



Measuring the Stock Liquidity Using a Market Microstructure Approach

Nastaran Hadi Doulabi, Mohammad Ali Rastegar*, Parastoo Mohammadi

Faculty of Industrial Engineering and Systems, Tarbiat Modares University, Tehran, Iran

Received: 03 June 2021, Revised: 11 June 2021, Accepted: 12 June 2021

© University of Tehran 2020

Abstract

The main objective of this research is to identify the most important liquidity measures and their behavior during the trading day. For this purpose, the intraday data of 7 stocks of the Tehran Stock Exchange have been used to calculate 27 liquidity measures selected from the literature. At the first step, the distribution features and the correlation structure of the liquidity measures are examined. Using the Principal Components Analysis method, these components are identified, and their intraday patterns are extracted. The results show that reducing the number of measures to four final measures that can describe all aspects of liquidity without eliminating helpful information is possible. Among the final measures, Relative Spread with mid quoted prices can be mentioned as the most practical microstructure component affecting liquidity. Based on this measure's intraday pattern, it can be said that this measure is minimized in the middle of the day. So, liquidity is high during these hours, and favorable conditions for trading are provided. In the end, stocks are ranked based on all 27 liquidity measures through two different methods. Therefore, this study helps traders make decisions about the liquidity of their stock portfolios using a comprehensive method.

Keywords:

Stock Liquidity Measures;
Market Microstructure;
Intraday Patterns;
Principal Component
Analysis;
TOPSIS

Introduction

One of the critical factors in deciding to invest in financial markets is the liquidity of assets. This concept can be defined in various markets; in this regard, this research focuses on the stock market and specifically on the Tehran Stock Exchange. In a recent study, Quah et al. [1], examined the association between stock liquidity and investment efficiency for companies with financial constraints and information asymmetry problems. They conclude that the effect of higher stock liquidity on lowering under-investment or over-investment is more pronounced for such companies.

Liquidity is a qualitative concept that means the ability to absorb buy and sell orders. Conducting studies on stock liquidity from a microstructural perspective is essential for improving financial markets' performance and stability. Many researchers have tried to quantify this concept in recent years and introduced several criteria for measuring it. However, liquidity is a multi-dimensional concept that cannot be measured by a single criterion. Therefore, researchers have defined four different aspects of liquidity: market depth (the effect of high volume orders on price), market width (difference between the bid and ask prices), resiliency (market's ability to bounce back from temporarily incorrect prices), and the speed of trades [2].

* Corresponding author: (M.A. Rastegar)
Email: ma_rastegar@modares.ac.ir

Stock liquidity is influenced by the market microstructure model. One of the theoretical market microstructure models is the limit order market model. With the development of information technology in recent decades, electronic trading systems have gradually replaced the designated roles of market makers, and many exchanges have established limit order markets. Today, more than fifty percent of stock exchanges worldwide have limit-order markets that make it possible for sellers and buyers to match orders through a Limit Order Book platform [3]. One of the advantages of having such a platform is that the system provides excellent transparency; quotes and transactions are visible to all participants, which generally improves the efficiency of the price discovery process [4]. Unlike the dealer market, there is no specific market maker who maintains market liquidity in this market type. The market participants play the roles of supplying and demanding liquidity, using either limit orders or market orders. This category of models focuses on analyzing how strategies for submitting orders and other aspects of trading affect an asset's price in limit order markets [5]. The current research has been done in limit order market models, knowing that the Tehran Stock Exchange is an order-driven market.

One issue to consider about liquidity measures is data frequency. Many of the liquidity measures in the literature are based on daily data, which are low-frequency. However, analyzing liquidity measures based on high-frequency data provides a better insight into the nature of liquidity change over time. In this regard, Korolev et al. [6] propose a micro-scale model studying the process of high-frequency order flow imbalance, which tracks the best bid and ask queues and changes much faster than prices. Therefore, it can describe the market dynamics at short time intervals in an efficient way. This order imbalance measure has also been used in Muranaga and Shimizu [7] and the current study. Also, according to many studies, such as Gabrielsen et al. [8], intraday liquidity measures can better reflect the internal characteristics of a market, such as the effect of new information reaching the parties to the transaction. In [9], the liquidity measures that can be calculated with high-frequency intraday data have been clustered, and the correlation between different measures to achieve similar behavior in each group of these proxies has been investigated. Therefore, more studies are needed to develop these measures. For this reason, in this paper, tick-by-tick data of limit orders and transactions have been used to calculate high-frequency liquidity measures.

Another field of literature deals with the extent to which these liquidity measures are correlated. This concept is called commonality in liquidity and can be captured by calculating the covariance between different liquidity measures within an almost long period. Będowska-Sójka [10] has explored this issue on the emerging order-driven market and considered few liquidity proxies, examining if there exists a commonality in liquidity measures of stocks quoted on the Warsaw Stock Exchange. For this purpose, they used the dataset for about ten years and calculated three illiquidity measures such as Amihud (2002) and some spread-based measures. They conclude that the commonality depends on the firm size is time-varying. Johann et al. [11] also addressed this issue in the German Electronic Exchange. They use a dataset covering the data of around 14 years and conclude that commonality in liquidity is highest during the financial crisis, which is bad news for asset managers because it implies that low returns, low market liquidity, and high commonality in liquidity tend to coincide.

There is a relatively extensive literature in the field of critique and review of various liquidity measures, comparing them with each other and estimating their power to reflect liquidity. These include the research of Minović et al. [12], which analyzes one-dimensional liquidity measures such as turnover and relative Spread. They state that the relative Spread, which results from dividing the absolute Spread by the midpoint quotes, is a better measure than the absolute Spread; because it is comparable between different stocks with different price levels. Although many researchers have proposed various proxies for quantifying liquidity in the market, the literature shows that many of them have serious shortcomings. For example, among price-

related measures, the Amihud (2002) measure evaluates the lack of liquidity by dividing daily returns by the daily dollar volume but cannot take non-trading days into account. In another study, Irvine et al. [13] tested the characteristics of a liquidity measure called Cost of Round Trip Trade, representing the approximate cost of transactions. They used the information on the Toronto Stock Exchange Limit Order Book and compared the performance of this measure with the quoted Spread and the effective Spread. Finally, they assess the ability of this proxy to predict future trading activities.

In the current study, 27 liquidity measures have been used to quantify the liquidity of 7 selected stocks from Tehran Stock Exchange. These measures are calculated based on the intraday data of the limit order book and the transaction data. The first contribution of this study is to find those liquidity proxies that, on the one hand, are simple to calculate based on available data and are comparable between different stocks, and on the other hand, have the most explanation of the concept of liquidity. The second contribution of this study is the ranking of stocks based on their liquidity characteristics so that they can be selected as candidates for investment portfolios.

The rest of the paper is organized as follows: [Section 2](#) introduces the liquidity measures used in this study and points out the methodologies for identifying the main components of stock liquidity and stocks ranking; [Section 3](#) outlines the discussion of results while conclusions are given in [Section 4](#).

Literature review

Problem description

Liquidity is an important issue in financial markets, and investors and regulators need to be aware of liquidity conditions in the stock market. Therefore, it is of importance to examine stock liquidity using a comprehensive method. In this paper, we attempt to examine liquidity from certain aspects and answer the following questions:

- 1- Which liquidity measures best explain the concept of liquidity and how they behave during a trading day?
- 2- What is the ranking order of stocks based on their liquidity characteristics through different ranking methods?

Liquidity measures

As mentioned before, liquidity is not an observable quantity and must be measured by various proxies. A list of 27 liquidity measures used in this study is shown in [Table 1](#), most of which have been used in [14] and a few of them in [15].

Table 1. Mathematical description of liquidity measures used in this study

Liquidity Measures	Formula		Liquidity Measures	Formula		In the above
Trading Volume	$TV_t = \sum_{i=1}^{N_t} q_i$	(1)	Relative Spread of Log Prices	$LogRelLogSp_t = \ln(\ln(p_t^A/p_t^B))$	(14)	
Trading Volume per Trade	$TVper_t = \frac{\sum_{i=1}^{N_t} q_i}{N_t}$	(2)	Effective Spread	$EffSp_t = p_t - p_t^M $	(15)	
Turnover	$Turnover_t = \sum_{i=1}^{N_t} p_i \cdot q_i$	(3)	Effective Spread with Last Trade	$RelEffSp_t = p_t - p_t^M /p_t$	(16)	
Number of Trades	N_t		Effective Spread with Mid Price	$RelEffSpMid_t = p_t - p_t^M /p_t^M$	(17)	
Volume Duration	$Vdur_t = \text{Min}(Vdur: TV_t + v_{dur} \geq TV_t + V)$	(4)	Quote Slope	$Qslope_t = p_t^A - p_t^B / \ln(q_t^A) + \ln(q_t^B)$	(18)	
Market Depth	$Depth_t = q_t^A + q_t^B / 2$	(5)	Log Quote Slope	$LogQslope_t = \ln(p_t^A) - \ln(p_t^B) / \ln(q_t^A) + \ln(q_t^B)$	(19)	
Log Depth	$LogDepth_t = \ln(q_t^A) + \ln(q_t^B) = \ln(q_t^A \cdot q_t^B)$	(6)	Adjusted Log Quote Slope	$AdjLogQslope_t = LogQslope_t \cdot (1 + \ln(q_t^B) - \ln(q_t^A))$	(20)	
Dollar Depth	$Ddepth_t = (q_t^A \cdot p_t^A + q_t^B \cdot p_t^B) / 2$	(7)	Composite Liquidity	$CL_t = RelSp/Ddepth$	(21)	
Near Depth	$Ndepth_t = \frac{(\sum_{j=1}^3 q_{t,j}^A) + (\sum_{j=1}^3 q_{t,j}^B)}{2}$	(8)	Liquidity Ratio 1	$LR1_t = Turnover_t / r_t $	(22)	
Near Depth Value	$NdepthV_t = \frac{(\sum_{j=1}^3 p_{t,j}^A \cdot q_{t,j}^A) + (\sum_{j=1}^3 p_{t,j}^B \cdot q_{t,j}^B)}{2}$	(9)	Liquidity Ratio 2	$LR2_t = \frac{\sum_{i=1}^{N_t} r_i }{N_t}$	(23)	
Absolute Spread	$AbsSp_t = p_t^A - p_t^B$	(10)	Flow Ratio	$FR_t = N_t \cdot Turnover_t$	(24)	
Log Absolute Spread	$LogAbsSp_t = \ln(p_t^A - p_t^B)$	(11)	Order Ratio	$OR_t = q_t^A - q_t^B / Turnover_t$	(25)	
Relative Spread with Mid Price	$RelSp_t = p_t^A - p_t^B / p_t^M$	(12)	Order Imbalance	$\%ORI_t = \frac{TV_t^{Sell} - TV_t^{Buy}}{TV_t} * 100$	(26)	
Relative Spread with Last Trade	$RelSpt_t = p_t^A - p_t^B / p_t$	(13)				

equations, p_i and q_i represent the price and volume of the transaction i ; p_t^A and p_t^B refer to the ask and bid prices, which are also called quotes; q_t^A and q_t^B are the quantity of quotes, and p_t^M is the mean of these quotes. while j indicates the level of the limit order book, $p_{t,j}^B$ and $p_{t,j}^A$ indicate the bid and ask prices of level j at time t , and so $q_{t,j}^B$ and $q_{t,j}^A$ indicate the quantity of these quotes, and finally, r_i is the stock return.

These liquidity measures can be classified into four groups:

1. Volume-related Liquidity Measures:

These proxies generally measure the frequency of a transaction using traded volumes directly or indirectly and are somehow related to the time dimension of liquidity; because in a market where larger volumes are traded, the time required to trade a certain number of shares is reduced. Another point is that the larger the measures of this group, the more liquid the market.

2. Depth-related Liquidity Measures:

These measures cover the depth dimension of liquidity by focusing on the volume and price of buy and sell orders. As in the previous group, the larger these measures, the greater the liquidity in the market.

3. Spread-related Liquidity Measures:

The difference between the bid and ask prices gives the approximate cost of the transaction. In addition to fees and taxes, the trader must pay the Spread to make a quick transaction. The smaller the criteria of this group, the more liquid the market.

4. Multi-dimensional Liquidity Measures:

The measures in this category are, in fact, a combination of the liquidity measures mentioned in the previous three groups. Some of these measures combine spread in the numerator and the volume in the denominator; Therefore, the smaller these proxies, the greater the stock liquidity.

Data collection

Due to the nature of the liquidity measures selected for this study, which are calculated based on intraday data, we need stocks with high-frequency trades. The data covers 77 trading days from September 22, 2016, until February 18, 2017, for seven stocks in Table 2, selected based on statistical reports available on the Tehran Stock Exchange website publishing the list of top 50 companies every three months. It should be noted here that due to restrictions imposed by the relevant organization in sharing transaction data in recent years, it was not possible to access newer data. However, since the results of this study are more influenced by the nature of the selected stocks in terms of the fundamental characteristics of the companies and the trading volume of their stocks, it is expected that the time period will not have a significant effect on the results.

The data used in this study includes the intraday data from transaction prices and volumes and also the data related to limit orders available on the Limit Order Book (LOB). Over the inhomogeneous time series, a 15-second grid was imposed to get homogeneous ones with a regular spacing from 9:03:30 to 12:30:00 pm using the previous tick method described in more detail in [16]. The reason for using the previous tick approach to linear interpolation is that the linear interpolation method uses future information. However, the previous tick method relies solely on information up to the present.

Table 2. List of stocks from the Tehran Stock Exchange used in this study

Company Name	English Symbol	Industry
Mobarakeh Steel	FOLD	Base Metals
INC Ind.	MSMI	Base Metals
Parsian Oil & Gas	PASN	Chemicals
Iran Khodro	IKCO	Automotive
Metals and Min	MADN	Metal Ore
Ir. Inv.Petr	IPTR	Chemicals
Saipa Inv	SSAP	Financing

PCA method

The principal component analysis method was first developed by Pearson (1901) to investigate the relationship between several variables and reduce their complexity. In this method, the variables in a multi-state correlated environment are summarized as a set of uncorrelated variables that can explain the dynamics of these variables. The obtained uncorrelated components are called principal components, each derived from the linear combination of n main variable.

$$PC_1 = b_{11} \cdot m_{11} + b_{12} \cdot m_{12} + \dots + b_{1n} \cdot m_{1n} \quad (27)$$

The first principal component, PC_1 , as shown in Eq. 27, explains the highest amount of data dispersion in the entire dataset. Also, the coefficients b_{1j} are the elements of Eigenvector b_1 , which is the Eigenvector with the highest eigenvalue of the variance-covariance matrix. When the first principal component is found, the calculations continue to find the second principal component. This component has two important features: first, it contains the largest variance of the data set that has not yet been calculated; second, it is uncorrelated with the first component. Other components extracted in this method also have the above two characteristics.

So far, this method has been developed in several ways to produce uncorrelated variables; for example, Baradaran et al. [17], considered the weights of initial criteria as well as their coefficients to determine the direction of the new components. Another noteworthy point is that the principal components can be extracted using the main dataset, and in case of lack of access to the main data, the variance-covariance matrix can be used. Also, correlation matrices can be used when variables have different measure units or there are different variables with different variances in the dataset, as is the case with this study.

Our goal in using this method is to reduce the size of the problem. To do this, because our liquidity proxies have different units of measurement, we use the correlation matrix as input data. After performing this analysis and obtaining outputs, to reduce the number of principal components, we consider only those with eigenvalues greater than one.

Stock ranking

There are different methods for ranking a set of items. For example, Peykani et al. [18] used the DEA technique to rank 18 stocks in the insurance industry based on certain parameters, such as stock liquidity. Here, two ranking methods are described using the averages of the liquidity measures. The first one simply ranks the stocks with the value of means obtained for the liquidity measures, and another method is TOPSIS (Technique of Order Preference Similarity to the Ideal Solution) which will be discussed later.

Simple ranking

In this method, first, for a liquidity measure, the means of stocks are compared with each other, and for those measures whose increase leads to an increase in liquidity, different stocks are ranked based on the greatness of their means. It means that the stock with the highest average in that specific measure is ranked first. Also, for measures inversely related to liquidity, the stock with the smallest value in that measure is ranked first. The same procedure is repeated for other liquidity measures, and eventually, the average of the ranks assigned to stock for different measures is calculated as its final rank.

TOPSIS

In recent decades, researchers have turned their attention to multiple criteria models for making complex decisions. There are several methods in this field, the most widely used of which is the TOPSIS method, presented by Hwang and Yoon (1981). TOPSIS is a technique for ranking and selecting some externally determined alternatives through distance measures [19]. There are many researches in various fields that have used this method to rank their desired criteria; for example, Sobhanifard and Shahraki [20], used a two-stage TOPSIS method with the combination of the neural network model and Monte Carlo simulation to analyze and compare the efficiency of banks. One of the essential advantages of this method is that the criteria can have different measurement units with a positive or negative nature. The TOPSIS process is carried out as follows:

Step 1. Create a decision matrix (D) consisting of m alternatives (stocks) and n criteria (liquidity measures): $D = [d_{ij}]_{m \times n}$

Step 2. Determine the weights of the criteria. There are generally two ways for weighing the criteria: quantitative method and qualitative method. Quantitative methods include the Shannon entropy method, eigenvalue vector, least squares, and logarithmic least squares. These methods are based on the decision matrix data. In qualitative methods, the criteria are weighed using a survey of experts in decision-making. In this study, due to having a decision matrix, the Shannon entropy method, which is quantitative, has been used to weigh the liquidity measures. The steps of this method are as follows:

1- Normalize the decision matrix (P):

$$P = [p_{ij}]_{m \times n} \quad , \quad p_{ij} = \frac{d_{ij}}{\sum_{i=1}^m d_{ij}} \quad j = 1, 2, \dots, n \quad \forall i, j \quad (28)$$

2- Calculate the entropy (E)

$$E = [e_j]_{1 \times n} \quad , \quad e_j = -\frac{1}{\ln(m)} \sum_{i=1}^m p_{ij} \ln(p_{ij}) \quad \forall j \quad (29)$$

3- Calculate the degree of diversification (DD), which states how much helpful information each criterion provides to the decision-maker.

$$DD = [dd_j]_{1 \times n} \quad , \quad dd_j = 1 - e_j \quad \forall j \quad (30)$$

4- Compute the weight of each liquidity measure (w_j):

$$W = [w_j]_{1 \times n} \quad , \quad w_j = \frac{dd_j}{\sqrt{\sum_{j=1}^n dd_j}} \quad \forall j \quad (31)$$

Step 3. Normalize the decision matrix (R), using the normalization method:

$$X = [x_{ij}]_{m \times n} \quad , \quad x_{ij} = \frac{r_{ij}}{\sqrt{\sum_{i=1}^m d_{ij}^2}} \quad j = 1, 2, \dots, n \quad \forall i, j \quad (32)$$

Step 4. Calculate the weighted normalized decision matrix (Y):

$$Y = [y_{ij}]_{m \times n} \quad , \quad \forall i, j: \quad y_{ij} = x_{ij} \times |w_j| \quad (33)$$

Step 5. Determine the best alternative ($A^+ = [a_j^+]_{1 \times n}$) and the worst alternative ($A^- = [a_j^-]_{1 \times n}$):

$$A^+ = \left\{ \left(\min (y_{ij} \mid i = 1, 2, \dots, m) \mid j \in J_- \right), \left(\max (y_{ij} \mid i = 1, 2, \dots, m) \mid j \in J_+ \right) \right\} \equiv \{ a_j^+ \mid j = 1, 2, \dots, n \} \quad (34)$$

$$A^- = \left\{ \left(\max (y_{ij} \mid i = 1, 2, \dots, m) \mid j \in J_- \right), \left(\min (y_{ij} \mid i = 1, 2, \dots, m) \mid j \in J_+ \right) \right\} \equiv \{ a_j^- \mid j = 1, 2, \dots, n \} \quad (35)$$

Where,

$J_+ = \{j = 1, 2, \dots, n \mid j\}$ associated with the criteria having a positive weight ($w_j > 0$) and,

$J_- = \{j = 1, 2, \dots, n \mid j\}$ associated with the criteria having a negative weight ($w_j < 0$).

Step 6. Calculate the Euclidean distance between the target alternative i and the best alternative (D^+) and also the distance between the target alternative i and the worst alternative (D^-):

$$D^+ = [d_i^+]_{m \times 1} \quad , \quad d_i^+ = \sqrt{\sum_{j=1}^n (y_{ij} - a_j^+)^2} \quad \forall i \quad (36)$$

$$D^- = [d_i^-]_{m \times 1} \quad , \quad d_i^- = \sqrt{\sum_{j=1}^n (y_{ij} - a_j^-)^2} \quad \forall i \quad (37)$$

Step 7. Calculate the similarity to the worst condition (C):

$$C = [c_i]_{m \times 1} \quad , \quad c_i = \frac{d_i^-}{d_i^- + d_i^+} \quad \forall i \quad (38)$$

Step 8. Compute the score of each alternative (S):

$$S = [s_i]_{m \times 1} \quad , \quad s_i = \frac{c_i}{\sum_{i=1}^m c_i} \quad \forall i \quad (39)$$

Step 9. Rank the alternatives according to s_i . Since each alternative's score is obtained based on the distance from the worst condition, and consequently, the greater the distance, the better the score, so the alternative with the highest score is ranked first.

Results and discussion

In this section, the results of correlation analysis and principal component analysis are presented; and finally, the intraday patterns are extracted.

Correlation analysis

In this section, the correlogram graphs between 27 different liquidity measures for different stocks are extracted to visualize the block structure of the correlations. Fig. 1 shows these graphs for each stock, which are calculated using all dataset information. In these graphs, the blue circles show a positive correlation, and the red circles show a negative correlation. Also, the color intensity of these circles is proportional to the magnitude of the correlation coefficient between the two measures. By paying close attention to the correlation matrix structure, we can see that the correlation matrix can be decomposed into multiple blocks.

As can be seen, three main blocks are formed in this matrix. In the larger block, 11 measures related to spread and quote slope have a high positive correlation. In the next block, the five market depth measures have relatively strong linear relationships with each other. There are four measures related to the volume and number of transactions in the smaller block, and some of them are highly correlated with each other.

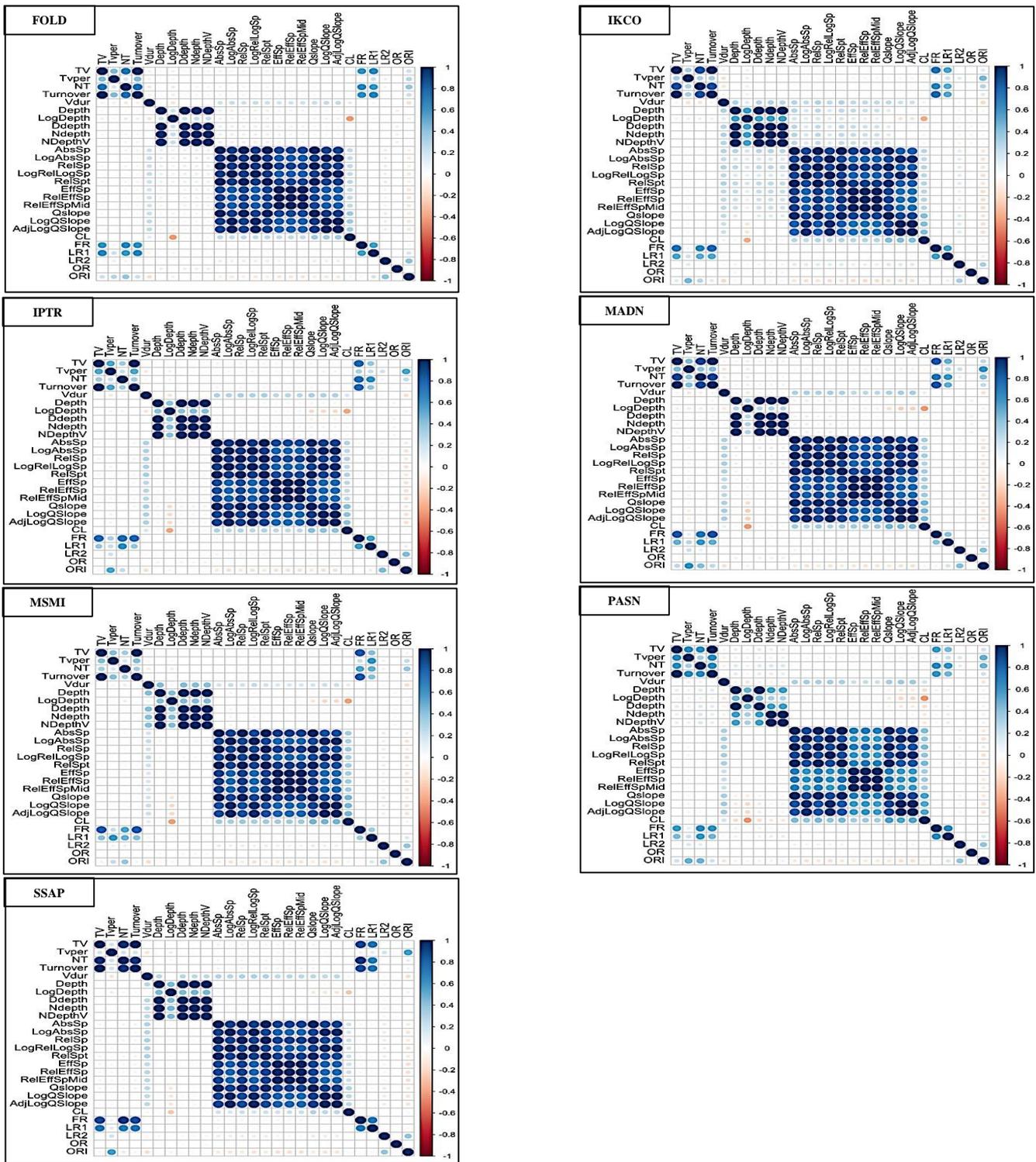


Fig.1. Correlation matrix of liquidity measures for each stock from September 22, 2016, to February 18, 2017

Reduction of the number of Liquidity Measures

According to the results of previous sections, the following ten measures are excluded from the study:

- 1- Trading Volume: It is highly correlated with turnover, and the latter measure is comparable among different stocks with different prices.

2- Number of Trades per Time Unit: It is directly used to calculate the Volume Duration measure. So, there is no need to study it separately.

3- Market Depth: It is highly correlated with Dollar Depth, and the latter measure is comparable across different stocks with different prices. Therefore, market depth is dropped.

4- Near depth: is highly correlated with Near Depth Value, and once again, because the latter measure is comparable among different stocks with different prices, it is retained.

5- Dollar Depth: It is highly correlated with Near Depth Value which is more precise due to using all three levels of limit order book data. So, it will be kept.

6- Absolute Spread: It is highly correlated with many spread measures, and because relative spreads are more comparable across stocks, Absolute Spread is eliminated.

7- Relative Spread with Last Trade: It is perfectly correlated with Relative Spread with Mid Price, and the latter is easier to calculate because of not depending on the last trade information, which may not always be available. Therefore, Relative Spread with Last Trade is dropped from the sample.

8- Effective Spread: It is highly correlated with Effective Spread with Last Trade, and once again, since Effective Spread with Last Trade is comparable among different stocks with different prices, it is retained.

9- Effective Spread with Last Trade: It is perfectly correlated with Effective Spread with Mid Price, which is easier to calculate.

10- Log Quote Slope: It is perfectly correlated with Adjusted Log Quote Slope, and the latter is a more comprehensive measure. So, Log Quote Slope is removed from further studies.

Therefore, after removing the last ten proxies, in the continuation of this research, PCA is performed using the following 17 measures:

- Volume-related liquidity measures: TVper, Turnover, Vdur
- Depth-related liquidity measures: NDepthV, LogDepth
- Spread-related liquidity measures: LogAbsSp, LogRelLogSp, RelSp, RelEffSpMid
- Multi-dimensional Liquidity Measures: QSlope, AdjLogQSlope, CL, LR1, LR2, FR, ORI, OR.

Principal component analysis

After removing ten measures in the last section, the principal components analysis will be performed on the remaining 17 measures. The output contains the eigenvalues, eigenvector coefficients, and percentages of variance explained by each component, arranged in descending order. For all the stocks, it is observed that only five principal components in each stock have eigenvalues greater than one, which explain about 70 to 75 percent of the total variance.

The results of this analysis include the eigenvalue, the percentage of variance explained, and the cumulative variance for each principal component and per stock. These for each stock are shown in [Table 3](#). As can be seen, the first five eigenvalues, which are greater than one, can explain more than 70% of the total variance.

Table 3. Principal Component Analysis of 17 liquidity measures for each stock

Stock	Output variables	b_1	b_2	b_3	b_4	b_5
FOLD	Eigenvalue	5.66	2.44	1.51	1.40	1.04
	Var. explained (%)	33.32	14.32	8.91	8.25	6.14
	Cum. Var. explained (%)	33.32	47.65	56.55	64.80	70.94
IKCO	Eigenvalue	5.61	2.36	1.82	1.54	1.00
	Var. explained (%)	33.02	13.86	10.69	9.03	5.90
	Cum. Var. explained (%)	33.02	46.88	57.57	66.60	72.50
IPTR	Eigenvalue	5.70	2.47	1.59	1.51	1.01
	Var. explained (%)	33.51	14.54	9.34	8.91	5.97
	Cum. Var. explained (%)	33.51	48.05	57.39	66.30	72.27
MADN	Eigenvalue	5.71	2.52	1.55	1.48	1.02
	Var. explained (%)	33.58	14.81	9.12	8.72	6.01
	Cum. Var. explained (%)	33.58	48.39	57.52	66.23	72.24
MSMI	Eigenvalue	5.635	2.423	1.701	1.426	1.075
	Var. explained (%)	33.14	14.26	10.00	8.39	6.33
	Cum. Var. explained (%)	33.14	47.40	57.40	65.79	72.12
PASN	Eigenvalue	5.47	2.49	1.50	1.47	1.01
	Var. explained (%)	32.18	14.65	8.83	8.64	5.95
	Cum. Var. explained (%)	32.18	46.82	55.65	64.29	70.24
SSAP	Eigenvalue	5.61	2.57	1.75	1.58	1.01
	Var. explained (%)	33.01	15.11	10.29	9.30	5.93
	Cum. Var. explained (%)	33.01	48.12	58.41	67.71	73.64

The eigenvectors related to each of these five components -denoted by b_j ; $j = 1, \dots, 5$ - are plotted and shown in [Fig. 2](#) for SSAP as an example.

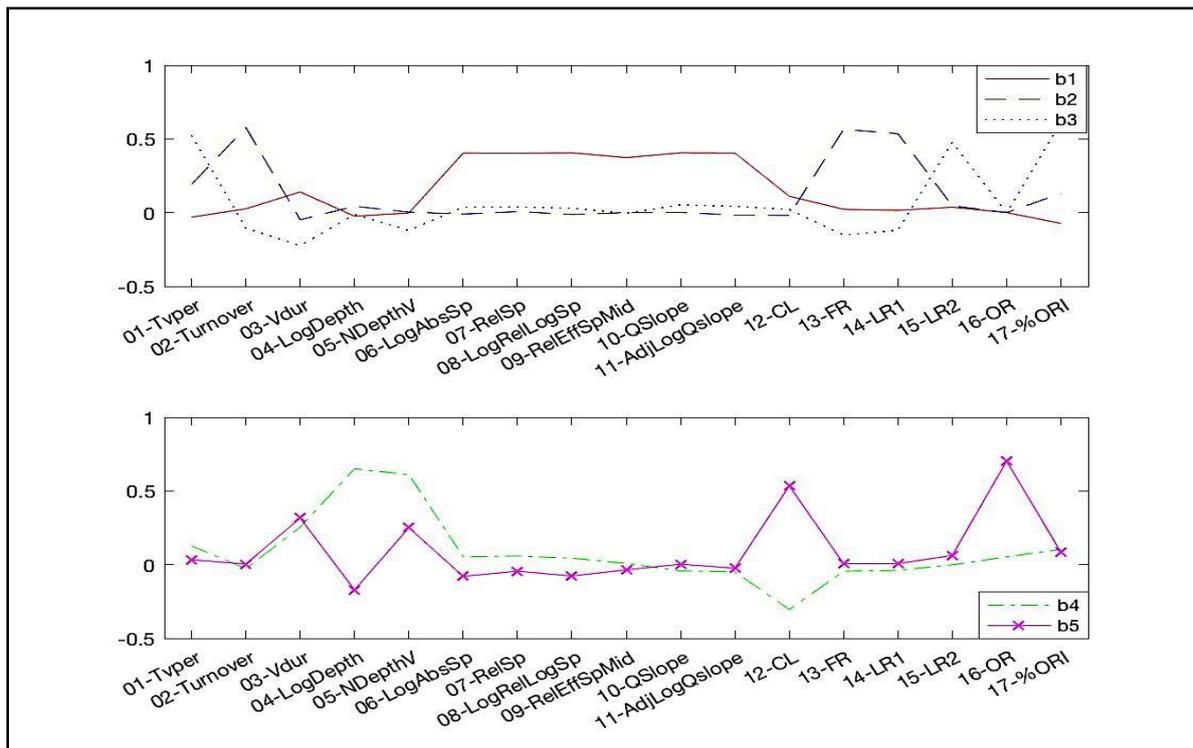


Fig. 2. First five eigenvectors from Principal Component Analysis of 17 liquidity measures for SSAP

Based on the principal component analysis method performed on the seven stocks, the following results are obtained:

1- One factor explains spread-related liquidity measures. This component with the highest percentage of variance explaining across all stocks - about 33% - covers the width dimension of liquidity. From this group, the relative Spread with mid-price is chosen due to the ease of calculation.

2- A second factor captures liquidity measures related to volume and the number of transactions. As a result, this component explains the volume and timing dimension of liquidity with a variance of about 14%. The turnover measure is considered representative of this group because of the ease of comparability between different stocks.

3- The next principal component explains the measures related to the market depth, such as Log Depth. This factor then describes the depth dimension of liquidity, explaining about 10% of the variance. The near depth value measure is chosen from this group because of using the information on all three Limit order book levels.

4- The fourth component shows the measure related to the market resiliency dimension, such as order imbalance and the liquidity ratio2, explaining about 9% of the variance. The Liquidity Ratio 2 is considered to represent this group because of its more superficial interpretation.

Therefore, by choosing one proxy from each group, all aspects of liquidity can be measured by these four selected measures without deleting any helpful information.

Intraday patterns

In this section, the Intraday pattern is extracted for the four selected liquidity measures, each chosen as a representative of one of the liquidity dimensions. Each graph is plotted based on the related liquidity measure's intraday data, which is homogenized at 15-second intervals. For example, for the average graph, which is drawn in red, the value obtained at each time point is the average data of 77 days at that point of time.

Relative Spread with Mid Price

This measure is one of the liquidity measures that can be seen in most studies because it is simple to calculate, and due to the use of the mid-price of quotes, it makes different stocks comparable with each other. Another advantage of this measure is that it does not use transaction data in the calculation, so there is no need for a transaction to have taken place. Various patterns, including U-shaped and inverted S-shaped patterns, have been reported in the literature for this measure. For instance, in a 2015 review article by Kumar [21], it is reported that Spread on the Shanghai and Istanbul stock exchanges has an inverse J-shaped behavior. While on the New York and Toronto stock exchanges, this pattern is U-shaped. This difference in spread behavior in different markets can be attributed to differences in their market structure and, consequently, differences in traders' behavior.

Here, as shown in Fig. 3, this measure has an inverted J-shaped pattern for most stocks, especially for IPTR and SSAP. It means that the Spread is very high at the beginning of the day, but over time and with the increase in market depth, this measure falls and rises slightly at the end of the day.

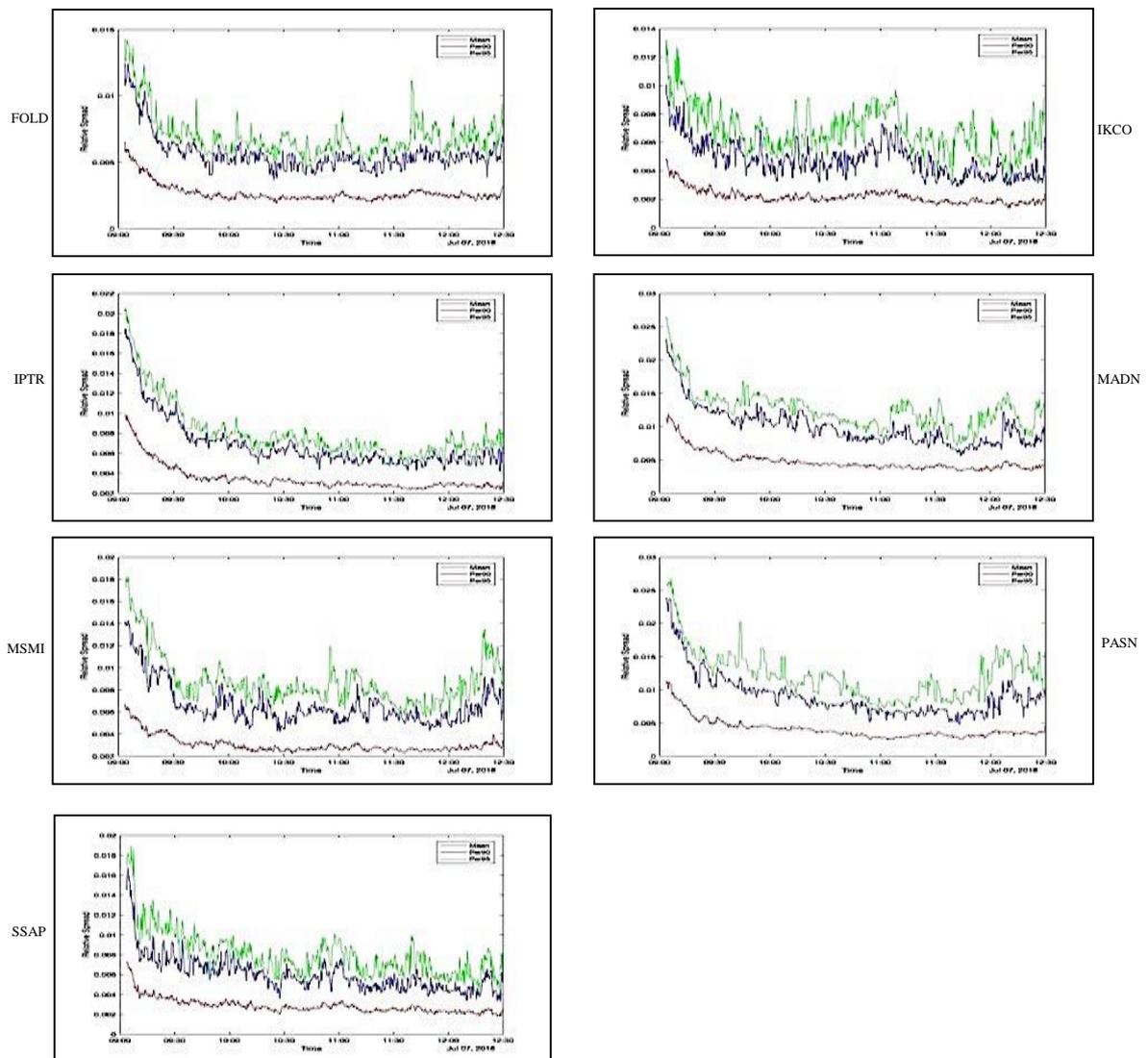


Fig. 3. Intraday patterns for Relative Spread with Mid Price

Turnover

This measure is calculated using transaction data in all available ticks over time. In addition to the trading volume, the transaction price is also included in the calculations, making it possible to compare liquidity between different stocks.

For this measure, in most studies, the U-shaped pattern is obtained. Krishnan [22], for example, examined the patterns of some liquidity proxies, using intraday data on the Indian Stock Exchange. They conclude that this measure shows a U-shaped behavior. A large amount of trading volume at the beginning and end of the market can be attributed to traders who take advantage of liquidity instability during these periods. Nevertheless, as can be seen in Fig. 4, only IKCO partially follows this pattern. Also, the stocks FOLD and MADN are almost J-shaped. There was no clear pattern in the SSAP stock case due to the sharp volatility of trading volume over the 15-second intervals. Finally, IPTR, MSMI, and PASN also fluctuate slightly above zero due to the low volume of trades in these short intervals, which indicates that these stocks are not liquid enough.

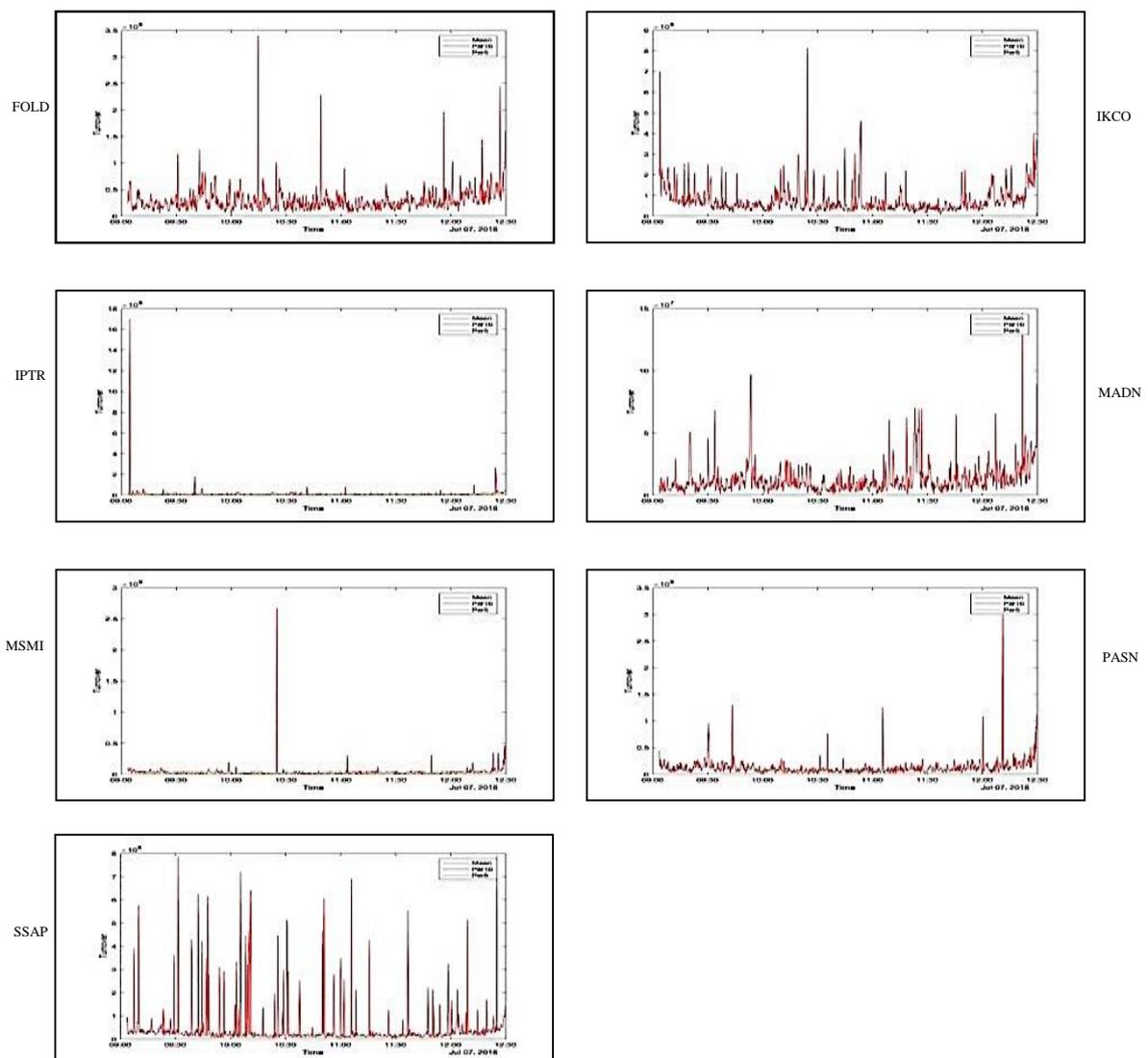


Fig. 4. Intraday patterns for turnover

Near depth value

In this measure, using all three levels of the limit order book data, the order prices are considered in addition to order quantities. As shown in Fig. 5, various patterns for this measure

have been obtained in different stocks, among which the FOLD and IPTR stocks have the S-shaped pattern. It means that at the beginning of the day, the market depth is low and relatively high at the end of the day. SSAP and IKCO have inverted U-shaped patterns.

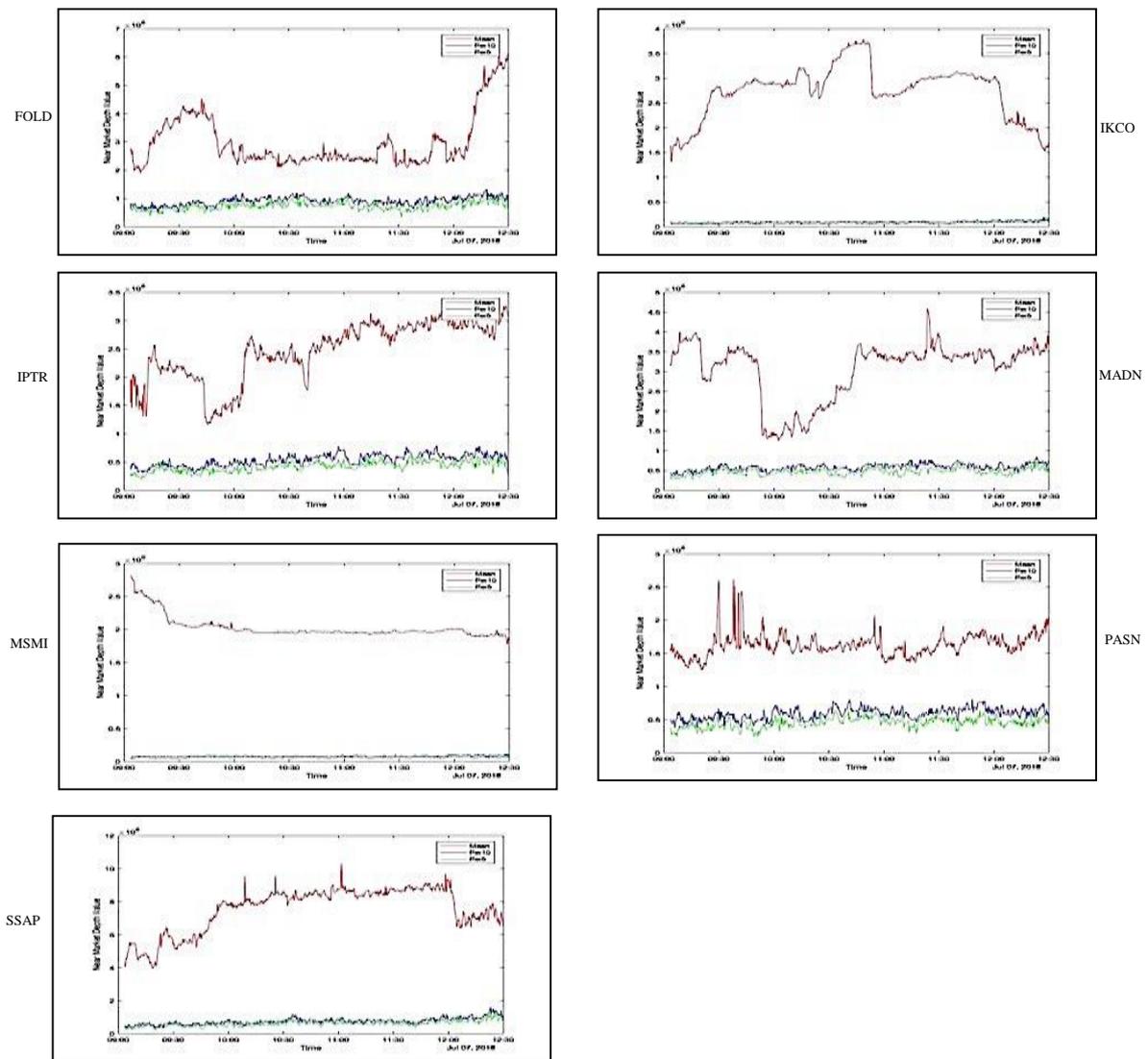


Fig. 5. Intraday patterns for Near Depth Value

Liquidity ratio 2

This ratio indicates the average percentage of price changes after each transaction. While the liquidity ratio 1 depends only on the absolute magnitude of changes in the price of a stock, in liquidity ratio 2, this problem is overcome by placing the number of transactions in the denominator. Also, the larger the value of this measure, the lower the stock's liquidity because it means that the average price changes over time are high. If the number of transactions in a given time interval is zero, this ratio is considered equal to zero.

As can be seen in Fig. 6, the mean graph for all stocks is slightly above zero. This is due to the nature of this measure which is highly dependent on trading volume, and it seems that the 15-second time interval is not long enough for a trade to take place. Therefore, no specific pattern is visible.

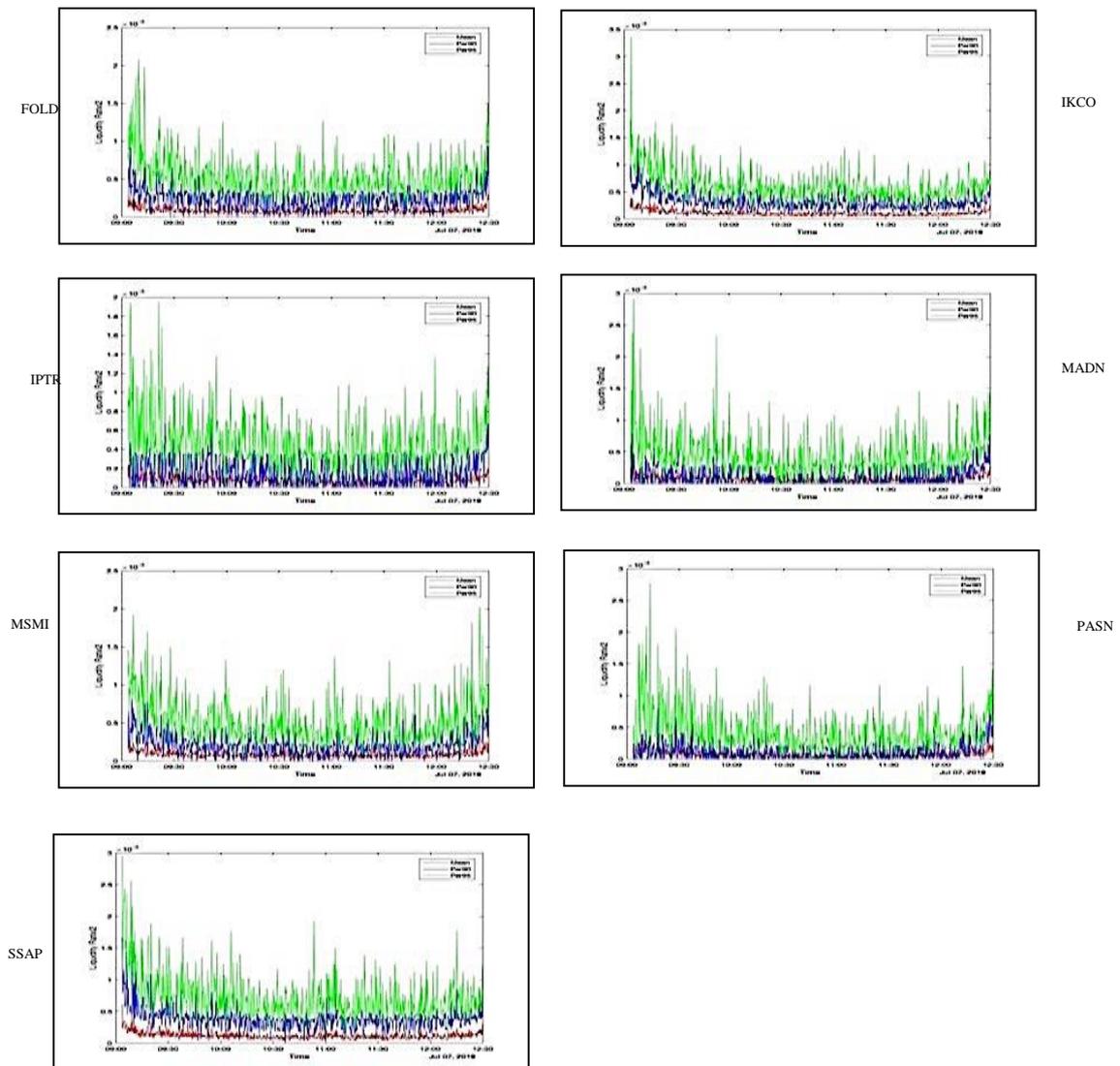


Fig. 6. Intraday patterns for Liquidity Ratio2

Stock ranking

This subsection presents the results of ranking the stocks based on 27 liquidity measures and using the two methods described in the previous section.

Simple ranking results

This ranking method, which works based on the means obtained for different liquidity measures, is shown in Table 4. Based on these ranking results, the order of different stocks in terms of having higher liquidity are SSAP, IKCO, FOLD, MSMI, IPTR, MADN, and PASN.

Table 4. Ranking of 7 stocks in terms of liquidity with the simple ranking method and using 27 liquidity measures

STOCK MEASURE	FOLD	IKCO	IPTR	MADN	MSMI	PASN	SSAP
TV	3	2	5	6	4	7	1
Typer	2	3	5	6	4	7	1
NT	3	1	5	6	4	7	2
Turnover	4	1	6	5	3	7	2
Vdur	2	1	6	7	4	5	3
Depth	6	2	5	4	1	7	3
LogDepth	2	6	4	5	3	7	1
Ddepth	6	1	5	4	2	7	3
Ndepth	4	2	6	5	1	7	3
NDepthV	5	1	6	4	2	7	3
AbsSp	2	5	3	6	4	7	1
LogAbsSp	2	5	3	6	4	7	1
RelSp	2	1	5	7	4	6	3
LogRelLogSp	2	1	6	7	3	5	4
RelSpt	2	1	5	7	4	6	3
EffSp	2	5	3	6	4	7	1
RelEffSp	2	1	5	7	4	6	3
RelEffSpMid	2	1	5	7	4	6	3
Qslope	2	5	3	6	4	7	1
LogQSlope	2	5	3	6	4	7	1
AdjLogQSlope	2	5	3	6	4	7	1
CL	2	1	6	7	3	5	4
FR	5	2	4	6	3	7	1
LR1	4	1	7	6	2	5	3
LR2	5	6	1	3	4	2	7
OR	5	6	3	2	7	1	4
%ORI	5	7	3	1	4	2	6
Average	3.15	2.89	4.38	5.48	3.48	5.96	2.56
Rank	3	2	5	6	4	7	1

The different rankings of the 27 liquidity measures are investigated for their correlation using the Spearman rank correlation, r_k :

$$r_k = 1 - \frac{6.R}{n.(n^2-1)} \quad (40)$$

R denotes the sum of the squared differences in ranks: $\sum_{i=1}^n di_i^2$. n is the number of observations which is seven in our case. To determine whether the correlations are different from zero, they are compared to the table of critical values for the Spearman rank correlation test. Table 5 shows the rank correlations produced by 27 liquidity measures for the seven stocks. There seem to be two different groups of liquidity measures that produce high correlations among the rankings of the different stocks:

Group one consists of trading volume, depth, log depth, the number of trades, the absolute and the log absolute spread, the effective Spread, the quote slope, the log quote slope, and the adjusted log quote slope. In general, these are the liquidity measures that depend on the absolute stock price. The second group, which leads to a similar stock ranking, consists of turnover, dollar depth, relative spread measures, and composite liquidity. In general, these measures are independent of the stock price.

Table 5. Spearman rank correlations after different liquidity measures. ***/*** denotes significance on a level of 10%/5%/1%

TOPSIS ranking results

In this method, based on the explanations provided in the previous section, each of the

	TV	Tvper	NT	Turnover	Vdur	Depth	LogDepth	Ddepth	Ndepth
TV	1								
Tvper	0.96***	1							
NT	0.96***	0.89***	1						
Turnover	0.89***	0.78**	0.93***	1					
Vdur	0.78**	0.75**	0.86**	0.75**	1				
Depth	0.53	0.39	0.57*	0.78**	0.286	1			
LogDepth	0.64*	0.78**	0.46	0.39	0.25	0.25	1		
Ddepth	0.60*	0.42	0.679*	0.86**	0.39	0.96***	0.14	1	
Ndepth	0.71**	0.64*	0.75**	0.89***	0.607*	0.89***	0.43	0.86**	1
NDepthV	0.67*	0.53	0.75**	0.93***	0.54	0.93***	0.21	0.96***	0.93***
AbsSp	0.75**	0.85**	0.61*	0.429	0.39	0.18	0.93***	0.14	0.36
LogAbsSp	0.75**	0.85**	0.61*	0.429	0.39	0.18	0.93***	0.14	0.36
RelSp	0.85**	0.82**	0.93***	0.78**	0.96***	0.36	0.36	0.46	0.64*
LogRelLogSp	0.67*	0.64*	0.78**	0.71**	0.96***	0.36	0.18	0.43	0.68*
RelSpt	0.85**	0.82**	0.93***	0.78**	0.96***	0.36	0.36	0.46	0.64*
EffSp	0.75**	0.85**	0.61*	0.43	0.39	0.18	0.93***	0.14	0.36
RelEffSp	0.85**	0.82**	0.93***	0.78**	0.96***	0.36	0.36	0.46	0.64*
RelEffSpMid	0.85**	0.82**	0.93***	0.78**	0.96***	0.36	0.36	0.46	0.64*
Qslope	0.75**	0.85**	0.60*	0.43	0.39	0.18	0.93***	0.14	0.36
LogQSlope	0.75**	0.85**	0.61*	0.43	0.39	0.18	0.93***	0.14	0.36
AdjLogQSlope	0.75**	0.85**	0.61*	0.43	0.39	0.18	0.93***	0.14	0.36
CL	0.67*	0.64*	0.78**	0.71**	0.96***	0.36	0.18	0.43	0.68*
FR	0.89***	0.78**	0.86**	0.86**	0.57*	0.75**	0.54	0.78**	0.75**
LR1	0.68*	0.57*	0.75**	0.86**	0.82**	0.68*	0.14	0.71**	0.86**
LR2	-0.89	-0.86	-0.86	-0.89	-0.79	-0.50	-0.50	-0.57	-0.71
OR	-0.68	-0.64	-0.75	-0.79	-0.68	-0.75	-0.43	-0.71	-0.93
%ORI	-0.93	-0.86	-0.96	-0.86	-0.93	-0.46	-0.39	-0.57	-0.68
	NDepthV	AbsSp	LogAbsSp	RelSp	LogRelLogSp	RelSpt	EffSp	RelEffSp	RelEffSpMid
NDepthV	1								
AbsSp	0.18	1							
LogAbsSp	0.18	1.00***	1						
RelSp	0.57*	0.54	0.54	1					
LogRelLogSp	0.57*	0.29	0.29	0.93***	1				
RelSpt	0.57*	0.54	0.54	1.00***	0.93***	1			
EffSp	0.18	1.00***	1.00***	0.54	0.29	0.54	1		
RelEffSp	0.57*	0.54	0.54	1.00***	0.93***	1.00***	0.54	1	
RelEffSpMid	0.57*	0.54	0.54	1.00***	0.93***	1.00***	0.54	1.00***	1
Qslope	0.18	1.00***	1.00***	0.54	0.29	0.54	1.00***	0.54	0.54
LogQSlope	0.18	1.00***	1.00***	0.54	0.29	0.54	1.00***	0.54	0.54
AdjLogQSlope	0.18	1.00***	1.00***	0.54	0.29	0.54	1.00***	0.54	0.54
CL	0.57*	0.29	0.29	0.93***	1.00***	0.93***	0.286	0.93***	0.93***
FR	0.75**	0.64*	0.643*	0.68*	0.500	0.68*	0.64*	0.68*	0.67*
LR1	0.82**	0.14	0.14	0.75**	0.86**	0.75**	0.14	0.75**	0.75**
LR2	-0.71	-0.50	-0.50	-0.75	-0.68	-0.75	-0.50	-0.75	-0.75
OR	-0.79	-0.43	-0.43	-0.75	-0.79	-0.75	-0.43	-0.75	-0.75
%ORI	-0.64	-0.57	-0.57	-0.96	-0.86	-0.96	-0.57	-0.96	-0.96
	Qslope	LogQSlope	AdjLogQSlope	CL	FR	LR1	LR2	OR	%ORI
Qslope	1								
LogQSlope	1.00***	1							
AdjLogQSlope	1.00***	1.00***	1						
CL	0.29	0.29	0.29	1					
FR	0.64*	0.64*	0.64*	0.50	1				
LR1	0.14	0.14	0.14	0.86**	0.64*	1			
LR2	-0.50	-0.50	-0.50	-0.68	-0.71	-0.66	1		
OR	-0.43	-0.43	-0.43	-0.79	-0.68	-0.85	0.57	1	
%ORI	-0.57	-0.57	-0.57	-0.86	-0.82	-0.85	0.82**	0.71*	1

liquidity measures is weighed first using the Shannon entropy method and the decision matrix data, which are the averages of different liquidity measures. These data are normalized and weighted, and then the best alternative and the worst alternative for each measure are calculated. In the next step, the Euclidean distance of each liquidity measure from the best and the worst condition is calculated. Finally, the stocks are ranked according to having the greatest distance

from the worst alternative. The result obtained after applying this method is as follows: SSAP, IKCO, MSMI, MADN, PASN, IPTR, and FOLD.

As can be seen, the result of the ranking using the TOPSIS method is almost the same as the previous method, with two differences: First, FOLD, despite having favorable liquidity characteristics compared to MADN, PASN, and IPTR, is ranked lower than these stocks. Second, PASN, despite having the weakest statistical results - which was examined in the previous sections - is located in fifth place. Therefore, after investigating, it was found that in the Shannon entropy method, the weight of Order Ratio (OR) measure is relatively high and equals 8.5% of the total weights among 27 liquidity measures. This liquidity measure indicates an imbalance in the volume of buying and selling orders, and its small amounts indicate high liquidity. As shown in Table 1, The denominator of this measure includes the turnover. When no transaction has taken place at time intervals, and therefore, the amount of turnover is zero, the value of OR measure is considered zero too contractually. It reduces the OR measure for illiquid stocks with low turnovers, such as MADN and PASN, and mistakenly shows the high liquidity for these stocks. Therefore, by removing this measure, weighting and ranking were done again, and more logical results were obtained, which are closely similar to the results of the simple ranking method. The ranking is now as follows: SSAP, IKCO, MSMI, FOLD, IPTR, MADN, and PASN. The result of these two methods is shown in Table 6.

Table 6. Ranking of 7 stocks in terms of liquidity with different methods

Rank	TOPSIS method without OR measure	TOPSIS method with 27 liquidity measures	Simple Ranking method
1	SSAP	SSAP	SSAP
2	IKCO	IKCO	IKCO
3	MSMI	MSMI	FOLD
4	FOLD	MADN	MSMI
5	IPTR	PASN	IPTR
6	MADN	IPTR	MADN
7	PASN	FOLD	PASN

Conclusions

The main purpose of this research was to conduct a comprehensive study on the concept of stock liquidity and how it occurs in the Tehran Stock Exchange. For this purpose, in the first step, seven stocks were selected from the top 50 stocks of this market. Then the intraday data of transactions and the limit orders related to 77 working days were collected, and the initial preprocessing of this data was done in Excel. The second step was to review the various aspects of liquidity that traditionally include transaction time, depth, breadth, and resiliency. These aspects can be measured by a wide range of liquidity proxies. After reviewing the existing literature in this area, 27 liquidity measures were selected.

In the next step, the investigation of behavioral similarities and correlations between different liquidity measures was targeted. For this purpose, the graphical structure of the correlation matrix based on the total data of 77 days was extracted separately for each stock in R software. It was observed that there are several sub-blocks in the correlation matrix. This analysis led to the elimination of 10 proxies that had a high correlation with others but were dropped due to inappropriate distribution and some other defects. Therefore, this analysis results helped us reduce the number of liquidity measures without losing any useful information.

In the next step, using the PCA technique, the correlation matrix of different stocks was given as input to MATLAB, and the eigenvalues, coefficients of eigenvectors, and variance explained by each component were received as output. According to the results obtained from

PCA and identifying four principal components explaining the highest amount of variance, it was observed that stock liquidity in the Tehran Stock Exchange could be measured by a limited number of liquidity measures that include all four dimensions of liquidity. The selected measures by the researcher were "Relative Spread with Mid Price", "Turnover", "Near Depth Value," and "Liquidity Ratio 2", which respectively include the breadth, time, depth, and resiliency aspects of liquidity. Based on these results, the liquidity measures describing market breadth – spreads and quote slopes - explain 33% of the total variance through the first component; Therefore, these proxies can be considered essential microstructural elements affecting market liquidity.

Using high-frequency data, we can assess the intraday patterns of liquidity measures more accurately. To this end, intraday patterns of four selected measures were extracted. For the RelSp measure, the observed pattern was inverted J-shaped, indicating that stock liquidity is low at the beginning of the day but increases over time and decreases again at the end of the day. So, in the middle of the day, the difference between buying and selling quotes is minimized, and favorable conditions are provided for trading. For the NDepthV measure, which represents market depth, various patterns were observed, the most repetitive of which is the S-shaped pattern, which states that market depth is low at the beginning of the day and high at the end of the day. In the other two selected measures, namely Turnover and LR2, no specific pattern was observed because the 15-second time interval for homogenizing time series was relatively short for these volume-related measures and in about half of the time points, no transaction occurred during this short period. Therefore, to achieve more accurate results, it is recommended to consider a longer time interval for this category of measures.

In the final part of this study, the stocks were ranked using two methods. At first, the result of the TOPSIS method was not compatible with the simple ranking method. After removing a liquidity measure, which falsely showed high liquidity for some stocks, the results were almost the same, and the first rank was given to the SSAP and the last rank to the PASN.

The feature reduction performed in this study can be useful in practice to reduce complexity for both investors and regulators to make the decision more easily based on a comprehensive method. Also, the obtained intraday patterns can be helpful for high-frequency traders to be aware of the liquidity behavior during a trading day and choose the best hours for making their trade. As it was observed, the liquidity was higher in the middle hours of the day, which is consistent with the findings of previous studies such as the results reported in two references [21,22]. Finally, the results of stock ranking in terms of liquidity can be useful for individuals to invest in stocks with higher liquidity.

To use the results of the present study in the future, we propose to develop a model predicting the selected liquidity measures and including all dimensions of liquidity. Intraday data related to specific time periods can also be used to examine the impact of critical economic and political conditions on various liquidity measures.

References

- [1] Quah, H., Haman, J., and Naidu, D. (2021). The effect of stock liquidity on investment efficiency under financing constraints and asymmetric information: Evidence from the United States. *Accounting and Finance*, 61, 2109-2150.
- [2] Ranaldo, A. (2000). Intraday trading activity on financial markets: The Swiss evidence (Doctoral dissertation, Université de Fribourg).
- [3] Jain, P. K. (2003). Institutional design and liquidity at stock exchanges around the world. Available at SSRN 869253.
- [4] Gould, M. D., Porter, M. A., Williams, S., McDonald, M., Fenn, D. J., and Howison, S. D. (2013). Limit order books. *Quantitative Finance*, 13(11), 1709-1742.

- [5] Cui, W., Hon, M. E., and Brabazon, A. (2012). An Empirical Investigation of Price Impact: An Agent-based Modelling Approach (Doctoral dissertation, University College Dublin).
- [6] Korolev, V. Y., Chertok, A. V., Korchagin, A. Y., and Zeifman, A. I. (2015). Modeling high-frequency order flow imbalance by functional limit theorems for two-sided risk processes. *Applied Mathematics and Computation*, 253, 224-241.
- [7] Muranaga, J., and Shimizu, T. (1999). Market microstructure and market liquidity. Bank of Japan.
- [8] Gabrielsen, A., Marzo, M., and Zagaglia, P. (2011). Measuring market liquidity: An introductory survey.
- [9] Salighehdar, A., Liu, Y., Bozdog, D., and Florescu, I. (2017). Cluster analysis of liquidity measures in a stock market using high frequency data. *Journal of Management Science and Business Intelligence*, 2(2), 1-8.
- [10] Będowska-Sójka, B. (2019). Commonality in liquidity measures. The evidence from the Polish stock market.
- [11] Johann, T., Scharnowski, S., Theissen, E., Westheide, C., and Zimmermann, L. (2019). Liquidity in the German stock market. *Schmalenbach Business Review*, 71(4), 443-473.
- [12] Minović, J., Stevanović, S., and Belopavlović, G. (2011). Survey of one-dimensional liquidity measures. *IASSIST quarterly*, 197-202.
- [13] Irvine, P. J., Benston, G. J., and Kandel, E. (2000). Liquidity beyond the inside spread: Measuring and using information in the limit order book. Available at SSRN 229959.
- [14] Von Wyss, R. (2004). Measuring and predicting liquidity in the stock market (Doctoral dissertation, Novidea di Luigi Hofmann).
- [15] Olbryś, J., and Mursztyn, M. (2017). Measurement of stock market liquidity supported by an algorithm inferring the initiator of a trade. *Operations Research and Decisions*, 27(4), 111-127.
- [16] Gençay, R., Dacorogna, M., Muller, U. A., Pictet, O., and Olsen, R. (2001). An introduction to high-frequency finance. Elsevier.
- [17] Baradaran, V., Kazem zadeh, R. B., Amiri, A. H., and Mogouie, H. (2012). A modified PCA approach for solving MADM problems with dependent criteria. *Advances in Industrial Engineering*, 46(2), 133-145
- [18] Peykani, P., Mohammadi, E., Rostamy-Malkhalifeh, M., and Hosseinzadeh Lotfi, F. (2019). Fuzzy data envelopment analysis approach for ranking of stocks with an application to Tehran stock exchange. *Advances in Mathematical Finance and Applications*, 4(1), 31-43.
- [19] Tsaour, R. C. (2011). Decision risk analysis for an interval TOPSIS method. *Applied Mathematics and Computation*, 218(8), 4295-4304.
- [20] Sobhanifard, F., and Shahraki, M. R. (2020). An Integrated Neural Networks and MCMC Model to Predicting Bank's Efficiency. *Advances in Industrial Engineering*, 54(1), 1-14.
- [21] Kumar, G., and Misra, A. K. (2015). Closer view at the stock market liquidity: A literature review. *Asian Journal of Finance and Accounting*, 7(2), 35-57.
- [22] Krishnan, R., and Mishra, V. (2012). Intraday Liquidity Patterns in Indian Stock Market Monash. Discussion Paper, Monash University, Malasya.



This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license.