Title: The relationship between volume imbalance and spread

Author: Minh Thi Hong Dinh

PII: S0275-5319(16)30166-0
DOI: http://dx.doi.org/doi:10.1016/j.ribaf.2017.03.003
Reference: RIBAF 615

To appear in: *Research in International Business and Finance*

Received date: 12-7-2016
Revised date: 6-12-2016
Accepted date: 11-3-2017

Please cite this article as: Minh Thi Hong Dinh, The relationship between volume imbalance and spread, *<![CDATA[Research in International Business and Finance]]>* (2017), http://dx.doi.org/10.1016/j.ribaf.2017.03.003

The relationship between volume imbalance and spread

Minh Thi Hong Dinh[1]
The School of Business and Economics, the Arctic University of Norway[2]

Abstract

This research examines the relationship between the ask-bid spread and the quoted volume imbalance. This relationship can be determined by the level of quoted volume imbalance and the level of stock liquidity. There are two types of samples using in this research: liquidity and low liquidity. For liquidity samples, the empirical results suggest that the relationship is negative if the imbalance is higher than 1. The relationship is positive if the imbalance is lower than -1. When the imbalance lies between -1 and 1, there is no obvious relationship between them. For low liquidity samples, the imbalance does not have a relationship with the spread if the imbalance is lower than -1 and the relationship between them varies from stock to stock if the imbalance is higher than 1.

Keywords: Imbalance, quoted volume, spread, intraday sample.
JEL classification: G120

[1]Email: minh.t.dinh@uit.no

[2]I would like to thank conference participants at the Doctoral Colloquium in Business Administration FIBE 2016 in Bergen, especially Professor Svein-Arne Persson, for helpful discussions and comments. Many thanks too to Associate Professor Espen Sirnes, Professor Oystein Myrland and Professor Derek John Clark at the School of Business and Economics at UIT-the Arctic University of Norway for providing me with the dataset, and for very helpful discussions and suggestions. I am also very grateful to Professor Andrew Mullineux of the Department of Finance at Birmingham Business School, Professor Hun-Tong Tan at NanYoung Technological University, and anonymous referees for helpful suggestions. All omissions and errors are the responsibility of the author.

1. Introduction

In a normal market, the imbalance between supply and demand determines the price of a commodity. Can this apply in the financial market? In Chordia et al. (2002, page 112) "prices and liquidity should be more strongly affected by more extreme order imbalances " meaning that order imbalances should be an important factor for price movements and liquidity. Given the importance of the imbalance between supply and demand, this research examines its relationship with the ask-bid spread, that could shed light on the movements of stock returns.

The supply and demand imbalance is differently defined on previous studies, but they are closely related (e.g., see Chordia et al., 2002; Hopman, 2007; Cartea et al., 2015; Louhichi, 2012). For example, the order imbalance in Chordia et al. (2002) is based on the differences between daily buyer-initiated and seller-initiated trading volumes in shares, trading volumes in dollars, and the number of trades on a quote driven market. The daily buyer-initiated and seller-initiated values are calculated according to the Lee and Ready (1991) algorithm. In Louhichi (2012), assigned volumes, which can be considered as order imbalance, are also based on initiated trading volumes for each five-minute interval on an order driven market. The buyer (seller) initiated trades are defined by comparing the time orders is submitted. The order imbalance in this research is defined as the quoted volume imbalance. It is simply calculated by taking log differences between ask and bid volumes at the best five levels of ask-bid prices in the order book. The main interest of the quoted volume imbalance in this research is the ask-bid volume imbalance at the first level because it can be considered as the potential supply and demand for stocks in the Norwegian financial market regarded as an order-driven market in which main liquidity providers are traders, not market makers.

This research is important and significantly contributes to the literature because of the following reasons.

First, the existing literature i) uses trading volume and ii) examines the relationship between returns and order imbalance, which depends on buyer (seller)

initiated trades. Therefore, initiated trades have to be decided before constructing imbalance variables. The dataset of buyer (seller) initiated trades is not easy to obtain although the formulae to generate order imbalance variables seems to be simple (e.g., see Chordia et al., 2002; Louhichi, 2012). However, the methods to decide initiated trades do not classify all transactions, as Louhichi (2012, page 628) states that "trades were left unclassified on cases where the two orders were submitted at the same time" in his method and the algorithm used by Chordia et al. (2002) " misclassifies 15% of the transactions". It is thus of great interest to investigate the relationship between the spread and the quoted volume imbalance because the first five levels of quoted values including bid-ask volumes and bid-ask prices are available in the order book, especially they are published in the "order depth" on the Norwegian market during the trading day. In addition, the method to calculate quoted volume imbalances is obviously simple, initiated trades do not need to be considered. Moreover, research in this relationship shows that the imbalance between the potential supply and demand in the financial market is informative as it is in a normal market.

Second, this research investigates the existence of a level of the quoted volume imbalance, at which there is a significant relationship between the spread and the quoted volume imbalance, helping investors in predicting the movements of spread and return.

Finally, this research relates to efficient trading and is thus of relevance to market participants such as traders, speculators, hedgers and arbitrageurs, as well as policy makers and regulators involved in the design of efficient trading markets.

The results of this research can be summarised as follows. First, for liquidity samples, there is a significant and non-linear relationship between the ask-bid spread and the quoted volume imbalance at the first level. If the quoted volume imbalance is higher than 1, the relationship is negative, and if the quoted volume imbalance is lower than -1, the relationship is positive. If the imbalance lies between -1 and 1, there is no obvious significant relationship. Second, for low liquidity samples, there is no significant relationship between the spread and the quoted volume imbalance when the imbalance is lower than -1. However, if the

imbalance is higher than 1, there can be a significant non-linear relationship for some low liquidity stocks. Finally, validity checks on the empirical findings for both liquidity and low liquidity samples suggest that these findings are robust.

The structure of the paper is as follows. In the following section, the literature is briefly presented. The third section briefly describes the Oslo Stock Market microstructure, and then variables, model, dataset and methodology are in section four. Section five presents the empirical results. In section six, the validity checks on the empirical findings are analysed, and then some discussions on the findings are presented in section seven. Section eight concludes.

2. Review of previous empirical work

There is a vast body of literature that is relevant to the topic of volume imbalance and returns, as well as the topics of trading volume, volatility and spread.

Chordia et al. (2002), Chordia and Subrahmanyam (2004), Louhichi (2012), Hopman (2007) and Cartea et al. (2015) research on the order imbalance and returns relationship which is the main motivation for this research. Chordia et al. (2002) and Chordia and Subrahmanyam (2004) define order imbalance variables as the difference between the number of buyer- and seller-initiated trades; the difference in the buyer-initiated shares purchased and the seller-initiated shares sold; and the difference between the buyer-initiated dollars paid and the seller-initiated dollars received. Chordia et al. (2002), working at market level, suggest that order imbalance variables strongly affect market returns. Chordia and Subrahmanyam (2004), working with individual stock returns, provide an intertemporal model in order to examine the relationship between the order imbalances and returns. Their empirical results show that imbalance-based trading strategies generate statistically significant returns, and this helps to explain the movements of daily individual stock prices in relation to the role of inventory effects. Louhichi (2012) using high frequency data (five-minute intervals) investigates the causal and dynamic relationship between stock returns and trading activities including raw volume metric (non-directional volume) and directional

volume. He highlights a strong relationship between returns and the directional one. The directional volume is calculated from buyer (seller) initiated trades determined by comparing the time submission of each order. A trade is seller (buyer) initiated one if the sell (buy) order was submitted after the buy (sell) order. The trade sign in Louhichi (2012) is differently decided from Chordia et al. (2002) using the Lee and Ready (1991) algorithm, because Louhichi (2012) recognizes that "this algorithm misclassifies 15% of the transactions" (page 628). His research indicates that there is a highly bidirectional relationship between returns and directional volume. Hopman (2007) finds that the order flow which is considered as the imbalance between buy and sell orders explains most of the changes in stock prices. He also indicates that this relationship is mainly driven by uninformed price pressure, not by private information. Cartea et al. (2015) use high frequency data to generate volume imbalances which are defined as the proportion of the differences between the best bid and ask volume to aggregate of them at a given time. They provide information on the predictive power of volume imbalance on the mid-price changes and suggest that their measure is a good predictor for price changes immediately after the arrival of a market order.

Early empirical research on the relationship between trading volume, volatility and spread can be grouped into three categories: first, the relationship between trading volume, spread, and price volatility; second, the changes in tick size or information impacting on spread and volume; and finally, the relationship between stock returns, stock prices, volume and volatility.

Empirical studies on the relationship between trading volume, spread and price volatility suggest that they are jointly determined. Wang and Yau (2000) employed the Generalized Method of Moments (GMM) procedure to estimate parameters and suggest that the relationship between trading volume and price volatility is positive, that between trading volume and spread is negative, and between price volatility and spread it is positive. Blume et al. (1994) examine the informational role of volume and develop an alternative equilibrium model to examine the behaviour of stock markets. They indicate that volume provides information and that sequences of volume can be informative. Findings by Fos-

ter (1995) seem to support Blume et al. (1994), showing that price volatility and volume are simultaneously determined. However, he suggests that volume by itself is not informative. The negative relation between spread and trading volume, and the positive one between spread and price volatility are also illustrated by Benston and Hagerman (1974), Berkman (1992), Roll (1984), French and Roll (1986), Glosten (1987), and Wang et al. (1994). The price volatility, trading volume and market depth relationship is also found in studies by Bessembinder and Seguin (1993) and Watanabe (2001). Bessembinder and Seguin (1993) investigate this relationship based on data from Futures Markets, and Watanabe (2001) on the Japanese stock index futures market. The findings by Watanabe (2001) support Bessembinder and Seguin (1993), showing that the relationship between volatility and volume is positive and negative between price volatility and expected open interest. However, Watanabe (2001) finds a flat relation between volatility, volume and open interest when the regulation changes gradually. Therefore, he suggests that this relationship may vary with regulation. However, the finding that there is no contemporaneous relationship between volume and volatility by Darrat et al. (2003) is not consistent with prior hypotheses.

The changes in tick size and information that impacts on spread and volume are studied by Ahn et al. (1996) and Lee et al. (1993). Ahn et al. (1996) investigate the relationship between tick size, spread and volume, and they find that the changes in tick size reduce both quoted and effective spread. However, they cannot find any evidence of a significant increase in volume. Spread, depth and the impact of earnings information are studied by Lee et al. (1993), who first show that wide spread is followed by low depth; and second, that when earnings announcements are released, spread grows dramatically within half an hour and becomes wider during announcement time. However, the quoted depth returns to normal level three hours after the announcement.

The relationship between stock returns, stock prices, volume and volatility is studied by Chen et al. (2001), Tay and Ting (2006), Ying (1966), and Gallant et al. (1992). Their studies indicate that stock prices and volume are

correlated. A positive correlation between price volatility and volume, and the fact that high volume follows large price movements, are found by Gallant et al. (1992). The fact that trading volume is informative to the returns process was discovered by Chen et al. (2001) and Tay and Ting (2006), who show that changes in prices increase with volume, but only for long duration trades. Ying (1966) confirms that if prices separate from volumes and vice versa, any model of the stock market will inevitably be incomplete if there are no errors in the results.

This brief summary of the literature indicates that all the studies mentioned above emphasize on a variety of topics relative to spread, volume, stock prices, order imbalance, returns and volatility, but all use trading volume to examine the relationships. On the Oslo stock market - an order driven market, the information of the five best levels of orders is published in the "order depth" through the trading day. This available information provides investors an opportunity to predict stock prices based on the quoted values. Therefore, this research investigate the relationship between the spread and the quoted volume imbalance, that can shed further light on the movement of returns.

3. Oslo Stock Market

3.1. Market microstructure

The Oslo Stock Exchange is known as an order-driven market. On this type of market, investors themselves provide liquidity and set price in a limit order or market order form. This market does not depend on market makers.

The Norwegian stock market allows two types of orders: limit orders and market orders. The most common one is the limit order. A limit order to buy, for example, is an order to buy stocks with specified volume for a specified maximum price (or "limit price"), while a market order to buy is an order to buy stocks with a specified volume without a specified maximum price meaning that the market order is at the current price in the limit order book. Hence, the market order is immediately executed at the best available prices and then

"walk to the book"'[1] until fully filled.

Opening time at the Oslo stock market is divided into four parts: pre-trading from 08:15 to 09.00, regular trading from 09:00 to 16:20, closing auction call from 16:20 to 16:25, and post close from 16:25 to 17:30. At the pre-trading time, brokers will register all trades happened after the close time on previous day. At the regular trading time, all orders registered in the order book are electronically matched to generate transactions. There can be more than one order at each level in the order book, in which there are five main levels. Priority for orders at each level is first by price and then by the time they were submitted to the market.

3.2. Information on spread and return on Norwegian data

There is an information of the relationship between return and spread applied to Norwegian data. This information shows at Table 1. In this table, five portfolios was sorted by spread, from the lowest to highest, during the period of 1980-2002. This information indicates that the portfolios with lowest spread have lowest average returns, while the portfolios with highest spread have highest average returns.

Table 1: The relationship between return and spread on the Oslo Stock Exchange (1980-2002)

Portfolio	*Return(%)*				
	Average	SD	Minimum	Media	Maximum
1(lowest spread)	1.67	6.8	-27.3	19.93	19.9
2	2.43	7.6	-26.7	2.72	30.7
3	2.45	6.9	-18.7	2.38	25.2
4	3.07	7.8	-17.7	2.13	41.3
5(highest spread)	3.55	7.7	-22.0	2.73	36.8

Source: Nas and Skjeltorp (2006)

This suggestion encourages to find factors might drive spread. The best five levels of bid-ask orders including volumes and prices are available in the "order depth" on the Norwegian market. In addition, liquidity is provided by

[1]It means that the market order will be executed at increasingly worse prices

investors themselves, not depended on market makers. This implies that the actual demand and supply stocks are shown at the first best orders. The quoted volume imbalance could thus be one of factors that we can explore.

4. Variables, model, dataset and methodology

4.1. Variables

The ask-bid spread is defined as the difference between the best ask-bid prices in the order book. It is calculated by formula (1) below:

$$ABSp_{i,j,t,1} = \frac{AP_{i,j,t,1} - BP_{i,j,t,1}}{(AP_{i,j,t,1} + BP_{i,j,t,1})/2} \quad (1)$$

where the subscripts of i, j, and t represent stock sample which is liquidity samples or low liquidity samples, minute and day. $ABSp_{i,j,t,1}$ denotes the ask-bid spread of stock sample i, at minute j and day t at the first level in the order book, $AP_{i,j,t,1}$ is the best ask price, and $BP_{i,j,t,1}$ is the best bid price in the order book. $(AP_{i,j,t,1} + BP_{i,j,t,1})/2$ is the mid-price at the best ask-bid prices.

The quoted volume imbalances at the best five levels of the ask-bid prices are denoted as DVolume1, DVolume2, DVolume3, DVolume4 and DVolume5. They are computed as in formula (2)

$$DVolume_{i,j,t,m} = log(AVolume_{i,j,t,m}) - log(BVolume_{i,j,t,m}) \quad (2)$$

where the subscript of $m = \{1, 2, 3, 4, 5\}$ denotes the best five levels of ask-bid prices. $AVolume_{i,j,t,m}$ represents the ask volumes at level m; and $BVolume_{i,j,t,m}$ is the bid volume at level m.

4.2. Model

The statistical test of the relationship between the ask-bid spread and the quoted volume imbalance is explained in the following model:

$$ABSp_{i,j,t,1} = \alpha_i + \sum_{m=1}^{5} \beta_{im} DVolume_{i,j,t,m} + \epsilon_{i,j,t,1} \quad (3)$$

Model (3) is a linear-log model showing the non-linear relationship between the spread and the quoted volume imbalance.

4.3. Dataset

This research uses intraday data with one minute intervals in 2006 from the Oslo Stock Exchange. The intraday data in 2006 is randomly chosen.

There are three stickers in this research: BIRD, STATOIL, and REC. The dataset of the stickers are in one minute intervals with different time horizons: one month and one year. Two different months are selected for sticker BIRD and sticker STATIOL: January and February. The reason for this choice is that in these months these stickers have a considerable difference in the number of observations which are considered as stock liquidity. Sticker BIRD has 8,743 observations in January, but only 5,694 in February. Sticker STATOIL has 9,261 observations in January, but only 7,938 in February.

There are two types of samples: liquidity and low liquidity. Samples are regarded as low liquidity because the number of observations is considerably lower than those of the liquidity samples. The intraday liquidity samples by one month are taken from about 8,700 to 9,300 observations, while the number of observations of the intraday low liquidity samples varies from 6,500 to 7,900. The intraday liquidity samples by one year have from 98,000 to 105,000 observations, while that of the intraday low liquidity samples is around 64,000.

The liquidity samples include sticker BIRD in January, sticker STATOIL in January, sticker BIRD in the one year horizon and sticker STATOIL in the one year horizon. Stickers: BIRD and STATOIL in the one year horizon have extremely large in size. BIRD in the one year horizon has 98,056 observations, and STAOIL in the one year horizon has 104,735. There is a problem with running a regression with an extremely large sample. Therefore, these samples are divided into quartiles before running the regression (3). The low liquidity samples include sticker REC in the one month horizon: May, sticker BIRD in one month horizon: February, sticker STATOIL in the one month horizon: February, and REC in the one year horizon. The liquidity samples and the low liquidity samples are summarised in Tables 2 and 3.

Table 2: LIQUIDITY DATA BY ONE MONTH AND ONE YEAR

Statistics	N	Mean	St. Dev.	Min	Max
BIRD DATA IN JAN					
Spread	8,743	0.009	0.004	0.007	0.052
DVolume1	8,743	−0.137	1.453	−5.159	3.970
DVolume2	8,743	−0.071	1.098	−4.043	4.234
DVolume3	8,743	0.071	1.020	−3.912	3.951
DVolume4	8,743	0.328	1.095	−3.761	10.240
DVolume5	8,743	0.643	1.328	−9.210	10.736
STATOIL DATA IN JAN					
Spread	9,261	0.003	0.0004	0.003	0.006
DVolume1	9,261	0.102	1.483	−8.149	8.632
DVolume2	9,261	0.040	0.485	−1.710	2.130
DVolume3	9,261	0.015	0.638	−3.815	3.164
DVolume4	9,261	−0.140	1.044	−6.197	5.915
DVolume5	9,261	−0.302	1.517	−7.960	5.821
BIRD DATA BY ONE YEAR					
Spread	98,056	0.011	0.006	0.001	0.097
DVolume1	98,056	−0.089	1.380	−5.273	5.050
DVolume2	98,056	−0.120	1.266	−4.997	4.700
DVolume3	98,056	−0.031	1.208	−9.903	9.210
DVolume4	98,056	0.075	1.290	−10.309	10.240
DVolume5	98,056	0.160	1.404	−10.840	10.897
STATOIL DATA BY ONE YEAR					
Spread	104,735	0.003	0.001	0.001	0.014
DVolume1	104,735	0.037	1.365	−9.427	9.376
DVolume2	104,735	0.111	0.642	−7.088	8.010
DVolume3	104,735	0.160	0.857	−6.894	7.671
DVolume4	104,735	0.198	1.236	−7.539	7.979
DVolume5	104,735	0.233	1.740	−12.891	8.468

Note: All data in 2006

Table 3: LOW LIQUIDITY DATA BY ONE MONTH AND ONE YEAR

Statistic	N	Mean	St. Dev.	Min	Max
REC DATA IN MAY					
Spread	6,579	0.004	0.002	0.002	0.048
DVolume1	6,579	−0.011	1.845	−6.619	6.583
DVolume2	6,579	0.336	1.886	−6.660	6.537
DVolume3	6,579	0.608	2.062	−7.313	7.438
DVolume4	6,579	0.515	2.190	−10.820	10.699
DVolume5	6,579	0.472	2.330	−10.820	6.924
BIRD DATA IN FEB					
Spread	5,694	0.012	0.006	0.008	0.051
DVolume1	5,694	−0.191	1.355	−3.784	4.331
DVolume2	5,694	−0.171	1.180	−3.497	3.490
DVolume3	5,694	0.041	1.044	−3.497	4.127
DVolume4	5,694	0.321	1.255	−9.473	4.107
DVolume5	5,694	0.433	1.570	−9.903	9.798
STAIOL DATA IN FEB					
Spread	7,938	0.003	0.0002	0.003	0.006
DVolume1	7,938	−0.015	1.297	−7.890	7.710
DVolume2	7,938	0.044	0.488	−1.814	2.690
DVolume3	7,938	0.019	0.638	−6.722	2.641
DVolume4	7,938	−0.044	0.897	−6.689	7.979
DVolume5	7,938	−0.135	1.153	−12.891	8.468
REC DATA BY ONE YEAR					
Spread	64,545	0.004	0.002	0.002	0.048
DVolume1	64,545	0.118	1.629	−6.619	6.852
DVolume2	64,545	0.114	1.714	−6.660	7.046
DVolume3	64,545	0.316	1.875	−7.313	7.438

Note: All data in 2006

4.4. Methodology

In order to estimate the regression (3), the Generalized Method of Moments (GMM) is applied. This method is chosen because it allows $\epsilon_{i,j,t,1}$ to be heteroskedastic and autocorrelated which is expected to exist in intraday data (Chausse, 2015; Bjornland and Thorsrud, 2015; Abhyankar et al., 1997; Wang and Yau, 2000; Jagannathan et al., 2002). Abhyankar et al. (1997) and Wang and Yau (2000) applied GMM when they regress intraday variables: the bid-ask spread, trading volume, and volatility. Jagannathan et al. (2002) state that authors who working with "market micro-structure" variables such as spread use GMM to examine hypothesis because of the presence of conditional heteroscedasticity and serial correlation in the data. Following the early empirical researchers, we employ GMM to examine the relationship between the ask-bid spread and the quoted volume imbalance [2].

The system in (3) is just identified, so the coefficients estimated by GMM are exactly the same as those from OLS, although the standard errors are higher. The intraday samples by one month and one year are constructed by first sorting the quoted volume imbalance at the first level (or DVolume1) and then dividing into quartiles[3]. All quartile samples and the intraday samples by complete one month, which are not sorted and not divided into quartiles, are used to run the regression (3). There are two intraday samples by one month and two by one year for the liquidity samples, and there are three intraday samples by one month and one intraday sample by one year for the low liquidity samples. Hence, there are five samples by one month and three samples by one year for both the liquidity and low liquidity samples. Because the samples by one month and year are divided into four quartiles, and all the samples by complete one month are also used, the total sample is 37, in which 18 samples are considered as liquidity samples and 19 samples are low liquidity samples.

[2]The paper uses the new GMM package in R introduced by Chausse (2015)
[3]The quartiles: 0%-25%, 25%-50%, 50%-75%, and 75%-100%

Table 4: The Hausman test for Endogeneity

Stickers	P values	Conclusion
BIRB in January	7.0340e-01	No correlation
REC in May	2.4342e-01	No correlation
STATOIL in January	6.9292e-02	No correlation at significant level of 5%
BIRB in February	7.7348e-01	No correlation
STATOIL in February	3.8587e-01	No correlation

H0: No correlation between explanatory variable DVolume1 and error terms
H1: Correlation between explanatory variable DVolume1 and error terms

5. Empirical results

5.1. Endogeneity test

Before running the model (3), the endogeneity problem is tested by using the Hausman test because the relationship between the ask-bid spread and the quoted volume imbalance at the first level could face this problem. The Hausman test for endogeneity examines whether the quoted volume at the first level is endogenous and it is applied for all samples in the one month horizon: BIRD in January, REC in May, STATOIL in January, BIRD in February and STATOIL in February. Basically, the test examines the correlation between the independent variable and the residuals. The Hausman test in Hill et al. (2012) infer that if the explanatory variable has a correlation with the error terms, then the explanatory variable is endogenous. The results of this test are presented in Table 4. These results indicate that there is no correlation between the quoted volume imbalance at the first level and the residuals estimated from the regression of the spread and the quoted volume imbalance at the first level. This means that there is no endogeneity problem when the model (3) is regressed.

5.2. Explaining tables

The empirical results of the relationship between the ask-bid spread and the quoted volume imbalance are presented in Table 5 for the liquidity samples, and in Table 6 for the low liquidity samples. The dependent variable is the ask-bid spread, and the independent variables are the quoted volume imbalances at

the best five levels. The quoted volume imbalance at the first level shows the imbalance between the potential supply and demand. Therefore, the estimate of the imbalance between the ask-bid volume at the first level (or DVolume1) is the main interest and is employed to analyse the relationship between the ask-bid spread and the potential supply and demand imbalance. In addition, the empirical results indicate that the significant relationship between the ask-bid spread and the quoted volume imbalances at other levels varies from stock to stock. Therefore, in order to save space, the slope of quoted volume imbalances at the first level is shown in Tables: 5 and 6, but not that of other levels of the quoted volumes. Numbers in parentheses show the standard errors; numbers in square brackets are the average values of the imbalance at the first level (or DVolume1). The asterisks (or the symbols "*") denote significance levels.

The average value of the imbalance at the first level will be written in what follows, omitting reference to at the first level. Similarly to the bid-ask volume imbalance at the first level.

5.3. Empirical results

In Tables 5 and 6, the intraday samples by one month and one year is constructed by sorting the quoted volume imbalance at the first level (or DVolume1) into quartiles. In columns: first and second quartiles, the average values of the imbalance are negative meaning that the ask volume is smaller than the bid volume. In columns: third and fourth quartiles, the average values of the imbalance take positive values meaning that the ask volume is higher than the bid volume. In these tables, it is obvious that the average values of the imbalance in the first quartile column is smaller than -1, in columns: second and third quartiles it lies between -1 and 1, and in the fourth quartile column it is higher than 1. This means that the first and fourth quartiles show a high imbalance between the ask volume and the bid volume while the second and third quartiles show a small imbalance.

For the liquidity samples, the empirical results are summarised in Table 5. Results of the samples with complete one month horizon show in columns

"Month" for both stickers: BIRD and STATOIL, and the first, second, third and fourth quartile columns contain results of quartile samples. The empirical results of the first and fourth quartiles suggest that the imbalance at the first level has a significant and non-linear relationship with the ask-bid spread. In the first quartile column, the imbalance has a positive relation with the spread. In other words, the liquidity stocks have a significantly positive relationship with the spread if the average value of the imbalance over the period is less than -1. In the fourth quartile column, the imbalance has a negative relationship with the spread. In other words, the liquidity stocks have a significantly negative relationship with the spread if the average value of the imbalance is more than 1. In the second and third quartile columns, it is not obvious whether there is a significant relationship between the spread and the imbalance because the empirical results of some samples suggest that there is a significant relationship, but others do not. In the Month columns, the average values of the quoted volume imbalance lie between -1 and 1, and the empirical results suggest that there is no significant relationship between them.

For low liquidity samples, the empirical results are presented in Table 6. The fourth quartile column suggests that there is a significant and non-linear relationship between the spread and the imbalance for most samples, except REC by one year. The other quartile columns and the Month columns indicate that there is no relationship between the spread and the imbalance.

Consequently, the non-linear relationship between the ask-bid spread and the quoted volume imbalance can be determined by the level of the imbalance and of stock liquidity.

Findings from this research imply that the ask-bid imbalance at the best ask-bid price is informative. In addition, the findings in this research are consistent with findings by Chordia et al. (2002), Chordia and Subrahmanyam (2004) and Hopman (2007) who find a significant relationship between returns and the imbalance. In the way that the ask-bid volume imbalance at the first best ask-bid price in this research can be regarded as one of the order imbalance variables in their research, and variations in the spread are associated to returns.

Table 5: LIQUIDITY DATA - REGRESSION

	Dependent variable: Spread				
Quartiles	First	Second	Third	Fourth	Month
BIRD IN JAN					
DVolume1	0.001^{***}	0.002^{**}	0.001	-0.001^{***}	0.00004
	(0.0002)	(0.001)	(0.001)	(0.0002)	(0.0001)
	[−2.0192]	[−0.5732]	[0.3735]	[1.6884]	[0.1021]
Constant	0.010^{***}	0.011^{***}	0.009^{***}	0.009^{***}	0.009^{***}
	(0.0004)	(0.001)	(0.0003)	(0.0004)	(0.0001)
Observations	2,200	2,173	2,200	2,171	8,743
STATOIL IN JAN					
DVolume1	0.00001^{**}	0.00001	−0.000004	-0.00001^{*}	−0.0000008
	(0.00001)	(0.00005)	(0.00003)	(0.000004)	(0.000002)
	[−1.7190]	[−0.2440]	[0.5227]	[1.8489]	[0.1021]
Constant	0.003^{***}	0.003^{***}	0.003^{***}	0.003^{***}	0.003^{***}
	(0.00001)	(0.00001)	(0.00002)	(0.00002)	(0.00001)
Observations	2,315	2,316	2,315	2,315	9,261
BIRD BY ONE YEAR					
DVolume1	0.001^{***}	−0.001	−0.001	-0.001^{***}	
	(0.0001)	(0.0004)	(0.001)	(0.0002)	
	[−1.8490]	[−0.5176]	[0.3378]	[1.6716]	
Constant	0.013^{***}	0.011^{***}	0.012^{***}	0.012^{***}	
	(0.0003)	(0.0002)	(0.0002)	(0.0003)	
Observations	24,514	24,514	24,515	24,513	
STATOIL BY ONE YEAR					
DVolume1	0.00003^{***}	0.0001^{***}	-0.0001^{***}	-0.00004^{***}	
	(0.000005)	(0.00002)	(0.00002)	(0.000004)	
	[−1.5857]	[−0.3190]	[0.3822]	[1.6696]	
Constant	0.003^{***}	0.003^{***}	0.003^{***}	0.003^{***}	
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	
Observations	26,184	26,184	26,184	26,183	

*Note: *p<0.1; **p<0.05; ***p<0.01*
Numbers in parentheses: std. error; numbers in square brackets: average values

Table 6: LOW LIQUIDITY DATA - REGRESSION

	Dependent variable: Spread				
Quartiles	First	Second	Third	Fourth	Month
REC IN MAY					
DVolume1	−0.0002	−0.0001	−0.0002	−0.0001***	−0.0001
	(0.0001)	(0.0002)	(0.0002)	(0.00005)	(0.0001)
	[−2.3471]	[−0.4897]	[0.5207]	[2.2730]	[−0.0109]
Constant	0.004***	0.004***	0.004***	0.004***	0.004***
	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Observations	1,645	1,645	1,645	1,645	6,579
BIRD IN FEB					
DVolume1	−0.0005	−0.0005	−0.001	−0.003***	0.00003
	(0.001)	(0.001)	(0.002)	(0.001)	(0.0003)
	[−1.8976]	[−0.6339]	[0.1938]	[1.5728]	[−0.1914]
Constant	0.010***	0.013***	0.014***	0.018***	0.013***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Observations	1,424	1,423	1,426	1,423	5,694
STATOIL IN FEB					
DVolume1	0.000002	0.00002	−0.0001	−0.00001***	0.000002
	(0.000003)	(0.00003)	(0.00003)	(0.000004)	(0.000002)
	[−1.5492]	[−0.4314]	[0.3207]	[1.6015]	[−0.0151]
Constant	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.00001)	(0.00002)	(0.00001)	(0.00001)	(0.000003)
Observations	1,985	1,986	1,985	1,983	7,938
REC BY ONE YEAR					
DVolume1	−0.00002	−0.00002	0.0004**	−0.0001	
	(0.0001)	(0.0001)	(0.0001)	(0.00005)	
	[−1.9089]	[−0.3563]	[0.5625]	[2.1764]	
Constant	0.004***	0.004***	0.004***	0.005***	
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	
Observations	16,136	16,137	16,136	16,137	

*Note: *p<0.1; **p<0.05; ***p<0.01*
Numbers in parentheses: std. error; numbers in square brackets: average values

However, this research also indicates that the ask-bid imbalance at the best ask-bid price has a strong and non-linear relationship with the ask-bid spread when the imbalance is high enough (lower than -1 and higher than 1). The relationship is positive or negative depending on the bid volume are higher or lower the ask volume. These findings significantly contribute to the literature because it shows a level of the imbalance between the ask and bid volumes, at which the relationship exists. That will help investors in predicting the movements of spread and stock prices.

6. Validity checks

In this section, we will re-check whether the empirical results for both the liquidity and low liquidity samples above are correct when the quoted volume imbalances take values: lower -1, higher 1, and between -1 and 1. The validity checks regress the ask-bid spread as a function of DVolume1 (or the quoted volume imbalance at the first level in the order book).

The results for liquidity samples can be seen in Table 7, where the dependent variable and the independent variable are the spread and DVolume1 respectively. The table has three parts: DVolume1 higher than 1, DVolume1 lower than -1, and -1 $<$ DVolume1 $<$ 1. In the first part, there are two liquidity samples: sticker BIRD in the one year horizon and sticker STATOIL in the one year horizon. These samples are filtered to collect samples in which DVolume1 is higher than 1 before running the regression. The test results indicate that DVolume1 has a significantly negative relationship with the spread for both samples. These results prove that for liquidity samples, when the imbalance is higher than 1, there is a significantly negative relationship between the spread and the imbalance. In the second part, there are two liquidity samples: sticker BIRD by one year and BIRD in January. These samples are also screened out to select DVolume1 lower than -1. The test results also suggest that DVolume1 has a significantly positive relationship with the spread for both samples. In the third part, there are two liquidity samples: sticker BIRD in January and by one year. They are screened out to select DVolume1 with values of between -1 and

1. The test results suggest that DVolume1 has no significant relationship with the spread for both samples. This illustrates that when the imbalance takes values between -1 and 1, it has no correlation with the spread.

Validity checks for low liquidity samples present in Table 8. There are two parts: the first shows DVolume1 higher than 1 and the second shows DVolume1 lower than -1. In the first part, four stickers: REC in May, STATOIL in February, BIRD in February, and REC by one year are screened out to select the values of DVolume1 higher than 1. The test results suggest that there is a significant relationship between the spread and the imbalance for two samples: REC in May and STATOIL in February at high significance level α, at 10% for STATOIL, but there is no significant relationship for the others. In the last part, two stickers: BIRD in February and REC by one year are filtered to obtain samples in which DVolume1 is lower than -1. The test results suggest that there is no relationship between the spread and the imbalance when the imbalance takes this value.

7. Discussions

The main finding for the liquidity samples can be summarized as follows. The relationship between the spread and the quoted volume imbalance is negative if the imbalance is higher than 1, and this relationship is positive if the imbalance is lower than -1. This statement is illustrated in Figure 1.

The positive side of Figure 1 shows a negative relationship between the ask-bid spread and the ask-bid imbalance, and the negative side of it shows a positive relationship between them.

On the positive side of the figure, the ask-bid volume imbalance takes positive values higher than 1, and the relationship between the ask-bid spread and the ask-bid volume imbalance is negative. The ask-bid volume imbalance takes positive values higher than 1 meaning that the ask volume at the best ask price (or potential supply stock) is higher than the bid volume at the best bid price (or potential demand stock). The relationship between the spread and the imbalance is negative at the current minute and if in the next minute the ask-bid

Table 7: VALIDITY CHECKS: LIQUIDITY DATA

	Dependent variable: Spread	
Part I	BIRD BY ONE YEAR: DVolume1 > 1	STATOIL BY ONE YEAR: DVolume1 > 1
DVolume1	-0.001^{***}	-0.00002^{***}
	(0.0002)	(0.00001)
Constant	0.012^{***}	0.002^{***}
	(0.0004)	(0.00005)
Observations	20,902	20,386
Part II	BIRD BY ONE YEAR : DVolume1 < -1	BIRD IN JAN: DVolume1 < -1
DVolume1	0.001^{***}	0.001^{***}
	(0.0002)	(0.0002)
Constant	0.013^{***}	0.010^{***}
	(0.0004)	(0.0005)
Observations	24,924	2,365
Part III	BIRD IN JAN: $-1 <$ DVolume1 < 1	BIRD BY ONE YEAR: $-1 <$ DVolume1 < 1
DVolume1	0.0002	0.0001
	(0.0003)	(0.0002)
Constant	0.010^{***}	0.012^{***}
	(0.0002)	(0.0002)
Observations	4,475	52,230
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 8: VALIDITY CHECKS: LOW LIQUIDITY DATA

	Dependent variable: Spread	
Part I	REC IN MAY: DVolume1 > 1	STATOIL IN FEB: DVolume1 > 1
DVolume1	-0.0001^{**}	-0.00001^{*}
	(0.0001)	(0.000004)
Constant	0.004^{***}	0.003^{***}
	(0.0002)	(0.00004)
Observations	1,759	1,517
	BIRD IN FEB: DVolume1 > 1	REC BY ONE YEAR: DVolume1 > 1
DVolume1	-0.001	-0.0001
	(0.001)	(0.0001)
Constant	0.012^{***}	0.005^{***}
	(0.002)	(0.0001)
Observations	1,011	17,379
Part II	BIRD IN FEB: DVolume1 < -1	REC BY ONE YEAR: DVolume1 < -1
DVolume1	0.0003	-0.000001
	(0.001)	(0.0001)
Constant	0.012^{***}	0.004^{***}
	(0.001)	(0.0002)
Observations	1,682	14,372
Note:	*p<0.1; **p<0.05; ***p<0.01	

Figure 1: Relationship between the ask-bid spread and the ask-bid volume imbalance

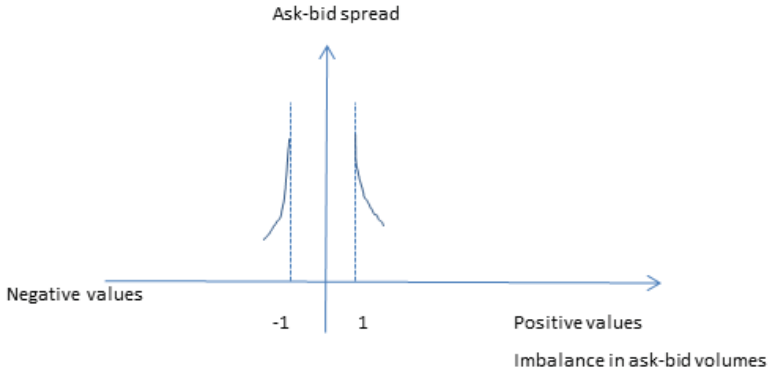
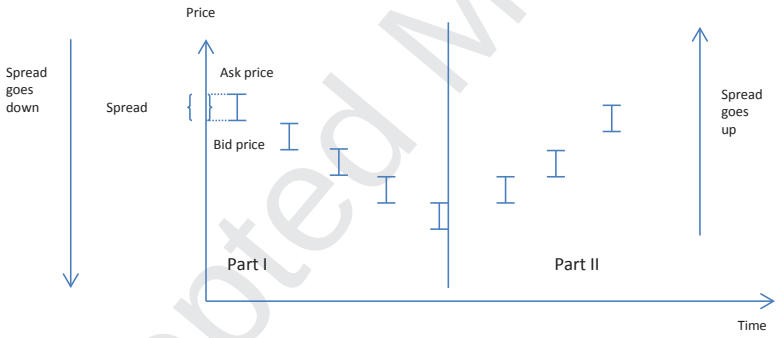


Figure 2: Movement of spread

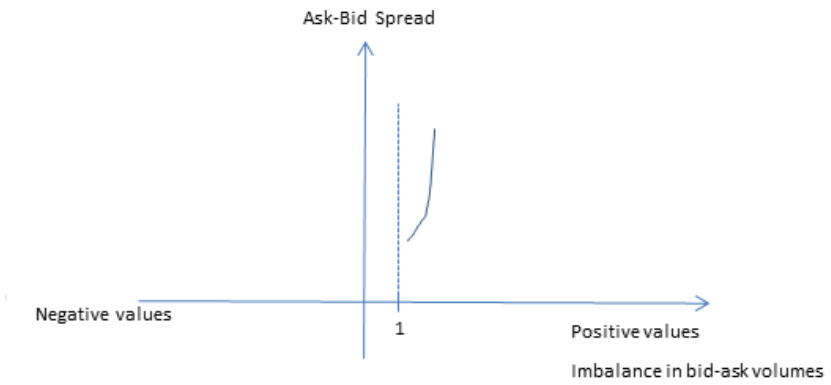


volume imbalance still increases meaning that the imbalance is extremely large, the ask-bid spread will decrease. Because the ask volume is much larger than the bid volume (or potential supply is much larger than potential demand), that causes the ask price to decrease and could reduce to the bid price; and the bid price also reduces (see part I on Figure 2). The transaction price could be at the bid price if a transaction happens during this time.

On the negative side of the figure, the ask-bid volume imbalance takes negative values lower than -1 and the relationship between the ask-bid spread and the ask-bid imbalance is positive. It could be more obvious if the ask-bid volume

imbalance is converted to the bid-ask volume imbalance. Ask-bid volume imbalance is equal to log(ask volume) minus log(bid volume), and bid-ask volume imbalance is equal to log(bid volume) minus log(ask volume). This means that the negative side of Figure 1 can be converted to the positive side of Figure 3. Figure 3 shows the positive relationship between the ask-bid spread and the bid-ask volume imbalance instead of the positive relationship between the ask-bid spread and the ask-bid volume imbalance. In the current minute, if the bid-ask volume imbalance is higher than 1, the ask-bid spread and the bid-ask volume imbalance has a positive relation. In the next minute of trading, if the bid-ask imbalance still increases, the ask-bid spread will move up, because the increase in the bid-ask imbalance means that the bid volume is much larger than the ask volume. In other words, the potential demand is much higher than the potential supply. That causes the bid price to go up and could reach to the ask price; and the ask price increases. In other words, the ask-bid spread increases (see part II on Figure 2). The transaction price could be at the ask price if a transaction happens during this time.

Figure 3: Relationship between the ask-bid spread and the bid-ask imbalance



Figures: 1 and 3 indicate that if the quoted volume imbalance in absolute value is high enough (say 1), it and the ask-bid spread have a significant relationship. In addition, the relationship between them at current time could

provide signals for movements of the spread in the next time period.

There is no discussion on the empirical results of low liquidity samples because it shows no obvious relationship between the spread and the quoted volume imbalance.

8. Conclusions

This paper first examines the relationship between the ask-bid spread and the quoted volume imbalance at the first level for both liquidity and low liquidity samples. Second, it provides the information on the level of the quoted volume imbalance at which there is a significant and non-linear relationship between the ask-bid spread and the imbalance.

Stocks with high trading are suggested that if the quoted volume imbalance at the first level in order book is high enough (say lower than -1, or higher than 1)[4], the spread and the imbalance have a significant and non-linear relationship. They have a positive relation if the imbalance < -1, and negative if the imbalance > 1. If the imbalance lies between -1 and 1, there is no obvious relationship between them, because the significant relationship can be found for some stocks, but not significant for others.

Stocks with low trading, the non-linear relationship between the spread and the imbalance is not clear. If the imbalance is higher than 1, the significant and non-linear relationship between them finds for some stocks, but not for the others. If the imbalance is lower than -1, there is no correlation between them.

These findings suggest investors, portfolio managers and hedgers that they can use quoted volume imbalance strategy for their investments. If the quoted volume imbalance is higher than 1, the spread goes down and the transaction price at the current time could be at the bid price in the previous time. If the quoted volume imbalance is lower than -1, the spread goes up and the transaction price at the current time could take the ask price in the previous time(see part 7).

[4]Lower than -1 or higher than 1 depends on bid volumes higher or lower than ask volumes

Market microstucture has long-term effects on trading, in the way that a market serves two important functions: providing liquidity for buyers and sellers, and ensuring that new information is reflected in stock prices. It thus could suggest that policy makers and regulators design an efficient platform of trading markets

It is possible to predict the next movement of the spread based on the relationship between the spread and the imbalance at the current time. Hence, future research could first focus on predicting variations in the spread, and second, extending its scope to the prediction of stock prices.

References

Abhyankar, A., Ghosh, D., Levin, E., Limmack, R., 1997. Bid-ask spreads, trading volume and volitility: intra-day evidence from the London Stock Exchange. Journal of Business Finance and Accounting 24, 343–362.

Ahn, H.J., Cao, C.Q., Choe, H., 1996. Tick Size, Spread, and Volume. Journal of Financial Intermediation 5, 2–22.

Benston, G., Hagerman, R., 1974. Determinants of bid-asked spreads in the over-the-counter market. Journal of Financial Economics 1, 353–364.

Berkman, H., 1992. The market spread, limit orders, and options. Journal of Financial Services Research 6, 399–416.

Bessembinder, H., Seguin, P.J., 1993. Price Volatility, Trading Volume, and Market Depth: Evidence from Futures Markets. The Journal of Financial and Quantitative Analysis 28, 21–39.

Bjornland, H.C., Thorsrud, L.A., 2015. Applied Time Series For Macroeconomics. Gyldendal Norsk Forlag AS.

Blume, L., Easley, D., O'Hara, M., 1994. Market Statistics and Technical Analysis: The Role of Volume. The Journal of Finance 49, 153–181.

Cartea, Á., Donnelly, R.F., Jaimungal, S., 2015. Enhancing trading strategies with order book signals. Available at SSRN 2668277 .

Chausse, P., 2015. Computing Generalized Method of Moments and Generalized Empirical Likelihood with R. Cran.r-project .

Chen, G.M., Firth, M., Rui, O.M., 2001. The Dynamic Relation Between Stock Returns, Trading Volume, and Volatility. The Financial Review 38, 153–174.

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

Chordia, T., Subrahmanyam, A., 2004. Order imbalance, and individual stock returns: Theory and evidence. Journal of Financial Economics 72, 485–518.

Darrat, A.F., Rahman, S., Zhong, M., 2003. Intraday trading volume and return volatility of the DJIA stocks: A note. Journal of Banking & Finance 27, 2035–2043.

Foster, A., 1995. Volume-volatility relationships for crude oil futures. Journal of Futures Markets 15, 929–951.

French, R.K., Roll, R., 1986. Stock return variances: The arrival of information and the reaction of traders. Journal of Financial Economics 17, 5–26.

Gallant, A.R., Rossi, P.E., Tauchen, G., 1992. Stock prices and Volume. The Review of Financial Studies 5, 199–242.

Glosten, R.L., 1987. Components of the bid-ask spread and the statistical properties of transaction prices. Journal of Finance 42, 1293–1307.

Hill, R.C., Griffths, E.W., Lim, C.G., 2012. Priciples of Econometrics. John Wiley & Sons.

Hopman, C., 2007. Do supply and demand drive stock prices? Quantitative Finance 7, 37–53.

Jagannathan, R., Skoulakis, G., Wang, Z., 2002. Generalized methods of moments: Applications in finance. Journal of Business & Economic Statistics 20, 470–481.

Lee, C., Ready, M.J., 1991. Inferring trade direction from intraday data. The Journal of Finance 46, 733–746.

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

Louhichi, W., 2012. Does trading activity contain information to predict stock returns? evidence from euronext paris. Applied Financial Economics 22, 625–632.

Nas, R., Skjeltorp, J., 2006. Is the maket microstructure of stock markets important? Economic Bulletin 77, 123–132.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. Journal of Finance 39, 1127–1139.

Tay, A.S., Ting, C., 2006. Intraday stock prices, volume, and duration: a nonparametric conditional density analysis. Empirical Economics 30, 827–842.

Wang, G.H.K., Yau, J., 2000. Trading volume, bid-ask spread, and price volatility in futures markets. Journal of Futures Markets 20, 943–970.

Wang, K.R., Michalski, J.R., Jordan, V.J., Moriarty, J.E., 1994. An intrady analysis of bid-ask spreads and price volatility in the S&P500 index futures market. Journal of Futures Market 14, 837–859.

Watanabe, T., 2001. Price volatility, trading volume, and market depth: evidence from the Japanese stock index futures market. Applied Financial Economics 11, 651–658.

Ying, C.C., 1966. Stock Market Prices and Volumes of Sales. Econometrica 34, 676–685.