Journal of Management Science and Business Intelligence, 2017, 2–2
Aug. 2017, pages 1-8
doi: 10.5281/zenodo.808452
http://www.ibii-us.org/Journals/JMSBI/
ISBN 2472-9264 (Online), 2472-9256 (Print)

CLUSTER ANALYSIS OF LIQUIDITY MEASURES IN A STOCK MARKET USING HIGH FREQUENCY DATA

Amin Salighehdar[1, *], Yang Liu[1] , Dragos Bozdog[1] and Ionut Florescu[1]

[1]Hanlon Financial System Laboratory, Financial Engineering Division, School of System and Enterprises, Stevens Institute of Technology, Hoboken, New Jersey, 07030

*Email: asalighe@stevens.edu

Received on February 28, 2017; revised on June 08, 2017; published on June 14, 2017

Abstract

Liquidity is one of the crucial factors in economy which reflects smooth operation of the markets. In a liquid market, traders are able to transact large quantities of security quickly with minimal trading cost and price impact. Many researchers have investigated the relationship between market liquidity and trading activity of a financial market. According to the existing literature, liquidity can measure different market characteristics such as trading time, tightness, depth, and resiliency. There is significant number of liquidity measures published in the literature. The main goal of this study is to use a hierarchical clustering algorithm to classify different liquidity measures. We examine the relationship between liquidity measures in order to detect commonality and idiosyncrasy among them. Then, we estimate the correlation among liquidity measures to quantify similarity between them and this quantity is used to develop a hierarchical clustering algorithm. At the end, we analyze the consistency in the structure of the clusters and we conclude that, clusters hold the same structure for almost 80% of the stocks in our sample. The data set that we are using for this study is NASDAQ High Frequency Trader (HFT) data. This data set contains trading and quoting activities of 26 HFT firms in 120 stocks on the Nasdaq exchange for various dates (in millisecond timestamp).

Keywords: Liquidity; High Frequency Trading; Correlation; Hierarchical Clustering.

1 Introduction

Liquidity is an essential feature of a financial market and it is used as an indicator for smooth operation of an economy. In a liquid market, traders are able to execute large quantities of security quickly with minimal trading cost and little price impact. A liquid market might be characterized as a continuous market that traders can buy or sell any amount of stock immediately (Black (1971), Kyle (1985)).

A significant amount of literature tries to shed light on the critical role of liquidity in finance. In the asset pricing literature, Acharya and Pedersen (2005) investigate how assets price is affected by liquidity risk. The authors show that the covariance of the security return and liquidity with the market return and market liquidity can affect a security's return.

Amihud and Mendelson (1986) investigate the relationship between stock bid and ask spread and stock return. They conclude that return increases with an increase in the spread. Amihud (2002) defines the daily ratio of absolute stock return to its dollar volume as an illiquidity measure, and it is shown that expected stock returns are an increasing function of expected illiquidity.

Chordia et al.(2001) study the relationship between liquidity and trading activity in order to analyze how liquidity and trading activity vary over different days in a week. The authors use daily data from 1988 to 1998, and they show that on Fridays we have lower liquidity and trading activity, while Tuesdays show opposite pattern. Bali et al. (2014) study how market reacts to liquidity shocks. The authors conclude that there is a positive and significant relation between liquidity shocks and future returns.

The existing literature shows that liquidity plays an essential role in the financial market worldwide. However, liquidity is not directly observable in the financial markets, and despite its importance, problems in quantifying liquidity still exist. Defining a globally accepted proxy for liquidity is an active area of research. O'Hara (2004) states that even though liquidity is a simple concept, it is hard to define it, and there exist a lot of different views of liquidity in literature. Furthermore, liquidity is a multidimensional variable that can capture different aspects of a market such as trading quantity, trading speed, trading cost and price impact (Liu (2006)). To the best of our knowledge, there are over 65 different measures of liquidity in literature developed using daily and monthly data. These measures may be classified into four categories based on the market feature they capture. The first category of the liquidity measures which includes liquidity measures such as number of trades, and turnover try to quantify *trading quantity* in the market. Amivest liquidity ratio, used by Cooper et al. (1985), Amihud illiquidity measure, Amihud (2002), Kyle lambda (λ), introduced by Kyle (1985), and Pastor reversal measure developed by Pastor and Stambaugh (2001) fall into a second category which captures *price impact*. The third category address the *trading cost* feature in the market.

For instance, difference between bid and ask price, known as spread, is one of the liquidity measures in this category. This group of liquidity measures has been studied in multiple papers (Roll (1984), Chordia et al. (2001), Acker et al. (2002), Dacorogna et al. (2001)). Finally, the last category of liquidity measures captures *trading speed*; e.g. Liu (2006) define a measure which falls into this category.

Although a lot of literature propose proxies to quantify liquidity in the market, some of the proposed measures suffer from serious shortcomings. For instance, among price impact measures, Amihud (2002) measures the lack of liquidity by dividing daily return over daily dollar volume. Although it is shown by Goyenko et al. (2009) that Amihud measure performs better than other measures, the measure cannot consider days without trading. Although there is no trading, these periods contain quotes which are important for a global liquidity measure. Amivest liquidity measure used by Cooper et al. (1985), and Amihud et al.(1997) suffer from the same issue. The quoted bid-ask spread is a noisy measure of illiquidity ((Lee (1993), Brennan and Subrahmanyam (1996)). Lee and Swaminathan (2000) show that trading volume, measured by the turnover ratio, is not a good indicator for liquidity. High trading volume does not necessarily indicate high liquidity in the financial market especially when we have significant volatility in the market, and this is proven during the *flash crash* on May 6, 2010 ((Van der Merwe (2015)).

Most of the measures existing in literature have been constructed using daily data (low-frequency). One of the reasons for using low-frequency based liquidity measures is saving in computational time compared to high-frequency based liquidity measures (Holden et al. (2014)). However, analyzing high-frequency based liquidity measures may provide more insight into time variation of liquidity. Therefore, more work is necessary to develop high frequency based liquidity measures.

Since limit orders play a crucial role in providing liquidity in the market, there are some works in the literature that propose liquidity measures based on information from a limit order book. For example, Cost of Round Trip (CRT), which is defined for any given transaction size, accumulates the status of the entire limit order book. This measure can capture the depth of a limit order book, and it is shown that CRT is correlated with other measures of liquidity such as quoted spread and effective spread (Irvine et al. (2000)). Kang and Zhang (2013) propose a liquidity measure which is capable of measuring the dispersion of a limit order book.

The main goal of this study is to classify the existing liquidity measures proposed in the literature into different groups. This classification have been done by considering the market feature that each liquidity measure captures. We are looking into the classification problem from different perspective. Correlation between individual stock liquidity and market liquidity, and co-movement between liquidity and other market factors such as return have been studied in literature (Von Wyss (2004), Sarr and Lybek (2002), Acharya and Pedersen (2005), Vu et al. (2015), and Amihud (2002)). In this study, we use the results obtained from the correlation analysis to develop a hierarchical clustering algorithm that is capable to detect a subset of liquidity measures that are similar to each other. We define the concept of the similarity of liquidity measures as a function based on Pearson correlation coefficient. Further, we analyze the consistency in the structure of the clusters. We are able to detect liquidity measures that introduce inconsistency. We call them *problematic liquidity measures*. We argue that our findings can provide more insight about the possible commonality and idiosyncratic feature between different liquidity measures. We perform all the analysis using a high frequency data set, so the first steps in our study is to investigate whether calculating the existing liquidity measures using a high frequency data is feasible.

This paper is organized as follows. In section 2, we provide a brief mathematical description for the liquidity measures used in this study. Also, we illustrate the correlation analysis and the hierarchical clustering algorithm used in this research. In section 3, we discuss the results including a discussion of the consistency in the structure of the clusters as well. We look into the consistency issue using an optimization concept. Finally, in section 4, we conclude the paper.

2 Methodology

In this section, we illustrate methodologies used in this study. First, we discuss the liquidity measures used in this study. Second, we explain the methodology used to analyze relation between different liquidity measures. Finally, we elaborate how liquidity measures can be classified into multiple groups based on the specific similarity function defined in this study.

2.1 Liquidity Measures

This section provides an overview of liquidity measures studied in literature. As we mentioned in section 1, there are more than 65 liquidity measures introduced in literature and most of them are developed using daily or monthly data. In this study, we are not proposing a new liquidity measure, but we are interested to analyze existing liquidity measures using high frequency data. We look into liquidity measures reviewed by Von Wyss (2004). Following same notation used by Von Wyss (2004), liquidity measures are divided into two groups: one-dimensional and multi-dimensional. The multi-dimensional liquidity measures are combination of multiple one-dimensional liquidity measures. In one-dimensional liquidity measures, only one market variable like price or size is used. Mathematical description of the liquidity measures used in this study is given in Table 6. We eliminate from the study all the liquidity measures that are not feasible to be calculated using our high frequency data set.

2.2 Correlation Analysis

In order to analyze the relation between different liquidity measures, we look into correlation among them. Correlation is a measure that quantifies the strength of linear relationship between two quantitative variables ((Moore and McCabe (1989)). Correlation between individual stock liquidity and market liquidity, and co-movement between liquidity and other market factors such as return have been studied in literature (Von Wyss (2004), Sarr and Lybek (2002), Acharya and Pedersen (2005), Vu et al. (2015), and Amihud (2002)). As we mentioned in the previous section, we are interested to use results from correlation analysis to classify liquidity measures into different groups. We believe that by this methodology, we can find a group of liquidity measures that provide same kind of information for the market. Therefore, we can reduce number of liquidity measures without loss of information. To achieve this goal, we build a correlation matrix for a set of liquidity measures.

Definition 1. *Let denote with* $X = \{x_i^j\}$ *where* $(i,j) \in \{1,\dots,n\} \times \{1,\dots,m\}$ *a set of* $n \times m$ *liquidity measures.* m *and* n *denote respectively the number of days in our sample data set and the total number of liquidity measures used in this study. So, the correlation matrix* C^j *for any given day* j *is a* $n \times n$ *symmetric matrix whose* p,q *entry is the correlation between a pair of liquidity measures like* x_p *and* x_q *where* $p,q \in \{1,\dots,n\}$*, and it is defined as following:*

$$C^{j}{}_{pq} = \begin{cases} 1 & , \ if\ p = q \\ \frac{Cov(x_p, x_q)}{\sigma_{x_p}\sigma_{x_q}} & , \ if\ p \neq q \end{cases} \quad (1)$$

where $Cov(x_p, x_q)$ *is the covariance between two liquidity measures* (x_p, x_q) *and is defined as following:*

$$Cov(x_p, x_q) = E[(x_p - E[x_p])(x_q - E[x_q])] \quad (2)$$

and in Equation (1), σ_{x_p} *and* σ_{x_q} *represent the standard deviation for* x_p *and* x_q *respectively.*

By employing correlation analysis, we are able to detect a set of liquidity measures that have common movement. We use the term *commonality* to describe this pattern[1]. The reason for using superscript j in our notation is that we are interested to analyze the structure of the correlation matrix during our sample period. More details are presented in section 3.4.

2.3 Clustering Algorithm

In order to reduce sample space of the liquidity measures without losing any information, we use a clustering algorithm to partition liquidity measures into multiple groups. Clustering algorithms have been used in different research areas. The main purpose of these algorithms is to partition data points into some groups so that the points within each group are similar to each other, and the points from different groups are dissimilar (Tan et al. (2006)). Various clustering algorithms have been developed based on how a similarity between data points is measured (Ketchen and Shook (1996), Jain and Dubes (1988), Xu et al. (2005)). In addition to defining a similarity measure, most of these algorithms require a prior knowledge of the number of clusters that we want to divide the data points. For instance, in *k-means* algorithm, we need to define *k* which is the number of clusters. On the other hand, there exist some algorithms that do not require any information about the number of clusters in advance. For example, hierarchical clustering algorithms have this property. The result of a hierarchical clustering algorithm can be presented by a dendrogram (tree-like visual representation). Since, in this study we use a hierarchical clustering algorithm, we provide more details about this algorithm in the following section.

2.3.1 Hierarchical Clustering

First step in implementing a hierarchical clustering algorithm is to define a measure that can quantify similarity or dissimilarity among data points. In general, the scientific question that researchers try to address is a key factor in determining the dissimilarity measure (Gareth (2013)). In this study, we use Pearson correlation coefficient to define a dissimilarity measure.

Definition 2. *Let* ρ_{pq} *to be Pearson correlation coefficient between liquidity measures* x_p *and* x_q*. Mathematically*

$$\rho_{x_p x_q} = \frac{\sum_{k=1}^{l}(x_{pk} - E[x_p])(x_{qk} - E[x_q])}{\sqrt{\sum_{k=1}^{l}(x_{pk} - E[x_p])^2 \sum_{k=1}^{l}(x_{qk} - E[x_q])^2}} \quad (3)$$

Where l *is the total number of observation in one day. Then, we define a dissimilarity measure as following:*

$$D_{x_p x_q} = (1 - |\rho_{x_p x_q}|) \quad (4)$$

It is clear that the higher the correlation coefficient between two liquidity measures, the smaller the dissimilarity between them. Xu et al. (2005) review different similarity and dissimilarity measures used in clustering algorithms.

Agglomerative and divisive methods are two main methods to implement a hierarchical clustering algorithm. The agglomerative method is a bottom-up approach. It assigns each data point into a separate cluster. So, for N data points, we will have N clusters in the beginning. Then, it merges two similar clusters until a certain stopping criteria is achieved. The divisive approach is a top-down approach. It puts all the data points in a unique cluster. Then, it divides this cluster into two clusters. This process will continue until all clusters are singleton clusters (Xu et al. (2005)). Since, divisive method is computationally expensive (Everitt et al. (2001)), we use the agglomerative approach in this study. The dissimilarity measure defined in Equation 4, can quantify dissimilarity between two liquidity measures. In order to quantify dissimilarity between two sub clusters that each one has multiple data points, we use the complete linkage which considers the dissimilarity value between all pairs of observations in two clusters and keeps the largest value. More detail about different type of linkage methods can be found in Gareth (2013).

3 Results

In this section, we illustrate our results. First, we explain the data set used in this study. Second, we discuss the results obtained from the correlation analysis of the liquidity measures. Third, we present the results from the hierarchical clustering analysis. Finally, we provide a discussion about the consistency in the structure of the clusters.

3.1 Data Set

For this study, we use a high resolution data set provided by NASDAQ. This data set have been studied in different literature (Brogaard (2010), Kearns et al. (2010), and Khashanah et al. (2014)). These researchers use this data set to analyze the role of High Frequency Traders (HFTs) in United States equity market. This data set contains trading and quoting activity of 26 HFT firms in 120 stocks on the NASDAQ exchange. The sample is organized by market capitalization and is evenly split by NASDAQ to three different categories, large cap, medium cap and low cap. The sample period covers the week of Feb 22 - 26, 2010. Timestamp for trades is millisecond, and each trade has a flag to indicate whether it is initiated by a buyer or a seller. Also, trade reports contains a field with the following codes: HH, HN, NH, or NN. H represents a HFT and N denotes a non-HFT. The first term in the pair classifies the liquidity seeking side, and the second character classifies liquidity supplier. For example, a trade labeled HN would mean an HFT took liquidity from a non-HFT. In addition to trade information, for the week of Feb 22 - 26, 2010, the data set contains the collective best HFT quote along with the collective best non-HFT quote.

3.2 Correlation Analysis Result

In this section, we present the results from the correlation analysis. First, we calculate 22 different liquidity measures within one second time interval for all 120 stocks in our data set. We repeat this process for each

[1] The commonality and idiosyncratic patters are two terms used in literature to analyze cross-sectional variation in liquidity. See Chordia et al. (2000), Mancini et al. (2013),Karolyi et al. (2012)

day. Then, we calculate the correlation matrix for that specific day. We set the correlation coefficients to zero if they are not statistically significant (Casella and Berger (2002)). One of these matrices is shown in Figure 1. In this figure, the blue color circles display positive correlation, and the red color circles shows negative correlation. Color intensity is proportional to the correlation coefficient. As we were expecting and documented in literature (Von Wyss (2004)), the economically similar liquidity measures are highly correlated to each other. For instance, high correlation between *Sabs*, *SrelM*, *Srelp* is shown in Figure 1.

Further, by paying close attention to the structure of the correlation matrix, we can see that the correlation matrix can be decomposed into multiple blocks. This special structure of the correlation matrix is quiet interesting, since it brings up the idea that there are some liquidity measures that are highly correlated to each other. Therefore, we can reduce the number of liquidity measures without losing any useful information. So, this analysis reveals the presence of commonality pattern among liquidity measures.

We do further analysis to examine whether this structure can be verified for all the stocks in our sample period. Figure 2 shows the correlation matrix for three stocks during five days. These stocks, *AMZN*, *SFG*, and *ROG* are chosen from high cap, medium cap, and low cap categories respectively. We can clearly see that, the correlation matrices have almost same structure during sample period. This result shows that there is a consistency in the structure of the correlation matrix. Block structure of the correlation matrix proves that there is a high linear relationship between some of these liquidity measures. This finding is used to classify liquidity measures into different groups.

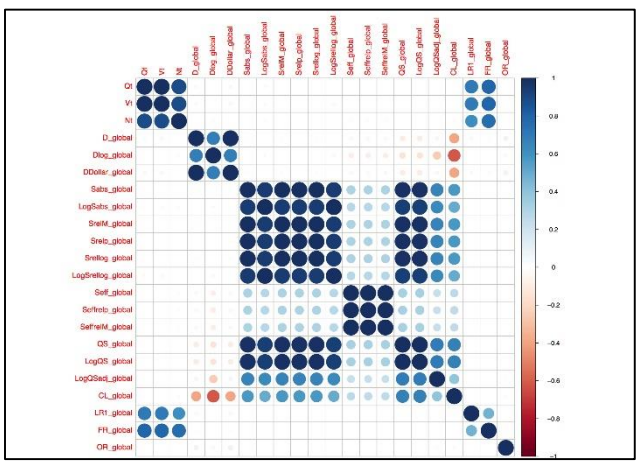

Fig 1: Correlation Matrix of Liquidity Measures for AAPL during Feb 22, 2010

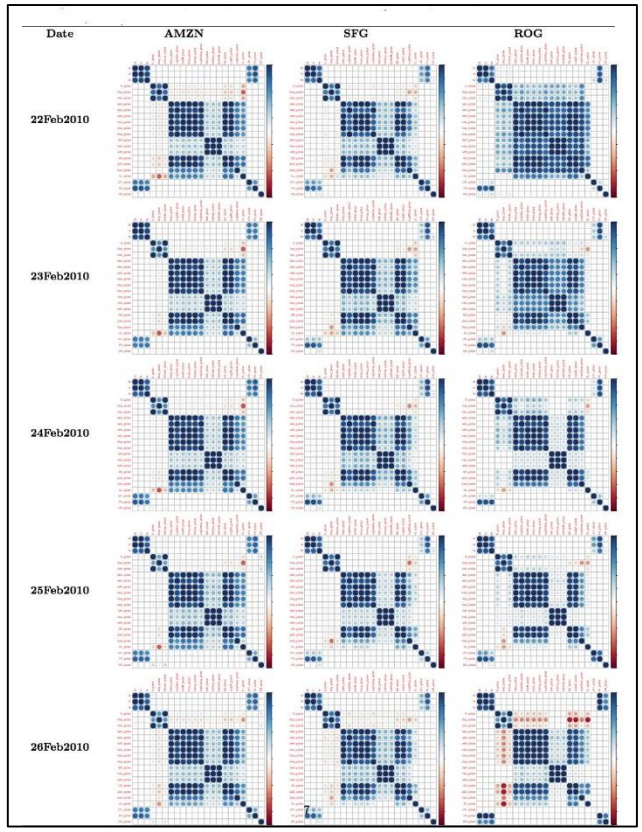

Figure 2: Comparing the structure of the correlation matrix for AMZ, SFG and ROG during a week from 22 Feb, 2010 - 26 Feb, 2010

3.3 Hierarchical Clustering Results

The main contribution of this study is to classify liquidity measures into different groups. This can help us to identify liquidity measures that are providing same kind of information to the market. It means that we can reduce the number of liquidity measures without losing any useful information. To achieve this goal, we employ a hierarchical clustering algorithm by using Pearson correlation coefficient to define our dissimilarity measure. The result of a hierarchical clustering can be presented by a dendrogram. For example, Figure 3 is obtained by employing hierarchical clustering on all 22 liquidity measures calculated in this study. By cutting the dendrogram at specific level (vertical axis), we will have different number of clusters. Table 1 summarizes the number of the clusters and the members of each cluster obtained from setting cutting level at 0.4. In this case, we are sure that the absolute value of the correlation coefficient among the liquidity measures is greater than or equal to 0.6. By changing the cutting level value, the structure of the clusters will be different. Figure 4 and Table 2 shows the dendrogram and the cluster structure obtained from choosing 0.2 as a cutting level. The presented results is just for one stock and it is considering only one day in our sample data set.

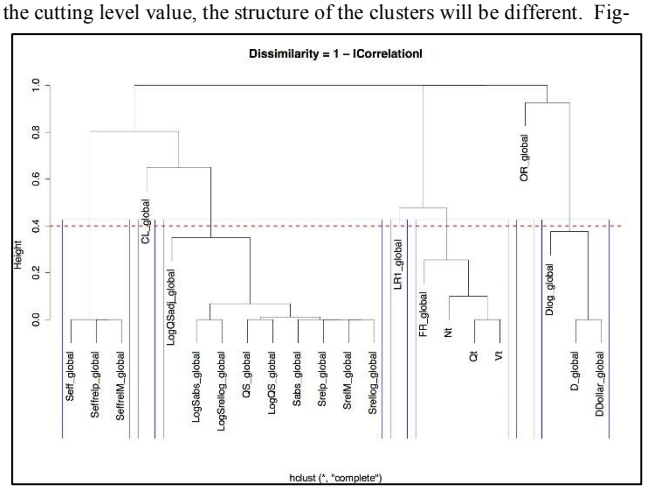

Fig 3: Hierarchical Cluster Analysis for AAPL during 26Feb, 2010, cutting level=0.4

Table 1: Cluster with their corresponding members . Ticker is AAPL, date is Feb 26, 2010. Cutting level of dendrogram is 0.4

Clusters	Liquidity Measures
1	Q_t, V_t, N_t, FR
2	$D, Dlog, D\$$
3	$Sabs, LogSabs, SrelM, Srelp, Srellog$ $LogSrellog, QS, LoggQs, LogQSadj$
4	$Seff, Seffrelp, SeffrelM$
5	CL
6	$LR1$
7	OR

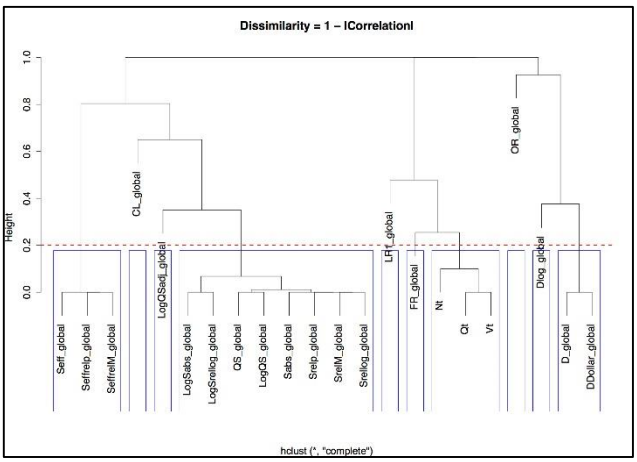

Table 2: Clusters with their corresponding members for AAPL during 26Feb, 2010, cutting level=0.2

Fig 4: Hierarchical Cluster Analysis for AAPL during 26Feb, 2010, cutting level=0.2

Clusters	Liquidity Measures
1	Q, V, N
2	$D, D\$$
3	$Dlog$
4	$Sabs, LogSabs, SrelM, Srelp$ $Srellog, LogSrellog, QS, LogQS$
5	$Seff, Seffrelp, SeffrelM$
6	$LogQSadj$
7	CL
8	$LR1$
9	FR
10	OR

As we mentioned in the above, the results we presented for the hierarchical clustering is only for one stock and one day. We need to investigate whether this structure is the same for all the stocks in our sample data set during all days. To do this, we repeat the process, and we count number of stocks that have same cluster structure for whole sample period. We find that this structure is not consistent. There are some liquidity measures that make inconsistency in the structure of the clusters. After doing analysis, we identified those liquidity measures to be *number of transactions per time unit* (N_t), *logDepth* ($DLog$), *composite liquidity* (CL_t), *liquidity ratio1*($LR1$) and *adjusted log quote slope* ($LogQSadj$). After removing these 5 problematic liquidity measures , we get significant improvement in the results. Table 3 shows the cluster structure obtained after removing the problematic liquidity measures. By removing the problematic liquidity measures, the consistency in the structure of the clusters is significantly improved. Table 4 shows the results for two cases. It is clear that 79.2% of the stocks in our date set keep having same cluster structure for all days in our sample period.

Table 3: Cluster structure after removing problematic liquidity measures. For AAPL during 26 Feb, 2010, cutting level=0.4

Clusters	Liquidity Measures
1	Q_t, V_t, FR
2	$D, D\$$
3	$Sabs, LogSabs, SrelM, Srelp, Srellog$ $LogSrellog, QS, LogQS$
4	$Seff, Seffrelp, SeffrelM$
5	OR

Table 4: Comparing consistency in the structure of cluster before and after removing problematic liquidity measures.

Stock Category	Considering all the liquidity measures	Removing the problematic liquidity measures
All stocks	2.5%	79.2%
Large cap	4.8%	76.2%
Medium Cap	0%	77.5%
Small Cap	2.6%	84.2%

3.4 How to choose a cutting level for the dendrogram?

In all the analysis and the results we discussed so far, we assume a fixed cutting level. Specifically, we set this value to be 0.4. In this section we provide a discussion about how we get this value. Generally, this value depends on the nature of the problem and the goal we want to obtain through a hierarchical clustering. In this research, we are using the following procedure. Let α to be the value of dissimilarity function which in our case is defined as following:

$$\alpha = 1 - |\rho| \tag{5}$$

where ρ is the Pearson correlation coefficient. The vertical axis in a dendrogram obtained from a hierarchical clustering (Figure 3) shows the range of α. Since $|\rho| \leq 1 \Longrightarrow 0 \leq \alpha \leq 1$.

- If $\alpha = 1$ ($|\rho|$ =0), then we have one big cluster.

- If $\alpha = 0$ ($|\rho|$ =1), then we have m clusters where m is the total number of liquidity measures.

So, let k to be the number of clusters we obtain by cutting the dendrogram at level α_i. \label{algo}The possible values of k are $1,2,\ldots,m,$ where m is the total number of liquidity measures. For any given stock s_i (in our data set), we cut the dendrogram at level α_i. Then, we calculate k and record members of ecah cluster as well. We repeat this process for all days (in our sample period). We say stock $s_i$has a consistent cluster structure if the number of clusters and the members of each cluster are the same for all days. We repeat this procedure for all the stocks, and count how many stocks have these properties. We need to do this procedure for different value of α_i. It is clear that in two cases we can have the best consistency. First case happens when $\alpha = 0$ which means $|\rho|$ =1 that implies $k = m$. Second case happens when $\alpha = 1$ which means that $|\rho|$ =0 that implies

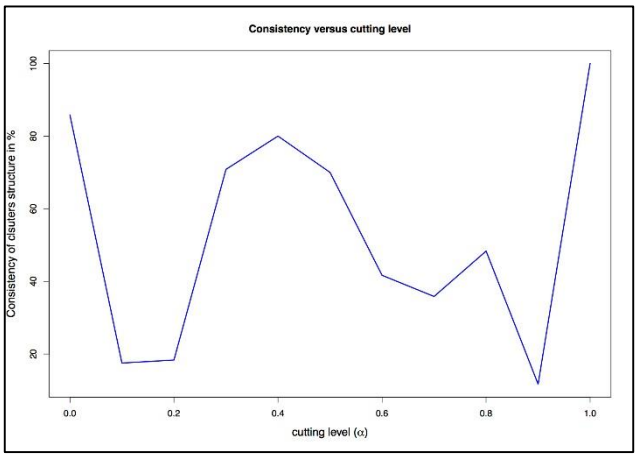

Fig 5: Consistency versus cutting level

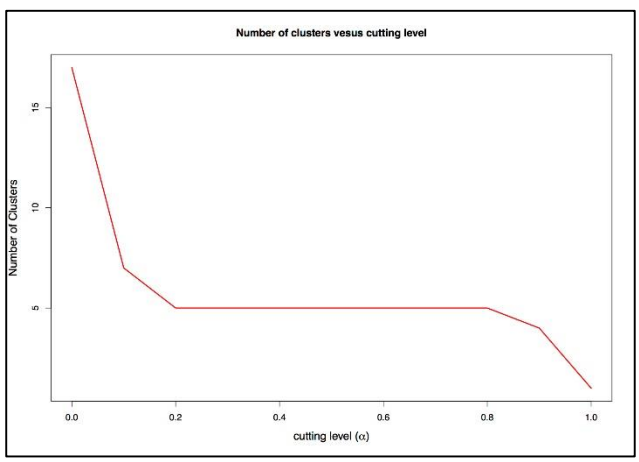

Fig 6: Number of clusters versus cutting level

$k = 1$. In other words, we can say that k is a decreasing function of α. Also, we need to take into account that the whole purpose of clustering is to partition similar liquidity measures into one groups. Therefore, we want to achieve the following while decreasing the value of dissimilarity function (α_i).

1. Increase consistency.
2. Decrease number of clusters.

We need to solve this optimization problem. We can look at the plot of the consistency versus α and the number of clusters versus α . Then, we can identify the region of interest (Figure 5, and Figure 6).

Numerically, by changing the value of α we record the consistency and the the total number of the clusters (Table 5). It is clear that number of clusters is a decreasing function of cutting level. And, For $\alpha = 0$ and $\alpha = 1$, we obtained highest consistency. We can see that by fixing the cutting level at 0.4, we can obtain the highest consistency in the structure of the clusters.

Table 5: Analyzing how cutting dendrogram obtained from hierarchical clustering at different levels will affect the consistency in the structure of the clusters

Cutting Level (α)	Consistency (Percentage)	Number of Clusters
0.00	85.83	17
0.10	17.50	7, 6
0.20	18.33	5,6
0.30	70.83	4,5
0.40	80.00	4,5
0.50	70.00	4,5
0.60	41.67	4,5
0.70	35.83	4,5
0.80	48.33	4,5
0.90	11.67	4
1.00	100.00	1

4 Conclusion

There exist a lot of liquidity measures in literature. These measures have been classified into different groups based on the market feature they capture. In this study we look into the classification of the liquidity measures from different perspective.

We perform a cluster analysis of the existing liquidity measures in the literature. Using a high frequency data assists us to investigate the intraday behavior of the liquidity measures more precisely. By analyzing the correlation matrix of the liquidity measures, we find that there is a block structure in the correlation matrix and further investigation reveals that this structure is consistent. We define a dissimilarity measure using a correlation concept. By using a hierarchical clustering algorithm, we are able to partition liquidity measures into different groups. We conclude that, there is a consistency in the the structure of the dendrograms obtained from the hierarchical clustering. This results obtained from the clustering can help us to reduce the number of the liquidity measures without losing any information. For future work, we are interested to use the results from the clustering procedure to develop a (ideally one) new liquidity index which is capable of capturing most of the market features.

Acknowledgements

The authors acknowledge the NASDAQ OMX Group for graciously providing us with a sample of 120 stocks with various levels of market capitalization listed on NYSE and NASDAQ where HFT traders are globally labeled under designation H.

Appendix

Mathematical description of the liquidity measures used in this study is presented in the following table. In this table, q_i and p_i denotes the number of shares of trade i and its price respectively. q_t^A and q_t^B refer to best ask volume and best bid volume respectively. Accordingly, p_t^A and p_t^B represent best ask price and best bid price respectively. p_t denotes the last paid price of the asset before time$t$.. r_t represents the return during the time interval for which we calculate liquidity proxies.

Table 6: Mathematical description of liquidity measures used in this study. The number of measures reviewed by Von Wyss (2014) is 31. We only present 22 of them which can be calculated using the high frequency data set that we analyzed for this study.

One Dimensional Liquidity Measures	
Liquidity Measures	Formula
Trading Volume	$Q_t = \sum_{i=1}^{N_t} q_i$
Turnover	$V_t = \sum_{i=1}^{N_t} p_i\, q_i$
Depth	$D_t = q_t^A + q_t^B$
LogDepth	$Dlog_t = \ln(q_t^A) + \ln(q_t^B)$
Dollar Depth	$D\$_t = \frac{q_t^A . p_t^A + q_t^B . p_t^B}{2}$
Number of Transaction per time unit	N_t
Absolute Spread	$Sabs_t = p_t^A - p_t^B$
Log Absolute Spread	$LogSabs_t = \ln(Sabs_t)$
Relative spread calculated with mid-price	$SrelM_t = \frac{p_t^A - p_t^B}{p_t^M} = \frac{2.(p_t^A - p_t^B)}{p_t^A + p_t^B}$
Relative spread calculated with last trade price	$Srelp_t = \frac{p_t^A - p_t^B}{p_t}$
Relative spread of log prices	$Srellog_t = \ln(p_t^A) - \ln(p_t^B)$
Log relative spread of log prices	$LogSrellog_t = \ln(Srellog_t)$
Effective Spread	$Seff_t = \lvert p_t - p_t^M \rvert$
Relative effective spread calculated with last trade price	$Seffrelp_t = \frac{\lvert p_t - p_t^M \rvert}{p_t}$
Relative effective spread calculated with mid-price	$SeffrelM_t = \frac{\lvert p_t - p_t^M \rvert}{p_t^M}$
Multi-Dimensional Liquidity Measures	
Quote Slope	$QS_t = \frac{Sabs_t}{Dlog_t}$
Log Quote Slope	$LogQS_t = \frac{Srellog_t}{Dlog_t}$
Adjusted Log Quote Slope	$LogQSadj_t = LogQS_t.(1 + \lvert\ln(\frac{q_t^A}{q_t^B})\rvert)$
Composite Liquidity	$CL_t = \frac{SrelM_t}{D\$_t}$
Liquidity Ratio 1	$LR1_t = \frac{V_t}{\lvert r_t \rvert}$
Flow Ratio	$FR_t = N_t.V_t$
Order Ratio	$OR_t = \frac{\lvert q_t^B - q_t^A \rvert}{V_t}$

Conflict of Interest: none declared.

References

Acharya, V. V. and L. H. Pedersen (2005). Asset pricing with liquidity risk. *Journal of Financial Economics 77 (2)*, 375-410.

Acker, D., M. Stalker, and I. Tonks (2002). Daily closing inside spreads and trading volumes around earnings announcements. *Journal of Business Finance & Accounting 29 (9-10)*, 1149-1179.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5 (1)*, 31-56.

Amihud, Y. and H. Mendelson (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics 17 (2)*, 223-249.

Amihud, Y., H. Mendelson, and B. Lauterbach (1997). Market microstructure and securities values: Evidence from the tel aviv stock exchange. *Journal of Financial Economics 45 (3)*, 365-390.

Bali, T. G., L. Peng, Y. Shen, and Y. Tang (2014). Liquidity shocks and stock market reactions. *Review of Financial Studies 27 (5)*, 1434-1485.

Black, F. (1971). Toward a fully automated stock exchange, part i. *Financial Analysts Journal 27 (4)*, 28-35.

Brennan, M. J. and A. Subrahmanyam (1996). Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of financial economics 41 (3)*, 441-464.

Brogaard, J. (2010). High frequency trading and its impact on market quality. *Northwestern University Kellogg School of Management Working Paper 66*.

Casella, G. and R. L. Berger (2002). *Statistical inference*, Volume 2. Duxbury Pacific Grove, CA.

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

Chordia, T., R. Roll, and A. Subrahmanyam (2001). Market liquidity and trading activity. *The Journal of Finance 56 (2)*, 501-530.

Cooper, S. K., J. C. Groth, and W. E. Avera (1985). Liquidity, exchange listing, and common stock performance. *Journal of Economics and Business 37 (1)*, 19-33.

Dacorogna, M., R. Gencay, U. Muller, R. Olsen, and O. Pictet (2001). An introduction to high frequency finance academic press. *San Diego*.

Everitt, B., S. Landau, and M. Leese (2001). Cluster analysis arnold. *A member of the Hodder Headline Group, London*.

Gareth, J. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.

Goyenko, R. Y., C. W. Holden, and C. A. Trzcinka (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics 92 (2)*, 153-181.

Holden, C. W., S. Jacobsen, A. Subrahmanyam, et al. (2014). The empirical analysis of liquidity. *Foundations and Trends in Finance* 8 (4), 263-365.

Irvine, P. J., G. J. Benston, and E. Kandel (2000). Liquidity beyond the inside spread: Measuring and using information in the limit order book. *Available at SSRN 229959* .

Jain, A. K. and R. C. Dubes (1988). *Algorithms for clustering data*. Prentice-Hall, Inc.

Kang, W. and H. Zhang (2013). Limit order book and commonality in liquidity. *Financial Review 48 (1)*, 97-122.

Karolyi, G. A., K.-H. Lee, and M. A. Van Dijk (2012). Understanding commonality in liquidity around the world. *Journal of Financial Economics 105 (1)*, 82-112.

Kearns, M., A. Kulesza, and Y. Nevmyvaka (2010). Empirical limitations on high frequency trading portability. *Available at SSRN 1678758* .

Ketchen, D. J. and C. L. Shook (1996). The application of cluster analysis in strategic management research: an analysis and critique. *Strategic management journal 17 (6)*, 441-458.

Khashanah, K., I. Florescu, and S. Yang (2014). On the impact and future of hft.

Kyle, A. S. (1985). Continuous auctions and insider trading. Econometrica: *Journal of the Econometric Society*, 1315-1335.

Lee, C. (1993). Market integration and price execution for nyse-listed securities. *The Journal of Finance 48 (3)*, 1009-1038.

Lee, C. and B. Swaminathan (2000). Price momentum and trading volume. *The Journal of Finance 55 (5)*, 2017-2069.

Liu, W. (2006). A liquidity-augmented capital asset pricing model. *Journal of financial Economics 82 (3)*, 631-671.

Mancini, L., A. Ranaldo, and J.Wrampelmeyer (2013). Liquidity in the foreign exchange market: Measurement, commonality, and risk premiums. *The Journal of Finance 68 (5)*, 1805-1841.

Moore, D. S. and G. P. McCabe (1989). *Introduction to the Practice of Statistics*. WH Freeman/Times Books/Henry Holt & Co.

O'Hara, M. (2004). Liquidity and financial market stability. *National Bank of Belgium Working Paper (55)*.

Pastor, L. and R. F. Stambaugh (2001). Liquidity risk and expected stock returns. Technical report, National Bureau of Economic Research.

Roll,R.(1984).A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance 39 (4)*, 1127-1139.

Sarr, A. and T. Lybek (2002). Measuring liquidity financial markets.

Tan, P.-N. et al. (2006*). Introduction to data mining*. Pearson Education India.

Van der Merwe, A. (2015). *Market Liquidity Risk: Implications for Asset Pricing, Risk Management, and Financial Regulation*. Palgrave Macmillan.

Von Wyss, R. (2004). *Measuring and predicting liquidity in the stock market*. PhD dissertation, University of St. Gallen.

Vu, V., D. Chai, and V. Do (2015). Empirical tests on the liquidity-adjusted capital asset pricing model. *Pacific-Basin Finance Journal 35*, 73-89.

Xu, R., D. Wunsch, et al. (2005). Survey of clustering algorithms. *Neural Networks, IEEE Transactions on 16 (3)*, 645-678.