ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20230801000842
Original scientific paper

CBISI-LSTM Deep Learning Model for Short-term Cross-border Capital Flow Prediction

Yuchen XIONG, Yihang CHU, Keyang ZHAN, Bixuan LIU, Gang XUE*

Abstract: With the drastic fluctuation of the international financial market in recent years, the cross-border capital flow between Shanghai and Hong Kong has become increasingly active. The lack of effective and timely tracking monitoring and scientific management of cross-border capital flow in the capital market will seriously affect the overall financial security of China's economy. This paper constructs the cross-border investor sentiment index *CBISI* based on principal component analysis and analyzes the impact of cross-border investor sentiment and cross-border capital flows by constructing the VAR model. In addition, *CBISI* is used as part of the input variable of *LSTM* to forecast the cross-border capital flow (*NF*). The findings of the study indicate that changes in cross-border investor sentiment will have a significant short-term impact on cross-border capital flows, and the addition of *CBISI* will improve the accuracy of cross-border flow estimates.

Keywords: cross-border funds; data mining; investor sentiment; operation management; *LSTM*

1 INTRODUCTION

The domestic securities market has grown rapidly in recent years, attracted and amassed a sizable amount of domestic capital, actively encouraged the development of the domestic financial market, provided capital security for the nation's economic construction, encouraged economic and social development, and initially served as an economic "barometer" [1]. Cross-border capital flow channels have expanded since the Shanghai-Hong Kong Stock Connect (SHKSC) opened in 2014, and cross-border money has progressively grown to significantly contribute to stock exchange investing. In addition to a significant inflow of foreign money into the domestic stock market, there have also been significant outflows of local stock market funds, which presents both possibilities and hazards. Uncertainty in the global stock market has dramatically risen since the global financial crisis of 2008. There is a possibility of a significant short-term outflow of cross-border equity capital in numerous nations due to the COVID-19 pandemic, as well as other recent occurrences [2]. To prevent financial risks on the global capital market, China must therefore further understand the factors that influence capital flows in its cross-border capital market. This is also a necessary condition for China to increase its level of opening up. Investor sentiment has grown to be a significant element influencing stock market capital flow with the quick growth of behavioral finance. The investor sentiment index was created by Yi et al. (2009) [3], Baker and Wurgler (2007) [4], and they discovered that it is effective in forecasting domestic stock market returns. There are, however, few studies on the influence of international investor attitudes on international capital flow in the capital market and international capital forecasting. Additionally, as smart devices become more and more popular, more and more people use them for investment-related activities, which increases the amount of data generated and the demand for it [5], both of which have a beneficial effect on the measurement of cross-border investor sentiment. Using this data as a starting point, this paper aims to quantify the sentiment of Chinese cross-border investors following the opening of the SHKSC, in order to construct the sentiment index of cross-border investment and conduct extensive research on the relationship between cross-border investor sentiment and cross-border capital flow as well as the prediction of cross-border capital flow, and observe the impact of cross-border capital sentiment index on the predictability of cross-border capital flow. The rest of this paper is structured as follows: the second part is literature review; the third part is index construction, which specifically explains the construction process of cross-border capital flow sentiment index; the fourth part is data analysis results; the fifth part is cross-border capital forecast analysis; the sixth part is research conclusions and policy recommendations; and the seventh part is research limitations and future research prospects.

2 LITERATURE REVIEW

The closed-end fund discount rate (CFD), which refers to the situation where the asset's cost is lower than its net asset value per unit (NAV), was initially used by Lee et al. (1991) [6] as a proxy indicator of investor mood in the study on the building of a complete investor sentiment index. This research had a big effect. Using the *IPO* first-day return rate, closed fund discount rate, *IPO* number, dividend premium index, total issued shares, and turnover rate as proxy variables, Baker and Wurgler (2006) [7] evaluated the effect of investor mood on stock exchange returns. This establishes the technical starting point for a significant amount of relevant future study. Local academics use this approach and have developed a few investor sentiment indices specific to the Chinese stock market. Utilizing the PCA method, Yi et al. (2009) [3] created the investor emotion index CICSI, which can more accurately reflect investor sentiment on the Chinese stock exchange, by combining the "consumer confidence index" and "number of new accounts opened by investors" indices, which reflect the sentiment of the Chinese market. Su (2011) [8] further developed Chinese investor mood indicators based on the evolution of the Chinese stock exchange from a PCA approach. Additionally, Liu (2019) [9] developed an inverted pyramid filter model open index optimization approach. The findings demonstrate that, compared to the CICSI index built with unoptimized indices, the optimized comprehensive investor sentiment index, or OISI, can better explain investor sentiment. Since the opening of the SHKSC, the research on the capital flow between the two stock exchanges mainly focuses on the liquidity of the two stock markets. Ba et al. (2014) [10] analyzed the potential impact of north-south fund flow, purchase quota, and capital flow in the preparation stage of the SHKSC, and found that although the correlation between the two markets will be strengthened, it is difficult to have a large number of one-way cross-border fund flows. Xu et al. (2016) [11] analyzed the liquidity and volatility of the stock exchange and concluded that the opening and gradual expansion of the Stock Connect would reduce the liquidity of the stock market to some extent. Guo (2018) [12] takes the illiquidity index related to stock yield as verification and finally gives the judgment that the SHKSC can improve the efficiency of stock trading of inefficient companies, but it is not clear about the liquidity improvement effect of efficient companies. Liu et al. (2020) [13] analyzed and compared the cash dividend, price-earnings ratio, and profit of the constituent stocks of the Shanghai Stock Connect, screened 53 constituent stocks as research samples, and used the fixed-effect model to find that the SHKSC trading has A enhancing effect on the efficiency and liquidity of the A-share market. The above-related literature introduces the methods of building an investor sentiment index and the impact of Stock Connect on the capital flow of the two markets. Most of these indexes take the domestic capital market as the research object, and there are few pieces of research on the measurement index of cross-border fund sentiment, and most of the data studied are monthly data, which is not consistent with the characteristics of daily changes of investor sentiment. Therefore, on the basis of previous studies, this paper selects the daily data of all Hong Kong stock Connect (HKSC) constituent stocks from November 2014 to August 2022, constructs the cross-border fund flow emotion index, and studies the relationship between cross-border investor sentiment and cross-border fund flow.

3 INDEX CONSTRUCTION

3.1 Source Index Selection

Based on the Baker-Wurgler index (BW index construction method) and CICSI index construction method [4], this paper selects indicators that can reflect cross-border investor sentiment, namely the turnover rate of all constituent stocks in the Hong Kong Stock Exchange, price-to-earnings ratio, price-to-book ratio, rise and fall, trading volume, *IPO* number, first-day earnings, and consumer confidence index. Using principal component analysis (PCA), the daily index of sentiment index of cross-border fund flows (*CBISI*) was constructed, with a time span from November 2014 to August 2022. Tab. 1 shows the illustrative statistical findings of variables, and the total observed sample is 1895. Data is from the Wind Financial database. The introduction of the indicators is as follows. Turnover rate (*TURN*) *TURN* is a common measure of stock liquidity and investor sentiment. *TURN* of the Hong Kong Stock Exchange (weighted average of current market value) selected in this paper can reflect the sentiment measurement of cross-border investors to a certain extent. Generally speaking, the higher the *TURN* is, the more enthusiastic the market trading will be.

Trading Volume (*VOL*)Turnover may partially indicate the liquidity of the stock exchange in addition to the engagement of investors. The desire to invest in those companies' equities is typically particularly strong when investor sentiment is high. The market size of HKSC is still growing right now, and as the total number of shares and market value are both continually rising, the size of transactions must also increase. In order to determine the elements influencing market size increase, we may divide the daily trading volume by the daily market value. The formula is as follows

$$VOL_i = \sum_{i=1}^{n} VOL_i \Big/ \sum_{i=1}^{n} MEV_i \tag{1}$$

where n is the trading days, the Hong Kong stock market's daily trading volume is referred to as *VOL* and its daily circulating market value is referred to as *MEV*.

The number of *IPO* (*IPON*) and the First-day earnings (*IPOR*). Generally speaking, in the stock exchange, the *IPON* and the *IPOR* are good indicators of investor enthusiasm. This paper adopts the *IPO* date data from the opening of the HKSC, and the *IPOR* adopts the weighted average form:

$$IPOR_i = \sum_{i=1}^{n} \left(P_i - P'_i\right) \cdot LSN_I \Big/ \sum_{i=1}^{n} LSN_i \tag{2}$$

where n is the number of new shares issued on that day, P_i is the closing price of new shares i on the first day of listing, P_i' is its issue price, and *LSN* is the number of outstanding shares issued.

Price/Earnings ratio (*PER*) and Price-to-book ratio (*PBR*). The *PER* is a popular measure of stock market valuation, calculated by dividing the price of a stock by net profit per share. The level of stock valuation can indirectly reflect the level of investor sentiment. When the *PER* is high, it shows that stock market participants are eager to purchase and sell companies at a premium., which means that investor sentiment is high. Conversely, when the *PER* is low, it means that investor sentiment is depressed. In this paper, the daily rolling *PER* (TTM) and *PBR* of individual shares in the Hong Kong Stock Exchange are used as indicators to reflect cross-border investor sentiment.

Percent change (*CHG*) The rise or fall partly reflect investor sentiment over time. Generally speaking, the bigger the gains in a day, the higher the sentiment, and the bigger the declines, the lower the sentiment. Consumer Confidence Index (*CCI*) The *CCI* is a measure of consumer confidence that takes into account all aspects of consumer perceptions of the economy, including their arbitrary perceptions of present economic conditions, income levels, income expectations, and psychological health. *CCI* is chosen in this article as the emotional proxy indicator as a result. The variables' descriptive statistics is displayed in Tab. 1.

Table 1 Descriptive statistics of variables

Variable	Obs	Mean	Std. Dev.	Min	Max
TURN	1895	0.26	0.09	0.06	0.85
VOL	1895	0.00	0.00	0.00	0.00
PER	1895	12.58	1.71	8.95	18.09
PBR	1895	1.05	0.23	0.65	1.83
IPOR	1895	−0.03	1.30	−38.00	15.50
IPON	1895	0.10	0.34	0.00	4.00
CCI	1895	114.92	9.65	86.70	127.60

3.2 The Selection of "Advance" and "Lag" Variables

PCA is a statistical analytic technique that reduces the number of original numerous variables into a small number of comprehensive indices. The specific process is as follows. There are n samples, each sample observes p indexes, and the standardized data is represented by a matrix:

$$X = \begin{bmatrix} \vdots & \ddots & \vdots \end{bmatrix} \tag{3}$$

And establish a variable correlation coefficient matrix:

$$R = \begin{bmatrix} r_{11} & \cdots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{p1} & \cdots & r_{pp} \end{bmatrix} \tag{4}$$

Among them:

$$r_{ij} = \frac{\sum_{k-1}^{n}\left(x_{ki} - \overline{x_i}\right)\left(x_{ij} - \overline{x_j}\right)}{\sqrt{\sum_{k-1}^{n}\left(x_{ki} - \overline{x_i}\right)^2 \sum_{k=1}^{n}\left(x_{ij} - \overline{x_j}\right)^2}} \tag{5}$$

Then, the characteristic root of R and corresponding unit eigenvectors are obtained, and the contribution rate of principal component:

$$\frac{\lambda_i}{\sum_{k=1}^{i}\lambda_k}\left(i = 1,\ 2,\ ...,\ p\right) \tag{6}$$

and the cumulative contribution rate are calculated.

$$\frac{\sum_{k=1}^{i}\lambda_k}{\sum_{k=1}^{p}\lambda_k}\left(i = 1,\ 2,\ ...,\ p\right) \tag{7}$$

Firstly, PCA was performed on the leading and lagging variables of the above eight indicators. A cross-border Investor emotion index (*cbisi*) with 16 variables was created after z-score was standardized to remove the impact of unit discrepancies before processing.

Then, correlation analysis was conducted between cbisi. 8 indexes with high correlation coefficients were chosen as the source indexes for creating *CBISI* after a correlation analysis between *CBISI* and the advance and lag of those 8 indexes was completed. Tab. 2 presents the outcomes.

Table 2 Correlation coefficients between cbisi and 16 variables

cbisi	$TURN_t$	VOL_t	PER_t	PBR_t
	0.72***	0.22***	0.56***	0.71***
	$TURN_{t-1}$	VOL_{t-1}	PER_{t-1}	PBR_{t-1}
	0.71***	0.23***	0.55***	0.70***
	CHG_t	$IPON_t$	$IPOR_t$	CCI_t
	0.19***	0.17***	0.07***	0.60***
	CHG_{t-1}	$IPON_{t-1}$	$IPOR_{t-1}$	CCI_{t-1}
	0.12***	0.19***	0.07***	0.60***

Note: *** indicates a significance level of 1%.

As can be seen from Tab. 1, cbisi is highly correlated with $TURN_t$, VOL_{t-1}, PER_t, PBR_t, CHG_t, $IPON_{t-1}$, $IPOR_{t-1}$, and CCI_{t-1}. Therefore, the above eight variables are used as the final index to construct *CBISI*.

3.3 The Construction of *CBISI*

Firstly, PCA (standardized) was performed on eight source indicators, namely, $TURN_t$, VOL_{t-1}, PER_t, PBR_t, CHG_t, $IPON_{t-1}$, $IPOR_{t-1}$, and CCI_{t-1} to obtain *CBISI*. The correlation analysis between *CBISI* and cbisi showed that the correlation between the two reached 87.6% (below the 1% significance level). The results indicate that the deletion of 8 variables has little effect on *CBISI*, and *CBISI* can still better reflect the changes in cross-border investor sentiment. Secondly, SPSS software was used to conduct the *KMO* test and Bartlett spherical test for the intended indicators, which showed that the *KMO* value of each indicator was higher, indicating the stronger commonality of variables. In this paper, $KMO = 0.586$, indicating that it is suitable for factor analysis. Bartlett loop test results were remarkably robust ($p < 0.01$). The factor load was rotated and its component score was calculated to obtain the results in Tab. 3 and Tab. 4.

Table 3 Table of factor loading coefficients

CV	1	2	3	4	5	6	7	8
$TURN_t$	0.47	0.74	−0.06	−0.12	−0.03	0.20	-0.42	0.06
VOL_{t-1}	−0.10	0.90	0.02	−0.01	0.00	0.05	0.41	−0.06
PER_t	0.79	0.13	−0.03	0.06	0.07	−0.58	0.06	0.15
PBR_t	0.93	−0.14	−0.04	0.01	0.01	0.03	0.01	−0.34
CHG_t	0.07	0.08	0.37	0.88	0.26	0.09	−0.04	0.01
$IPON_{t-1}$	0.07	0.01	0.73	−0.07	−0.68	−0.06	−0.02	−0.01
$IPOR_{t-1}$	0.08	−0.04	0.64	−0.48	0.62	0.03	0.01	0.01
CCI_{t-1}	0.80	−0.28	-0.04	−0.03	−0.08	0.41	0.24	0.20

Table 4 Component score coefficient matrix

CV	1	2	3	4	5	6	7	8
$TURN_t$	0.20	0.50	−0.06	−0.12	−0.03	0.35	−1.04	0.34
VOL_{t-1}	−0.04	0.61	0.02	−0.01	0.00	0.08	1.02	−0.34
PER_t	0.33	0.09	−0.03	0.06	0.07	−1.03	0.14	0.79
PBR_t	0.39	−0.10	−0.04	0.01	0.01	0.06	0.02	−1.83
CHG_t	0.03	0.05	0.34	0.88	0.28	0.16	−0.11	0.05
$IPON_{t-1}$	0.03	0.01	0.68	−0.07	−0.73	−0.10	−0.04	−0.04
$IPOR_{t-1}$	0.03	−0.03	0.59	−0.45	0.68	0.05	0.02	0.04
CCI_{t-1}	0.34	−0.19	−0.04	−0.03	−0.09	0.74	0.59	1.09

According to the study in Tab. 3 and Tab. 4, the investor emotion index is created by employing the weighted ratio of the variance contribution rates of each key component retrieved to the overall variance contribution rates. Eq. (8) presents the ultimate outcome.

$$\begin{aligned} &CBISI = 0.10TURN_t + 0.15VOL_t + 0.079PER_t + \\ &+0.058PBR_t + 0.213CHG_t + 0,001IPON_{t-1} - \\ &-0.111IPOR_{t-1} + 0.153CCI_{t-1} \end{aligned} \tag{8}$$

It is evident that *CBISI* has the following advantageous characteristics: Statistics show that the following factors have a positive correlation with cross-border investor sentiment: *TURN*, *VOL*, *PER*, *PBR*, *CHG*, *IPOR*, and *CCI*. Although it has a tiny impact, the amount of ipos will cause a slight downturn in investor mood. Second, investor sentiment is predicated on the *VOL*, *IPON*, *IPOR*, and *CCI*. Accordingly, the higher the trading volume, first-day earnings, and consumer confidence index in the early period, the higher the *IPON* in the early period, and the higher the *CCI* in the early period, the more the investor sentiment will be impacted by the later period.

4 EXPERIMENTAL ANALYSIS

4.1 Index Selection

Daily data of North-South net flow from the opening of HKSC to 2021 are selected. Northbound funds represent inflows, which is positive. Southbound capital represents the outflow of capital, which is negative. All data are standardized. Data from the wind database.

4.2 Data Analysis

In this paper, ADF was used for the unit root test, and SPSS software was used to test the stationarity of the data. The results showed that the P values of the *CBISI* sequence and fund flow sequence *NF* were both less than 0.01, and the null hypothesis could be rejected at the significance level of 1%. Both the *CBISI* sequence and *NF* sequence were stationary time series and could be analyzed for time series.

This paper studies the influence of *CBISI* on the cross-border capital market (*NF*) by constructing the VAR model.

$$\begin{pmatrix} NF_{t-k} \\ CBISI_t \end{pmatrix} = a_0 + A_1 \begin{pmatrix} NF_{t-1} \\ CBISI_{t-1} \end{pmatrix} + \ldots + A_k \begin{pmatrix} NF_{t-k} \\ CBISI_{t-k} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \quad (9)$$

Before building the model, the hysteresis coefficient should be determined, and the specific results are shown in Tab. 5.

Table 5 Lag order of VAR model

Lag	Log*L*	*LR*	*FPE*	*SC*	*HQ*
0	−3449.49		0.18	3.95	3.95
1	−2832.96	1230.94	0.09	3.26	3.25
2	−2697.83	269.47	0.08	3.13	3.11
3	−2648.96	97.36	0.07	3.09	3.06
4	−2633.82	30.12	0.07	3.09*	3.05*
5	−2627.5	12.56*	0.07*	3.1	3.05

Note: * refers to the lag order selected by the standard.

It can be seen from the results in Tab. 5 that the *SC* value is the minimum when the lag order is 5, that is, the lag order of the VAR model is 4. Fig. 1 shows the test results of model stability. All points fall within the unit circle, indicating that the model is stable.

Based on the fact that both *CBISI* and *NF* sequence data are stationary time series, the Granger causality test was conducted for them. The results show that the change in fund flow can cause a change in investor sentiment at a significant level of 1% ($p < 0.01$). Changes in the investor emotion index *CBISI* can lead to changes in *NF* at a significant level of 1% ($p < 0.01$), indicating the Granger causality between the two.

Table 6 Granger causality test results

Null hypothesis	*F*	*P*
NF does not Granger Cause *CBISI*	16.562	0.000***
CBISI does not Granger Cause *NF*	7.876	0.000***

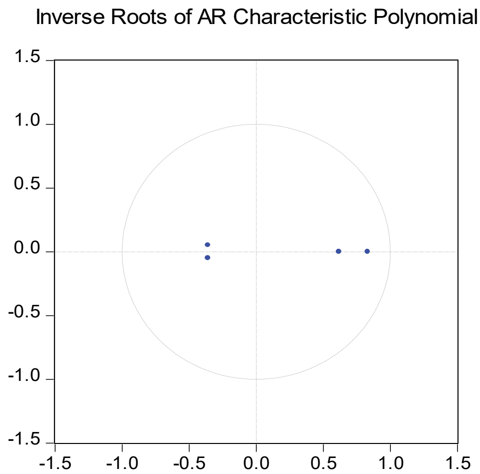

Figure 1 Stability test results of the VAR model

Fig. 2 shows the impulse response function of the *CBISI* and cross-border capital flow (*NF*). According to the findings, *NF* will see a negative impact from a positive standard deviation of investor mood, leading to a capital outflow. The impact does not respond in the first period and then reaches its maximum impact in the second period, and then gradually weakens. This indicates that *CBISI* has a lag effect on *NF*, and the higher the *CBISI* in the mainland, the greater the outflow of capital. Secondly, when a positive standard deviation impact is given to investor sentiment, investor sentiment will decline immediately after a brief rise and turn negative. This suggests that when money flows into the mainland, cross-border investor sentiment briefly rises, but then quickly wanes within a few days before levelling off. Therefore, we can draw a preliminary conclusion: cross-border investor sentiment will have a great impact on *NF* in the short term, but this emotional impact will gradually fade in the long term.

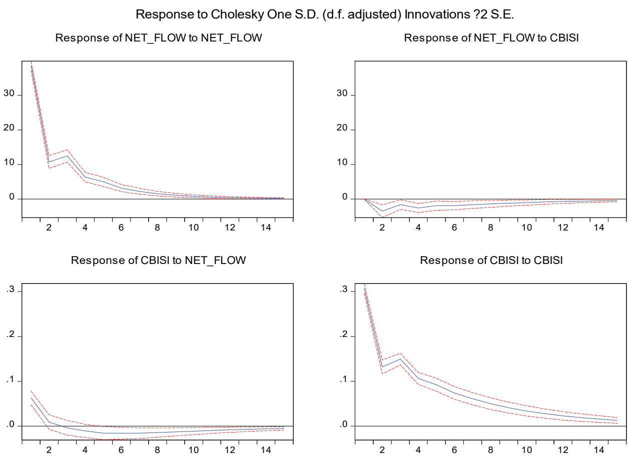

Figure 2 *CBISI* and *NF* impulse response

4.3 Robustness Test

Here, the daily data of the Shanghai composite index (SSEC) is used as the instrumental variable instead of the daily data of the net flow of funds from the North to the south and investor sentiment for regression analysis to evaluate the reliability of the findings of the impact of the aforementioned investor mood on the international stock market. First, do the Winsorize processing on the sequence within 1% up or down, and then the ADF test was performed. The results were significant at the significance level of 1% ($p < 0.01$), which is a stable sequence. Then, the lag order of the VAR model is determined to be 2, and Granger causality test is conducted. The results shown in Tab. 7 show that they are mutually causal. The robustness test results are consistent with the above test results, SSEC has an impact on investor sentiment, that is, the change of SSEC can cause the change of investor sentiment, while investor sentiment can also affect the change of SSEC, the two are Granger causality relationship, therefore, the regression results above are robust.

Table 7 Granger causality test results

Null hypothesis	*F*	*P*
CNY does not Granger Cause *CBISI*	17.483	0.000***
CBISI does not Granger Cause *CNY*	9.872	0.000***

5 FORECAST ANALYSIS OF CROSS-BORDER CAPITAL FLOW

The constructed *CBISI* will be used as part of the input of *LSTMS* to construct new *LSTMS*, and perform the task of predicting cross-border capital flows. Then, compared with the *LSTMS* without *CBISI*, its effect in predicting *NF* will be observed. Here, the daily data of the North-South net flow of funds from the opening of the HKSC to July 29, 2022 is selected for prediction, with a total of 1815 pieces of data. The northbound fund represents the inflow of funds, which is positive. Southbound capital represents the outflow of capital, which is negative.

5.1 *LSTM*

The Long Short-Term Memory Network (*LSTM*) is a particular type of temporal cyclic neural network that was created to address the long-term reliance issue with the ordinary RNN (Recurrent Neural Network). It is a kind of RNN that is widely used for the classification/prediction of time-series data types with a high correlation between previous and current values [14]. For time series data, the *LSTM* network may successfully maintain the data's long-term influence [15, 16], and this structure has the capability to remember the values over time intervals [17]. Unlike normal RNN, *LSTM* has an extra channel that stores the neuron's memory between now and the previous moment.According to the new input and the output from the prior time, the *LSTM* will first select which memories to forget when a new input is received. It will next decide whether to incorporate the new data into the unit state through the memory gate. Finally, the *LSTM* unit calculates the output value of the current time through the output gate.

5.2 *LSTM* with *CBISI* (*CBISI-LSTM*)

Ordinary *LSTM* takes the data of t moments before the current moment as input, in order to forecast the present moment's data. Here, the *CBISI* at the current moment and t previous moments and the data of cross-border fund flows at t previous moments are input to observe the effect of *CBISI* on the prediction of cross-border fund flows. The overall structure is shown in Fig. 3

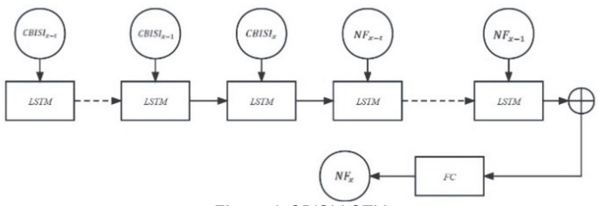

Figure 3 *CBISI-LSTM*

5.3 Network Structure

One input layer, two hidden layers, and one output layer make up its four levels of network structure. Each layer has 11, 6, 6 and 1 neurons, respectively. The four-layer network in the control *LSTM* without *CBISI* had the same network architecture as the *CBISI-LSTM* and had 5, 6, 6, 1 neurons in each layer. The table below displays a few of the model's hyperparameters. Among the 1815 pieces of data, the first 1267 pieces of data are selected as the training set, and the last 543 pieces of data are selected as the test set. The ratio of data volume between the training set and the test set is about 7:3, and $t = 5$ is taken, as shown in Tab. 8

Table 8 Hyper parameters

Hyper Parameter	Value
epochs	10
number of hidden layers	2
number of hidden layer units	6
activation function	Relu
batch-size	2
optimizer	ADAM

5.4 Results

Fifty experiments were carried out for each model, and the average of the 50 experiments was taken as the result. Compared with *LSTM*, *CBISI-LSTM* has a better prediction effect, with a lower MSE of 0.007. This suggests that appropriate use of *CBISI* to forecast cross-border flows can lead to more accurate forecasts and thus help to better identify the risks of cross-border flows. The mean square error between the predicted value and the real value of the two models on the test set is shown in Tab. 9.

Table 9 MSE

Model	MSE
LSTM	0.5830
CBISI-LSTM	0.5760

6 CONCLUSIONS AND POLICY SUGGESTIONS

This paper selects the *TURN*, *PER*, *PBR*, *CHG*, *VOL*, *IPON*, *IPOR*, and *CCI* of all the constituent stocks of the HKSC, and constructs the daily index of *CBISI*. The results show that *TURN*, *VOL*, *PER*, *PBR*, *CHG*, *IPOR*, and *CCI* are positively correlated with the cross-border investor sentiment index. With the increase in the *IPON*, investor sentiment will be slightly depressed, its influence factor is small. *NF* will be impacted by *CBISI*. The quantity of capital outflow will increase in direct proportion to the level of investor mood, and this influence has a delayed effect. In order to investigate how the constructed cross-border capital sentiment index *CBISI* affects the prediction of cross-border capital flow, this study also includes it as an input variable for *LSTM*. 50 experiments were performed on each model, and the average value of the 50 experiments was taken. The results demonstrate that the right incorporation of *CBISI* in specific *LSTM* input variables can increase prediction accuracy, allowing *CBISI* to be used to further investigate the cross-border capital early warning problem and forecast the risk of cross-border capital flow.

Policy suggestion:

In the increasingly complex international capital market, a variety of information in the market will greatly affect the emotions and behaviours of cross-border investors, and the blindness of the market will have a significant influence on the financial market (Dinger et al., 2020) [18]. Therefore, relevant departments can issue a relevant early warning through various kinds of media when the market risks are large, increase the publicity of relevant domestic financial policies, prevent the spread of panic in the market through various ways, guide reasonable market expectations, keep the domestic capital market stable and reduce the likelihood of a significant outflow of local funds.

Relevant departments should also do a good job in the statistics, monitoring, and analysis of the net outflow of cross-border capital (Ye et al., 2020) [19], constantly update and improve the cross-border investor sentiment index, and increase the early warning indicators affecting cross-border capital outflow by incorporating it into the cross-border capital early warning system, and set up scientific risk early warning lines. Thus, it expands the toolbox of finance security early warning. The government should also focus on coordinating the implementation of the financial emergency response work plan, the adjustment of the national fiscal policy, monetary policy and the adjustment of the national macro-financial prudential management policy, and firmly grasp the bottom line of systemic and regional financial risks effectively at the same time.

Research limitations and future research prospects:

This paper uses the data from the establishment of the HKSC to 2022, but does not use the data after 2022, which may lead to the lack of objectivity in the research and have a certain impact on the results. In this paper, PCA is used to construct the sentiment measurement index of cross-border funds. The construction method is relatively traditional, and some new methods can be used to construct the index, and the effect of indexes constructed by different methods on reflecting cross-border investor sentiment is compared. Update the data in a timely manner, and use the data from the establishment of the HKSC to the present for research to improve the objectivity of the research. The comments of investors in the forum can be used as the source data to construct the sentiment measurement index of cross-border investors, and there is innovation in the data source. The index can be constructed using deep learning methods and compared with other methods to compare its effectiveness in reflecting cross-border investor sentiment.

7 REFERENCES

[1] Bathia, D., Bouras, C., Demirer, R., & Gupta, R. (2020). Cross-border capital flows and return dynamics in emerging stock markets: Relative roles of equity and debt flows. *Journal of International Money and Finance*, *109*. https://doi.org/10.1016/j.jimonfin.2020.102258

[2] Hao, H., Guo, L., & Dong, J. (2022). How does venture capital cross-border syndication spur corporate innovation? Evidence from China. *Frontiers in Psychology*, *13*, https://doi.org/10.3389/fpsyg.2022.921168

[3] Yi, Z. & Mao, N. (2009). Measurement of investor sentiment in Chinese stock market: Construction of CICSI. *Journal of Financial Research*, *11*, 174-184.

[4] Baker, M. & Wurgler, J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*, *21*, 129-152.

[5] Šutić, D. & Varga, E. (2022). Scaling industrial applications for the Big Data era. *Computer Science and Information Systems*, *19*(1), 117-139. https://doi.org/10.2298/CSIS200531039S

[6] Lee, C. M. C., Shleifer, A., & Thaler, R. H. (1991). Investor sentiment and the closed-end fund puzzle. *The journal of finance*, *46*, 75-109.

[7] Baker, M. & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The journal of finance*, *61*, 1645-1680.

[8] Wei-dong, S. (2011). Construction of BRAR of Investors Based on PCA. *Taxation and Economy*, *3*, 42-47.

[9] Liu, X. W. (2019). Study on the Optimization of Investor Sentiment Measure Indicators in Chinese Stock Market. *Chinese Journal of Management Science*, *27*(1), 22-33.

[10] Ba, S. & Zhang, X. (2014). The impact of SHKSC on cross-border capital flows. *China Finance*, *18*, 48-49.

[11] Xu X. C. & Chen Z. J. (2016). The Impact of SHKSC Program on the Volatility and Liquidity of the Stock Market. *Journal of Zhejiang Gongshang University*, *6*, 76-83.

[12] Guo, Y. S. (2018). On the Role of SHKSC Mechanism in Enhancing Firm Value. *Journal of Guangdong University of Business Studies*, *33*, 77-88.

[13] Liu, Y. & Li, J. (2020).The effect of SHKSC on the Liquidity of Shanghai's Stock Market. *Journal of Chongqing University of Arts and Sciences (Social Sciences Edition)*, *39*, 38-48.

[14] Kim, K., Lee, J., Lim, H., Oh, S. W., & Han, Y. (2022) Deep RNN-Based Network Traffic Classification Scheme in Edge Computing System. *Computer Science and Information Systems*, *19*(1), 165-184. https://doi.org/10.2298/CSIS200424038K

[15] Wang, T. (2021). A K-means Group Division and LSTM Based Method for Hotel Demand Forecasting. *Tehnički vjesnik*, *28*(4), 1345-1352. https://doi.org/10.17559/TV-20210507172841

[16] Wu, P. J. & Yang, D. (2021). E-Commerce Workshop Scheduling Based on Deep Learning and Genetic Algorithm. *Int. Journal of Simulation Modelling*, *20*(1), 192-200. https://doi.org/10.2507/IJSIMM20-1-CO4

[17] Nelson, C., S., Kumar, M. T., & Prakash, G. L. (2022). A Novel Optimized LSTM Networks for Traffic Prediction in VANET. *Journal of System and Management Sciences*, *12*(1), 461-479. https://doi.org/10.33168/JSMS.2022.0130

[18] Dinger, V. & Kaat, D. M. (2020). Cross-border capital flows and bank risk-taking. *Journal of Banking and Finance*, *117*. https://doi.org/10.1016/j.jbankfin.2020.105842

[19] Ye, Z., Zhuo, R., & Han, S. (2020). Cross-Border Capital Flow, Measurement Method Selection and Financial Risk Prevention. *Test Engineering and Management*, *83*, 25879-25885.

[20] Sfenrianto, Q. G. A. & Rianto, L. (2023). Tuberculosis treatment failure classification based on electronic medical records using PCA-ANN. *Journal of System and Management Sciences*, *13*(1), 133-153. https://doi.org/10.33168/JSMS.2023.0108

Contact Information:

Yuchen XIONG
School of Economics and Management, Southeast University,
Nanjing 211189, China
E-mail: 213200910@seu.edu.cn

Yihang CHU
School of Economics and Management, Southeast University,
Nanjing 211189, China
E-mail: 213200624@seu.edu.com

Keyang ZHAN
School of Economics and Management, Southeast University,
Nanjing 211189, China
E-mail: zhankeyang@seu.edu.cn

Bixuan LIU
School of Mathematics, Southeast University,
Nanjing 211189, China
E-mail: weifeng_0911@qq.com

Gang XUE
(Corresponding Author)
School of Economics and Management, Tsinghua University,
Beijing 100084, China
E-mail: xuegang@sem.tsinghua.edu.cn