Phase Report on Market Liquidity Model

Shuqi Wang,
April 30 2023

1. Introduction

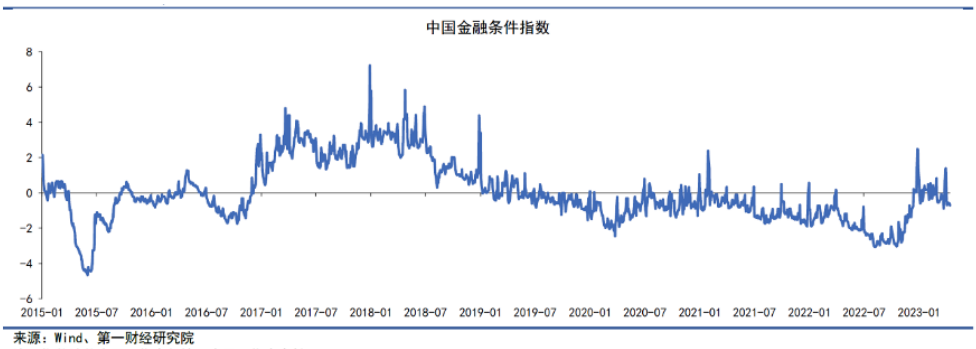
Market liquidity monitoring is crucial for commercial banks to mitigate liquidity risks, which are highly complex and uncertain due to various influencing factors. Even well-capitalized banks with sound risk management can face liquidity crises, challenging liquidity risk regulation. The global financial crisis exposed significant shortcomings in financial regulatory systems for market liquidity monitoring and prediction, leading to a need for research into relevant market liquidity monitoring models.

2. Problem Statement

A market liquidity monitoring system is established by combing through external market data (such as interest rates and exchange rates) and internal financial asset data (such as asset size and financial instrument types) to provide a warning analysis of fair value fluctuations.

The current project focuses on two aspects: finance and modeling. In terms of finance, we have found that market liquidity indicators are a comprehensive set of indicators influenced by numerous factors. From the literature reviews, the financial research direction is divided into three main aspects: monetary market factors, bond market factors, and stock market factors. These three factors can be further divided into six aspects: the trading volume and interest rates in the monetary market, open market operations of the central bank, bond issuance, bond yields and trends, the overview of the primary stock market, and the overview of the secondary stock market. The model specifically involves 16 financial indicators, which will be elaborated on in the data description section.

The China Financial Condition Index (CFCI) is selected as a predictive variable for the market liquidity model. CFCI is an index that measures China's financing conditions, financing availability, and overall macro-financial environment's looseness or tightness. It quantifies factors such as interest rates, yield curves, credit spreads, and asset price indices.



来源：Wind、第一财经研究院
注：指数读数大于 0 代表紧缩，小于 0 代表宽松。

The fluctuation of the CFCI is shown in the following figure (the source of the figure is from the Chinese website Wind, and no corresponding English version is available). The title of the image is the Chinese name of the China Financial Condition Index. The horizontal axis represents time, and the vertical axis represents the specific index. The first row in the lower left corner indicates the source of the image, and the second row provides an interpretation of the index: an index greater than 0 indicates tightening, and less than 0 indicates loosening.

The modeling part will be put in the model selection section.

3. Data Description

3.1. Data Sources

The original model incorporates 24-dimensional indicator data from both internal data sources and external data sources. The confidential data have been removed and the existing model contains a 16-dimensional indicator. External data sources include Wind data, Reuters data, Bloomberg data, and various Chinese financial associations (the National Bureau of Statistics of China, etc). As this report is a phased report, the dimensional indicators may be updated in the future.

3.2. Data Preprocessing

Since the data sources are reliable, data preprocessing is mainly focused on handling missing data, which mainly comes from local holidays. The following are three methods for handling missing values:

- Imputation methods (upper bound substitution) if a few data points are missing
- Linear interpolation is used to fill in the missing values within a small range
- Deletion of continuous missing values reaches 30%

3.3. Data Description

Category	Sub-Category	Indicators
Currency Market	Money Market Volume and Interest Rates	Seven-day Moving Average Repo Rate (R007)
		Inter-bank Pledged Repurchase Weighted Rate (GC001)
		7-day Repurchase Rate (DR007)
		Shanghai Interbank Offered Rate (SHIBOR)
		Interbank Certificate of Deposit: Issue Rate: 1 Month
	Central Bank Open Market Operations	Reverse Repurchase Quantity: 7 days
		China: Reverse Repo: Amount at Maturity
		CNY Deposit Reserve Ratio: Large Deposit-Based Financial Institution
		CNY Deposit Reserve Ratio: Small and medium-sized depository financial institutions
Bond Market	Bond Yield Trends	ChinaBond Treasury Bond Yield to Maturity: 1 Year
		ChinaBond Treasury Bond Yield to Maturity: 10 Year
Stock Market	Primary Market	China: Margin Balance

		Stock Pledged Repurchase: Initial Transaction Amount: Shanghai and Shenzhen Stock Exchange
		Stock Pledged Repurchase: Initial Transaction Amount: Shanghai Stock Exchange
	Secondary market	Shanghai Composite Index: A Shares Turnover
		Shanghai Composite Index: A-share average price-earnings ratio

4. Feature Engineering

4.1. Date Time Feature Selection

For the 16 selected financial metrics, the data with the longest historical record dates back to 1980. Due to the significant impact of time and policy on financial data, the data selected for model building is sourced from the period between 2016 and 2023. In order to achieve precise target results, the selected data granularity is daily.

4.2. Data Accumulation Evaluation

Due to the potential impact of long-term environmental factors on financial data, we calculated the ratio between the current value and the mean value over a given time period as an indicator of the cumulative nature of the data within that period. This was done prior to model building in order to better understand the trends in the data over time.

4.3. Data Differential Features

To better understand the predictive and modeling objects and address time delay issues in prediction, we employ the computation of the difference in indicator data within a time period to describe the feature characteristics of the time series cycles and map their changing trends. This approach helps account for the cumulative impact of long-term environmental changes on financial data, thereby enhancing the accuracy and reliability of our analysis.

4.4. Data Fluctuation Evaluation

To capture the significant changes in certain indicators that may affect the predicted outcome, we compute the logarithmic difference between the current value and the previous value. Then, we calculate the standard deviation of the difference over a specific period to reflect the volatility of the value.

4.5. Feature Selection

The feature selection process involves analyzing the correlation, volatility, and number of singular values of the target variable curve to filter out dimensions with poor data quality and eliminate feature noise.

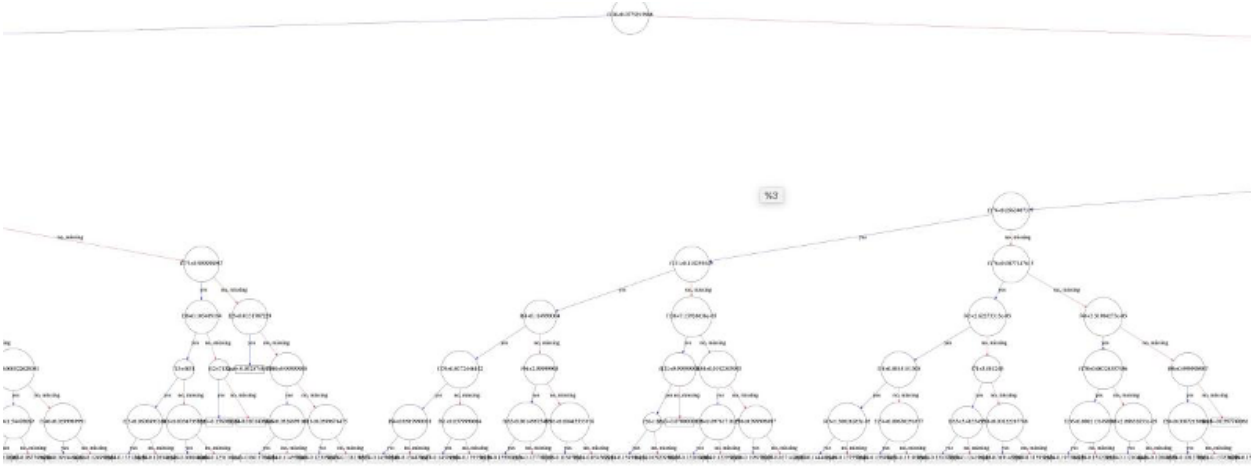
After preprocessing and feature engineering, each original data is expanded to ten times its quantity, and the data is filtered to retain only those within a threshold value before being sent to the model for training and learning.

5. Model Selection

The predictive variables (CFIC) is a weekly updated dataset. However, the data granularity of the research subject is daily, so we will perform result fusion at the end. The following three models have shown good performance.

5.1. XGBoost

XGBoost utilizes gradient boosting to train multiple weak learners (usually decision trees) and combine them into strong learners. The algorithm iteratively improves the prediction accuracy by fitting a new model on the residuals in each iteration and then summing the weighted predictions of all models to obtain the final prediction.



5.2. Light GBM

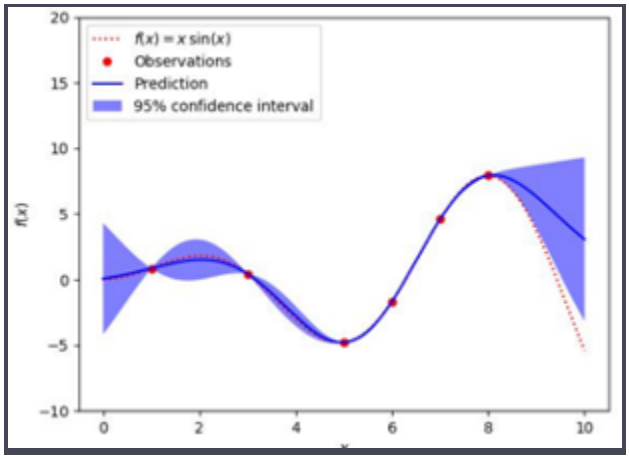
LightGBM is also a gradient-boosting framework based on decision trees. However, it differs from XGBoost in its focus on tree-building methods, using both leaf-wise and level-wise tree growth strategies, while XGBoost primarily uses level-wise methods. Additionally, the regularization rules employed by the models also differ. Overall, these differences enable the models to learn different aspects of the data features in the field of tree-based machine learning models, leading to more comprehensive feature learning.

5.3. TabNet

TabNet is a recently developed deep neural network model based on attention mechanisms. Its design is inspired by the self-attention mechanism in the Transformer model. TabNet learns the relationships between input features and extracts the most useful information by alternately performing feature selection and feature importance evaluation. In each feature selection step, TabNet uses a self-attention-based masking mechanism to select important features, and effectively combines and compresses features through the GLU gating mechanism, reducing overfitting and improving model generalization performance.

5.4. Bayesian Hyperparameter Optimization

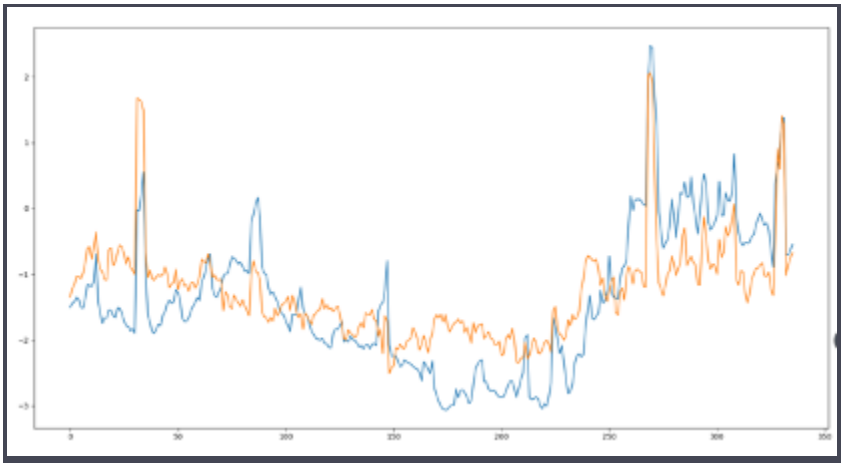
Bayesian model parameter optimization is a model parameter tuning method based on the Bayesian optimization algorithm. The main idea is to use the existing model running results in the model parameter space, estimate the distribution of the objective function under different parameter combinations through the probability model, and select the next parameter combination. This can find the optimal model parameter combination as quickly as possible with fewer running times.



6. Model Performance

During the model fusion process, we aim to improve the overall accuracy of the model by assigning optimal weight allocation. To achieve this, we have developed a dynamic scaling weight factor that can dynamically reduce the weight of easily distinguishable samples during the training process. This allows the focus to quickly shift to the features of samples that are difficult to distinguish, thus improving the accuracy of the model fusion.

At the current stage, the model can predict liquidity status 1 to 1.5 weeks in advance, and the fitting performance for market liquidity is shown in the figure below:



The orange curve represents the model's output for the market liquidity indicator, and the blue curve represents the actual data. The correlation between the two curves is 0.81, and the mean absolute error (MAE) is 0.73.

7. Next Step

From literature reviews, it shows that the indicators representing market liquidity may vary from period to period. Therefore, we aim to increase the model's flexibility by implementing a self-selection process for the indicators. Additionally, we hope to advance the prediction time from 1-1.5 weeks to provide banks with more response time.