Expected Bond Liquidity in Thailand Bond Market

Nirun Aramratanapan

School of Development Economics, National Institute of Development Administration, Bangkok, Thailand

Corresponding author: nirun.phd@gmail.com

Yuthana Sethapramote

School of Development Economics, National Institute of Development Administration, Bangkok, Thailand

Abstract

Liquidity in the stock market is crucial for investment decisions, as it enables investors and issuers to meet their needs for investment, financing, and hedging. This, in turn, reduces investment costs and the overall cost of capital, contributing to a more efficient market. The aim of this paper is to explore modern techniques using computer science to address liquidity challenges and develop machine learning and deep learning models for liquidity prediction, particularly in the bond market, where liquidity is a key issue. Liquidity plays a significant role in investment strategies and investor decisions in the securities market. It also affects bond prices and returns on investment through the cost of trading and liquidity by transferring bond ownership. The nature of liquidity in the bond market is complex—volatile, unpredictable, dynamic, and non-linear making it difficult to predict due to the many influencing factors, such as trading volume and value. This study focuses on the Thailand bond market from 2015 to 2024, characterized by limited market tightness, depth, and modest size. We introduce a forward-looking illiquidity measure, Expected Illiquidity (EI), represented by the bid-ask spread. To develop robust liquidity prediction models, we use Multilayer Perceptron, Mixed Deep Learning, Long Short-Term Memory (LSTM), Multiple Linear Regression, and Extreme Gradient Boosting (XGB) algorithms. These models are evaluated by comparing their out-of-sample forecasting errors to a naïve estimate using current daily illiquidity. The XGB models demonstrate superior predictive capabilities, with the lowest Mean Absolute Percent Error, Mean Absolute Error, and Mean Squared Error when compared to other methods in relevant literature. The results of this research have practical significance, providing a foundation for the development of decision support tools tailored to the unique dynamics of the bond market. By leveraging the predictive insights generated by the XGB models, stakeholders can improve decision-making processes, thereby enhancing market efficiency. This study makes a significant contribution to portfolio management, investor decisions, and policymaker. Regulators and central banks can closely monitor expected liquidity developments and take timely action, particularly in terms of government policy. For example, regulators might revise legislation to stimulate capital market liquidity before a crisis and introduce supportive measures. As seen in mutual fund policies, allowing the cost of redemptions to be passed on to redeeming shareholders before or during a crisis could help prevent fire sales that impact the entire market. Regulators could incentivize funds to adopt new policies and alternatives to mitigate liquidity crises.

Keywords: deep learning, machine learning, liquidity prediction, bond market, liquidity risk, liquidity premium, liquidity risk management

1. Introduction

In financial economics, a key principle is the expectation of future market conditions, which heavily influences investor decisions. Liquidity, in particular, is a complex and elusive concept. The literature suggests that liquidity has multifaceted relationships with other market factors, making it challenging to form reliable expectations about future liquidity. Issuers of corporate bonds respond to anticipated liquidity changes, particularly in times of deterioration, to avoid distress from unfavorable refinancing conditions. Regulators and central banks, who closely monitor expected liquidity, are interested in identifying signs of worsening conditions to implement timely countermeasures. However, due to the lack of a universally accepted forecasting method, researchers have often inferred insights into the bond market's dynamics by assuming that a bond's current liquidity is the best predictor of its future liquidity [1]. However, expectations about future bond liquidity are crucial for determining today's bond prices and influencing the behavior of market participants. As a result, our current understanding of liquidity's role in the bond market may be limited or biased.

In this paper, we design and build the predictive model which able to forecast the bond's liquidity, its lack of liquidity in the upcoming day. Based on this predictive model (Distribution), we drive the forward-looking illiquidity measure: Expected Illiquidity (EI). The EI is computed by mean the distribution, providing a measure for the expected transaction cost of the bond. In our prediction model is based on modern technique and state-of-the-art machine learning model or deep learning model which allowing the incorporation of the non-linear relations between a rich set of illiquidity predictors and future illiquidity. We selection of predictions is motivated by boards stand of the bond market illiquidity literature. In our model, we train the deep learning model (MLR, MLP, MDL, LSTM, XGB) based on the information available. We implement the prediction models on the Thailand Bond Market using end of day (EOD) transaction data for the January 1, 2015, to January 24, 2024. In the part of target illiquidity measure within our analysis, we choose the size-adapted bid- ask spread (Reichenbacher, 2022), which measure a bond for a trade with representative size. However, we applied to any illiquidity measure by simple average bid-ask spread (Hong, An Empirical Study of Bond Market Transactions, 2000).

In part of the performance of our predictive concept, the expected illiquidity measure, EI with the literature's approach to use illiquidity in the current daily and we compare the forecasting errors in out-of-sample for the EI with naïve measure as benchmark prediction including compare between model using mean absolute error (MAE), mean absolute percent error (MAPE) and mean square error (MSE).

Our expected contribution of paper we offer an easy-to-implement application that can help practitioners improve their decision-making in financial markets. Second, and more importantly, we contribute to the literature on the role of liquidity in corporate bond prices. The conclusions regarding liquidity in existing studies are based on the current level of a bond's illiquidity. Third, the implementation of this studies which study on liquidity is a significant contribution to improve portfolio management or investor' decision and Regulators and Central banks monitor the expected development of liquidity very closely to take timely countermeasures. Especially, the policy of Government and regulators may change act actively to stimulate capital market liquidity before occurring the crisis as well as support affect some industries. Moreover, the policy of mutual fund may change the regulatory by allowing to pass on the cost of redemptions to the redeeming shareholders before or during crisis. Because the fire sales struggling funds might also impact the market as a whole, the regulator may incentivize funds to broadly use this new possibility.

2. Literature Review

This paper addresses three strands of literature. First, it contributes to the growing body of research on machine learning approaches related to enhancing our understanding of financial markets. Gu, Kelly, and Xiu (2020) served as an initial inspiration for our work, and Bali, Beckmeyer, Mörke, and Weigert (2023) helped us refine the interpretation of our results. Our contribution to this strand lies in providing a practical application that may be useful for practitioners in improving their decision-making in financial markets. Second, and more importantly, we advance the literature on the role of liquidity with respect to corporate bond prices. The conclusions drawn in related studies (e.g., Bao, Pan, and Wang, 2011; Friewald, Jankowitsch, and Subrahmanyam, 2012; Dick-Nielsen, Feldhütter, and Lando, 2012; Bongaerts, de Jong, and Driessen, 2017) rely on the current level of illiquidity of bonds. Our approach refines these findings and strongly emphasizes the significance of liquidity for corporate bond prices. We also build on Kelly, Palhares, and Pruitt (2023), who developed a factor model for the corporate bond market. To the best of our knowledge, we are the first to develop a forecasting model for individual bond liquidity. While Boyarchenko, Giannone, and Shachar (2019) found that autoregressive models excel in forecasting liquidity at the market level, we demonstrate that this result does not apply to individual bond liquidity. Third, we contribute to the literature on market fragility in corporate bond mutual funds. Unlike the convex flow-return relationship observed in equity mutual funds (Chevalier and Ellison, 1997), corporate bond mutual funds exhibit a concave flow-return relationship (Chen, Goldstein, and Jian, 2010; Goldstein, Jiang, and Ng, 2017). Existing empirical analyses in this literature typically rely on the current illiquidity of a fund's corporate bond portfolio to gauge the magnitude of the first-mover advantage. We enhance this literature by showing that the effect of illiquidity concerns on fund outflows approximately doubles when incorporating our measure of expected illiquidity.

2. Methodology

2.1 Data description and preparation

The analysis is based on end of day bond transaction from Thailand Bond Market Association (TBMA) from January 1, 2015, to January 24, 2024. The daily transaction data of bonds issuing and appear in TBMA between 2015 to 2024. The bonds which issue by issuers cover every sector in market. Moreover, including the Bond characteristics, rating histories, and outstanding amounts are from Refinitiv Eikon and Bloomberg. We implement our illiquidity forecast for the size-adapted average bid-ask spread measure of Reichenbacher and Schuster (2022), which overcomes the strong dependence of standard transaction cost measures on the observed trade sizes in a bond. The size-adapted bid-ask spread quantifies the cost to trade a “representative” volume, i.e., a trade size for which transaction costs equal their volume-weighted average across all trades. The measure is based on a market-wide functional form of transaction costs depending on trade size, which is estimated daily using the full cross-section of bonds. The bond market liquidity is represented by measure following,

Life time (Term to maturity), Despite bond illiquidity being persistent [3] [4], a large body of empirical literature shows that it varies predictably with certain characteristics of a bond over its lifetime. As a bond approaches its maturity date, its price volatility typically decreases, leading to a narrower bid-ask spread. Early in its life, a bond might have a wider spread due to greater uncertainty and higher perceived risk. As maturity nears, the certainty around the bond's payout increases, reducing the risk for dealers and narrowing the spread.

Age, the studies about the age of bond find that bonds are typically most liquid directly after issuance and get more illiquid when they age or in other words bonds are liquid and trade frequently. However, liquidity is usually highest immediately after issuance and tends to decline significantly as the bonds age. [5] [2]. When a bond is first issued, there is usually a significant amount of interest from investors. The initial distribution process involves many market participants, including institutional investors, which leads to a high volume of trading. This increased trading activity means that there is a large number of buy and sell orders, resulting in a narrower bid-ask spread due to higher liquidity.

Outstanding, Bonds with a higher outstanding amount and bonds that trade more frequently have lower transaction costs (Edwards, 2007) (Bao, 2011) (Jankowitsch, 2011) which meaning there are more buyers and sellers in the market. This higher liquidity reduces the bid-ask spread, which is a key component of transaction costs. When a bond is more liquid, it's easier to buy and sell without significantly affecting the price.

Credit rating, riskier bonds with more credit risk are typically less liquid than comparable bonds with lower risks (Mahanti, 2008) (Hotchkiss, 2017). Bonds with greater credit risk are perceived as more likely to default. This higher risk makes investors more cautious, reducing the number of willing buyers and sellers in the market. This lack of market participants leads to lower liquidity and wider bid-ask spreads, as dealers and investors demand a higher premium to compensate for the increased risk. Regarding credit risk,

we use the average numerical bond rating of the two rating agencies Fitch and TRIS (local credit rating agency) which some of bonds have change between period of data in case of downgrade.

Based on this literature, we build our set of predictors to forecast for each bond next period's illiquidity measure. Given the high persistence of individual and market illiquidity, we naturally include a bond's illiquidity measure in the current daily t (Current Illiquidity).

Next, we include a bond's age, its duration, and its (log-transformed) outstanding amount. We capture trading activity with the logarithms of average trade size and total trading volume. Following Chordia, Sarkar, and Subrahmanyam (2005), we incorporate a bond's monthly return and order imbalance as possible predictors.

We measure bond order imbalance as the difference between a bond's buying and selling dollar volume normalized with total trading volume which capture and represent the number of trades (Pastor, 2003).

We are using the Trading value as a predict as a predictor variable (X) in a predictive model, as the liquidity proxy which trading Value can be used as a proxy for liquidity. Higher Trading Values might indicate more liquid markets, where assets can be easily bought and sold without significantly impacting prices. Lower Trading Values may suggest illiquidity, where transactions could be more challenging or costly. Moreover, the trading value have indicated the market sentiment which might reflect market sentiment or investor behavior. For instance, a sudden surge in Trading Value might suggest increased interest or confidence in the market, while a decline could indicate caution or uncertainty. (Brennan, Alternative factor specifications, security characteristics, and the cross-section of expected stock returns, 1998)

We are using the Turnover Ratio (TO) as a predict as a predictor variable (X) in a predictive model, as the liquidity proxy which Turnover Ratio can serve as a proxy for liquidity. Higher turnover ratios may suggest greater liquidity, making it easier for investors to buy and sell shares without significant price impact. Moreover, the turnover ratio indicated the price movement which changes in the Turnover Ratio might also reflect movements in the stock's price. For instance, a high turnover ratio accompanying a price increase could suggest bullish sentiment, while a high turnover ratio accompanying a price decrease could indicate bearish sentiment. We defined total trading values of a stock over a specified timeframe divided by market capitalization. (Amihud, Market microstructure and securities values: Evidence from the Tel Aviv Exchange: Evidence from the Tel Aviv Exchange, 1997)

We are using the Illiquidity Ratio (AMH) as a predict as a predictor variable (X) in a predictive model, measure is ratio of the absolute bond returns over the trading value or the other word measure occurs to using price which sensitivity to gauge the impact or influence of order flow on price. Measure, an illiquidity. The metric for the resilience dimensions encompasses the return, incorporating the trading value, for $stock_i$ on a given day. (Amihud, Illiquidity and Stock Returns: Cross-Section and Time Series Effects, 2002)

We using Bid-Ask Spread (SPRD) as the dependence or precited variables which the values of difference between the best ask and best bid price (Amihud Y. , 1980). Bid-ask spread is commonly used as a proxy for liquidity due to its direct relationship with the ease and cost of trading a security. The bid-ask spread represents the immediate transaction cost to trade a security. A narrower spread indicates lower trading costs, reflecting higher liquidity, as it implies that buyers and sellers can trade the security at prices closer to each other. So, the bid-ask spread is a valuable proxy for liquidity because it encapsulates various aspects of market activity, including trading costs, market depth, competition, information efficiency, risk, trade frequency, dealer activity, and overall market conditions. By measuring how closely bid and ask prices align, the spread provides a clear and immediate indication of a security's liquidity.

Table 1: Summary variables in model

Type	Data	Variables
Independent (predictor)	Transaction	AMH, Credit rating, Issue size, Issue term, Trading value, Term to maturity, Turnover ratio, Volume
Dependent (predicted)	Transaction	Spread

We summaries the descriptive statistic which AMH The variable shows a mean value of 0.03 with a standard deviation of 1.33, indicating some variability around the mean. The range -4.28 to 4.78 suggests a wide spread of data points. Credit rating the mean credit rating is 7.31, with a relatively low standard deviation of 0.76, indicating that ratings are clustered closely around the mean. The range 6.26 to 8.26 shows a moderate variation. Issue size and Issue term: Both variables have no variation deviation of 0.00, indicating that all observations are identical at 9,363 and 14.64, respectively. Trading value, Term to maturity, Volume, and Spread: These variables show varying degrees of dispersion as indicated by their standard deviations. Trading value SD 8.72, Term to maturity SD 1.34, Volume SD 0.0847, and Spread SD 27.02 all demonstrate considerable variability in their respective data ranges. The turnover ratio has a mean of 0.00098 and a standard deviation of 0.0012, indicating low variability relative to its mean. In conclusion, the dataset encompasses a range of variables with varying levels of dispersion and centrality. Some variables exhibit minimal variation (like Issue size and Issue term), while others show broader ranges and higher standard deviations (such as Spread and Trading value). Understanding these statistics helps in interpreting the distribution and characteristics of each variable within the dataset.

Table 2: Summary descriptive statistic

Variables	Mean	Min	Max	SD	25%	50%	75%
AMH	0.03	-4.28	4.78	1.33	-0.42	0.01	0.40
Credit rating	7.31	6.26	8.26	0.76	6.83	7.32	7.79
Issue size	9,363	9,363	9,363	0.00	9,363	9,363	9,363
Issue term	14.64	14.46	14.46	0.00	14.64	14.64	14.64
Trading value	7.62	0.13	55.43	8.72	2.61	5.22	10.05
Term to maturity	12.21	10.02	14.55	1.34	11.11	12.17	13.31
Turnover ratio	0.00098	0.00002	0.00723	0.0012	0.00029	0.0006	0.0012
Volume	0.0745	0.0013	0.5475	0.0847	0.0257	0.0510	0.0985
Spread	75.87	36.46	141.97	27.02	55.04	71.97	91.76

The objective of the Ordinary Least Square (OLS) in the table3 regression is to estimate the relationships between a dependent variable which is bid-ask spread and more independent variables. So, we also check the importance factors for liquidity prediction model which statistically significant dependence variables y which is bid-ask spread. conclusion, the consistent significance of all variables like AMH, Credit rating, Issue size, Issue term, Trading value, Term to maturity, Turnover ratio, and the volume all of them can explain the bid-ask spread which represent the liquidity expectation. So, overall, the model has a high R-squared value, suggesting that variables can explains a large proportion of the variance in the bid-ask spread which is dependent variable. Most of the coefficients are statistically significant, as indicated by their low p-values. Which we will take all of independence variables input into the machine learning and deep learning models for predict the liquidity

Table 3: the summary table of the Ordinary Least Square (OLS) regression

Variables	coef	Std err	t	P>|t|	[0.025	0.975]
Intercept	2.589e-05	6.53e-07	39.614	0.000	2.46e-05	2.72e-05
AMH	-7.6797	2.392	-3.210	0.001	-12.385	-2.975
Credit rating	0.0005	1.18e-05	39.614	0.000	0.000	0.000
Issue size	0.0518	0.001	39.614	0.000	0.049	0.054
Issue term	5.177e-05	1.31e-06	39.614	0.000	4.92e-05	5.43e-05
Trading value	-0.0029	0.001	-3.218	0.001	-0.005	-0.001
Term to maturity	52.2724	1.460	35.815	0.000	49.402	55.143
Turnover ratio	-1.436e-12	4.46e-13	-3.217	0.001	-2.31e-12	-5.58e-13
Volume	0.2846	0.089	3.212	0.001	0.110	0.459
R-squared (average)	0.792					
Adjusted R-squared (average)	0.79					

We also check the correlation between variables including correlation between lagged variables (1 until 7 days). The correlation coefficients are very weak for daily liquidity proxy data.

2.2 Prediction Model

The first step before developing the neural network. We try to be rescaling the numerical data of variables to a common scale. This is done to eliminate any inherent differences in the scales of the variables, which could otherwise result in biased models. After that, using the min-max normalization scales the data to a range between 0 and 1, where the minimum value of the data is mapped to 0 and the maximum value is mapped to 1. The formula for min-max normalization is:

$$Xnorm = \frac{(X - Xmin)}{(Xmax - Xmin)}$$

And spilt the data into train and test are 80% for train data and 20% for test data. We developed the neural network which is the regression model. The optimal structure of network based on experimental networks. We try to tune the hyper parameter of networks. The first, multi-layer perceptron fully connected neural network (MLP) (Khang, 2020) was developed. By the parameters of MLP are as follows:

- The input layer has 60 neurons by using reLU activation.
- The two hidden layers have from 30 and 20 neurons by using reLU activation.
- The output layer: 1 neuron by using reLU activation.

The architecture of the MLP is illustrated in Figure 1.

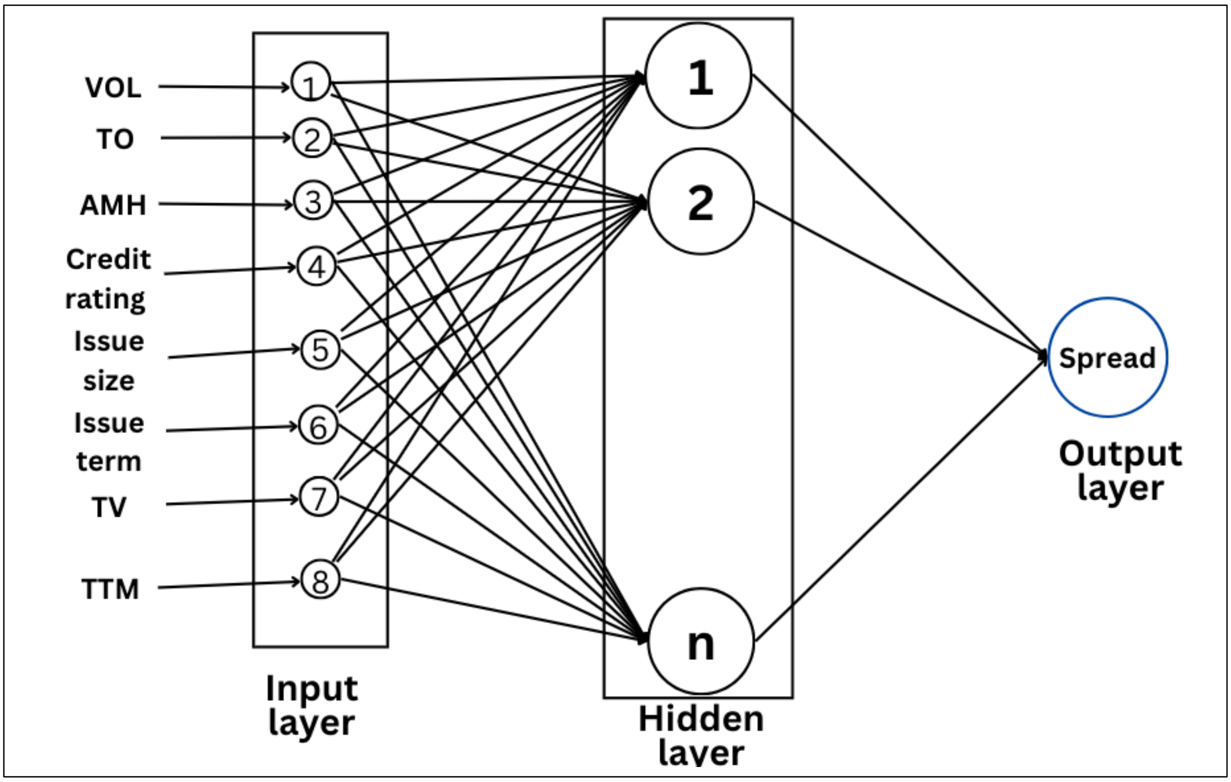

Figure. 1 Architecture of the MLP

The second state of model development, we try to develop the process of the model by mixed the architecture of deep learning is mixed deep learning (MDL) [20]. We tune parameter as follow:

- The input layer: 60 neurons by using reLU activation.
- The recurrent layer: 300 neurons by using reLU activation.
- The three fully-connected hidden layers with dropout's: from 300 and 20 neurons by using reLU activation.
- The Output layer: 1 neuron by using reLU activation.

The architecture of the MDL is illustrated in Figure 2

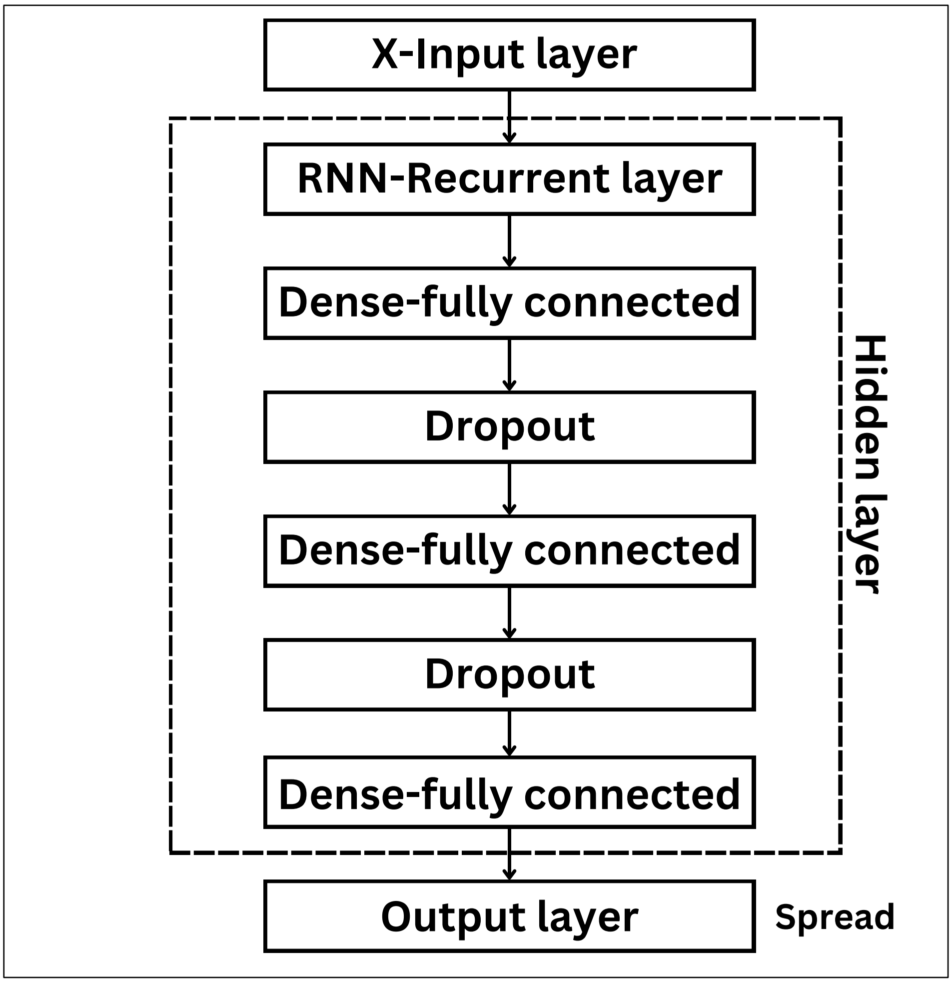

Figure. 2 Architecture of the MDL

The third state of the model development, we try to develop the Long Short-term memory (LSTM) is one of machine learning or artificial type be recurrent neural network used in the deep learning. The network of LSTM compensates addressing the challenges of vanishing gradients and short-term memory in traditional recurrent neural networks. The LSTM architecture comprises the cell state, which acts as the memory unit of the network, along with its regulators. The cell state holds information that can be stored, written to, or read from a preceding cell state through gates. So, the third stage of model development process is LSTM. We tune parameter as follow:

• The input layer: 64 neurons by using reLU activation.

• The recurrent layer: 300 neurons by using reLU activation.

• The three fully-connected hidden layers with dropout's: 0.2 by using reLU activation.

• The output layer: 1 neuron by using reLU activation.

The architecture of the LSTM is illustrated in Figure 3.

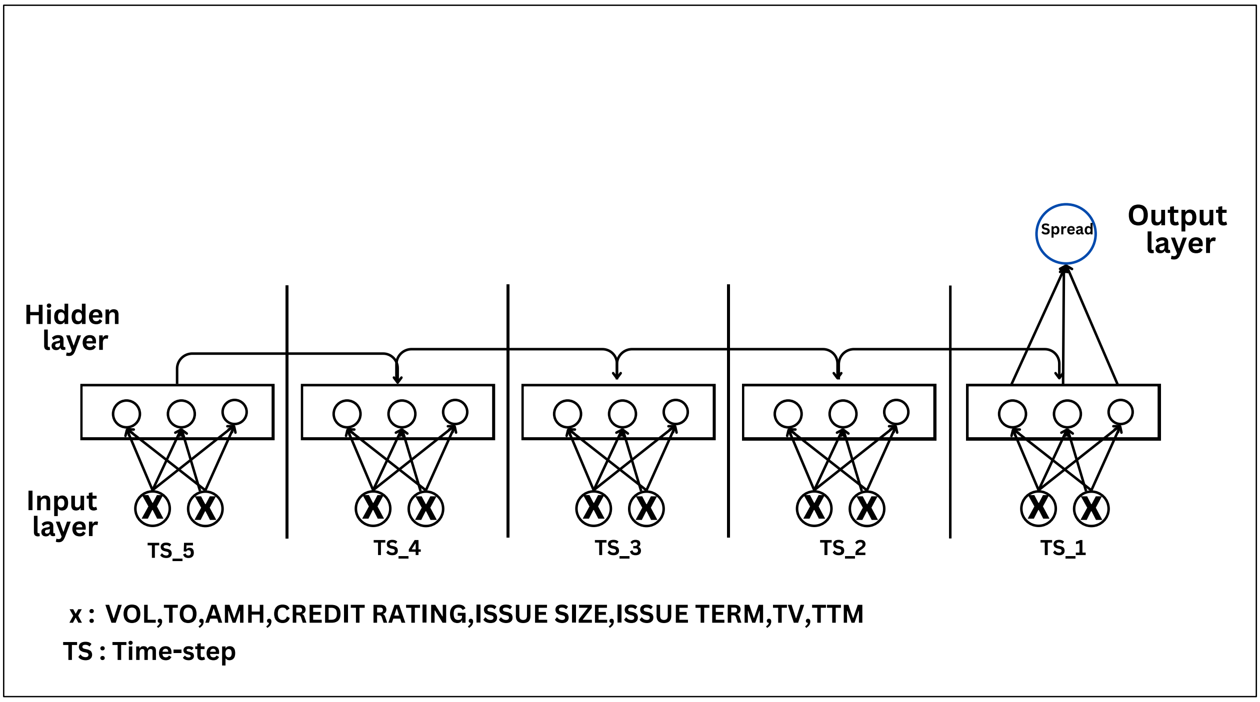

Figure. 3 Architecture of the LSTM

The fourth state, we try to develop the XGB

$$EI_{i,t} \equiv E_t[Iliquidity_{i,t+1} = g_t(Predictorss_{i,t})]$$

Which estimate g_t, we use gradient boosted regression trees (GBRT). This machine learning algorithm allows for a flexible shape of g_t and is more efficient than methods based on neural networks, which is important, since we re-estimate the model on a rolling basis. The GBRT algorithm approximates g_t via building a regression tree based on a set of training observations, $Train_t$, and a set of validation observations, V alt. Once we have trained the tree according to the algorithm described below, we will evaluate its predictive performance via a test set $Test_t$. Figure 4, The output of architecture the XGB

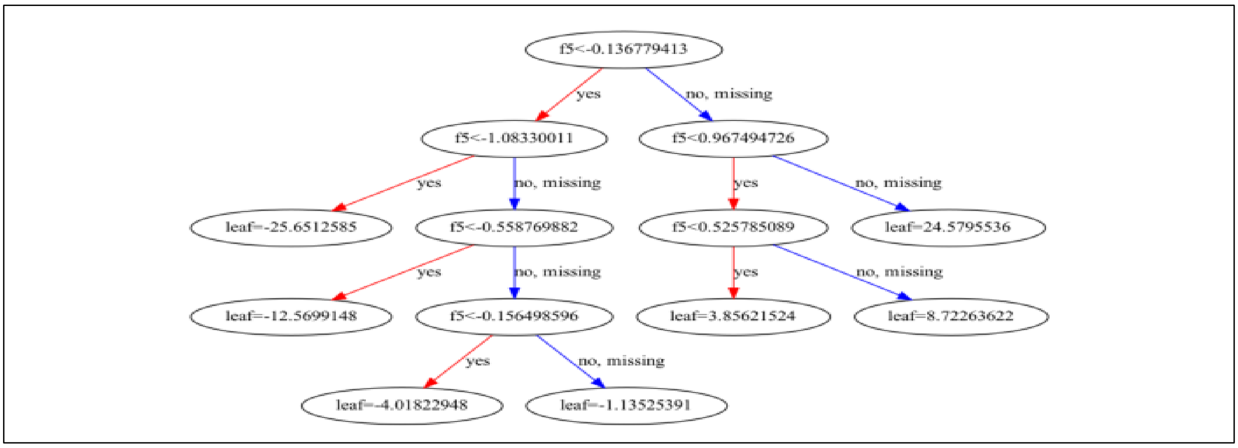

Figure. 4 Architecture of the XGB

Bid-Ask Spread (SPRD) the values of difference between the ask and bid price. (Amihud Y. , 1980) So, we set he input data is defined as follows:

$$X = X^{ttm}, X^{age}, X^{out}, X^{rating}, X^{TV}, X^{TO}, X^{AMH}, X^{VOL}$$

Where

X_{ttm} $represent\ age\ (term\ to\ maturity)$

X_{age} $represent\ lifttime\ (age)$

X_{out} $represent\ outstanding$

X_{rating} $represent\ credit\ rating$

X_{TV} $represent\ trading\ value$

X_{TO} $represent\ turnover\ ratio$

X_{AMH} $represent\ illiquidity\ ratio$

X_{VOL} $represent\ volumn$

For deep learning network has a mixed architecture (MLP, MDL,LSTM). The output of models is spread (bid-ask spread) defined as follows:

$$y(output) = f(y^1, y^2, y^3, y^4, \dots, y^L)$$

Where :

y^l denotes a layer in which l = 1, 2,…..L (L denotes a number of layers)

$f()$ denotes the output activation function.

Let dense denotes the layer composed of certain number of neurons which denotes a number of neurons (RNN represents a recurrent layer), *relu* denotes rectified linear unit activation function defined as follows:

$$relu\ (z) \begin{cases} z\ if\ z \geq 0 \\ 0\ \ 0, otherwise \end{cases}$$

The initial (input) layer is specified as follows:

$$y = dense(X^{ttm}, X^{age}, X^{out}, X^{rating}, X^{TV}, X^{TO}, X^{AMH}, X^{VOL}$$

The second layer is the RNN layer in the case of MLP, MDL and LSTM, defined as follows:

$$y^2 = RNN(y^1, \text{neurons relu})$$

For, the full equation of MLP, MDL and LSTM model will input into

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$A^{[1]} = ReLU(Z^{[1]})$$

$$Z^{[2]} = A^{[1]}W^{[2]} + b^{[2]}$$

$$A^{[2]} = ReLU(Z^{[2]})$$

$$Z^{[3]} = A^{[2]}W^{[3]} + b^{[3]}$$

$$A^{[3]} = ReLU(Z^{[3]})$$

$$\hat{y} = A^{[3]}W^{[out]} + b^{[out]}$$

Where $Z^{[1,2,3]}$: the equation of linear function

$A^{[1,2,3]}$: Activation function of hidden layer 1,2,3 which is the equation of non-linear function

$W^{[1,2,3]}$: Matrix weight output of hidden layer 1,2,3

$W^{[out]}$: Matrix weight output

$b^{[out]}$: Bias output

X : Matrix of input variables

2.2 Model performance metrics

In the part of Model performance metrics, we implement multiple liner regression (MLR), multilayer perceptron (MLP), mixed deep learning (MDL) and long-short term memory (LSTM) architecture to predict the bid-ask spread. Within each of these models, several options are considered with different number of neurons. Prediction accuracy and reliability of these models are assessed by calculating three different performance metrics Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Square Error (MSE). The analytical form of these metrics is defined as follows

$$\boldsymbol{MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{(y_i - \hat{y}_i}{y_i}\right|}$$

$$\boldsymbol{MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

Where, y_i : refers to the original bid-ask spread. $\hat{y}_i$: refer to the predicted bid-ask spread.

N : Number of observation.

Smaller values of MAPE, MAE and MSE, better the performance model. Moreover, performance scores are calculated after applying the inverse transformation in the predictions obtains from the normalized data. Each model is executed multiple times independently to address the stochastic behavior.

3. Results

The objective of the experiment was to evaluate the developed liquidity prediction model for fixed income in Thailand Bond Market. We employed mean absolute percent error (MAPE), mean absolute error (MAE) and mean square error (MSE) as metrics. The experiments pertaining to MLR MPL, MDL, LSTM and XGB which are deep learning and machine learning model compare with traditional model and naïve benchmark. Hyper parameters and parameters in all models were determined based on related works and experimentation. The results are presented in the table 1 and the Fig 4 illustrates the predicting procedure with continuous data values.

In comparison, the bid-ask spread which represent EI from our XGB which are the machine learning model that prediction model only has a forecasting error of 2.88 basis points, corresponding to a 0.24% reduction in error. On top of the XGB which have GBRT algorithm approach. we also assess a set of alternative models to forecast EI. We estimate MLP, MDL, LSTM which are deep learning model have percent error 4.27%, 10.33%, 17.99% and have forecasting error of 36.02, 252.46, 369.40 and 2,930.17 basis point respectively. Compare with naïve benchmark which have percent error 11.97% and forecast error 154.47 basis point. Based on a Diebold and Mariano (1995) test, shows that GBRT significantly outperforms. Moreover, the best value of performance metric in MLP, MDL and LSTM model has been achieved in 100 epochs. we also estimate a model that only uses predictors that are directly related to illiquidity. Table 4 shows that the XGB model, although using a much smaller number of predictors, still out- performs the naïve benchmark. our XGB prediction model generates an error that is 2.88 basis points lower than the error of the naïve benchmark model, leading to a 0.24% error reduction. In comparison with the naïve benchmark. Again, MLP is better than MDL is better than MLR which the linear approach and LSTM without variable selection.

Table 4: Comparative analysis of average MAPE, MAE and MSE on individual bonds in Thailand Market Bond Association which value obtained using MLR, MLP, MDL, LSTM and XGB models **(**Performance of prediction models).

Models	TEST (average)		
	MAPE	MAE (bsp)	MSE (bsp)
MLR	0.1799	18.59	369.40
MLP	0.0427	4.40	36.02
MDL	0.1033	10.64	252.46

Models	TEST (average)		
	MAPE	MAE (bsp)	MSE (bsp)
LSTM	0.5261	54.12	2,930.17
XGB	0.0024	0.27	2.88
NAÏVE BENCHMARK	0.1197	12.31	154.47

In comparison aspect duration, For Short-Term Predictions (1-5 years and 5-10 years), LSTM which are deep learning model have performed well is the most reliable model that prediction model only has a forecasting error of 28.5301, 33.5584 basis points, corresponding to a 18.01%, 13.05% reduction in error respectively. For Long-Term Predictions (more than 10 years), XGB is the most reliable model that prediction model only has a forecasting error of 8.1743, 13.8052 basis points, corresponding to a 0.7%, 2.82% reduction in error respectively. For alternative models to forecast EI. We estimate MLP, MDL which are deep learning model have high percent error and forecasting error. Also, naïve benchmark which have high percent error and high forecast error. Table 5 still shows that the LSTM, XGB model, although using a much smaller number of predictors, still out- performs the naïve benchmark. our LSTM, XGB prediction model generates an error that is lower than the error of the naïve benchmark model. In comparison with the naïve benchmark. Again, LSTM, XGB is better than MDL is better than MLR which the linear approach and LSTM, XGB without variable selection.

Table 5: Comparative analysis of average MAPE, MAE and MSE on individual bonds classify by duration which value obtained using MLR, MLP, MDL, LSTM and XGB models (Performance of prediction models).

Models		TEST (Average) classifies by durations (years)			
	Metrics	1-5	5-10	10-15	> 15
MLR	MAPE	1.7576	0.6215	0.0192	0.0522
	MAE	45.1363	19.4861	3.3705	5.5580
	MSE	5700.8941	849.9031	25.8292	36.6509
MLP	MAPE	0.4966	0.8785	0.1325	0.2015

Models	Metrics	TEST (Average) classifies by durations (years)			
		1-5	5-10	10-15	> 15
	MAE	12.3732	27.4828	23.5942	21.4592
	MSE	415.3636	2750.9444	672.7231	474.4174
MDL	MAPE	1.6153	0.5161	0.4172	0.4245
	MAE	39.8025	17.2539	72.0472	44.2189
	MSE	1964.7712	542.4732	5190.8047	2803.2707
LSTM	MAPE	0.1801	0.1305	0.9184	0.9277
	MAE	4.8816	4.6767	165.0243	102.2540
	MSE	28.5301	33.5584	27233.0837	10484.1633
XGB	MAPE	0.3800	0.2252	0.0070	0.0282
	MAE	9.6084	7.2380	1.2101	3.0610
	MSE	105.1098	64.0236	8.1743	13.8052
NAÏVE BENCHMARK	MAPE	1.4010	0.7085	0.0650	0.1730
	MAE	36.5732	23.2184	11.6448	18.5415
	MSE	1350.4441	571.8879	142.4078	349.2697

In comparison aspect credit rating, For AAA and AA rating, MLP and LSTM which are deep learning model have performed well is the most reliable model that prediction model only has a forecasting error of 67.6726, 109.2670 basis points, corresponding to a 30.13%, 19.24% reduction in error respectively for MLP, forecasting error of 62.8964, 53.4266 basis points, corresponding to a 34.76%, 14.42% reduction in error respectively for LSTM. For A rating, LSTM outperforms other models forecasting error of 26.2699 basis points, corresponding to a 15.18% reduction in error respectively. For BBB rating, XGB shows excellent performance, forecasting error of 22.2224 basis points, corresponding to a 1.36% reduction in error respectively. For non-investment grade, XGB has relatively low MAPE, indicating better percentage accuracy. For alternative models to forecast EI. We estimate MDL which are deep learning model have high percent error and forecasting error. Also, naïve benchmark which have high percent error and high forecast error. Table 6 still shows that the LSTM, MLP, XGB model, although using a much smaller number of predictors, still out- performs the naïve benchmark. our LSTM, MLP, XGB prediction model generates an error that is lower than the error of the naïve benchmark model. In comparison with the naïve benchmark. Again, LSTM, MLP, XGB is better than MDL is better than MLR which the linear approach and LSTM, MLP, XGB without variable selection.

Table 6: Comparative analysis of average MAPE, MAE and MSE on individual bonds classify by credit rating which value obtained using MLR, MLP, MDL, LSTM and XGB models **(**Performance of prediction models).

Models		TEST (Average) classifies by durations (years)				
	Metrics	AAA	AA	A	BBB	Non-investment grade
MLR	MAPE	0.3921	0.1432	0.3078	0.1536	3.7131
	MAE	8.7506	6.1353	6.0684	26.2520	21.9435
	MSE	106.6033	59.8573	55.5594	708.5144	1059.4885
MLP	MAPE	0.3013	0.1924	1.4296	0.0565	1.3795
	MAE	6.6651	8.3561	16.7018	9.4879	11.8654
	MSE	67.6726	109.2670	378.9988	144.6566	242.3360
MDL	MAPE	1.2567	0.7255	2.3218	0.2546	1.6353
	MAE	27.6543	31.5420	25.0551	43.8210	17.2069
	MSE	1352.4046	1489.2201	837.9330	2948.7107	391.9197
LSTM	MAPE	0.3476	0.1442	0.1518	0.5997	0.6230
	MAE	7.3494	6.0412	3.9816	103.0142	29.7549
	MSE	62.8964	53.4266	26.2699	10624.1374	923.3857
XGB	MAPE	0.1877	0.3376	0.6530	0.0136	2.2795

Models	TEST (Average) classifies by durations (years)					
	Metrics	AAA	AA	A	BBB	Non-investment grade
	MAE	5.1978	14.5132	8.3325	2.2223	15.2455
	MSE	46.3033	245.4450	93.4824	22.2224	426.3757
NAÏVE BENCHMARK	MAPE	1.1499	0.8151	1.9474	0.1621	2.0353
	MAE	25.8425	36.0396	25.2251	27.6069	14.5179
	MSE	693.6692	1333.1195	726.0078	779.4268	334.2332

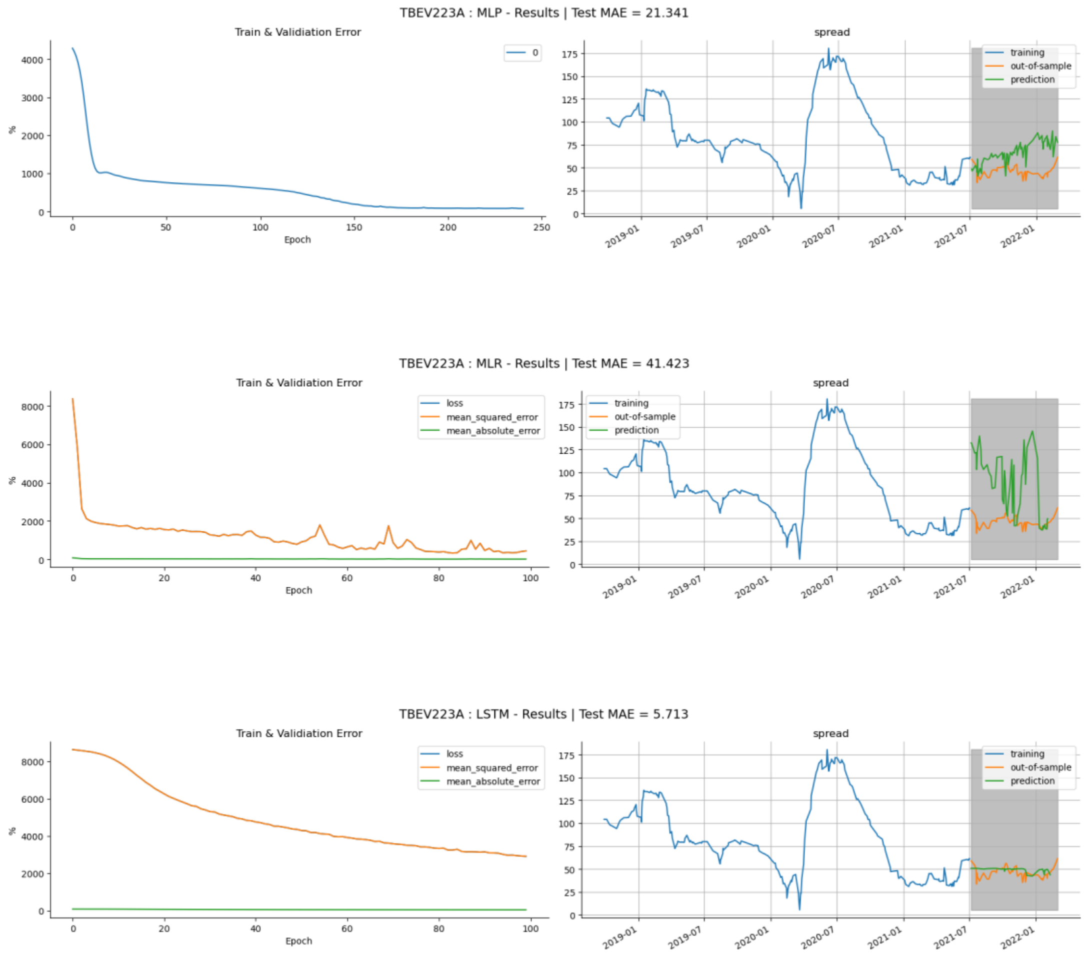

Figure. 5 The performance metric MPL, MDL and LSTM

The table 7 presents the weights or coefficients of various financial and market variables used in a regression model, along with the average values of these coefficients. Intercept (Bias). The implication shows the model indicates that Trading Value, Turnover Ratio, and Volume are the most significant factors positively influencing the outcome. In contrast, AMH and Credit Rating have a slight negative impact. Issue Size and Issue Term do not influence the outcome in this model. The base value for the prediction starts at 80, with adjustments made according to the coefficients of the respective variables. So, the equation of MLR model will input into

$$\hat{y} = w_0 + w_1 x_1 + w_2 x_2 + \ldots\ldots + w_p x_p$$

Table 7: Average bias and weight of the equation MLR model for predicting the spread.

Weight output / coefficient of Variables	Average
Bias (interception (w_0))	80
AMH (w_1)	-0.64
Credit rating (w_2)	-0.50
Issue size (w_3)	0
Issue term (w_4)	0
Trading value (w_5)	147.57
Term to maturity (w_6)	7.80
Turnover ratio (w_7)	147.57
Volume (w_8)	295.83

The table 8 presents the weights or coefficients of variables for different layers of three types of neural network models: MLP, MD, and LSTM. Each layer's weights and biases are given specific values. From the weight values provided, MLP places significant importance on both its hidden layers and output layer, as indicated by the larger weights and biases. MDL generally has smaller weights and biases, suggesting a more uniform or distributed approach to handling data across layers. LSTM shows strong reliance on the input layer and significant biases in hidden layers, indicating a strong initial influence and adjustment capacity. In summary, the MLP model seems to emphasize deep layer learning and final output computation, the MDL model takes a balanced approach across layers, and the LSTM model focuses on initial data input and internal adjustments for sequential data handling. So, the equation of MLP, MDL and LSTM model will input into

$$\hat{y} = A^{[3]}W^{[out]} + b^{[out]}$$

Table 8: Average bias and weight of the equation MLP, MDL and LSTM model for predicting the spread.

Weight / coefficient of Variables	MLP	MDL	LSTM
Input layer Hidden layer 1 ($W^{[1]}$)	-0.0004125	-0.000353906	0.535931719
Hidden layer 1 (BIAS) ($b^{[1]}$)	0.168631875	0.0035125	0.000346582
Hidden layer 2 ($W^{[2]}$)	0.027387083	0.016515	0.358102271
Hidden layer 2 (BIAS) ($b^{[2]}$)	0.016515	0.003870792	-0.008585352
Hidden layer 3 ($W^{[3]}$)	0.050803125	0.028845479	0.076608617
Hidden layer 3 (BIAS) ($b^{[3]}$)	0.063284688	0.013106875	0.025435506
Output layer ($W^{[out]}$)	0.119960313	0.02665107	0.138594431
Output layer (BIAS) ($b^{[out]}$)	0.32075	0.0358125	0.23485

4. Discussion

In this section, we endeavor to compare the developed XGB MLP, MDL, and LSTM models with alternative approaches. Table 2 present the outcomes of comparing the developed model with other methods from the relevant literature, all focusing on the common goal of liquidity prediction, are presented in this section.

Table 9: Comparative Analysis of Difference Models for Fixed income (Bond)

Authors	Years	Models	MSE (%)
Marcel Müller, Michael Reichenbacher, Philipp Schuster,	2023	Extreme Gradient Boosting	0.604

Marliese Uhrig-Homburg **Invalid source specified.**			
this research	2024	Extreme Gradient Boosting	0.0288
		Multi-layer perceptron	0.3602
		Mixed Deep Learning	2.5246
		Long-term Short-term Memory	29.3

The techniques exhibit a low diversity in terms of MSE. The methods proposed in this paper (XGB/MLP) attained a lower Mean Squared Error (MSE) compared to the others approach and demonstrated a lower Mean Squared Error (MSE) than the presented approach, and the results are predominantly derived from simulated data. Hence, the MSE may be influenced by the simulation methods. The results obtained by XGB/MLP are based on real data. In general, it is challenging to create a model that is universally applicable for predicting liquidity or illiquidity across various bond characteristics or markets in different countries. For instance, liquidity characteristics in emerging markets and developed markets may vary.

5. Conclusion

In this study, we propose a prediction model for individual expected bond illiquidity, EI. We apply a machine learning and deep learning methodology that accounts for the information available to a contemporary forecaster and that allows for changing structural non-linear relations between variables which have effect to liquidity and illiquidity which represent by bid-ask spread over time. Our approach can generally be implemented for an arbitrary illiquidity measure. Examining the effects of our predictive illiquidity measures on corporate bond prices by looking at yield spread changes. Using our predictive measures for illiquidity, both alphas and risk premiums of illiquid bonds are substantially higher than previously reported in the literature. Finally, we use EI to examine the impact of rising illiquidity concerns in poorly performing corporate bond funds on investors’ redemption decisions. Consistent with the implications of strategic complementarities among corporate bond fund investors, our findings extend the insights of Goldstein, Jiang, and Ng (2017), They highlight that historically badly performing funds are especially sensitive to illiquidity due to the fear of fund runs. When we add expected illiquidity deterioration (based on EI) on top of current illiquidity of a fund. Our results provide important insights regarding the discussion of a proposal by the Security Exchange and Commission

(Thailand) to oblige mutual funds to apply swing pricing together with an earlier closing time for orders of mutual fund shares that receive the end-of-the-day price from the same trading day. Swing pricing provides a way for the fund manager to alter the end-of-the-day NAV of a fund depending on flows and the fund's (expected) liquidity via a so-called swing factor. This way, anticipated liquidation costs can be charged from redeeming investors instead of being passed on to the investors that remain in the fund. To this date, swing pricing is voluntary and, in practice, not implemented by any mutual fund. Additionally, an adaption of our forecasting model could be used to calibrate the swing factor by mutual fund managers. Essentially, expected illiquidity is exactly what fund managers need to estimate when setting this factor.

6. Policy Implication

The complexity and non-linear relationships involved in liquidity prediction and liquidity risk in capital markets necessitate the use of advanced models to develop tailored strategies. Policymakers can implement targeted measures to mitigate liquidity risks and enhance financial stability across various sectors. Effective liquidity risk management relies on predictive models that can accurately anticipate and manage these risks. Advanced models like Extreme Gradient Boosting (XGB), which have demonstrated lower error metrics, should be prioritized, especially in sectors with higher liquidity risks. These models offer more accurate forecasts, aiding in better liquidity management. For economic stability, accurate predictive models are crucial for informed decision-making. Policymakers should focus on enhancing the accuracy of models like XGB to ensure reliable forecasts and better economic outcomes through improved liquidity management in capital markets. The integration of advanced predictive models into liquidity risk management processes offers a transformative opportunity to enhance economic stability and support informed policy decisions. By adopting sector-specific interventions, improving data quality, and providing robust regulatory support, policymakers can significantly reduce liquidity risks. Effective implementation of these strategies will lead to more resilient economic systems, better anticipation of financial challenges, and sustainable growth. This approach underscores the critical role of predictive analytics in shaping future economic policies and ensuring financial stability.

Using Predictive Models for Regulators to Regulate the Capital Market. The deployment of predictive models for regulating capital markets presents a transformative approach for enhancing market stability, improving risk management, and fostering investor confidence. By incorporating advanced machine learning techniques such as MLP and XGB, regulators can more accurately predict liquidity risks and take preemptive actions to mitigate potential market disruptions.

Enhanced Market Surveillance:

- Utilize predictive models to continuously monitor liquidity conditions and identify early warning signals of market stress. This allows for timely interventions to prevent market crises.

- Improve the identification of systemic risks within different market segments, enabling more targeted regulatory actions.

Risk Management Frameworks:

- Implement tailored regulatory measures for sectors exhibiting high liquidity risk, such as Financials, Industrials, and Services. Predictive models can guide the formulation of specific capital and liquidity requirements.
- Incorporate predictive insights into stress testing scenarios to evaluate the resilience of financial institutions and markets under adverse conditions.

Optimized Regulatory Policies:

- Leverage model predictions to inform policy decisions, ensuring that regulations are grounded in robust data analysis and predictive accuracy.
- Adapt regulatory policies dynamically based on ongoing predictive analysis, allowing for more flexible and responsive regulation.

Improved Transparency and Confidence:

- Enhance communication with market participants regarding the regulatory use of predictive models, promoting transparency and building confidence in the regulatory framework.
- Strengthen investor protection mechanisms by predicting and mitigating liquidity risks that could adversely impact market stability and investor interests.

Strategic Actions for Regulatory Implementation:

- Encourage the adoption of advanced predictive models within regulatory agencies to enhance surveillance and risk management capabilities.
- Develop and enforce sector-specific regulations based on predictive model insights, addressing unique liquidity risks and vulnerabilities.

The application of predictive models in the regulation of capital markets represents a significant advancement in ensuring market stability and protecting investor interests. By leveraging the capabilities of advanced machine learning techniques, regulators can enhance their surveillance, risk management, and policy-making processes. This approach not only enables more accurate prediction and mitigation of liquidity risks but also fosters a more resilient and transparent financial system. The strategic actions outlined provide a roadmap for regulators to effectively integrate predictive models into their regulatory frameworks, thereby promoting sustainable market development and economic stability.

Acknowledgements

We extend our sincere thanks to all who contributed to preparing the instructions. List here your funding agency and those individuals who provided help during the research (e.g., providing language help, writing assistance, or proofreading the article, etc.).

Author Contributions

S.C.: conceptualization, investigation, reviewing and editing; E.P.: investigation, methodology, writing an original draft; S.K.: research design, data analysis; M.S.: conceptualization, data curation, writing—reviewing and editing, funding acquisition, project administration. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

Amihud, Y. H. (2015). *The illiquidity premium: International evidence* (Vol. 117). Journal of Financial Economics.

Haroon, O. a. (2020). *Flatten the curve and stock market liquidity–an inquiry into emerging economies.* (Vol. 56). Emerging Markets Finance and Trade.

Foucault, T. P. (2013). *Market Liquidity: Theory, Evidence, and Policy.* Oxford Scholarship Online.

Kang, W. a. (2013). *Measuring liquidity in emerging markets.* Available at SSRN: https://ssrn.com/abstract=2326380.

Amihud, Y. a. (2000). *The Liquidity Route to a Lower Cost of Capital.* (Vol. 12). Journal of Applied Corporate Finance.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective.* MIT press.

Owoc, M. L. (2017). *Towards better understanding of context-aware knowledge transformation.* Federated Conference on Computer Science and Information Systems (FedCSIS).

Sałabun, W. K. (2018). *Handling data uncertainty in decision making with COMET.* IEEE Symposium Series on Computational Intelligence (SSCI).

Informatyczne, P. T. (2017). Towards better understanding of context-aware knowledge transformation. *Proceedings of the 2017 Federated Conference on Computer Science and Information Systems, 11*, 1123-1126.

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

Brennan, M. J. (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics, 49*(3), 345-373.

Amihud. (1997). Market microstructure and securities values: Evidence from the Tel Aviv Exchange: Evidence from the Tel Aviv Exchange. *Journal of Financial Economics, 45*, 365–390.

Amihud, Y. (1980). Asset pricing and the bid-ask spread. *Journal of Financial Economics, 17*(2), 223-249.

Chung, K. H. (2014). A simple approximation of intraday spreads using daily data. *Journal of Financial Markets, 17*, 94-120.

Amihud. (2002). Illiquidity and Stock Returns: Cross-Section and Time Series Effects. *Journal of Financial Markets, 5*, 31-56.

Bharath, S. (2009). Does Asymmetric Information Drive Capital Structure Decisions? *Review of Financial Studies, 22*(8), 3211-3243.

Diamond. (1991). Disclosure, Liquidity, and the Cost of Capital. *Journal of Finance, 46*, 1325-1360.

Acharya, V. V. (2005). Asset pricing with liquidity risk. *Journal of Finance, 77*, 375-410.

Brennan, M. J. (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics, 49*(3), 345-373.

Frieder, L. (2006). On Capital Structure and the Liquidity of a Firm's Stock. *SSRN*.

Kale, J. R. (2011). Product market power and stock market liquidity. *Journal of Financial Markets, 14*(2), 376-410.

Rhee. (2009). Foreign Institutional Ownership and Stock Market Liquidity: Evidence from Indonesia. *Journal of Banking & Finance, 33*, 1312-1324.

Whaley. (1983). Transaction costs and the small firm effect. *Journal of Financial Economics, 12*(1), 57-79.

Narayan, P. K. (2010). Gold and oil futures markets: Are markets efficient? *Applied Energy, 87*(10), 3299-3303.

Keene, M. (2007). The Importance Of Liquidity As A Factor In Asset Pricing. *Journal of Financial Research, 30*(1), 91-109.

Al-Nai. (2014). Stock Liquidity Determination: Evidence from Amman Stock Exchange. *Asian Economic and Financial Review, 5*(8), 1894-1905.

Ding, X. (. (2016). Free float and market liquidity around the world. *Journal of Empirical Finance, 38*.

Khang, P. Q. (2020). Liquidity prediction on Vietnamese stock market using Deep learning. *Procedia Computer Science, 176*, 2050-2058.

Balaji. (2018). Applicability of deep learning models for stock price forecasting an empirical study on Bankex. *Procedia computer science, 143*, 947-953.

Addo. (2018). Credit Risk Analysis Using Machine and Deep Learning Models. *UniversityCa'Foscari of Venice, Dept. of Economics Research Paper Series, 18*.

Liu. (2019). Pricing Options and Computing Implied Volatilities using NeuralNetworks. *Risks, 7*.

Khanga, P. Q. (2021). Machine learning for liquidity prediction on Vietnamese stock market. *Procedia computer science*, *192*, 3590-3597.

Reichenbacher, M. a. (2022). Size-Adapted Bond Liquidity Measures and Their Asset Pricing Implications. *Journal of Financial Economics, 146*, 425-443.

Hong, G. a. (2000). An Empirical Study of Bond Market Transactions. *Financial Analysts Journal, 56*, 32-46.

Chordia, T. A. (2005). An Empirical Analysis of Stock and Bond Market Liquidity. *Review of Financial Studies*, *18*, 85-129.

Acharya, V. V. (2013). Liquidity Risk of Corporate Bond Returns: Conditional Approach. *Journal of Financial Economics, 110*, 358-386.

Warga, A. (1992). Bond returns, liquidity, and missing data. *Journal of Financial and Quantitative Analysis, 27*, 605-617.

Hong, G. a. (2000). An Empirical Study of Bond Market Transactions. *Financial Analysts Journal*, *56*, 32-46.

Edwards, A. K. (2007). Corporate Bond Market Transaction Costs and Transparency. *Journal of Finance, 62*, 1421-1451.

Bao, J. J. (2011). The Illiquidity of Corporate Bonds. *Journal of Finance, 66*, 911-946.

Jankowitsch, R. A. (2011). Priced Dispersion in OTC Markets: A New Measure of Liquidity. *Journal of Banking and Finance, 35*, 343-357.

Mahanti, S. A. (2008). Latent Liquidity: A New Measure of Liquidity, with an Application to Corporate Bonds. *Journal of Financial Economics, 88*, 272-298.

Hotchkiss, E. a. (2017). Determinants of Corporate Bond Trading: A Comprehensive Analysis. *Quarterly Journal of Finance, 7*.

He, Z. a. (2012). Rollover Risk and Credit Risk. *Journal of Finance, 67*, 391-430.