized market volatility, $\overline{\sigma^{M}_{\tau}}$ is the mean value of the market volatility up to time τ, and $\widehat{\sigma^{M}_{\tau+h}}$ is the fitted market volatility obtained by regressing volatility on the variable of interest, such as *RLIQ* or other liquidity measures.

We evaluate the comparison by using two different metrics: the difference in mean-squared errors (ΔMSE) and the out-of-sample R^2. If the proposed measure has superior out-of-sample forecasting ability relative to the average of past data, then both of these measures will be positive. We employ the Diebold and Mariano (1995) predictive ability test (DM) to test the significance of ΔMSE. Finally, the out-of-sample R^2 is calculated as follows:

$$R^2=1-\frac{\sum_{\tau=1}^{L_{\text{test}}-h}\varepsilon^2_{1,\tau+h}}{\sum_{\tau=1}^{L_{\text{test}}-h}\varepsilon^2_{2,\tau+h}}. \tag{4}$$

Table 5 reveals that forecasts based either on relative liquidity, relative spread, number of trades, or slope of the best quotes increase the predictive power relative to forecasts based only on the sample mean of past volatility. Moreover, the predictive power of the variables is decreasing almost monotonically with the prediction horizon. When $L_{\text{train}}=400$, $RLIQ^{\text{buy}}$ delivers out-of-sample R^2 from 12.9% when forecasting one-period-ahead market volatility up to 5.5% when predicting 90-minutes-ahead market volatility.

On the other hand, the results are stronger for both relative spread and log quote slope when $L_{\text{train}}=350$. Note that both $\overline{relSPR}$ and $\overline{\log QS}$ are the variables that capture the liquidity at the best quotes, in other words, the tightness dimension of liquidity only. Thus as a further analysis, we examine whether including relative liquidity, in addition to the tightness dimension of liquidity, produces better forecasts. To do so, the first forecast errors are calculated from the model where $RLIQ^{\text{buy}}$ and $\overline{relSPR}$ ($\overline{\log QS}$) are the explanatory variables, whereas the second (benchmark) forecast

Table 4
Predictability regressions – further horizons.

This table presents the estimated coefficients of the regression model defined in (2). The dependent variable is the market volatility, $\sigma^M_{\tau+h}$, calculated as the mid-quote volatility of the value-weighted index via the TSRV estimator (multiplied by 100) in period $\tau+h$, for $h=1,2,\ldots,11$. $RLIQ^{\text{buy}}$ ($RLIQ^{\text{sell}}$) is the first principal component of the aggregate limit order book distribution for the buy (sell) side, as outlined in Section 3.2. All of the control variables are constructed as the cross-sectional average of the corresponding individual stock measures. *SLOPE* is the slope of the limit order book, *relSPR* is the relative spread, *NT* is the number of trades, and *AQ* is the average trade size. *AMR* is the Amihud (2002) illiquidity measure. The log*QS* is the log quote slope, introduced by Hasbrouck and Seppi (2001). Finally, *DHW* is the Domowitz, Hansch, and Wang (2005) illiquidity measure. In Panel A, for every time horizon, we report the "simple" regressions, where relative liquidity measures along with the lagged volatility and interval dummies are used as regressors. On the other hand, Panel B reports the results when all of the control variables are included in the regression equation. All of the explanatory variables are standardized. *t*-Statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals and reported in parentheses. ***, ** and * indicate significance at the 0.1%, 1%, and 5% levels, respectively. For the sake of brevity, the estimated coefficients of the intraday dummies are omitted.

Dep. var.: $\sigma^M_{\tau+h}$	Panel A: "simple" regressions						Panel B: multiple regressions					
	0–15	**15–30**	.	**120–135**	**135–150**	**150–165**	**0–15**	**15–30**	.	**120–135**	**135–150**	**150–165**
$RLIQ^{\text{buy}}_\tau$	−0.035***	−0.032***	.	−0.030***	−0.028**	−0.025**	−0.027***	−0.023***	.	−0.030***	−0.029***	−0.017*
	(−7.47)	(−5.97)	.	(−3.95)	(−3.21)	(−2.77)	(−5.70)	(−5.33)	.	(−3.87)	(−3.92)	(−1.98)
$RLIQ^{\text{sell}}_\tau$	−0.022***	−0.025***	.	−0.016	−0.014	−0.015	−0.010	−0.011*	.	0.00	−0.003	0.003
	(−4.14)	(−3.78)	.	(−1.92)	(−1.66)	(−1.85)	(−1.73)	(−2.26)	.	(0.16)	(−0.38)	(0.33)
$\overline{SLOPE}_\tau$			.				0.014*	0.021**	.	0.01	0.012	0.010
			.				(2.38)	(3.04)	.	(1.44)	(1.39)	(1.34)
$\overline{relSPR}_\tau$			.				0.012	0.015*	.	0.03**	0.045***	0.010
			.				(1.54)	(2.44)	.	(3.06)	(3.79)	(1.07)
$\overline{NT}_\tau$			.				0.007	0.010	.	0.00	0.001	0.011
			.				(1.28)	(1.59)	.	(0.25)	(0.08)	(1.42)
$\overline{AQ}_\tau$			.				0.002	0.004	.	0.00	−0.009	0.011
			.				(0.36)	(0.89)	.	(0.04)	(−1.46)	(1.54)
$\overline{AMR}_\tau$			.				0.001	0.007***	.	0.00	−0.001	0.001
			.				(0.45)	(5.98)	.	(0.20)	(−0.22)	(0.69)
$\overline{\log QS}_\tau$			.				0.025*	0.030***	.	0.00	−0.017	0.028*
			.				(2.16)	(4.13)	.	(0.29)	(−1.23)	(2.30)
$\overline{DHW}_\tau$			.				0.001	0.001	.	0.020*	0.015	0.013*
			.				(0.20)	(0.21)	.	(2.08)	(1.94)	(2.14)
σ^M_τ	0.023***	0.016***	.	0.005	0.011	0.011	0.012*	0.001	.	0.00	0.007	−0.006
	(5.04)	(4.17)	.	(1.00)	(1.66)	(1.62)	(2.51)	(0.32)	.	(−0.76)	(1.02)	(−0.88)
Adj. R^2(%)	25.15	20.93	.	12.61	13.36	14.01	29.53	28.03	.	19.20	20.00	21.81

Table 5
Out-of-sample forecasting evaluation.

This table presents out-of-sample forecasting experiment results. The h-period-ahead forecast error is obtained as the difference between the realized volatility at $\tau+h$ and the fitted value of the predictive regression estimated up to time τ with the variable of interest listed in the first column used as a regressor. On the other hand, the competing error is calculated from the sample mean volatility up to time interval τ. The dependent variable is the market volatility, $\sigma^{M}_{\tau+h}$, calculated as the mid-quote volatility of the value-weighted index via the TSRV estimator (multiplied by 100) in period $\tau+h$, for $h=1,2,\ldots,6$. $RLIQ^{\text{buy}}$ is the first principal component of the aggregate limit order book distribution for the buy side, as outlined in Section 3.2. $\overline{relSPR}$, $\overline{NT}$, and $\overline{\log QS}$ are the cross-sectional averages of the relative spread, number of trades, and log quote slope, respectively. Although we examine the out-of-sample forecasting power of all of the control variables introduced in Section 4.3, for the sake of brevity we report only the ones with significant forecasting power. The out-of-sample R^2(%), the difference in mean-squared errors (ΔMSEx1000), and the Diebold and Mariano (1995) predictive ability test (DM) are reported. Panels A and B report the results when the training period is set to 400 and 350 observations, respectively. ***, ** and * indicate significance at the 0.1%, 1%, and 5% levels, respectively.

Forecasting variable		0–15 minutes	15–30 minutes	30–45 minutes	45–60 minutes	60–75 minutes	75–90 minutes
Panel A: Training period: 400 obs.							
$RLIQ^{\text{buy}}_{\tau}$	Out-of-sample R^2(%)	12.86	10.13	9.20	8.25	7.08	5.50
	ΔMSE	1.91**	1.50*	1.36**	1.23*	1.06*	0.83
	DM t-stat	2.68	2.45	2.74	2.46	2.35	1.86
$\overline{relSPR}_{\tau}$	Out-of-sample R^2(%)	10.85	9.95	8.94	9.13	9.48	10.57
	ΔMSE	1.61*	1.47*	1.32	1.36	1.42*	1.59*
	DM t-stat	2.24	2.18	1.94	1.94	1.99	2.30
$\overline{NT}_{\tau}$	Out-of-sample R^2(%)	12.89	8.32	0.69	−0.54	0.38	2.16
	ΔMSE	1.91*	1.23*	0.10	−0.08	0.06	0.32
	DM t-stat	2.46	2.00	0.31	−0.49	0.25	0.87
$\overline{\log QS}_{\tau}$	Out-of-sample R^2(%)	14.83	12.59	10.57	10.54	10.04	11.83
	ΔMSE	2.20**	1.86**	1.56*	1.57*	1.50*	1.78*
	DM t-stat	2.92	2.66	2.50	2.43	2.24	2.41
Panel B: Training period: 350 obs.							
$RLIQ^{\text{buy}}_{\tau}$	Out-of-sample R^2(%)	11.23	8.70	7.95	7.19	6.32	4.93
	ΔMSE	1.80**	1.40*	1.28**	1.16*	1.03*	0.80
	DM t-stat	2.76	2.44	2.69	2.46	2.37	1.90
$\overline{relSPR}_{\tau}$	Out-of-sample R^2(%)	13.35	12.38	11.66	11.68	11.89	12.59
	ΔMSE	2.14**	1.99**	1.88**	1.89**	1.93**	2.05**
	DM t-stat	2.98	2.93	2.75	2.73	2.74	3.03
$\overline{NT}_{\tau}$	Out-of-sample R^2(%)	11.11	6.91	0.73	−0.41	0.03	1.24
	ΔMSE	1.78*	1.11*	0.12	−0.07	0.00	0.20
	DM t-stat	2.43	1.97	0.35	−0.44	0.03	0.60
$\overline{\log QS}_{\tau}$	Out-of-sample R^2(%)	15.67	12.89	11.10	10.66	11.34	12.56
	ΔMSE	2.51***	2.07**	1.79**	1.73**	1.84**	2.05**
	DM t-stat	3.60	3.22	3.12	2.92	2.95	3.00

errors are calculated from the model, in which relative spread (log quote slope) is the only explanatory variable. Similarly, we repeat the analysis for two different estimation window sizes: 350 and 400 observations. The results show that including $RLIQ$ in the analysis increases the out-of-sample R^2 by

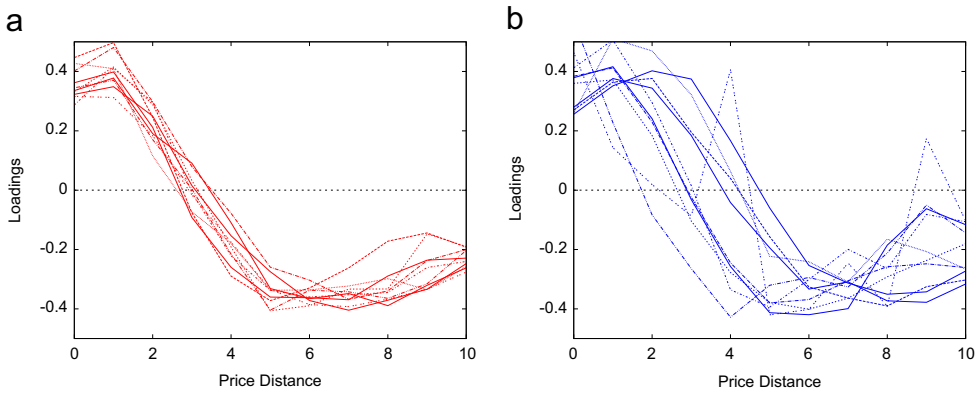

Fig. 3. Individual stock loadings. This figure plots the loadings of the first principal component for a given stock's limit order book distribution (*indPDF*) for the buy side of the market (Panel A) and the sell side of the market (Panel B). The loadings of a stock are presented only if the relative liquidity of the given stock is a significant predictor of 15-minutes-ahead volatility.

Table 6
Individual stock predictability regressions.

This table presents the estimated coefficients of the regression model defined in (5). $RLIQ^{\text{ind,buy}}$ ($RLIQ^{\text{ind,sell}}$) is the first principal component of the individual stock limit order book distribution, for the buy (sell) side. In a given trading interval τ, *SLOPE* is the slope of the limit order book, *relSPR* is the relative spread, *NT* is the number of trades, and *AQ* is the average trade size. *AMR* is the Amihud (2002) illiquidity measure. The log*QS* is the log quote slope, introduced by Hasbrouck and Seppi (2001). Finally, *DHW* is the Domowitz, Hansch, and Wang (2005) illiquidity measure. All of the explanatory variables are standardized. The dependent variable is $\sigma_{\tau+1}$, which is the volatility calculated using the mid-quotes of the orders originated in interval $\tau+1$ via the TSRV estimator (multiplied by 100). Panel A shows the results for the pooled regression. *t*-Statistics based on cluster robust standard errors on stock level are reported in parentheses. Panel B summarizes the results when the model is estimated for each stock separately. The cross-sectional median of the estimated significant coefficients at the 5% level is reported. In brackets, first, the percentage of the stocks with a significant coefficient at the 5% level and second, the percentage of the positive estimates, are reported. ***, ** and * indicate significance at the 0.1%, 1%, and 5% levels, respectively. For the sake of brevity, the estimated coefficients of the intraday dummies and stock fixed effects are omitted.

Dep. var.: $\sigma_{s,\tau+1}$	**Panel A: Pooled regression**				**Panel B: Summary of individual regressions**			
	I	II	III	IV	V	VI	VII	VIII
$RLIQ_\tau^{\text{ind,buy}}$	−0.057***	−0.058***	−0.058***	−0.055***	−0.065	−0.057	−0.055	−0.052
	(−13.07)	(−12.95)	(−12.60)	(−11.94)	[87/0]	[87/0]	[87/0]	[80/0]
$RLIQ_\tau^{\text{ind,sell}}$	−0.024***	−0.026***	−0.019***	−0.019***	−0.039	−0.046	−0.048	−0.042
	(−4.99)	(−5.36)	(−3.56)	(−3.66)	[37/9]	[33/0]	[27/0]	[27/0]
$SLOPE_\tau$		−0.019*	0.031***	0.022***		−0.057	0.054	0.039
		(−2.47)	(4.47)	(3.58)		[43/8]	[30/78]	[20/67]
$relSPR_\tau$		0.040***	−0.001	−0.007		0.088	−0.104	−0.164
		(4.09)	(−0.14)	(−0.67)		[37/91]	[30/33]	[23/14]
NT_τ		0.037***		0.040***		0.048		0.057
		(9.70)		(11.18)		[57/100]		[57/100]
AQ_τ		−0.005		0.001		0.023		0.036
		(−1.02)		(0.20)		[17/60]		[17/100]
AMR_τ			0.000	0.003			0.007	0.011
			(−0.02)	(0.93)			[20/50]	[20/50]
$\log QS_\tau$			0.091***	0.094***			0.112	0.126
			(10.13)	(10.22)			[53/100]	[50/100]
DHW_τ			0.013**	0.014**			0.055	0.055
			(2.61)	(2.79)			[33/90]	[33/90]
σ_τ	0.082***	0.055***	0.064***	0.042***	0.083	0.068	0.065	0.060
	(16.19)	(9.59)	(14.29)	(7.62)	[97/100]	[50/100]	[77/100]	[37/100]
Adj. R^2(%)	13.24	14.75	15.50	15.95	10.27	13.13	14.08	14.95

Table 7
One-period-ahead predictive power-principal components.

This table presents in-sample predictive power of other principal components over 15-minutes-ahead market volatility, $\sigma^M_{\tau+1}$, calculated as the mid-quote volatility of the value-weighted index via the TSRV estimator (multiplied by 100). *PCi* is the *i*th principal component of the aggregate limit order book distribution (*aggPDF*) for the buy (sell) side, as outlined in Section 3.2. The rest of the variables are defined in Table 2. All of the explanatory variables are standardized. *t*-Statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals and reported in parentheses. ***, ** and * indicate significance at the 0.1%, 1%, and 5% levels, respectively. For the sake of brevity, the estimated coefficients of the intraday dummies are omitted.

Dep. var.: $\sigma^M_{\tau+1}$	I	II	III	IV	V	VI
$PC1^{buy}_\tau$ ($RLIQ^{buy}_\tau$)	−0.035***	−0.034***	−0.035***	−0.027***	−0.028***	−0.029***
	(−7.47)	(−7.27)	(−7.42)	(−5.70)	(−5.63)	(−5.85)
$PC1^{sell}_\tau$ ($RLIQ^{sell}_\tau$)	−0.022***	−0.022***	−0.022***	−0.010	−0.010	−0.012*
	(−4.14)	(−3.73)	(−3.77)	(−1.73)	(−1.71)	(−2.03)
$PC2^{buy}_\tau$		0.003	0.002		0.000	−0.001
		(0.86)	(0.73)		(−0.13)	(−0.42)
$PC2^{sell}_\tau$		−0.003	−0.003		0.001	0.001
		(−0.78)	(−0.81)		(0.21)	(0.26)
$PC3^{buy}_\tau$		−0.002	−0.002		0.001	0.001
		(−0.54)	(−0.57)		(0.40)	(0.25)
$PC3^{sell}_\tau$		0.007*	0.008*		0.003	0.004
		(2.14)	(2.50)		(0.89)	(1.17)
$PC4^{buy}_\tau$			0.005			0.004
			(1.51)			(1.13)
$PC4^{sell}_\tau$			0.000			−0.002
			(−0.07)			(−0.50)
$PC5^{buy}_\tau$			−0.008*			−0.007
			(−2.09)			(−1.83)
$PC5^{sell}_\tau$			−0.002			−0.003
			(−0.59)			(−0.82)
$\overline{SLOPE}_\tau$				0.014*	0.014*	0.016**
				(2.38)	(2.37)	(2.69)
$\overline{relSPR}_\tau$				0.012	0.013	0.013
				(1.54)	(1.63)	(1.68)
$\overline{NT}_\tau$				0.007	0.007	0.007
				(1.28)	(1.26)	(1.35)
$\overline{AQ}_\tau$				0.002	0.001	0.001
				(0.36)	(0.34)	(0.15)
$\overline{AMR}_\tau$				0.001	0.001	0.001
				(0.45)	(0.42)	(0.34)
$\overline{\log QS}_\tau$				0.025*	0.024	0.022
				(2.16)	(1.94)	(1.85)
$\overline{DHW}_\tau$				0.001	0.000	0.001
				(0.20)	(0.07)	(0.25)
σ^M_τ	0.023***	0.022***	0.021***	0.012*	0.012*	0.011*
	(5.04)	(4.82)	(4.54)	(2.51)	(2.48)	(2.33)
Adj. R^2(%)	25.15	25.38	25.76	29.53	29.27	29.42

13.9% and 11.7%, in addition to using only relative spread and the log quote slope, respectively, for $L_{\text{train}} = 400$. Three variables together deliver an out-of-sample R^2 of over 24% when forecasting one-period-ahead market volatility relative to forecasts based only on the sample mean of volatility. We conclude that capturing both the tightness and the depth dimension of liquidity significantly increases the out-of-sample forecasting power.

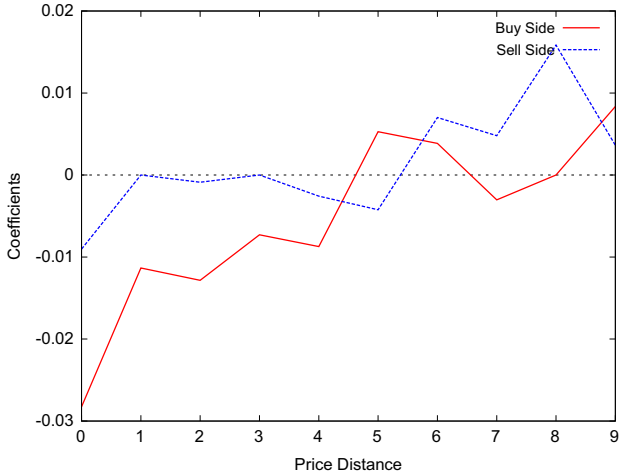

Fig. 4. Estimated LASSO coefficients. This figure presents the estimated least absolute shrinkage and selection operator (LASSO) coefficients with respect to price distances for buy and sell sides of the market.

5.5. Predicting individual stock volatilities

In this section, we examine the in-sample predictive power of a limit order book distribution over future volatility on an individual stock level. To this end, we first run the following predictive regression in a pooled data with stock fixed effects:

$$\sigma_{s,\tau+1} = a_0 + a_1\sigma_{s,\tau} + a_2 RLIQ_{s,\tau}^{\text{ind,buy}} + a_3 RLIQ_{s,\tau}^{\text{ind,sell}} + \sum_{j=1}^{20} b_j D_{j,\tau} + \sum_{s=1}^{30} c_s FE_s + \text{controls} + \varepsilon_{s,\tau+1}, \tag{5}$$

where for stock s and interval τ, $\sigma_{s,\tau}$ is the mid-quote TSRV, $RLIQ_{s,\tau}^{\text{ind,buy}}$ and $RLIQ_{s,\tau}^{\text{ind,sell}}$ are the first principal components of the individual stock limit order book distributions (*indPDF*) for the buy and sell sides of the market, respectively. $D_{j,\tau}$ is the intraday dummy that equals to 1 if $j=\tau$, and FE_s are stock-specific dummy variables allowing for stock fixed effects.

Fig. 3 reveals that the loadings of the first principal components differ slightly from one stock to another and they are similar to the loadings of the first principal component of the aggregate limit order book distribution presented in Fig. 2.

In Table 6 Columns I–IV, we report the estimated coefficients for (5) with the corresponding t-statistics. To take into account the possible cross-sectional variations that cannot be captured by the stock fixed effects, we also estimate the predictive regressions for each stock s separately. The summary of these results is presented in Columns V–VIII.

Our main result is confirmed in these individual volatility regressions. *RLIQ* is negatively related to future volatility for 87% of the stocks for the buy side of the market at the 5% level of significance. We conclude that the time series relationship between the aggregate liquidity and market volatility is not driven by variations in a particular stock or industry, but rather it is shared by the majority of the stocks. The results reveal the asymmetry between the buy and sell sides of *RLIQ* at the individual stock level as well. The sell side of the market is informative only for 37% of the stocks in the individual regressions at the 5% level of significance, in contrast to the informativeness for 87% of the stocks on the buy side. Similarly, both $RLIQ^{\text{ind,buy}}$ and $RLIQ^{\text{ind,sell}}$ are significant in the pooled regression, but the estimated coefficients of the buy side are at least two times greater than the sell side.

Table 8
Robustness – volatility.

This table presents the results when alternative volatility measures are employed to approximate the 15-minutes market volatility. In Columns I to II, volatility is calculated as the realized volatility (RV) of 30 seconds sampled squared mid-quote returns. Columns III through VI show the results when Barndorff-Nielsen and Shephard (2004)'s realized bipower variation (BPV) and tripower variation (TPV) are employed, respectively. Columns VII through X present the results when minRV and medRV of Andersen, Dobrev, and Schaumburg (2012) are used as volatility proxies, respectively. To account for the noise, the "pre-averaging" approach of Podolskij and Vetter (2009) is employed in all of the volatility estimators. All of the explanatory variables are standardized. In all of the specifications, *t*-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals. ***, ** and * indicate significance at the 0.1%, 1%, and 5% levels, respectively. For the sake of brevity, the estimated coefficients of the intraday dummies are omitted. All of the explanatory variables are defined in Table 2.

	$\sigma^M_{\tau+1,\mathrm{RV}}$		$\sigma^M_{\tau+1,\mathrm{BPV}}$		$\sigma^M_{\tau+1,\mathrm{TPV}}$		$\sigma^M_{\tau+1,\mathrm{minRV}}$		$\sigma^M_{\tau+1,\mathrm{medRV}}$	
	I	II	III	IV	V	VI	VII	VIII	IX	X
$RLIQ^{\mathrm{buy}}_\tau$	−0.047***	−0.039***	−0.041***	−0.035***	−0.042***	−0.037***	−0.037***	−0.032***	−0.041***	−0.035***
	(−6.66)	(−5.43)	(−6.03)	(−4.68)	(−6.06)	(−5.00)	(−5.58)	(−4.10)	(−6.09)	(−4.79)
$RLIQ^{\mathrm{sell}}_\tau$	−0.028***	−0.011	−0.028***	−0.010	−0.028***	−0.011	−0.027***	−0.009	−0.029***	−0.011
	(−3.76)	(−1.41)	(−3.52)	(−1.22)	(−3.57)	(−1.44)	(−3.47)	(−1.07)	(−3.77)	(−1.42)
$\overline{SLOPE}_\tau$		0.024*		0.022*		0.022**		0.018		0.024**
		(2.37)		(2.45)		(2.64)		(1.91)		(2.85)
$\overline{relSPR}_\tau$		0.023*		0.022		0.026*		0.019		0.024*
		(2.07)		(1.90)		(2.29)		(1.62)		(2.12)
$\overline{NT}_\tau$		0.008		0.008		0.011		0.008		0.010
		(1.04)		(1.00)		(1.33)		(0.94)		(1.23)
$\overline{AQ}_\tau$		0.005		0.001		−0.002		0.001		0.002
		(0.74)		(0.21)		(−0.43)		(0.15)		(0.33)
$\overline{AMR}_\tau$		0.006		0.001		0.002		−0.001		0.001
		(0.97)		(0.26)		(0.77)		(−0.42)		(0.29)
$\overline{\log QS}_\tau$		0.029		0.026		0.021		0.023		0.026
		(1.65)		(1.41)		(1.26)		(1.24)		(1.49)
$\overline{DHW}_\tau$		0.007		0.013		0.013		0.015		0.014*
		(1.01)		(1.68)		(1.82)		(1.85)		(1.96)
σ^M_τ	0.020***	0.007	0.025***	0.010	0.028***	0.014*	0.025***	0.012*	0.028***	0.012*
	(3.56)	(1.14)	(4.81)	(1.78)	(5.02)	(2.20)	(5.02)	(2.21)	(5.26)	(2.15)
Adj. R^2(%)	22.16	26.81	23.14	28.00	26.24	30.98	21.17	25.42	23.54	28.68

Table 9
Other robustness tests.

Columns I through IV present the results when *RLIQ* is re-calculated by using the information up to the 20th and 30th best quotes, respectively. Columns V and VI show the results when the sampling period is 30 minutes instead of 15 minutes. Columns VII through X report the results when the explanatory variables are aggregated via value-weighted and trade-weighted cross-sectional averages. Finally, Columns XI and XII present the estimated coefficients for the log-transformed variables. All of the explanatory variables are standardized. In all of the specifications *t*-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals. ***, ** and * indicate significance at the 0.1%, 1%, and 5% levels, respectively. For the sake of brevity, the estimated coefficients of the intraday dummies are omitted. All of the variables are defined in Table 2.

	$\Delta_c = 20$		$\Delta_c = 30$		30-minutes sampling		Value-weighted		Trade-weighted		Log transform.	
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
$RLIQ_\tau^{buy}$	−0.033***	−0.026***	−0.032***	−0.025***	−0.053***	−0.043***	−0.034***	−0.025***	−0.038***	−0.028***	−0.031***	−0.024***
	(−7.71)	(−5.65)	(−7.43)	(−5.50)	(−6.26)	(−6.22)	(−7.23)	(−5.09)	(−7.33)	(−5.30)	(−7.33)	(−5.98)
$RLIQ_\tau^{sell}$	−0.017**	−0.007	−0.016**	−0.006	−0.036***	−0.012	−0.023***	−0.012*	−0.027***	−0.016**	−0.022***	−0.008
	(−3.04)	(−1.12)	(−2.83)	(−0.91)	(−3.35)	(−1.67)	(−4.53)	(−2.32)	(−4.81)	(−2.65)	(−3.87)	(−1.55)
$\overline{SLOPE}_\tau$		0.015*		0.014*		0.034**		0.019*		0.021*		0.015**
		(2.51)		(2.39)		(3.20)		(2.34)		(2.23)		(2.70)
$\overline{relSPR}_\tau$		0.012		0.013		0.025*		−0.017		−0.006		0.011
		(1.58)		(1.67)		(2.49)		(−1.69)		(−0.60)		(1.43)
$\overline{NT}_\tau$		0.006		0.006		0.006		0.010*		0.007		0.006
		(1.06)		(1.15)		(0.62)		(2.30)		(1.57)		(0.99)
$\overline{AQ}_\tau$		0.003		0.003		0.011		−0.003		−0.004		−0.001
		(0.74)		(0.77)		(1.52)		(−0.84)		(−1.06)		(−0.26)
$\overline{AMR}_\tau$		0.002		0.002		0.009		0.001		0.002		0.001
		(0.65)		(0.58)		(1.71)		(0.26)		(0.77)		(0.44)
$\overline{\log QS}_\tau$		0.022		0.022		0.056***		0.046***		0.041**		0.029*
		(1.89)		(1.86)		(4.03)		(3.55)		(2.96)		(2.49)
$\overline{DHW}_\tau$		0.001		0.001		0.005		−0.001		−0.003		−0.002
		(0.21)		(0.16)		(0.72)		(−0.16)		(−0.59)		(−0.33)
σ_τ^M	0.022***	0.012*	0.023***	0.012**	0.033***	0.004	0.024***	0.013**	0.023***	0.015**	0.025***	0.015**
	(4.83)	(2.56)	(4.90)	(2.58)	(4.25)	(0.51)	(5.09)	(2.68)	(4.94)	(2.95)	(5.45)	(3.15)
Adj. R^2(%)	25.82	29.49	25.55	29.33	31.21	41.71	25.32	28.56	25.61	28.58	24.55	28.81

5.6. Robustness

5.6.1. Relative liquidity: alternative measures

We define relative liquidity as the first principal component of the limit order book distribution. In this section, we examine the sensitivity of the findings when the definition of the proposed measure, *RLIQ*, is changed. First, instead of using the first principal component, we consider the first three and five principal components of the aggregate limit order book distribution introduced in (1). The first three and five components explain 65% and 77% of the variation in the distribution, respectively. Table 7 shows that the first principal components for both the buy and sell sides have the leading explanatory power for future market volatility. Including the first three and five principal components, in addition to the first component, increases the adjusted R^2 by only 0.23% and 0.61%, respectively. Moreover, they do not have significant predictive power over volatility when other control variables are included.

One could easily argue in favor of using the empirical frequencies of orders waiting at each price distance instead of summarizing this information. Hence, as a second robustness test we consider the volume distribution for $\Delta_c = 10$ separately for each bin as predictors of volatility and exclude the last bin to avoid multicollinearity. In unreported results, we find that, by including 20 variables, instead of using only $RLIQ^{buy}$ and $RLIQ^{sell}$, the adjusted R^2 increases only by 1.01%.

Finally, we adopt another variable selection technique, least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996), to reduce the dimensionality instead of employing principal component analysis. LASSO finds the coefficients of a model by minimizing the sum of squared residuals plus an I1-norm penalty function. We determine the sparse model, which corresponds to minimum mean squared errors, by employing a 10-fold cross validation technique. Fig. 4 presents the estimated LASSO coefficients, from which several conclusions arise. First, the estimated coefficients are negative for the bins closer to the best quotes, suggesting that higher liquidity around the best quotes is associated with lower subsequent volatility. On the other hand, the coefficients switch sign after the five best quotes, similar to the loadings of the first principal component plotted in Fig. 2. Note that the signs of the loadings of *RLIQ* and those of the estimated LASSO coefficients are opposite as expected: on the one hand, the loadings summarize the liquidity provision in the limit order book; the higher *RLIQ*, the higher the liquidity around the best quotes. On the other hand, the LASSO coefficients summarize the relationship between liquidity and volatility; the higher the liquidity around the best quotes, the lower the volatility.

Second, we see that the estimated coefficients of the sell side distribution are smaller in absolute terms compared to the buy side, indicating that the buy side is more informative of volatility compared to the sell side of the market, in line with the results presented in Section 5.

5.6.2. Alternative volatility measures

In this section, we investigate whether the predictive power of relative liquidity is driven by the volatility estimator employed. To this end, instead of calculating volatility via TSRV, we start by employing the realized volatility (RV), calculated as the square root of the sum of the squared mid-quote returns of the value weighted index.

Moreover, for such high frequencies, the return variation caused due to discontinuities or jumps can dominate the continuous part of volatility. Hence, we further examine whether the effect of relative liquidity is on jumps or on the continuous part of volatility. To this end, we consider four jump-robust volatility estimators: the minimum RV (minRV) and the median RV (medRV) of Andersen, Dobrev, and Schaumburg (2012), the realized bipower variation (BPV) and tripower variation (TPV) of Barndorff-Nielsen and Shephard (2004). The first two estimators use the squared returns of minimum and median of the two or three consecutive returns, which are obtained by using nearest neighbor truncation. In BPV and TPV, volatility is estimated through the cumulative sum of products of adjacent absolute returns.

However, as market microstructure noise is still a concern, we employ the “pre-averaging” approach introduced by Podolskij and Vetter (2009) prior to the application of both RV and the jump-robust volatility estimators. Pre-averaging enables us to produce a set of non-overlapping noise-

reduced returns obtained via a kernel function. Table 8 shows that the predictive power of *RLIQ* is robust to alternative volatility measures and jumps do not significantly affect the liquidity–volatility relationship we are investigating.

5.6.3. Further robustness tests

We perform four sets of additional robustness tests. The results presented so far consider the orders to be traded up to the 10th best quotes by setting $\Delta_c = 10$ in (1). First, *RLIQ* is re-calculated when the information up to the 20th and 30th best quotes (considering the whole book) is used. Second, instead of sampling the trading day using the 15-minute intervals, we test the predictive power of the limit order book distribution over volatility using 30-minute intervals.

In the analysis, we calculate *RLIQ* as the first principal component of the aggregate limit order book distribution function (*aggPDF*), which is calculated as the equally-weighted cross-sectional average of individual stock limit order book distribution functions (*indPDFs*). We calculate all of the control variables for each stock and get the cross-sectional averages. The next robustness check includes the re-calculation of the aggregate measures by using value-weighted and number-of-trades-weighted cross-sectional averages. The former weights are calculated by using the market capitalization values of the individual stocks at the end of the sample period, whereas we calculate the latter weights using the daily average number of trades.

Finally, we perform a robustness test on the specification of the regression model. We re-estimate the benchmark specification in (2) with the log-transformed variables to allow the left-hand side of the equation to include potentially both positive and negative numbers.

Table 9 confirms the robust relationship between relative liquidity and future volatility. We observe that by considering the whole book instead of the first 10 best quotes, the sampling frequency, and the regression specification do not change the results. Interestingly, we find that the sell side of the market turns to be significant when the aggregate sell side *RLIQ* is approximated as the value-weighted or trade-weighted average of the individual stocks. This indicates that bigger and more actively traded stocks are the ones that are informative of future volatility.

6. Conclusion

Most of the equity and derivatives exchanges around the world are either pure order-driven or at least allow limit orders in addition to the on-floor market making. The role of limit orders in trading processes expanded progressively over the last decade. This paper contributes to the literature on the informativeness of a limit order book on future volatility. However, we are the first to examine the predictive power of aggregate liquidity for intraday market volatility. We identify a strong and robust link between volatility and the relative depth provision, rather than the absolute volume of orders submitted.

To measure the relative depth provision, we propose a new way of summarizing the distribution of liquidity in a limit order book, while taking into account the relative magnitude and the location of quoted depth. Our summary measure, relative liquidity (*RLIQ*), accounts for how liquidity is distributed in the whole book and assigns weights to the information provided at different quotes. By using high-frequency data from the Istanbul Stock Exchange, we show that *RLIQ* has a strong in-sample and out-of-sample predictive power with respect to market volatility, where the relationship is significant for up to 75 minutes ahead.

Our findings show that the state of a limit order book contains non-negligible information about short-term volatility. In a market microstructure context, information on future volatility is important because the execution probability of a limit order increases with volatility. Put differently, the probability that the current price hits the pre-determined limit price increases when volatility is higher. Hence, the presented relationship can be used to design trading strategies that may allow market participants to submit less aggressive orders and reduce execution costs.

Appendix A. Calculation of relative liquidity

Suppose that the limit order book for stock X at 11:00am is given in Table A1.

The first step in the calculation of relative liquidity ($RLIQ$) involves the calculation of the tick-adjusted price distance Δ of each limit order in the given book relative to the best limit price:

$$\Delta_{i,\tau}^{\text{buy}} = (p_\tau^B - p_i^{\text{buy}})/\text{tick},$$

$$\Delta_{i,\tau}^{\text{sell}} = (p_i^{\text{sell}} - p_\tau^A)/\text{tick},$$

where p_τ^B (p_τ^A) is the best bid (ask) price at the end of interval τ. In this example $p_\tau^B = 8$ and $p_\tau^A = 8.05$. On the other hand, p_i^{buy} (p_i^{sell}) is the limit price of the ith order. Say the tick size is 0.05. Then for a given order, price distances are calculated in Table A2.

Next, we obtain of the percentage of total volume supplied/demanded at a given Δ for $\Delta = 0, 1, 2, \ldots, 30$. This way, we reach the limit order book probability density function for stock X and time interval τ (*indPDF*), presented in Table A3.

By repeating the procedure for each time interval τ, we end up a time series of frequencies for each price distance Δ. Hence, we obtain the *indPDF* of stock X for a given day, for 21 trading intervals, and for both sides of the market. Table A4 presents the *indPDF* for the sell side of the market.

The *aggPDF* is obtained as the equally-weighted cross-sectional average of the *indPDF*s. In other words, we repeat the steps to obtain the *indPDF*s (Table A4) for all of the stocks in our sample and

Table A1
Limit order book for stock X at 11:00am.

Order type	Volume	Limit price	Time	Best bid	Best ask
Sell	50,000	8.4	09:30:00	–	8.2
Buy	10,000	7.6	09:30:01	7.9	8.2
Sell	1,800	8.3	09:30:02	7.9	8.2
.					
.					
.					
Sell	3,334	8.05	10:58:17	8	8.05
Buy	25,000	8	10:58:20	8	8.05
Buy	50,000	7.9	10:58:38	8	8.05
Sell	1	8.1	10:58:50	8	8.05

Table A2
Price distances Δ of the orders waiting at the limit order book for stock X at 11:00am.

Order type	Volume	Limit price	Time	Best bid	Best ask	Δ
Sell	50,000	8.4	09:30:00	–	8.2	7
Buy	10,000	7.6	09:30:01	7.9	8.2	8
Sell	1,800	8.3	09:30:02	7.9	8.2	5
.					.	.
.					.	.
.					.	.
Sell	3,334	8.05	10:58:17	8	8.05	0
Buy	25,000	8	10:58:20	8	8.05	0
Buy	50,000	7.9	10:58:38	8	8.05	2
Sell	1	8.1	10:58:50	8	8.05	1

Table A3
The frequency of orders waiting at each price distance for stock *X* at 11:00am.

	Buy side		**Sell side**	
Δ	Total volume	Frequency	Total volume	Frequency
0	78,500	0.270	68,400	0.186
1	52,575	0.181	71,602	0.195
2	58,440	0.201	54,588	0.148
3	45,579	0.157	62,068	0.169
.		.		
.		.		
.		.		
29	0	0.000	0	0.000
30	0	0.000	0	0.000
Total	290,740	1	367,742	1

Table A4
Sell side limit order book probability density function for stock *X* for a given day.

	Frequencies								
Time/Δ	**0**	**1**	**2**	**3**	.	.	.	**29**	**30**
10:00	0.212	0.259	0.182	0.133	.	.	.	0.000	0.000
10:15	0.214	0.249	0.183	0.120	.	.	.	0.000	0.000
10:30	0.180	0.243	0.184	0.122	.	.	.	0.000	0.000
10:45	0.194	0.230	0.160	0.124	.	.	.	0.000	0.000
11:00	**0.186**	**0.195**	**0.148**	**0.169**	.	.	.	**0.000**	**0.000**
.					.	.	.		
.					.	.	.		
.					.	.	.		
16:30	0.213	0.223	0.146	0.112	.	.	.	0.000	0.000
16:45	0.213	0.224	0.156	0.122	.	.	.	0.000	0.000
17:00	0.188	0.240	0.171	0.118	.	.	.	0.000	0.000

calculate the cross-sectional averages of frequencies to have a distribution of the market for a given time interval τ. Finally, *RLIQ* is the summary measure of this aggregate limit order book distribution calculated as the first principal component of the *aggPDF*.

References

Ahn, H., Bae, K., Chan, K., 2001. Limit orders, depth, and volatility: evidence from the stock exchange of Hong Kong. J. Finance 56, 767–788.

Ait-Sahalia, Y., Mykland, P., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Rev. Financ. Stud. 18, 351–416.

Ait-Sahalia, Y., Mykland, P., Zhang, L., 2011. Ultra high frequency volatility estimation with dependent microstructure noise. J. Econom. 160, 160–175.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. J. Financ. Markets 5, 31–56.

Andersen, T.G., Dobrev, D., Schaumburg, E., 2012. Power and bipower variation with stochastic volatility and jumps. J. Econom. 169, 75–93.

Bandi, F.M., Russell, J.R., 2009. Microstructure noise, realized variance, and optimal sampling. Rev. Econ. Stud. 75, 339–369.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. J. Financ. Econom. 2, 1–37.

Bollerslev, T., Domowitz, I., 1993. Trading patterns and prices in the interbank foreign exchange market. J. Finance 48, 1421–1443.

Burdett, K., O'Hara, M., 1987. Building blocks: an introduction to block trading. J. Bank. Finance 11, 193–212.

Coppejans, M., Domowitz, I., Madhavan, A., 2001. Liquidity in an automated auction. Working paper. BlackRock; Barclays Global Investors.

Delattre, S., Robert, C.Y., Rosenbaum, M., 2013. Estimating the efficient price from the order flow: a Brownian Cox process approach. Stochastic Processes Appl. 123, 2603–2619.
Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econ. Stat. 13, 253–263.
Domowitz, I., Hansch, O., Wang, X., 2005. Liquidity commonality and return co-movement. J. Financ. Markets 8, 351–376.
Duong, H., Kalev, P., 2008. Order book slope and price volatility. Working paper. Monash University.
Ekinci, C., 2008. Intraday liquidity in the Istanbul stock exchange. In: Stock Market Liquidity: Implications for Market Microstructure and Asset Pricing. John Wiley & Sons, Hoboken, New Jersey, pp. 77–94.
Foucault, T., Moinas, S., Theissen, E., 2007. Does anonymity matter in electronic limit order markets? Rev. Financ. Stud. 20 1707–1747.
Ghysels, E., Sinko, A., 2011. Volatility forecasting and microstructure noise. J. Econom. 160, 257–271.
Goettler, R., Parlour, C., Rajan, U., 2005. Equilibrium in a dynamic limit order market. J. Finance 60, 2149–2192.
Goettler, R., Parlour, C., Rajan, U., 2009. Informed traders and limit order markets. J. Financ. Econ. 93, 67–87.
Griffiths, M., Smith, B., Turnbull, A., White, R., 2000. The costs and determinants of order aggressiveness. J. Financ. Econ. 56, 65–88.
Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. J. Bus. Econ. Stat. 24, 127–161.
Hasbrouck, J., Seppi, D.J., 2001. Common factors in prices, order flows, and liquidity. J. Financ. Econ. 59, 383–411.
Jones, C., Kaul, G., Lipson, M., 1994. Transactions, volume, and volatility. Rev. Financ. Stud. 7, 631–651.
Koksal, B., 2012. An analysis of intraday patterns and liquidity on the Istanbul stock exchange. Working Paper. No. 12/26. Central Bank of Turkey.
Marshall, B.R., 2006. Liquidity and stock returns: evidence from a pure order-driven market using a new liquidity proxy. Int. Rev. Financ. Anal. 15, 21–38.
Naes, R., Skjeltorp, J., 2006. Order book characteristics and the volume-volatility relation: empirical evidence from a limit order market. J. Financ. Markets 9, 408–432.
Pascual, R., Veredas, D., 2010. Does the open limit order book matter in explaining informational volatility? J. Financ. Econom. 8 57–87.
Podolskij, M., Vetter, M., 2009. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. Bernoulli 15, 634–658.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. 58, 267–288.
Valenzuela, M., Zer, I., 2013. Competition, signaling and non-walking through the book: effects on order choice. J. Bank. Finance 37, 5421–5435.
Zhang, L., Mykland, P., Ait-Sahalia, Y., 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. Am. Stat. Assoc. 100, 1394–1411.
Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. J. Bus. Econ. Stat. 14, 45–52.